Department of Embedded Technology
School of Electronics Engineering
Vellore Institute of Technology, India

Arun Manoharan
Hemprasad Patil
Sujatha Rajkumar
Usha Rani Seshasayee

Emerging Trends in Artificial Intelligence for Internet of Things

ISBN 9789389640083 Vol 1, Dec 2019

Arun Manoharan
Hemprasad Patil
Sujatha Rajkumar
Usha Rani Seshasayee

Department of Embedded Technology
SENSE, Vellore Institute of Technology
Vellore, Tamil Nadu, India.

Emerging Trends in

Artificial Intelligence for Internet of Things

ISBN 9789389640083 Vol 1, Dec 2019

Preface

Advances in science, engineering and technology have paved way to solve many challenging problems targeting societal needs. Artificial Intelligence (AI) and Internet of things (IoT) are some of the areas which have a vast potential to provide tangible benefits in leaps and bounds. The role of the scientific community is to reap the maximum benefits of the existing technology and grow parallel with it. The book, Emerging Trends in Artificial Intelligence for Internet of Things presents the chapters of researchers and academicians who contribute their work in the areas of AI and IoT.

Chapters of this book are presented in the Virtual Conference of Artificial Intelligence for Internet of Things which are research outcomes and technological advances at the crossroads of artificial intelligence, industrial automation and network security, all of them tailored for IoT. This book would be a resource for engineers and scientists to refer theoretical, experimental, and simulation results targeting diverse applications of AI for IoT.

Editors have organized the publication of the original work, be it simulation, hardware implementation and review material in these sixteen chapters, pertaining to AI and IoT. The editors would like to thank each and every one for their time and effort towards the successful publication of the work.

Editors

TABLE OF CONTENTS

Malicious URL Detection by Combining Machine Learning and Deep Learning Models

A. RAKOTOASIMBAHOAKA*, I. RANDRIA, N. R. RAZAFINDRAKOTO

EDMI University, BP 1264 – Campus Universitaire d'Andrainjato, Fianarantsoa, 301, Madagascar

Abstract. Nowadays, research on efficient methods for detecting malicious website by URL is still an open issue despite the different approaches proposed in the literature. Malicious URL has become a serious and common threat to cyber security. Our goal is to address this issue with a reliable, generic and flexible system. So in this paper, an hybrid approach based on Machine Learning and Deep Learning methods (CNN-LSTM-RF and RF-CNN-LSTM) is proposed to solve the problem: the one performs a manual extraction of features with RF model, and the other an automatic extraction with CNN_LSTM. CNN_LSTM_RF holds an interesting result, with convergence after three epochs and an accuracy rate of malicious URL detection of 96%. RF-CNN-LSTM algorithm precision rate is 50%, which tends to overfit.

Keywords: cyber-security; deep learning; machine learning; malicious URLs.

1. Introduction

The digital word has much evolved these last years, specifically in using distributed systems such as Internet, which is vital for communication and information exchange, online selling, etc, web being exchange and communication's main source. Yet it appeared that more used is web more insecurity increases because of hackers' innovative techniques. This major concern focuses our interest to solve problem related to automatic detection of web vulnerability relevant from Y. Liu et al, 2015 [1] and K. Soska et al, 2014 [2].

There are several basic clues of web vulnerability detection such as http data, URLs data and web page content, code source of web page. In this paper, data URL is used to solve web insecurity's problem. In fact, malicious URLs have become a common threat for online services such as business, social networks, banking services, etc. And furthermore, according to Sahoo's work, 2017 [3], despite extensive studies in recent

* Corresponding author.
E-mail address: rakotoas.cyprienna@gmail.com, iadaloharivola.randria@emit.mg

years, malicious URLs detection using machine learning remains an open problem, for the future of web applications security.

Methods of malicious URL detection can be classified as follows: blacklist, machine learning and deep learning. It turns out that blacklist method is limited by impossibility of detecting newly generated URLs, which constitutes our interest in machine learning and deep learning. Several works compared performances of these two approaches, as J. Zhao, N. Wang, Q. Ma, et Z. Cheng's work, 2018 [4] and F. Douksieh Abdi, L. Wenjuan's work, 2017 [6]. J.Zhao colleagues have suggested Random Forest (RF) method with manual extraction of features as well as gated recurrent units (GRU) method, with 150 000 legitimate URLs and 240000 malicious URLs, GRU obtain 98, 5% of accuracy whereas 96, 4% for RF model.

However, combining these two approaches proved to be effective in several domains such as image processing by G. Ca, 2013 [7], text classification by D. Gordeev, 2016 [8], speech recognition by L. Zheng, 2018 [9]. Fang's colleagues [5] have used it in phishing websites recognition case. Comparing RF, LSTM and LSTM+RF, the latter significantly improves the accuracy with 98, 52% rate. The result of this work gives us a motivation to combine the three models CNN, LSTM and RF. This model has not been processed yet in malicious URL detection case.

It is important to mention, that these models have limitation: i) Random Forest (RF) is a machine learning model, which is limited by the manual extraction of features and still requires human intervention; ii) the Long Short Term of Memory (LSTM) algorithm cannot determine the phishing URL whose functionalities are similar to those of normal URLs [5]; iii) Convolutional Neural Network (CNN) do not have the ability to capture data synchronization and long-term dependencies. To overcome these limitations, a hybrid architecture of these three models will be proposed using supervised learning since: i) RF combines other characteristics to compensate LSTM performance deficiencies [5]; ii) LSTM presents the characteristics of memory selectivity [5]; iii) CNN can automatically learn the features of the inputs.

In order to have a better performance on malicious URLs detection, compared experiments in an offline environment are conducted. In Section 2, materials and methods are used to detect malicious URLs, as in section 3 for experiments and results' discussions, and finally, in section 4, for conclusion and perspectives.

2. Materials and methods proposed

2.1. Materials and software

The experimental environment and the configuration information are as follows:

- Computer configuration: Ubuntu 18.04.3;
- Memory: 20 Go;
- TensorFlow version: 1.13.1, keras version: 2.2.4, Python version: 3.7;
- Framework: Eclipse and anaconda with pydev plug-in;
- 50% malicious URLs, 50% legitimates URLs from Kaggle in all types of data;
- Duration of training: about 10 to 20 min per model for automatic extraction; about 6h for manual extraction.

2.2. Features extraction

Four different experiments are performed: in the first experiment, the most popular models of machine learning are compared. In the second experiment, a comparison of the most used deep learning models has been done. In the third experiment, the most efficient models in experiments 1 and 2 are combined (CNN_LSTM and RF); CNN_LSTM plays a role of features extractor and RF plays a role of classifier. In the fourth experiment, the same models as in the experimentation 3 are used except that RF model extract the features and the CNN_LSTM model classifies the URL. Figure 1 shows the mapping of URLs features to detect whether a URL is malicious or not.

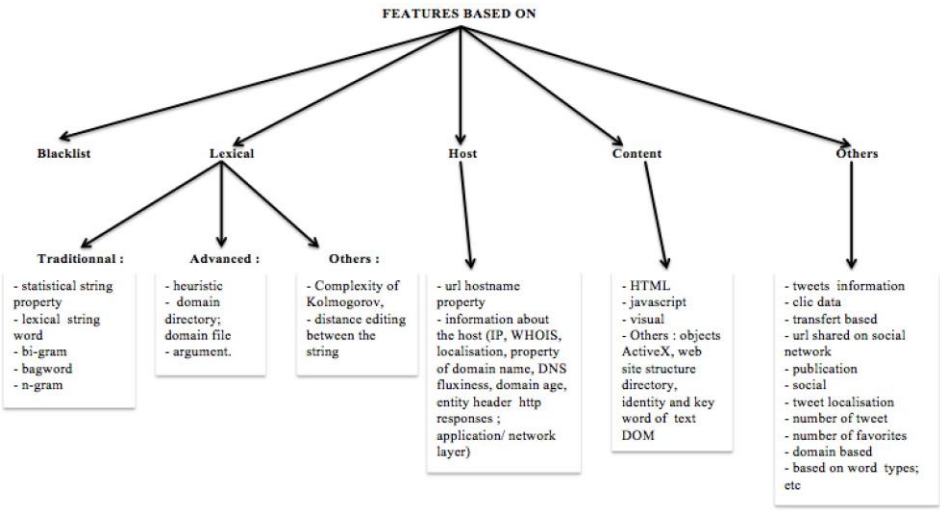


Fig. 1. Different URLs features available in malicious URL detection

The existing features of malicious URL detection methods can be classified in five groups. In this paper, lexical and host features were chosen because their combination would give the best prediction. In addition, features based on host allow identifying the hacker IP address. In lexical features, features based on heuristic can withstand the obfuscation method used by hackers to confuse web users. Twelve features will be used during the experiment: number of dots; number of ';', '_', '? ','=','&'; IP address presence; presence of '-'; presence of '@'; presence of '//'; presence of '/'; URL extension; length of subdomain; query presence.

2.3. Proposed methods

Architecture: RF+CNN_LSTM and CNN_LSTM+RF

Two models for malicious URL detection will be proposed: CNN_LSTM_RF with automatic extraction of features and RF_CNN_LSTM with manual extraction of features. Figure 2 shows the architecture of CNN model combined with LSTM as features extractor and RF as classifier.

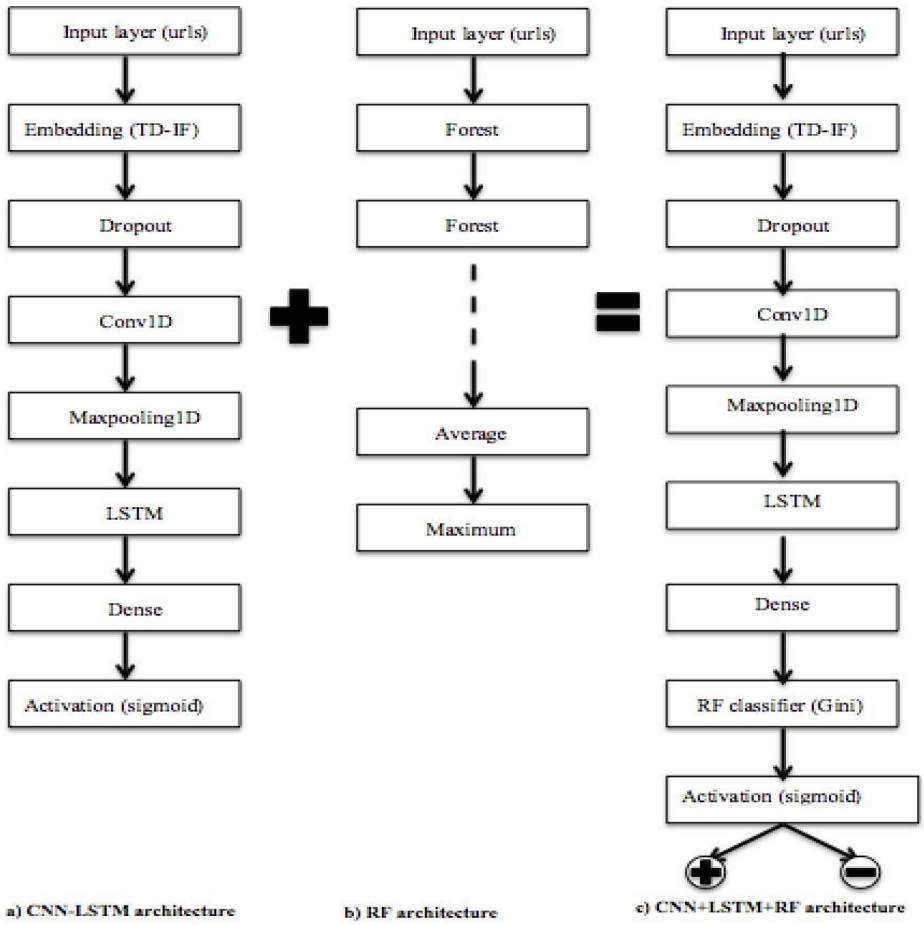


Fig. 2. Malicious URL detection model with CNN_LSTM_RF

Figure2 represents CNN_LSTM_RF architecture, the steps of detection are as follow: CNN-LSTM model extract the feature automatically. But since the model needs to be able to process the URL data, these latter must be converted into feature vectors. Then, convolution layer and maxpooling of CNN process the matrix data and the output of this layer becomes the input of LSTM model. Finally, the two layers are combined and connected in order to help Random Forest model predicts and classifies the values of output layer. The probability of this model is used to predict whether the new URL is malicious or not.

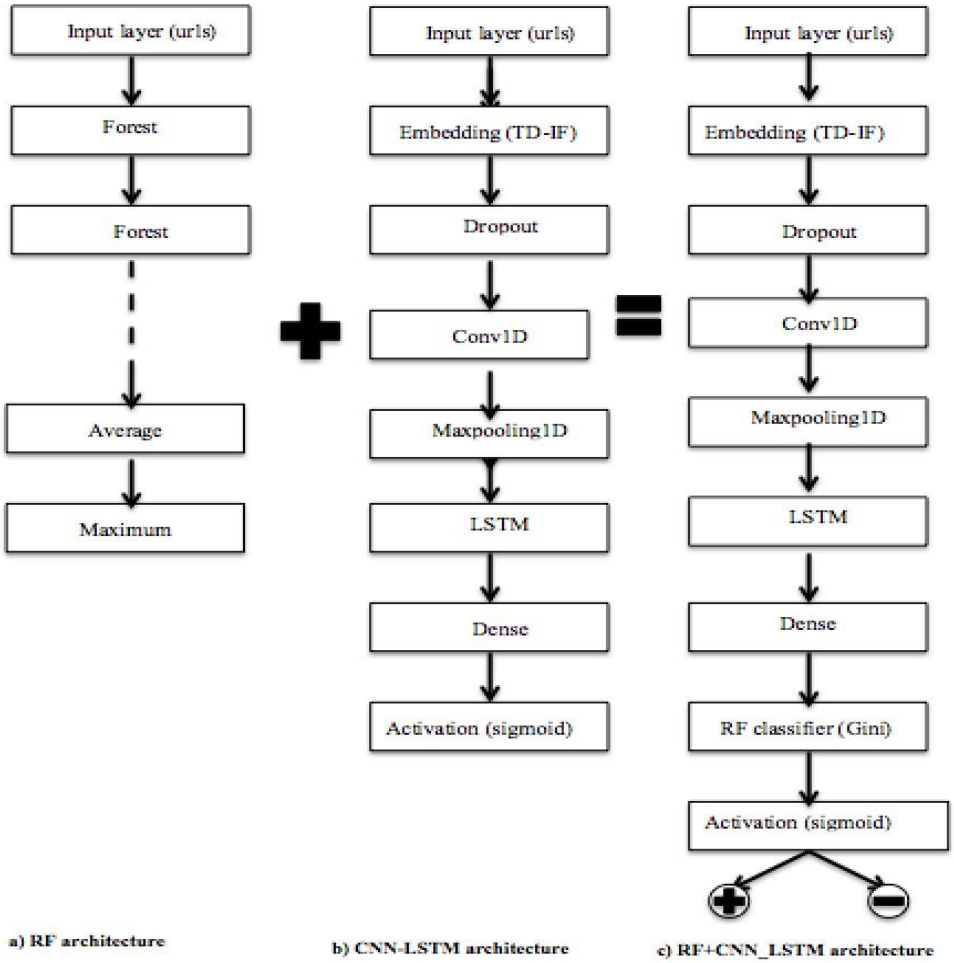


Fig. 3. Malicious URL detection model with RF_CNN_LSTM

Figure.3 uses a manual extraction of features with RF and a classifier with CNN_LSTM model. First, the RF model performs an extraction of features according to the features mentioned in section 3.1.This would be trained to CNN_LSTM model in such way, that the values of the output layer predicts and classifies if the URL is malicious.

Algorithms:

Our models combine CNN and LSTM to extract the features of URLs, the RF model predicts if the URL is malicious or not. Loss and accuracy are used as metrics.

- *Algorithm1: CNN*

Input: features vector of URL $x(L \times d), F x(L \times d), F$ filter

Output: the most important features

For each filter *do*

$$f \in \{1, \dots, F\} f \in \{1, \dots, F\}$$
$$w_j = [\, x_j + x_{j+1} + \cdots + x_{j+k-1} \,]$$
$$c_j = \mathrm{ReLU}\left(w_j\ m + b\right) c_j = \mathrm{ReLU}\left(w_j\ m + b\right)$$
$$w = [\, c_1 + c_2 + \cdots + c_n w = [\, c_1 + c_2 + \cdots + c_n$$

End for

- *Algorithm2: LSTM Input*: output of CNN

Output: output from the hidden state at the last time step of LSTM

For every time step do

$$t \in \{1, \dots, T\} t \in \{1, \dots, T\}$$
$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right)$$
$$f_t = \sigma\left(W_f x_1 + U_f h_{t-1} + b_j\right)$$
$$o_t = \sigma\left(W_o\, x_1 + U_o h_{t-1} + b_o\right)$$
$$\check{c} = \tanh(W_c\, x_1 + U_c h_{t-1} + b_c) \check{c} = \tanh(W_c\, x_1 + U_c h_{t-1} + b_c)$$
$$c_t = i_t * \check{c}_t + f_t * c_{t-1}$$
$$h_t = o_t * \tanh_{(c_t)}$$

End for

Where

- σ: is the logistic sigmoid function that has an output in [0,1]
- Tanh denotes the hyperbolic tangent function that has an output in [−1, 1].

- *Algorithm3: CNN+LSTM*

The output of the hidden state at the last time step of LSTM is used as CNN_LSTM output with sigmoid activation, Adam optimiser and loss minimisation with binary cross-entropy.

Loss is defined as follow:

$$L\left(y^i, \ddot{y}_j^i\right) = -\frac{1}{n}\sum_{i=0}^{n}\left(y^i * \log\left(\ddot{y}_j^i\right) + (1 - y^i) * \log\left(1 - \ddot{y}_j^i\right)\right)$$
$$L\left(y^i, \ddot{y}_j^i\right) = -\frac{1}{n}\sum_{i=0}^{n}\left(y^i * \log\left(\ddot{y}_j^i\right) + (1 - y^i) * \log\left(1 - \ddot{y}_j^i\right)\right)$$
$$L\left(y^i, \ddot{y}_j^i\right) = -\frac{1}{n}\sum_{i=0}^{n}\left(y^i * \log\left(\ddot{y}_j^i\right) + (1 - y^i) * \log\left(1 - \ddot{y}_j^i\right)\right)$$

- *Algorithm 4: CNN+LSTM+RF*

Once learning the feature model CNN_LSTM is finished, the features extracted by the model form a random forest feature to train.

Step1: let $K = \{k_1, k_2, \dots, k_n\}K = \{k_1, k_2, \dots, k_n\}$ $M_K = \{m_1, m_2, \dots, m_i, \}M_K = \{m_1, m_2, \dots, m_i, \}$ the composition where $n \geq 1n \geq 1$ and

where ii the features extracted by CNN_LSTM. Step 2: On i selected features; the splitting metric of node is determined by Gini coefficient as follow:

$$\mathrm{Gini(M)} = 1 - \sum_{i=1}^{n} p_i^2 \ \mathrm{Gini(M)} = 1 - \sum_{i=1}^{n} p_i^2 \mathrm{Gini(M)} = 1 - \sum_{i=1}^{n} p_i^2$$

3. Results and experiments

3.1. Experiments

To evaluate model's efficiency and accuracy, accuracy rate, loss rate, false negative rate and true negative rate are widely used as metrics.

- *True Negative (TN)* is the correct identification of the benign URL as benign;
- *False Negative (FN)* is incorrectly identifying the malicious URL as benign;
- *False Positif (FP)* is the incorrect identification of the benign URL as malicious.

$$FNR = \frac{FN}{FN + TP} \quad (7)$$

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

$$accuracy = \frac{TP + TN}{2TP + TN + FP + FN} \quad (9)$$

Experiment 1 and results:

Our first experiment is comparison of popular and powerful machine learning models in malicious URL

detection case such as« Adaboost », « Decision Tree », « Gradient Boosting », « Logistic Regression »,

« Random Forest ». There are three types of data to evaluate the detection performance according to the number of URLs: i) dataURL.csv: 20000 labelled URLs, ii) dataURL2.csv: 40000 labelled URLs, iii) dataURL3.csv: 20000 labelled URLs and 204840 Unlabelled URLs. These data are extracted from Kaggle.

Table 1. Comparison of machine learning models

Number of URLs	20 000	40 000	224 480
	Algorithms and Accuracy		
Adaboost (ADB)	0,70	0,77	0,90
Decision Tree (DT)	0,81	0,83	0,97
Gradient Boosting (GB)	0,80	0,82	0,97
Logistic Regression (LR)	0,71	0,75	0,95
Random Forest (RF)	0,82	0,84	0,98
FNR (Random Forest)	14,13	12,02	0,00
TPR (Random Forest)	16,16	16,39	0,00

Table 1 shows the results concerning the comparison of some models of machine learning. From 20 000 URLs to 224 840, the performance of detection increases when the number of URLs increases and false positive rate, false negative rate decreases.

Of this result, the number of data used has an impact on the detection performance; this means that it takes a lot of URLs to have a better prediction.

In all types of data, the Random Forest model has the best performance and the rate ranges from 82% to 98% with 0% false positive rate and 0% false negative rate. The model, in addition to having the best performance, detects better the new URL test. But extracting the URL manually takes a long time and the training data is about 6h.

Experiment 2 and results:

The second experiment is a comparison of five different models of deep learning such as CNN, CNNLSTM, FullCNN, LSTM and MLP. We chose these models because they prove to be popular and powerful in malicious URL detection case. There are three types of data to evaluate the detection performance according to the number of URLs: i) dataURL.csv: 20000 labelled URLs, ii) dataURL2.csv: 40000 labelled URLs, iii) dataURL3.csv: 20000 labelled URLs and 204840 Unlabelled URLs. The training is performed with3 epochs.
(Loss (L), Accuracy (A), Validation-Accuracy (V)).

Table 2. Comparison of deep learning models

Number of URLs	20 000	40 000	224 480
Algorithms and Accuracy			
CNN	L = 0,38; A = 0,83; V =	L = 0,35; A = 0,84; V =	L = 0,03; A = 0,98; V =
CNNLSTM	L = 0,35; A = 0,84; V =	L = 0,28; A = 0,88; V =	L = 0,02; A = 0.99; V =
FullCNN	L = 0,49; A = 0,77; V =	L = 0,48; A = 0,78; V =	L = 0.05; A = 0.97; V =
LSTM	L = 0,40; A = 0,82; V =	L = 0,38; A = 0,82; V =	L = 0,04; A = 0,97; V =
MLP	L = 0,73; A = 0,53; V =	L = 0,53; A = 0,73; V =	L = 0,07; A = 0,97; V =

As shown in Table 2, CNN model combined with LSTM achieved the best detection rate 84% to 99% for each type of data, and this performance is due to large number of URLs used. Here, extraction is done automatically with CNN model and which has not taken much time.

But even if CNN_LSTM (99%) got the best accuracy, sometimes, it's has a trouble to detect URL, unlike Random Forest model (96%) above.

Hence our interest in combining CNN, LSTM and RF model in experiment 3, because CNN_LSTM model facilitates the extraction of features that is done automatically and doesn’t take much time. And the RF model classifies and predicts better the newly URL formed. Figure 4 represents the summary of our model CNN_LSTM.

```
Layer (type)                 Output Shape              Param #
=================================================================
main_input (InputLayer)      (None, 75)                0
_________________________________________________________________
embedding_1 (Embedding)      (None, 75, 32)            3200
_________________________________________________________________
dropout_1 (Dropout)          (None, 75, 32)            0
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 75, 256)           41216
_________________________________________________________________
elu_1 (ELU)                  (None, 75, 256)           0
_________________________________________________________________
max_pooling1d_1 (MaxPooling1 (None, 18, 256)           0
_________________________________________________________________
dropout_2 (Dropout)          (None, 18, 256)           0
_________________________________________________________________
lstm_1 (LSTM)                (None, 32)                36992
_________________________________________________________________
dropout_3 (Dropout)          (None, 32)                0
_________________________________________________________________
output (Dense)               (None, 1)                 33
=================================================================
Total params: 81,441
Trainable params: 81,441
Non-trainable params: 0
```

Fig. 4. Summary of the architecture of our CNN_LSTM

Experiment3 and results:

The third experiment compares the models of machine learning, deep learning, and machine learning combined with deep learning that got the best accuracy in experiment 1 and 2 (CNN_LSTM and Random Forest). Our choice is from the fact that, their accuracies are the best on experimentation 1 and

2 (CNN_LSTM and Random Forest). Here, CNN_LSTM model is used to extract URL features and the Random Forest model for classification and prediction. There are three types of data to evaluate the performance of detection according to the number of URL: i) dataURL.csv: 20000 labelled URLs, dataURL2.csv: 40000 labelled URLs; dataURL3.csv: 20000 labelled URLs and 204840 Unlabelled URLs. The data is train with 3 epochs.

Table 3. Comparison of our CNNLSTM + RF model with the others models of machine learning and deep learning

Number of URLs	20 000	40 000	224 480
	Algorithms and Accuracy		
CNN	L = 0,38; A = 0,83; V =	L = 0,35; A = 0,84; V	L = 0,03; A = 0,98; V
CNNLSTM	L = 0,35; A = 0,84; V =	L = 0,28; A = 0,88; V	L = 0,02; A = 0,99; V
LSTM	L = 0,40; A = 0,82; V =	L = 0,38; A = 0,82; V	L = 0,04; A = 0,97; V
RF	A = 0,81; FNR = 14,13;TPR = 16,16	A = 0,83; FNR = 12,02; TPR = 16,39	A = 0,98; FNR = 0,00; TPR = 0,00
CNNLSTM+RF	A = 0,51	A = 0,54	A = 0,96
CNN+RF	A = 0,53	A = 0,52	A = 0,958
LSTM+RF	A = 0,56	A = 0,53	A = 0,95

A comparison of one model of machine learning (RF), two models of deep learning (LSTM, CNNLSTM) and three models of machine learning and deep learning combined such as CNN+RF, LSTM+RF, CNNLSTM+RF is showed in table 3.CNN_LSTM model with dataURL3.csv converges after 3 epochs and the malicious URL detection accuracy is 99% with a low loss rate of 0,2%.

Of course, CNN_LSTM model has a better rate of precision and the extraction of the features is done automatically but sometimes, it has a difficult to detect the new URL test. And it may be the same problem in [5] that LSTM algorithm can't determine the phishing URL whose functionalities is similar of the normal URL.

It's the reason of the combination of CNN_LSTM with RF. And here, our CNN_LSTM + RF model detects better the new URL test even if its detection rate is 96% lower than that of CNN_LSTM (99%). Because Random Forest (RF) combines other characteristics to compensate LSTM performance deficiencies [5].

Experiment 4 and results:

Same experiment as in experiment 3 is used with manual extraction (RF) and classifier (CNN_LSTM) for RF + CNN, RF + LSTM, RF + CNNLSTM models.

Table 4. Comparison of our RF+ CNNLSTM model with others models of machine learning and deep learning

Number of URLs	20 000	40 000	224 480
	Algorithms and Accuracy		
CNN	L = 0,38; A = 0,83; V =	L = 0,35; A = 0,84; V	L = 0,03; A = 0,98; V
CNNLSTM	L = 0,35; A = 0,84; V =	L = 0,28; A = 0,88; V	L = 0,02; A = 0,99; V
LSTM	L = 0,40; A = 0,82; V =	L = 0,38; A = 0,82; V	L = 0,04; A = 0,97; V
RF	A = 0,81;	A = 0,83;	A = 0,98;
RF+CNN	L = 11,5; A = 0,25; V =	L = 11,5; A = 0,25; V	L = 0,69; A = 0,25; V
RF+ LSTM	L = 0,69; A = 0,50; V =	L = 0,69; A = 0,50; V	L = 0,69; A = 0,50; V
RF+CNNLSTM	L = 0,69; A = 0,75; V =	L = 0,69; A = 1; V =	L = 0,69; A = 0,50; V

During the experimentation of our model RF + CNNLSTM, even if the epoch is set to 100 or the number of URLs is increased, the result is almost unchanged and the loss is above the accuracy.

The reason may be the model tends to overfit. This over-fitting comes from the fact that our RF model has as output a probability according to Bernoulli's law and CNN_LSTM takes as input this probability to generate another one that follows the same distribution.

In this architecture, CNN_LSTM play no role because RF directly classifies the data. Therefore, the combined model is biased because CNN_LSTM contains a lot of parameters for a single input variable that promotes over-fitting. Over-fitting makes the model imprecise during classification.

Here the summary of these 4 experiments according to dataURL3.csv: three types of URL data were used to show that the number of URLs increases the detection performance.

Table 5. Summary of the 4 experiments according to dataURL3.csv

Experiment 1	Experiment 2	Experiment 3	Experiment 4
Algorithms and Accuracy			
Adaboost = 0,90	CNN =	CNN = 0,98	CNN = 0,98
Decision Tree = 0,97		CNNLSTM = 0,98	CNNLSTM = 0,99
Gradient Boosting = 0,97	FullCNN	LSTM = 0,97	LSTM = 0,97
Logistic Regression = 0,95	LSTM =	RF = 0,98	RF = 0,98
RF= 0,98	MLP =	CNNLSTM+RF = 0,96	RF+CNNLSTM = 0,50
		CNN+RF = 0,956	RF+CNN = 0,25
		LSTM+RF = 0,95	LSTM+RF = 0,50

Given the results in table 5, our proposed model CNN_LSTM+RF (96%) performs less than CNN_LSTM (99%), but it detects URL better, due to LSTM's propriety for detecting newly formed URL and also not requiring a lot of time by automatic extraction.

The second model RF+CNNLSTM tends to over-fit; changing the learning process of Neural Network. In addition, using deep learning gives difficulty to explain the generated result, so, the hackers could also have trouble in understanding how the result was generated.

Our next work is to use CNN_LSTM model in an online environment by using the Hedge Backpropagation of Online Deep Learning with CNN_LSTM architecture,

in order to detect millions of URLs online that evolve over time and perform accuracy and loss.

References

[1] Y. Liu *et al.*, « Cloudy with a chance of breach: Forecasting cyber security incidents », in *24th ${$USENIX$}$ Security Symposium (${$USENIX$}$ Security 15)*, 2015, p. 1009–1024.

[2] K. Soska et N. Christin, « Automatically detecting vulnerable websites before they turn malicious », in *23rd ${$USENIX$}$ Security Symposium (${$USENIX$}$ Security 14)*, 2014, p. 625–640.

[3] D. Sahoo, C. Liu, et S. C. Hoi, « Malicious URL detection using machine learning: a survey », *ArXiv Prepr. ArXiv170107179*, 2017.

[4] J. Zhao, N. Wang, Q. Ma, et Z. Cheng, « Classifying Malicious URLs Using Gated Recurrent Neural Networks », in *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2018, p. 385–394.

[5] Y. Fang, X. Long, C. Huang, et L. Liu, « Research on Classifying Phishing URLs Using Hybrid Architecture of LSTM and Random Forest », *Gongcheng Kexue Yu JishuAdvanced Eng. Sci.*, vol. 50, p. 196‑201, sept. 2018.

[6] F. Douksieh Abdi et L. Wenjuan, « Malicious URL Detection Using Convolutional Neural Network », *Int. J. Comput. Sci. Eng. Inf. Technol.*, vol. 7, no 6, p. 01‑08, déc. 2017.

[7] G. Cao, S. Wang, B. Wei, Y. Yin, et G. Yang, « A hybrid CNN-RF method for electron microscopy images segmentation », *Tissue Eng. J Biomim Biomater Tissue Eng*, vol. 18, p. 2, 2013.

[8] D. Gordeev, « Detecting state of aggression in sentences using CNN », in *International Conference on Speech and Computer*, 2016, p. 240–245.

[9] L. Zheng, Q. Li, H. Ban, et S. Liu, « Speech emotion recognition based on convolution neural network combined with random forest », in *2018 Chinese Control And Decision Conference (CCDC)*, 2018, p. 4143–4147.

IoT Plant Monitoring System

Sonalika Awasthi, Raghav Sharma, Yepuganti Karuna*
School of Electronics Engineering (SENSE), Vellore Institute of Technology (VIT), Vellore

Abstract

In this paper, we have discussed about enhancing the experience of personal gardening as a method of therapy for mental-health patients given its suspected role on a person's mood and general positivity, using an Internet of Things (IoT) approach. The developed prototype continuously senses and monitors the state of an indoor plant through different sensors which the user is notified about through channels over cloud in-real time, of the plant's needs for water, sunlight etc through generated notifications. Thus, we were able to successfully create a humanised experience of recovery where the user can accessibly 'talk' with it's plant through a smartphone.

Keywords: Internet of Things, Raspberry Pi, Mental health, user interface design, Smart monitoring

1. Introduction

Mental healthcare is an increasingly evolving field that is undergoing discoveries of the increasing role of one's surroundings and atmosphere on one's mental state. To that end, there is an increasing attention towards the anticipated link between gardening and its effect on mental health[1]. Various people claim that the activity contributes to a patient's mood, attention span, self-esteem and can help inculcate a sense of responsibility.

This paper aims to build a prototype for a channel to make that gardening experience more interactive for the user by establishing a means of communication through the application of Internet of Things. Through continuously sensing the plant's state (soil moisture, temperature, etc) and generating an interactive mechanism for the user to interact with that data - via triggered notifications on the user's smartphone making requests to water the plant/move its location, etc - the paper aims to make the whole personal gardening experience fulfilling and wholesome for the user.

Through the added characterisation of the plant as someone who you can talk with and enquire about out of care adds to the healing experience. Periodically, the plant also offers the user some light jokes to up-turn a bad mood. Overall 'Mr. Plant' aims

* Corresponding author.
E-mail address: karun@vit.ac.in,

at providing a humanised, interactive experience to help a user pass through a low-phase towards mental healing.

To seek guidance for our paper we referred few research articles that were oriented towards two important aspects, the technical side and the therapeutic side of a smart plant. The technical papers discussed about ways to automate the processes required for taking care of the plant incorporating the usage of different and experimental setup and electrical components. One discussed about an embedded system using Arduino Nano, Node MCU, line following robot system[2] and sensors while the other used IoT[3], Raspberry Pi as Processor[4], sensors of CO2, humidity, luminosity, temperature, detection of plants and a hybrid application for its remote monitoring connected to a local area network[5]. We also came across an article illustrating the benefits of using Iot in the Raspberry Pi platform to provide comfort to the people by reducing the manual work and to improve the overall performance of any system without user interaction[6][7]. Ultimately we finished our literature survey with papers that emphasised on the therapeutic effects of plants on patients and people in general, providing evidence of substantial human health benefits[8].

2. Methodology

The prototype runs via a chain of events which include first sensing the data from sensors (as specified in the figure below), uploading them to the cloud platform, and connecting user-interacting applications over cloud through triggers and responses. The technical operation flow as shown in Fig 1 represents the components involved and the connections between them, while Fig 2 represents the chain of events that occur while the application runs:

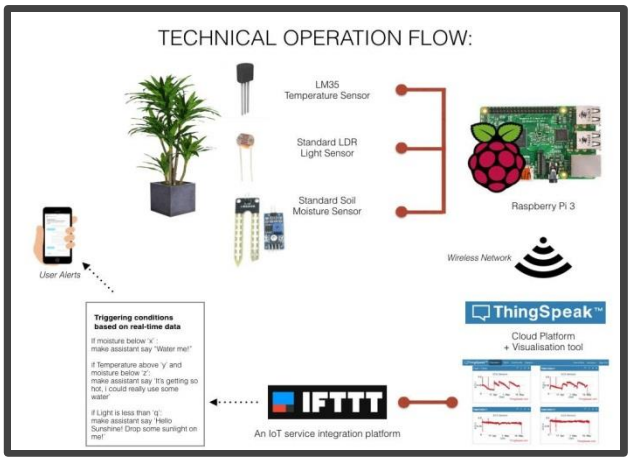


Fig. 1. Technical Operation flow

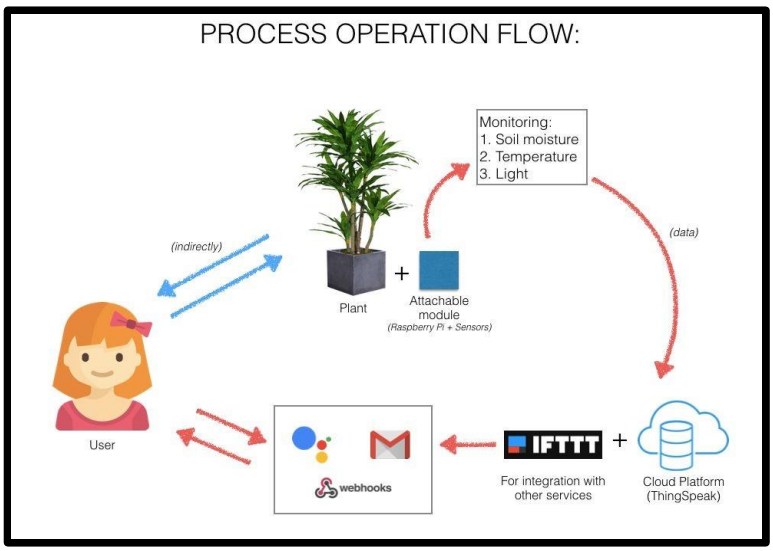


Fig. 2. Process Operation flow

The developed prototype overall can hence be divided into three stages of operations:

2.1 Setting up hardware

Throughout this stage, we accumulated all the listed sensors into a compact circuit, connected with Raspberry Pi 3, and the sensors appropriately onto the plant pot as shown in Fig 3. For the prototype, we made use of a breadboard however the same can be eliminated in further iterations of the refined model.

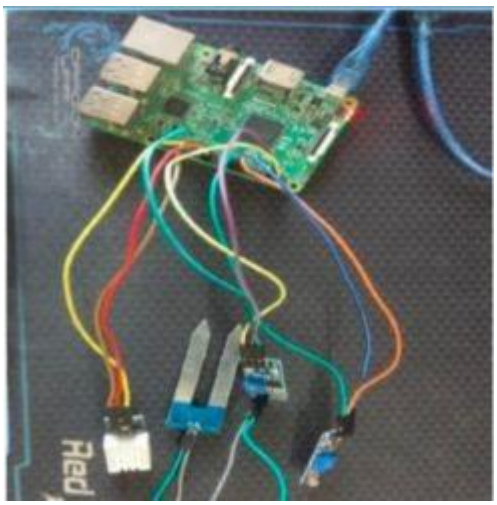


Fig. 3. Hardware configuration

2.2 Data transmission and visualisation on cloud

Connected to a power source and wifi, we ran a python script on the Raspberry Pi to collect and upload sensor readings to the cloud. A private channel on ThingsSpeak platform was generated, real-time data was uploaded and visualised. The figure below represents readings of the temperature sensor from test runs done across three days for varying temperatures, as shown in Fig. 4 below.

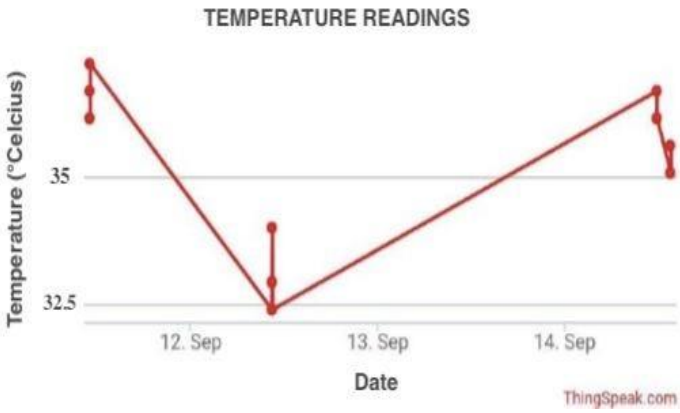


Figure 4: Temperature readings across three days (in °C)

2.3 User interaction interface

Thresholds for notification triggers were set on the data being uploaded to the cloud. Note that these threshold values for water, temperature and light requirements may differ from species to species of plants and they were simply assumed for the purpose of this experimentation. Based on generated triggers via Webhooks, the Integromat platform allowed us to push notifications on a connected android user device. Further, we integrated an inquiry system for the user so it may ask about the plant's status and get a response - via Google Assistant, with an IFTTT applet running in the background. The response message from the plant for specific instances were pre-set and an added feature of the plant offering the user a light joke on inquiry was also added. The following figure represents the chain of trigger and response functions that was setup on the Intergormat (Fig. 5) and IFTTT platform (Fig. 6) respectively:

Figure. 5: The developed 'scenario' on integromat platform

The method of the request e.g. GET, POST, DELETE

Figure 6. On the IFTTT platform: (a) Connection status of the activated applet (b) Web request details of the applet

3. Simulation Results

The prototype was repeatedly ran and tested across development and continuously refined. The final simulation of the running application from the user's phone screen (on an Android smartphone) was as illustrated below in Fig. 7 in sequence:

Figure 7. (a) User makes a voice request using Assistant (b)A push notification pops (c)The notification displays the message from 'Mr Plant', requiring to be watered. Additionally, a joke featured is offered on tap. (d) The user is redirected to a joke online

4. Future Aspects

4.1 Refined 'plug-and-play' hardware module

The hardware module can be refined further so as to make it compatible with more plants with an easy method of installation into the plant pot like a simple 'plug and play' instead of separate sensor modules.

4.2 More conversational abilities via Natural Language Processing

The plant could be given its own conversational abilities by employing natural language processing combined with artificial intelligence concepts. The conversations hence can go beyond a predefined set and be more humane.

4.3 More mental-health specific support

Added abilities may include feeding the plant with professional-recommended and approved therapy-like responses post adding some diagnostic abilities to the plant via picking up patterns in conversations, noticing changes in routine, etc.

4.4 As a help-tool to amateur gardeners:

The device may also help people beyond the scope of mental health in terms of making personal gardening a lot more easier and accessible to the wider public that could guide them and help keep their plants alive[9][10].

5. Conclusion

Mental health often revolves around therapeutic exercises that may contribute to one's personal development across a journey of healing. By using IoT to make the exercise of maintaining a plant easier and wholesome, this paper was successfully able to set-up and run an interactive experience for potential users. In line with the set aims of this paper, the prototype made successfully ran across all simulations and was able to monitor plant-data in real time over ThingSpeak cloud platform, while all the corresponding messages sent to the user via push-notifications were accessible through a smartphone, given internet connectivity.

There were no issues faced across repeated runs, given the applet on the IFTTT platform Integromat scenario were running. The time delay between the user generating the said request via an audio command and the push-notification appearing via Integromat remained under 1.5 seconds and can run across all android devices supporting Google Assistant. In the wider conception of this paper, we were successful in setting up the conceptualised user experience.

Acknowledgements

The authors gratefully acknowledge and thank the Head of Department (HoD), Professors and all technical staff of SENSE, VIT Vellore, India for their support in completing this paper.

References

[1] Masashi Soga, Kevin J. Gaston, Yuichi Yamaur, 2016. "Gardening is beneficial for health: a meta-analysis", Preventive Medicine Reports.

[2] Bandari Theja, P.Sathvik Prasad, A.Pavan Kalyan, N.Harshitha, P.Pooja, 2019. "Smart Plant Pot",International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 05

[3] A.M.Ezhilazhahi, P.T.V.Bhuvaneswar, 2017. "IoT Enabled Plant Soil Moisture Monitoring Using Wireless Sensor Networks", IEEE 3rd International Conference on Sensing, Signal Processing and Security (ICSSS).

[4] Ramkumar.E, Nagarani.S, Roger Rozario A. P, Arjuman Banu.S, 2018. "Tomen:A Plant monitoring and smart gardening system using IoT", International Journal of Pure and Applied Mathematics, Volume 119 No. 7, p703-710.

[5] Nishath Shireen, B.Vasundhara Devi, 2018. "Smart Garden Management System", International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 5, Issue 3.

[6] Anusha k , Dr. U B Mahadevaswamy, 2018. "Automatic IoT Based Plant Monitoring and Watering System using Raspberry Pi", I.J. Engineering and Manufacturing, MECS, p55-67

[7] Asst. Prof. Kawale Jayashri, Sanjay More, Akshay Bankar, Ganesh Dongre, Pooja Patil, 2018. "IoT Based Smart Plant Monitoring System" International Journal of Advance Research in Science and Engineering, Volume 7, Issue No.3

[8] Richard Thompson, 2018. "Gardening forn health: a regular dose of gardening ", Clinical Medicine Vol 18, No 3: 201–5

[9] K. Krishna Kishore, M. H. Sai Kumar, M. B. S. Murthy, 2017. "Automatic Plant Monitoring System", International Conference on Trends in Electronics and Informatics ICEI.

[10] Gabriela Carrion, Monica Huerta, Boris Barzallo, 2018 "Internet of things (IoT) applied to an urban garden", IEEE 6th International Conference on Future Internet of Things and Cloud.

A Comprehensive Study on Aspect Based Sentimental Analysis Framework and its Techniques

Sindhu Ca, Asha Gutlapallib*

aAssistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai 603203, India

bDepartment of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai 603203, India

Abstract

Over the years, there has been a rapid influx of big data and a steep increase in the speed as well as the volume of data generation comprising online customer reviews on E-commerce websites and social media. It has bestowed consumers with the opportunity of making a thoughtful decision about purchasing a product or subscribing to a service based on opinions surfacing on the Internet. Aspect based sentimental analysis makes the process of gathering, analyzing, summarizing, and categorizing all the elements and views available in these multimodal sources easier. While making a choice, people are deliberately looking for specific details they require and compare them against others rather than just the verdict. The classes of traits, the desirability, and the disadvantages of those characteristics and differentiation between other commodities are illustrated. In this paper, a comparative study is performed between methods and algorithms used for data pre-processing, feature extraction, aspect detection, and emotion categorization. The various approaches towards aspect based sentimental analysis are delineated. Their performance and functionalities are explored, and conclusions are established.

Keywords: Aspect Based Sentimental Analysis, Natural Language Processing, Machine Learning, Information extraction, Opinion Mining, Sentimental Analysis, Text Mining.

1. Introduction

In the modern age of advanced technology, it is not hard to search anything online that we are not aware of and educate ourselves. The Internet enables data sharing between places located across the globe. Vast amounts of data are produced every moment in this day and age because of the improvement in devices that allow it. Discussions on numerous topics are open on multiple platforms. People visiting such websites are free to share their opinion on them publicly. Others utilize this

* Corresponding author.
E-mail address: ashagutlapalli0316@gmail.com

information as tools in making a decision accordingly or in using this in other fashions. The context and the emotion expressed in the outlook must be recognized correctly to do it efficiently.

This paper was put together to examine the extensive procedures of aspect-based sentimental analysis. Primarily, to assist internet users by providing them with a relative breakdown of the entity's facets, without having to read every comment on it. It will reduce their time and effort vastly. They might even come across facts that they wouldn't have otherwise. It enables them to be productive and highly decisive. Exploring competent structures for carrying out these tasks has been the motive of this paper. It guides the construction of such structures that would require low human interaction, extensible, and receptive to instantaneous queries. It has numerous applications that would enrich human lives on a day to day basis.

Natural language processing and machine learning make it possible today to build robust models with the storage capacity, processing power available to us now. Natural language processing concepts deal with processing text and discovering patterns, relationships and, semantics between words. Machine learning is a field of computer science where a machine is trained with input and output data, and the machine determines the program or the algorithm needed to improve its performance to predict unseen data and eventually become intelligent systems.

Buying products after reading reviews about them is a popular way of choosing the right item before purchase. This helps the user to make an informed decision. Advertisements are customized for the consumer facilitated by the feedback they provide on their merchandise. Campaigning for political parties and leaders is becoming easier as the views of the citizens about them are recorded in social media, which can be used for their advantage. Directors and actors working on films can get instant comments on the trailers of the movies, which assist them in rendering to the taste of the audience. Songs and videos are recommended to an individual based on their preferences. Restaurants and tourist spots are suggested in the opinion of the client.

1.1 Sentimental Analysis

Sentimental analysis plays a significant role in determining the polarity of a review or a comment. It is used to know whether the person is talking about something in the right way or a wrong way. It can be classified broadly into positive, negative, and neutral. For example, on a tourism website, a person leaves a remark stating, “There are beautiful tourist spots in Switzerland. “The word ‘beautiful’ is positive as it describes something as pretty.

1.2 Aspect Based Sentimental Analysis

Aspect based sentimental analysis is a part of sentimental analysis where emotion mining is performed on detailed text or information. It increases the comprehension of the attitude (i.e.) the way people feel about an article or a subject. For example, in the review, the story of the film was boring. The word “story” is labelled as an aspect, and the word "boring" is labelled as sentiment in Fig. 1. This outlook is beneficial while considering watching the movie.

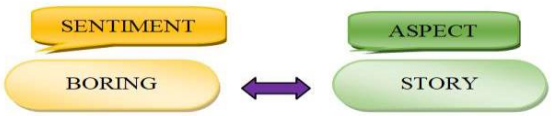


Fig. 1. Aspect – Sentiment Pair

2. Related Work

Paper	Concept	Algorithm	Mathematical Tools	Enhancement	Pros	Cons
[1]	Emotion mining on tourism reviews	Naïve Bayes and Decision Tree	Probability, differentiation, F-measure, Precision and, Recall	Increase scalability	High accuracy	low response time
[2]	For Indonesian reviews using RNN	GRU and CNN	Probability, differentiation, integration, F-measure, Precision and, Recall	Better aspect matrix	High performance	Weak aspect matrix and convolutions filters
[3]	To detect emotion on online reviews using RNN	LDA	Probability, performance visualization	Implicit expressions to be identified	High emotion analysis	Sarcasm reviews not tagged accurately

Paper	Concept	Algorithm	Mathematical Tools	Enhancement	Pros	Cons
[4]	Frequent itemset mining in product reviews	Naïve Bayes	Precision, Recall, F1-score	Aspects of relative importance to be determined	High accuracy	Interesting aspects are not identified
[5]	A-P Collocations for Chinese dataset	Clustering Algorithm	Algorithmic, syntactic and compression error	An enhanced compression model must be developed	Low error	Cannot be translated to sentimental sentences directly
[6]	Rule-Based Technique using lexicon resource	SentiWordNet Lexicon	Accuracy	Optimize algorithm to extend to more semantics	Low Bias	High Variance
[7]	Sentiment Lexicon to adapt shallow parsing	Ontology	Accuracy	To make it more scalable	Faster	Overfits the data
[8]	Ontology for cognizing customer's feelings	SentiWordNet Lexicon	Precision, Recall, F1-score	To improve efficiency	High knowledge of the domain	Takes time and effort
[9]	WEKA tool for student's opinions	Naïve Bayes and clustering	Accuracy, precision, recall and F1 score	Train on more negative data	Good for positive reviews	Bad for negative reviews
[10]	Twitter data sentimental analysis	Support vector machine (SVM)	Accuracy	Extend to more languages	Reduces storage and stages on real-time data	Only word-level analysis and sentence level
[11]	The ratio information of positive and negative sentiment	Double propagation, Extract Transform Load	Dependency relation, cosine similarity formula	Chosen features from the analysis result to improve sales	Uses initial opinion lexicon to extract aspects and extraction results to expand the opinion lexicon	Low performance
[12]	An aspect related to QoS parameters obtained from user reviews	Support vector machine (SVM) and Gradient boosting decision tree (GBDT)	Naïve Bayes, Logistic Regression, SVM and LDA	Can be effectively utilized for evaluation of quality features	Significant rise in performance in terms of both accuracy and F1- score	Quality of reviews and the credentials of users not considered
[13]	Help the users to analyze the film criticism	SVM and CRF	Normalization, probabilistic models	Decrease skewing to improve performance for the minority class	Based on the emotion rather than a reason	Wrong analysis of sentiments with repeated words
[14]	Automated extraction of sentiment words	Aspect Term Extraction using filtration	Pointwise Mutual Information	Extraction of implicit aspect terms	High precision and recall	PMI values do not perfectly filtrate aspect terms

Paper	Concept	Algorithm	Mathematical Tools	Enhancement	Pros	Cons
[15]	A novel co-attention based network to capture the correlation between aspect and contexts	Co-attention based network	Low-rank Bilinear Pooling using Hadamard product	Bi-directional attention mechanism to attend aspect information and context information simultaneously	The best model accuracy on two datasets	Low training efficiency
[16]	Retrieves review data from the social media networks and gives a rating to an institution	Data Gathering and Processing Module	ANN	Fine-tune as a commercial product	Recommendation system results sharing and saving functions	Limited number of main features.
[17]	To identify prominent aspects using contextual and domain-specific information	Contextual and domain-specific information	Frequency and Similarity Pruning	Discovering methods to correlate sentiment and behavior	Contextual opinion lexicon approach improved the precision value	Accuracy can be improved
[18]	A comprehensive overview of recent and past research on sentiment analysis and text summarization	Hopfield Network algorithm	Naive Bayes Classifier and Support Vector Machines	To discover the best combination of sentences for sentence scoring methods	High efficiency and performance	Improvement of scope of negation to classify
[19]	To produce a ranked list of the most representative product aspects	MCDM approach	Decision matrix normalization	Ranking the aspects outperformed	Can identify the most relevant product aspects efficiently	No comparative analysis with other MCDM techniques
[20]	To use a text mining approach to obtain hotel user opinion in the form of sentiment	RNTN algorithm	RNTN Classifier model for the Confusion matrix	Future research may limit the upload time of the reviews to gain more in-depth analysis	The model works best in categorizing aspects to positive sentiment	Low accuracy due to size of text fragments, training data and, classification features
[21]	To assist user generated content	WSAM-OWA Operator	Cardinality and CRF	To consider the credibility of the reviewers	It considers majority opinion and degree of importance of information	The rank of the reviewers is not taken into account
[22]	To summarize the sentiments from aspect based reviews	ASS System	ASS parameters	The performance of the system can be improved	Precise and compact than extractive summarization	Less number of functions in the system

3. Methodology

3.1 Data Pre-processing

3.1.1 Sentence tokenization

The input is given as text or a corpus. The output generates a list of sentences. For example, in the text, "I love dogs. I have a dog", the output is ["I love dogs," I have a dog"].

3.1.2 Word tokenization

Word tokenization is the same as sentence tokenization, but rather than applying it to sentences and it is used to words so that individual words are separated as items in a list. For example, in the sentence, "Chennai is humid," the result is ["Chennai," "is," "humid"].

3.1.3 Delimiter Removal [1],[2]

Delimiters are removed to reduce the size of the dataset as they do not supply any vital information. Some delimiters are question marks (?), full stop (.), and exclamation (!). The workflow of the methodology is illustrated in Fig. 2, and the different pre- processing modules are shown in Fig. 3.

3.1.4 Redundancy Removal[1]

Redundant sentences can occur in a dataset or a document and occupy unnecessary space. Two people might have mentioned the same remark about the same thing as they had similar observations.

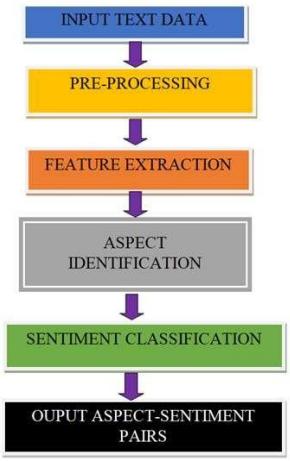


Fig. 2. Workflow of Aspect based Sentimental Analysis

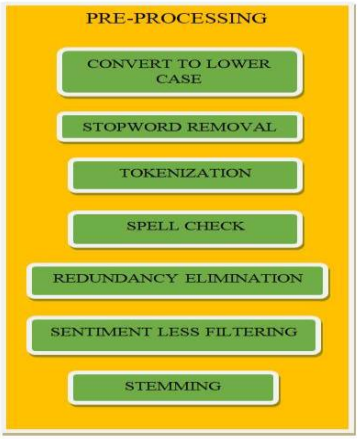


Fig. 3. Pre-processing Module

3.1.5 Spell Check[1]

While communicating online, people use shorthand, urban dictionary words, and slang in their language Standard dictionaries do not hold these words, so lookups into these dictionaries lead to inconsistency in data. Therefore it is essential to correct the spelling.

3.1.6 Sentiment less sentence filtering[1]

Sentences that do not contain any emotion can be discarded as they do not help the model in sentimental analysis. Unnecessary sentences are useless to the system and do not need to be given space in memory.

3.1.7 Stop word removal [1] , [4], [9], [12]

The dataset may contain words like ‘after,’ ‘every’ and ‘I.’ These words are not relevant to the sentiment detection process. Thereby, these words can be deleted to minimize the burden on the system.

3.1.8 Negation Resolution[1]

Negative emotion is not stated explicitly. It can be prefixed by words like “not" and "no." The word must be converted to the opposite polarity to the existing polarity in the attendance of such words. For example, in the phrase, “not nice,” it can be replaced by “unpleasant.”

3.1.9 Stemming [4], [9], [12], [18]

Stemming applies algorithmic rules to extract the stem out of the derived word simply. The words produced by this step do not have any essential meaning, but they are simply a bunch of letters put together without affixes. For example, the word “beautiful” is stemmed to “Beauti.”

3.1.10 Sentence Compression [5]

Sentences consisting of colloquial forms or words are compressed as it takes many words to state one emotion of a thing. Perception words are also eliminated, as these can be put in more straightforward sentences. For example, in the phrases, "Besides that,” “feel like,” the words “besides” and “like” can be deleted.

3.1.11 Removal of tags [8]

The data is scraped from the web pages residing on the website, and they contain HTML tags. These tags do not provide any necessary information and hence, can be erased. For example, tag like “<br>” - Break is deleted.

3.2 Feature Extraction

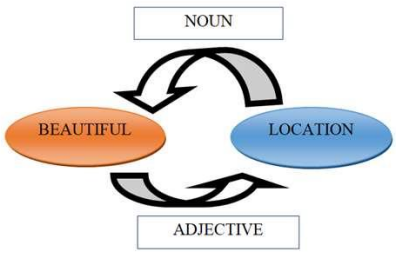


Fig. 4. Feature Extraction Module

3.2.1 POS tagger [1], [2], [4 -9], [11-13], [16-19]

POS tagger is parts of speech tagger that is an in-built function found in a standard library. It tags the words in the sentences according to the language of the grammar. For example, in the text, “The pizza was disgusting but the location was beautiful”, the result after implementing POS tagger will be [“The [DT]”, “pizza [NN]”, “is [VB]”, “disgusting [VBG]”, “but [CC]”, “the [DT]”, “location [NN]”, “was [VBD], “beautiful [JJ]] in Fig. 4.

3.2.2 N-Gram [1], [10], [19]

N-gram is a language model widely used in NLP and is applied to statistical problems involving text and audio. It is a probabilistic model that predicts the next series of words. For example, in the sentence, "The movie was boring." Unigram processes the text as ["The," "movie," "was," "boring"]. Bi-gram processes the text as ["The movie," "movie was," "was boring"]. Tri-gram processes the text as ["The movie was," "was boring"].

3.2.3 Word Embedding [2], [5], [15]

Word embedding is the process of converting text into numerical data. Since computers work on only numbers and cannot accurately compute data in the form of strings, words are embedded into a unique integer or numerical code. For example, the word "Severe" is fixed as a number "3214".

3.2.4 Bag of words [2], [3], [10], [21]

The bag of words carries out sentence tokenization and word tokenization. After that, it counts the number of appearances of each word. For example, in a sentence, "It is nice but horrid, and that's not a nice thing." The word "nice" is extracted and countered with two occurrences.

3.2.5 Presence [1]

Presence is described as a parameter for measuring the significance of a word in a document or a dataset. It is a feature method that works with a statistical measure. For example, the word "Gadget" has a higher incidence than the word "Mobile," so the former word is assigned a greater weight.

3.2.6 TF [1],[22]

TF – Term frequency is described as the number of times that a term occurs in a document. It considers all the terms of equal importance. For example, the word "Fruit" appears five times in a document of 100 words, then the TF for "Fruit" is 5/100 = 0.05.

3.2.7 TF-IDF [1], [11]

TF-IDF – Term frequency-inverse document frequency is described as the importance of a word in a document, which is proportional to the number of times

the word appears in the document. For example, the word "Fruit" appears in 100 of 10000 documents and the term frequency is 5 then the TF-IDF is 0.05 * log(10000/100) = 5 * 2 = 10.

3.2.8 Dependency parser[1], [5], [7], [8], [11], [16-17]

Stanford dependency parser establishes the relationship between entities in the language using grammatical rules. The output of the parser is a tree structure that is annotated. For example, in the sentence "The funny boy joked," "funny" is an adjective for the noun "boy."

3.2.9 MapReduce Hadoop [10]

The MapReduce function in the Hadoop platform takes text as input and output words with the words with their positive and negative scores. It has two modules map and reducer. It has three steps, namely mapping, shuffling, and reducing. For example, in a sample, the word "abandoned" has a positive score of 29 and a negative score of 228.

3.3 Aspect Establishment

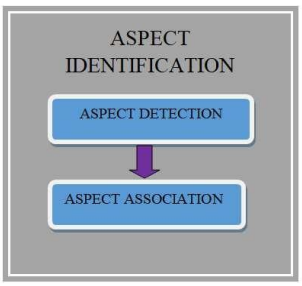


Fig. 5. Aspect Identification Module

3.3.1 Explicit Aspect Identification

3.3.1.1 Decision Tree [1], [9]

Decision trees are famous in Data mining, Machine learning, and Statistics. It depicts the decisions and the decision-making process of themes. Metrics evaluate the amount of information gained from the existence of a branching factor. For example, in the text, "The laptop has a very high speed; however, it has a less storage space." Here the words "Speed" and "Storage" are tagged as nouns, and hence the

decision tree opts for these words. The aspect identification modules are depicted in Fig. 5.

3.3.1.2 Ontology [8], [17]

Ontology presents the concepts involved in the domain and shows how they are related. The keywords in the domain help the algorithm to be aware of the fundamental entities and the correlated entities. For example, in the domain of restaurants, there are entities and co-entities as follows [["Food": "Burger," "Pizza"], ['Service": "Waiter," "Hospitality"], ["Ambience": "Surrounding," "Mood"]].

3.3.2 Implicit Aspect Identification

3.3.2.1 WordNet – Synsets [1], [18-19]

WordNet is a database that can be thought of as an arrangement of a dictionary and a thesaurus. The WordNet database has a component called Synsets. Synsets assemble words of similar meanings or synonyms into one set. For example, in the corpus, let's say it is declared as "The food is cold. By the time, edibles were served, it had lost heat". Here, with aspect association, the words "edibles," which is a subclass, is mapped to the superclass "food."

3.3.2.2 Clustering Algorithm [5], [9]

The brown clustering algorithm is an algorithm that creates clusters of words based on the cluster head. If the distance between the word and the clustering head is lowest, then the word is a member of the cluster. There are multiple clustering programs like Brown, clustering, which is a hierarchical clustering, and K-means clustering and PCA. For example, when the cluster head is "Ambience," it has a member "Environment" in the cluster. In Table1, the aspect terms are associated seed words are grouped.

3.3.3. Hybrid Approach

3.3.3.1 Frequent Item-set Mining [2], [4]

Frequent itemset mining is an unsupervised learning algorithm used frequently in data mining. It is utilized for mining the continually existing words that are pointed to features. An association can be built to tags as well to enhance the

exactness of the system. For example, the words "Cost" and "Price" are present collectively, so these are thought as one aspect.

3.3.3.2 Topic Modelling [3], [12]

Topic modeling is employed for taking out the main features so that aspect-based sentimental analysis can be exploited thoroughly. It can be built upon document distributions or the word distributions. The several topic models are LDA, MG-LDA, STM, and Local LDA. There are types in LDA, too, such as PLDA, LCTA, LGTA, LCARS, GEOPFM, JST, and FLDA. For example, in a biology text, some of the topics and the words in it are [["Gene": "DNA," "Genetics," "RNA"], ["Brain": "Neurons," "cerebral," "Nerves"]].

Table 1. Aspects of various domains

DOMAIN	ASPECT	SEED WORDS
RESTAURANT REVIEWS	FOOD	BURGER, PIZZA, FRIES
	AMBIENCE	SURROUNDING, MOOD, ENVIRONMENT
MOVIE REVIEWS	STORY	SCRIPT, TALE, NARRATIVE
	ACTING	PORTRAYAL, DRAMA, PERFORMANCE
SOCIAL MEDIA COMMENTS	PHYSIQUE	BUILT, FORM, FRAME
	CHARACTER	PERSONALITY, NATURE, PERSONA
PRODUCT REVIEWS	SPEED	RATE, PACE, MOMENTUM
	STORAGE	SPACE, AREA, DIMENSIONS
TOURISM REVIEWS	LOCATION	PLACE, SPOT, POSITION
	COST	PRICE, AMOUNT, FEE

3.4 Opinion Categorization

Table 2. Sentiments in various domains

DOMAIN	SENTIMENTS	SEED WORDS
RESTAURANT	POSITIVE	TASTY, DELICIOUS, ATTRACTIVE
	NEGATIVE	COSTLY, DIRTY, BLAND
	NEUTRAL	FINE, SATISFACTORY, ADEQUATE
SOCIAL MEDIA	POSITIVE	GOOD, PRETTY, CHARMING
	NEGATIVE	BAD, UGLY, DARK
	NEUTRAL	HUMBLE, INSTRUCTIVE, DETACHED
MOVIES	POSITIVE	ENTERTAINING, EXCITING, FUN
	NEGATIVE	DULL, BORING, MONOTONOUS
	NEUTRAL	AVERAGE, NORMAL, PAR

3.4.1 Supervised learning

3.4.1.1 Naïve Bayes Classifier [1], [4], [9], [12]

Naïve Bayes is a machine learning algorithm that is used for classification problems. It is a supervised learning approach to classifying words in a corpus. It is a probabilistic model that sorts the word instance into the classes available. It

determines the probability that a word can be assigned to each class and selects the class with the highest probability to the instance. For example, say there is the word "Irritate"; the classifier estimates the chances of it being a positive or a negative word. The probability of that word being a negative word is more likely than it being a positive word. In this way, the Naïve Bayes classifier can be implemented for sentiment categorization. In Table2, Sentiment terms and their seed words are displayed.

3.4.1.2 Support Vector Machine [1], [9], [10], [12-13], [18]

Support Vector Machine is a non-probabilistic model. It portrays the features of a multidimensional space. Hyperplanes are the boundaries on the space that decide the class of the feature. These Hyperplanes divides the classes with a large gap that distinguishes the categories in the output. For example, let's take the word "cunning." With the help of the labeled training data, each word or feature is plotted in the dimensional space for its class, and this facilitates the algorithm to estimate the maximum- margin hyperplane, which is the distance between the two classes. Therefore, the word "cunning" is appropriated towards the negative side.

3.4.1.3 Maximum Entropy [1]

Maximum entropy is a classifier that is an extension of the logistic regression for multiclass classification. Maximum entropy of the features ensures that only the features with maximum entropy or information are used in the model. Maximum entropy makes the least assumption at the beginning of the training period, but as it progresses, the algorithm stabilizes its uncertainty towards the certainty of a possibility. For example, in the word "brilliant," it is found that the probability of the word belonging to a negative class is higher as that word is recurrently mapped to a negative assemblage.

3.4.1.4 Random Forest [1]

Random forest is a collection of decision trees. The primary motive of random forests is to decrease the overfitting of data that is one of the major drawbacks of a decision tree. Random forest randomly nominates a division of features while splitting is done at the decision node. This is the reason behind accomplishing such low variance. For example, the decision node has to make a split between positive and negative labels in sentiment categorization in sentimental analysis. The word "Bright" has to be labeled, and hence it is split as a positive label as it is a positive word.

3.4.1.5 Fuzzy lattice reasoning [1]

Fuzzy lattice reasoning has a function with attributes lattice and valuation function. An uppercase valuation function is a ubiquitous group of data objects of every type. The lattice has many components where each component belongs to each class in the valuation function. The fuzzy lattice Reasoning algorithm defines an inclusion measure. The inclusion measure is described as the similarity measure which decides the membership of an object in a class. For example, the lattice rule derived from training is "Dull->Negative," which is a pair. The word in the test data such as "Blunt" is classified as negative using the existing rule.

3.4.2 Lexicon and Rule-based

3.4.2.1 Statistical Rule-based Method [5], [20]

The rule-based method is the method in which specific laws or regulations are composed to constraint the activity of the model customized by the developer. Parameters are instantiated that mines A-P collocations in the dataset and is known to be valid only if they are in par with the threshold. For example, in the axiom, "Lenovo has a wide camera," the A-P collocation is "Camera- Wide."

3.4.2.2 SentiWordNet [6], [8], [16]

SentiWordNet is derived from the WordNet database. It is used to classify a word into a polarity that can be positive, negative, and neutral. It has numerical scores for each polarity of a word where the object score is equal 1- (Positive score + Negative Score). For example, the word "Good" has a positive score of 0.875 and a negative score of 0.

3.4.2.3 Sentiment Lexicon [7], [10], [14]

The Sentiment Lexicon is a database of opinion or emotional words of different polarities. The model does look to this lexicon to classify the aspect of its sentiment. For example, the sentiment lexicon consists of words "Pleasant," "Pretty," "High," and "Fast" as a positive set and words "Dirty," "Ugly," "low" and "Slow" as a negative set.

3.4.3 Hybrid Approach

3.4.3.1 Recurrent Neural Networks [2], [15]

Gated Recurrent Unit is a type of recurrent neural network that has two gates, which are update and reset gate. The memory gate is additional in LSTM. The update gate selects the words important features, and the reset gate connects the aspects to its superclasses. For example, the word "story" is projected near the class "Script" of the embedding projector by maintaining a cell state.

3.4.3.2 Convolution Neural Networks [18]

Convolution neural network includes an aspect matrix. It executes word embedding to transform the text into numbers and then lays padding upon it so that the input vector has an even distribution across the layers. It has two major steps convolution and max pooling. The word "fluent" is extrapolated towards the positive space in the dimensional area.

4. Evaluation

4.1 Evaluation metrics

4.1.1 Accuracy [1-4], [6], [9], [10], [16], [20]

The accuracy metric is an important parameter to determine the loss between the actual and the predicted output. For example, accuracy is depicted as 90% for an artificial neural net.

4.1.2 Precision [1-4], [8], [9], [14], [17]

Precision is defined as the ratio of the number of relevant objects to the number of available objects. For example, the precision is 3% in which 3 are true negatives, 97 are false negative, and there are 0 true positives and 0 false positives.

4.1.3 Recall [1-4], [8], [9], [14], [17]

Recall is defined as the ratio of the number of corrected predicted instances to the number of instances predicted. For example, the recall is 100% in which 3 are true negatives, 97 are false negative, and there are 0 true positives and 0 false positives.

4.1.4 F1 – measure [1-4], [8], [9], [14], [17], [20]

The F1-score is defined as a weighted average of precision and recall where the best value is '1', and the worst value is '0'. For example, the F1-score is 0.714 for a Precision of 0.625 and a Recall of 0.833.

4.1.5 Perplexity [3]

Perplexity is the measure of uncertainty in predicting a feature's catalogue. Lower complexity means that it is good at predicting. For example, the perplexity of Unigram in a sample is 962, while the perplexity of Bigram is 170.

4.1.6 Algorithmic error [5]

Algorithmic error is defined as the error caused by the procedure that is employed where some are qualified than the rest. For example, Naive Bayes is more qualified than SVM in specific domains, like restaurant reviews.

4.1.7 Syntactic parsing error [5]

Syntactic error is described as the error when a word and the relationship between the words are mislabeled. For example, the proper noun "Sruthi" is misjudged as an adjective.

4.1.8 Compression error [5]

Compression error termed as the error when a phrase is compressed by leaving out the essential words. For example, the phrase "but the keyword is huge" is shortened as "but the keyword," leaving out the emotion "huge."

4.2 Comparative Analysis

Among the supervised classifiers used for sentiment classification, the Naïve Bayes reaches the highest accuracy, precision, and recall than SVM Classifier, Random Forest, Maximum entropy, and Fuzzy lattice reasoning. All the classifiers give better results when it is trained with more training data. Unigram attains a higher accuracy compared to the POS tagger, bi-gram, and tri-gram for feature extraction types in order. The feature extraction methods are recommended as Presence, TF, and TF-IDF for feature extraction methods in sequence. The issue with supervised classification is that it overfits the data and encounters high variance. Moreover, it is not extensible to multiple domains and applications [1]. In the models used for

deep neural networks, GRU conquers enhanced accurateness than convolution neural networks and Fachrina. GRU is faster and has a high response time as it does not a memory gate [2]. Local LDA model results prove to be superior as it capable of recognizing sentence-level aspects. CRATS achieves the lowest perplexity in comparison to other types of LDA models because it relies upon both provinces and opinions with topics in communities, unlike the rest [3]. With the inclusion of frequent itemset mining, the accuracy of the model jumps to 80%, and the sentiment is sorted with 92% accuracy [4]. Sentence compression reduced the algorithmic, syntactic, and the parsing error than the ones that are not compressed [5]. The SentiWordNet methodology has an accuracy of 75% - 100%, where it either has a high bias, high variance, or both [6]. The sentiment lexicon practice is much faster and also notices difficult words, too [7]. Developing ontology for each domain takes a lot of time and effort, as it is not scalable [8]. The results are evaluated with the aid of a tool called 'WEKA' for analysis and visualization [9]. The Twitter model has its application with real- time data and reduces storage [10]. The performance of all algorithms is projected as a metric of accuracy in Fig. 6.

Fig. 6. Performance – Accuracy of algorithms.

5. Conclusion

Supervised and unsupervised machine learning practice aspect based sentimental analysis. Though supervised learning methods accomplish high accuracy, semi-supervised learning gives the best result and output when encounters new environments and challenges. There are few concerning liabilities embedded in this perspective that cause damage in terms of sustainability. One of them is the lack of adaptability of the model to new situations. The model is constrained and in a way, customized to a particular application but cannot get accustomed to novice scenarios. The response time of the system has not come up to excellent levels. Semi-supervised learning makes a transition to progress into entirely unsupervised learning. Semi-supervised learning has shown that it can make resolve these difficulties and ameliorate the portability of the system. The future of aspect-based sentimental analysis aspires to look up to more unsupervised learning methods without assistance, which takes up a lot of time and effort. The most popular algorithms are expected to evolve to become self-sufficient, independent, and self-sustainable. New methods

are anticipated to draw inspiration and derivation from traditional programs. Aspect based sentimental analysis decreases manual discourse and makes it user-friendly to comprehend the synopsis of documents, articles, and reviews with attractive analysis and representation.

References

[1] Muhammad Afzaal, Muhammad Usman, and Alvis Fong, "Tourism Mobile App With Aspect-Based Sentiment Classification Framework for Tourist Reviews," IEEE Transactions On Consumer Electronics, Vol. 65, No. 2, May 2019, published by IEEE in 2019.

[2] AfrindaIlmania, Abdurrahman, Samuel Cahyawijaya, AyuPurwarianti, "Aspect Detection and Sentiment Classification using Deep Neural-Network for Indonesian Aspect-Based Sentiment Analysis," 2018 International Conference on Asian Language Processing (IALP), IEEE, published by IEEE in 2018.

[3] Jia-Dong Zhang and Chi-Yin Chow, "CRATS: An LDA-Based Model for Jointly Mining Latent Communities, Regions, Activities, Topics, and Sentiments from Geosocial Network Data," IEEE Transactions on Knowledge and Data Engineering, Vol. 28, No. 11, November 2016, published by IEEE in 2016.

[4] A.Jeyapriya, C.S.KanimozhiSelvi, "Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm," IEEE Sponsored 2nd International Conference On Electronics And Communication Systems (ICECS '2015), published by IEEE in 2015.

[5] Wanxiang Che, Yanyan Zhao, Honglei Guo, Zhong Su, and Ting Liu, "Sentence Compression for Aspect-Based Sentiment Analysis," IEEE/ACM Transactions On Audio, Speech, And Language Processing, Vol. 23, No. 12, December 2015, published by IEEE in 2015.

[6] Alexandra Cernian, Valentin Sgarciu, Bogdan Martin, "Sentiment analysis from product reviews using SentiWordNet as a lexical resource," ECAI 2015 - International Conference – 7th Edition Electronics, Computers, and Artificial Intelligence, 25 June -27 June 2015, Bucharest, ROMÂNIA, published by IEEE in 2015.

[7] Mr. Jayraj M. Desai, Swapnil R. Andhariya, "Sentiment analysis Approach to adapt a shallow parsing based sentiment lexicon," IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems, ICIIECS'15, published by IEEE in 2015.

[8] Anh-Dung Vo, Quang-Phuoc Nguyen, And Cheol-Young Ock, "Opinion Aspect Relations in Cognizing Customer Feelings via Reviews," published by IEEE in 2018.

[9] M.Sivakumar, Dr.U.Srinivasulu Reddy, "Aspect Based Sentiment Analysis of Students Opinion using Machine Learning Techniques," Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017), published by IEEE in 2017.

[10] M.Trupthi, Suresh Pabboju, G.Narasimha, "Sentiment Analysis On Twitter Using Streaming API," 2017 IEEE 7th International Advance Computing Conference, published by IEEE in 2017.

[11] Sean Wijayanto, MasayuLeyliaKhodra, "Business Intelligence According to Aspect-Based Sentiment Analysis using Double Propagation," published by IEEE in 2018.

[12] Jaspreet Singh, Gurvinder Singh, "Sentiment Analysis of Social Media Reviews using QOS Parameterization," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), published by IEEE in 2018.

[13] Deepu S. Nair, Jisha P. Jayan, Rajeev R.R, Elizabeth Sherly, "Sentiment Analysis of Malayalam Film Review Using Machine Learning Techniques," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), published by IEEE in 2015.

[14] Elena Brunova, Yuliya Bidulya, "Aspect Extraction and Sentiment Analysis in User Reviews in Russian about Bank Service Quality," 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (ACIT), published by IEEE in 2017.

[15] Peiran Zhang††, Hongbo Zhu†, Tao Xiong†, Yihui Yang‡, "Co-attention Network And Low-Bank Bilinear Pooling For Aspect-Based Sentiment

Analysis," †Alibaba Group, Hangzhou, China, ‡Afterpay, Melbourne, Australia, ICASSP 2019, published by IEEE in 2019.

[16] Luxchippiriya Balachandran, AbarnahKirupananda, "Online Reviews Evaluation System for Higher Education Institution: An Aspect Based Sentiment Analysis Tool," 2017 11th InternationalConference on Software, Knowledge, Information Management and Applications (SKIMA), published by IEEE in 2017.

[17] Ganpat Singh Chauhan, Yogesh Kumar Meena, "Prominent Aspect Term Extraction in Aspect Based Sentiment Analysis," 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering, 22-25 November 2018 (IEEE Conference Record # 43534), published by IEEE in 2018.

[18] Pankaj Gupta, Ritu Tiwari, and Nirmal Robert, ”Sentiment Analysis and Text Summarization of Online Reviews: A Survey," International Conference on Communication and Signal Processing, April 6-8, 2016, India, published by IEEE in 2016.

[19] Saif A. Ahmad Ali Alrababah, KengHoon Gan, Tien-Ping Tan, "Product aspect ranking using sentiment analysis and TOPSIS," 2016 Third International Conference on Information Retrieval and Knowledge Management, published by IEEE in 2016.

[20] PuteriPrameswari, IstiSurjandari, Enrico Laoh, "Opinion Mining from Online Reviews in Bali Tourist Area," 2017 3rd International Conference on Science in Information Technology (ICSITech), published by IEEE in 2017

[21] Hossein Abbasimehr, Mostafa Shabani, “A Sentiment Aggregation System based on an OWA Operator”, 2019 5th International Conference on Web Research (ICWR).

[22] Tuan Anh Tran, Jarunee Duangsuwan, Wiphada Wettayaprasit, “A Novel Automatic Sentiment Summarization from Aspect- based Customer Reviews”, 2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE).

State-of-the-Art: Blockchain Security, Data Access Control and Consensus

Mythili R[a]*, Revathi Venkataraman[b]

[a]*Assiatant Professor/IT, SRM Institute of Science and Technology, Ramapuram, Chennai, India*
[b]*Professor/CSE, SRM Institute of Science and Technology, Kattankulathur, India*

Abstract

Blockchain (BC) technology becomes the most promising security mechanisms that has attracted the researchers and academicians in its true dimension of trust and immutability. Most emerging domain of healthcare and other industries experience the non-linear growth of data explosion and hence by unauthorized access issues between various parties involved. While access solutions are formulated using the decentralized ledger of transactions (BC), in which the consensus plays a vital role. Consensus in BC decides the performance of BC transaction system. This paper elaborates the detailed insights of BC technology, access control applications mainly healthcare and also the significance of consensus. The study also includes the recent article contributions that could guide blockchain researchers in the right path of developing an advanced information access control system.

Keywords: Blockchain; Data Security; Access Control; Consensus Algorithms.

1. Introduction

Looking at the current information era, the communication among individuals or organizations include mainly the pervasive values of information/knowledge. This digital revolution expects the privacy of information access as obligatory. Computer and internet applications ranging from bill payments, banking, insurance, healthcare, online learning and other businesses require the information access control as the integral one. Generally, information access control includes both authentication and authorization. Presently existing access control mechanisms are more flexible in allowing the malicious/unauthorized users to access the valuable information and also lagging with the technology. Hence, it is the right time to upgrade to an advanced

* Corresponding author.
E-mail address: mythilir2@srmist.edu.in. revathin@srmist.edu.in

access control mechanism which can incorporate current business requirements along with the user's expectations of privacy.

Blockchain (BC) the foundation of Bitcoin, is on the rise at an astonishing swiftness over the past few years. Nakamoto et al.,2008 [30] invented the first digital cryptocurrency bitcoin. Ethereum [16] is released as second public blockchain not only as currency, can record other assets because of its decentralized, time-stamped digital trust. Now-a-days, Bitcoin [27] is just one of the several hundred applications that use Blockchain technology. Although, blockchain is originated from cryptocurrency, the application extensions are not limited to banking, insurance, inventory, energy, agriculture, logistics and healthcare etc., most possibly for achieving access control and privacy.

As Blockchain is decentralized, trust-based ledger, integrity and non-repudiation [35] are the strengths of the technology. The Blockchain system can then be prepared as the most secured one by enforcing confidentiality and privacy. It is recommended to find the suitable access control system that uses the minimum resources. The study analyzes the recent articles to help regarding this research.

The paper is structured as Section 1, blockchain overview which gives the basics about Blockchain and its applications, types, basic requirements and the trust system. Section 1.2 explains the detailed BC architecture. Section 2 gives the major articles involving BC based access control applications, Section 3 lists the specific Blockchain based Healthcare access control applications and Section 4 gives the implementation of BC in other domains. Section 5 lists the major challenges in the scope. Consensus algorithms and its significance are reviewed in Section 6. The paper is concluded and the future scope is mentioned as in Section 7.

1.1. Overview of Blockchain

Blockchain is described as a distributed ledger of cryptographically linked growing records (Blocks) list. Each block has the cryptographic parameter of the previous hash, block timestamp and transaction details. In this section, the Blockchain applications overview, types of chains, basic requirements and the trust system involved are explained to understand the Blockchain basics.

Applications of Blockchain

Blockchain is being widely used in various domains of Internet of Things (IoT), Big data, financial services, transportation, reputation system etc., for different security

purpose. Blockchain technology has the embedded benefits of time saving, cost removal, lower risk with increased trust. Presently blockchain is being used by almost all industries like financial, public sector, retail, insurance and manufacturing. Some of the relevant applications are cross currency payments, mortgages, asset transfer, health record systems and supply chain management.

Table 1. Blockchain-Applications

Domain	BC-Applications
Agriculture	Soil processing, Product shipping, sales and marketing, growth, yields etc.,
Business	Digital transaction processing
Energy	Generation of Energy, supplier-demand records, tariff maintenance
Food and Nutrition	Ordering through online and transaction data, Quality check data
Medical Healthcare	Genomic information, electronic medical information, Case reports, prescription records
Manufacturing	Product assurance, guarantee and warranty information, robotics, packaging & delivery transaction
Smart city	Smart services, energy, water & pollution management, smart digital data & transactions
Transport, logistics	Toll maintenance, Logistics identifiers, Good delivery & shipping, Tracking of vehicle & container

Types of Blockchain

Blockchains are of three different types [1, 15] depending on the data management, data availability, and user actions. These include:

Public permission less Blockchains: Public chain is the blockchain in which anyone can join to act either as a node or as a miner, without the permission of third-parties, (by Consensus protocol). Some of the popular examples are Bitcoin, Ethereum and Litecoin.

Private permissioned Blockchains: Private chain is the one in which one can't join unless he is invited by the administrator. Restricted participant(node) and validator(miner) access is applicable. Since it is permissioned, smart contracts are mandatory for transaction processing. Examples are Hyperledger Fabric and Ripple.

Consortium/Public permissioned Blockchains: A Consortium chain is known to be semi-decentralized and permissioned, can be controlled by multiple organizations unlike single one as in private chains.

1.1.3. Basic Requirements of Blockchain

Blockchain, the centralized ledger that is distributed to every peer node in the network. BC needs the following basic requirements mandatory for its performance essence.

Shared Ledger: Shared Ledger is the distributed system of transaction records, shared across the network; Shared ledger is capable of **append-only**. In permissioned chains, participants can see only appropriate transactions.

Smart Contract: Smart contracts are tiny codes written and deployed in the nodes. Smart codes are verifiable, signed rules/policies for transactions.

Privacy: While ledger is shared, ensuring appropriate visibility among subsets of participants is achieved by central cryptography actions.

Trust: Always ledger is a considered to be a trusted source of information. Relevant participants endorse the transactions, and then added to the ledger with proper confidentiality. Chain transactions are verifiable by an audit trail. The same is achieved through consensus, provenance, immutability and finality. Next section explains in detail the trust system.

1.1.4. Trust System

It is well-known that Blockchain networks are basically trust based systems [18]. Trust of the network is concurred with the following four essential features as per Fig.1. Consensus, Provenance, Immutability and Finality.

Consensus: Consensus is defined as the group decision making among participants (e.g.) shared reference data(i.e.) metadata. It results in consolidated reduced errors and consistent dataset, also near real-time reference data access. It naturally supports code edit and routing code transfers between participants.

Provenance: Provenance denotes mainly the source of information like complete node details. It gives increased trust because no authority owns provenance and improved system utilization. Also recalls specific than cross fleet.

Immutability: Immutability ensures user authorization with tamper proof and privacy. It benefits by lowering the audit and regulatory costs. It also provides seek and find access to auditors and regulators.

Fig. 1. Trust System of Blockchain

Finality: Finality offers reward based common ledger to participants, same validated transaction and increased execution speed, reduced cost and risk along with some value-added services.

1.2. Blockchain Architecture

In a Blockchain network, during the data communication, all the participating nodes should have coordination with the decentralized public ledger called as ”CHAIN [18]”. For every transaction, the chain is appended with the new node based on the consensus. As in Fig. 3, the miner nodes verify and validate the transaction details for appending in the Blockchain after endorsement. The miner node is decided as any special node either a cluster head or the node having the higher reputation value. Blockchain performance is determined on the basis of the trust [18] about the nodes involved in the transaction.

Distributed consensus/mining [1,12] justifies and improves the trust of the BC network. During mining, if any node is found to be malicious then the reputation/recommendation of that node will not be encouraged by the miners. Node transaction will be added as a new block only after successful consensus by all respective miners.

Each block has unique ID, time stamp, previous block hash, hash of present block and other transaction details [18]. The transaction details could be any relevant information of data communication. The size of Blockchain decides the mode of storage, either local server or cloud (delegated), hence the performance of the Blockchain.

Selected recent surveys on Blockchain shows the need of research works majorly focusing the factors of scalability [36], consensus and BC-oriented network attacks [13,26] as mentioned in Table 2.

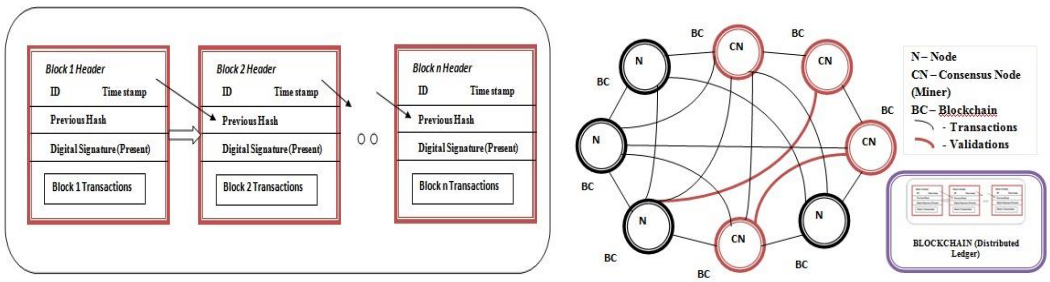


Fig. 2. (a) Blockchain Structure; (b)Blockchain with Consensus Nodes

A lightweight and scalable Blockchain namely LSB is proposed by Dorri et al., 2019[13] to achieve security and privacy in IoT networks. The system proves that blockchain is the effective technology to provide security and privacy in IoT through user-friendly consensus algorithm. Fernandez et al., 2018 [15] have worked on significant improvements in the area of consensus.

Table 2. Bibliometric summary of recent articles on Blockchain

Reference	Author(S)	Year	Publication Type	Publisher	Highlights
[26]	Liang Liu, Budong Xu	2018	Conference	IEEE	DoS & DDoS attacks
[13]	Ali Dorri, Salil S. Kanhere	2017	Conference	IEEE	Lightweight Consensus algorithm & Attacks
[15]	Tiago M. Fernández-Caramés, Paula Fraga-Lamas	2018	Journal	IEEE	Consensus & mining mechanisms
[23]	Nallapaneni Manoj Kumar, Pradeep Kumar Mallick	2018	Conference	Elsevier	Large Scale IoT-Advantages
36]	Ana Reyna, Cristian Martín	2018	Journal	Elsevier	Performance in IoT
[45]	Zibin Zheng, Shaoan Xie	2019	Conference	IEEE	Selfish mining
[14]	Dmitry Efanov and Pavel Roschin	2018	Conference	Elsevier	Pervasiveness

2. BC Based Access Control

Data access control protocols are used to regulate the sensitive data under the specific set of permissions expressed in terms of access policies. Fine grained access control can be possible through ABE encryption techniques (Attribute-Based Encryption) [3]. Attribute specific access policies [35] define the combinations of set of conditions and attributes together preserve the data privacy.

Blockchain has the feature of smart contracts which help to achieve the attribute specific access policies [35]. Table.3. shows the selected recent state-of-art Blockchain articles that includes access policies and henceforth smart contracts. The literature gives the new concepts of multi-rule policies [11], distributed hash function [32], RBAC [10] and CBAC [42] contracts and the consensus mechanism [21] are some of the popular Blockchain technology mechanisms to be strengthened.

Table 3. Bibliometric summary of recent articles on Blockchain applied in Access Control

Reference	Author(S)	Year	Publication Type	Publisher	Highlights
[11]	Damiano Di Francesco Maesa, Paolo Mori	2019	Conference	IFIP	Self-enforcing policies, Multi-Rule Policies
[38]	Tara Salman, Maede Zolanvari	2018	Journal	IEEE	Encryption & Authentication
[40]	Licheng Wang, Xiaoying Shen	2019	Journal	Elsevier	Hash functions & Signatures
[32]	Aafaf Ouaddah, Anas Abou Elkalam	2017	Journal	Wiley	Authorization-Distributed Hash Table
[33]		2017	Journal	Springer	Decentralized, privacy preserving authorization management
[34]	Rafael Pass, Elaine Shi	2017	Journal	IEEE	Mining rewards

Reference	Author(S)	Year	Publication Type	Publisher	Highlights
[10]	Ja,Son Paul Cruz, Yuichi Kaji	2018	Journal	IEEE	RBAC+Smart Contract, PoW
[42]	Ronghua Xu, Yu Chen	2018	Journal	IEEE	CBAC+Smart Contract
[29]	Bing Mo, Kuiren Su	2019	Journal	IEEE	Smart Contract for IoT
[44]	Yuanyu Zhan, Shoji Kasahara	2018	Journal	IEEE	Multiple Contracts
[21]	Emanuel Ferreira Jesus, Vanessa R. L. Chicarino	2018	Journal	Wiley	Stalker Attacks, Consensus & Mining

3. BC Based Healthcare Data Access Control

Due to the voluminous data explosion, healthcare data storage demands the third-party data centers. On- demand healthcare which is also digital and eco-friendly is the state-of-art scope of Blockchain technology. Health organizations maintain the health records in the form of databases internally or sometimes with external delegation. EHRs (Electronic Health Records) [1] suit to be this category where healthcare providers maintain the medical information of patients. HIPAA, Health Insurance Portability and Accountability Act [2] claim set of regulations for secure access of patient data which should not be compromised by external third-party data providers. In the other category of PHRs (Personal Health Records) [37], patients health records are managed by themselves and do not have access control over their data providers' databases. Hence, it is envisaged two such PHR context challenges as (1) Unified view by patients about their records, (2) Up-to-date patients' data access by healthcare providers irrespective of any changes.

Gordon et al., 2018 [17] presented a healthcare system by enforcing data liquidity, immutability and patient identity. Biomedical evidence data version [22] with smart digital contracts is introduced by Kleinaki et al., 2018 [22] . And Zhang et al., 2018

[43] developed a decentralized Healthcare system "FHIRChain" in which he used smart contract codes to achieve secure sharing of clinical data.

Table 4. Bibliometric summary of recent articles on Blockchain in Healthcare Access Control

Reference	Author(S)	Year	Publication Type	Publisher	Highlights
[19]	Marko Holbl, Marko Kompara	2018	Journal	Symmetry (MDPI)	Systematic Review of BC uses in Healthcare
[7]	James Brogan, Immanuel Baskaran	2018	Journal	Elsevier	Masked Authenticated Messaging (MAM) using IoTA protocol
[6]	Shaimaa Badr,Ibrahim Gomaa	2018	Conference	Elsevier	Multi-tier BC using Pseudonym Based Encryption with Different Authorities
[37]	Alex Roehrs, Cristiano André da Costa	2017	Conference	Elsevier	omniPHR-Distributed PHR, Elasticity & Scalability
[17]	William J. Gordon, Christian Catalini	2018	Conference	Elsevier	Data liquidity & Immutability, Patient Identity
[22]	Athina-Styliani Kleinaki , Petros Mytis-Gkometh	2018	Conference	Elsevier	Biomedical evidence data versioning, Smart digital contracts
[43]	Peng Zhang, JulesWhite	2018	Journal	Elsevier	FHIRChain: Decentralization, Scalability

4. BC Based Data Access Control in Other Domains

Blockchain technology is not only applied in eHealthcare but also in all trending sectors of construction, manufacturing [24] and production, education, energy management [4], supply chain management, social networking, sales and marketing, insurance and many more.

This paper surveys some of such recent domain specific articles which uses the Blockchain technology as listed below: Casadovara et al., 2018 [8] implemented

Blockchain in supply chain management with multi-agent system. Information sharing during Injection Mould Redesign (IMR) is achieved through the K-Nearest Neighbors algorithm along with Blockchain technology [24].

Table 5. Bibliometric summary of recent articles on Blockchain in Access Control of Other Domains

Reference	Author(S)	Year	Publication Type	Publisher	Highlights
[20]	Žiga Turk and Robert Klinc	2017	Conference	Elsevier	BC Technology for BIM transactions - provenance tracking
[25]	Chiehyeon Lim, Kwang-Jae Kim	2018	Journal	Elsevier	Smart City challenges using ICT & Big data using relevant Policies
[8]	Roberto Casado-Vara, Javier Prieto	2018	Conference	Elsevier	BC based circular economy for supply chain using Multi agent system
[24]	Zhi Li, Layne Liu	2018	Conference	Elsevier	Knowledge sharing using BC with K-nearest Neighbors for Injection Mould Redesign
[4]	Merlinda Andoni, Valentin Robu	2019	Journal	Elsevier	BC based Energy Management system-challenges & use cases with tamper-proof transactions

4. Challenges in the Scope

The outstanding Blockchain technology is facing enlisted challenges to be broadly sought out to derive the solutions. The research scope of this article is encircled to the specified areas of Blockchain based access control using consensus algorithms. Researchers can focus on the solutions for below said challenges and contribute to the society. Future direction in this regard helps to achieve the efficient privacy-preserving Blockchain implementations in the scope with novel solutions.

Table 6. Major Challenges in the Scope

Access Control	**Blockchain Technology**
• Data management & sharing	Storage
• Data replication	Scalability
• Data volume	Selfish Mining (Consensus)
• Data value	Transaction Costs
• Access policies	Source of Trust
• Smart Contract Access Rules	Sybil & DoS attacks
Healthcare	**Consensus Algorithm**
Data Ownership	Scalability
Privacy	Delegation
Financial Benefits	Processing location
Health Benefits	Processing Time Energy consumption
Secure Data Access Transfer	

4. Consensus Algorithms and Significance

Consensus protocols [9] are the core of Blockchain technology. The performance measures of blockchain such as processing time, energy consumption and scalability are significantly decided by consensus procedure followed in the respective context. In Blockchain network of distributed computing with system reliability, all the participating nodes have to agree and accept decisions on certain data value for new block inclusion. This goal is implemented by proper context of consensus and mining. There are many consensus protocols as per literature survey which differs in performance, scalability, delegation, processing location, time and energy consumption [41] etc. Probably, the consensus protocol contributes the significant improvement in the performance of Blockchain projects.

Selfish mining [12] is a special consensus that needs more attention. In this type of unusual mining, the miners collude themselves and hide the original miner who had succeeded already and hence by improving the revenue. Consensus procedures should always adhere the key characteristics of safety, liveness and fault-tolerance [39]. More and more research focus in this domain of consensus improvements delivers the more efficient Blockchain implementations.

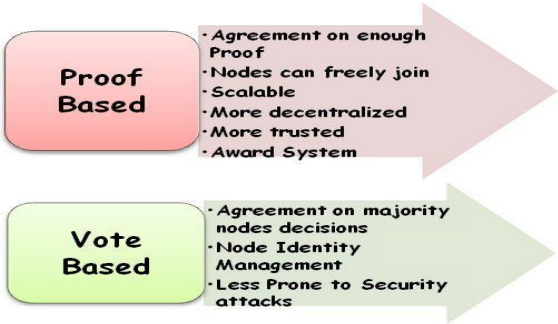


Fig. 4. Consensus Scenario **Fig. 5. Advantages of Consensus algorithms (Groupwise)**

Consensus protocols are generally grouped as two categories, Proof-based and Voting-based [5]. Proof of Work, Proof of Stake, Proof of Space , Proof of Luck and Proof of Burn are some of the popular proof based consensus algorithms. Special Vote-based [31] algorithms include Byzantine Fault Tolerance (BFT), Crash Fault Tolerance (CFT). Wang et al., 2019 [41] had listed the detailed list of consensus algorithms in practice. He also described mining management in Blockchain networks. The advantages of consensus groups as follows:

Proof-based consensus has the advantage of free flow of nodes to join/leave from the network when compared to Voting-based consensus where in the nodes should be adjustable during verification.

Table 7. Popular Consensus algorithms [31,41]

Proof-Based	Voting-Based
Proof-of-Work	Byzantine Fault Tolerance (BFT)
Proof-of-Stake	Crash Fault Tolerance (CFT)
Proof-of-Burn	Delegated Byzantine Fault Tolerance (dBFT)
Proof-of-Elapsed Time	Practical Byzantine Fault Tolerance (PBFT)
Proof-of-Luck	
Proof-of-Importance	
Proof-of-Space	
Proof-of-Activity	
Proof-of-Ownership	

Table 8. lists the recent the survey article works which focuses on the blockchain consensus algorithms and its power to rule. In-depth sights of these articles will give the researchers the functional flow of the consensus algorithms, advantages and

disadvantages etc.

Table 8. Bibliometric summary of recent articles on Consensus Protocols of Blockchain

Reference	Author(S)	Year	Publication Type	Publisher	Highlights
[46]	S. Zoican, M. Vochin, R. Zoican, and D. Galatchi	2018	Conference	IEEE	Performance comparison of PoW, Practical Byzantine Fault Tolerance, Binary Consensus,
[9]	Nutthakorn Chalaemwongwan	2018	Journal	IEEE	Survey & Applications on CAP, BGP,PBFT,PoW,PoS, PoA,PoP,PoB,PoR,PoET, PoC,PoI,PoO,PoSpc, DPOS along with BC Platforms
[5]	L.M.Batch,M.Mihaljevic	2018	Conference	IEEE	Comparison of BFT,dBFT,PoW, RPCA,PoS,dPoS,PoI on Scalability, Power
[28]	Du Mingxiao, Ma Xiaofeng	2017	Conference	IEEE	Popular consensus algorithms in different scenarios
[31]	Giang-Truong Nguyen, Kyungbaek Kim	2018	Journal	JIPS	Proof Based & Voting Based Consensus protocols
[41]	Wenbo Wang, Dinh Thai Hoang	2018	Journal	IEEE	Survey on Proof of Concept consensus algorithms

4. CONCLUSIONS AND FUTURE DIRECTIONS

Blockchain becomes trending now-a-days due to the extensive access control applications. The exhaustive study comprises of the favorite Blockchain technology including architecture and applications, with emphasis on consensus. Specifically, the study includes a comprehensive survey in the domain of access control and identified the various challenges as in Table 6. It is found that consensus protocols played very important role in Blockchain optimization. Table 7. lists the popular consensus algorithms in practice and can be improvised further.

As future extension, the introduction of attribute specific policies still can lead to fine-grained access control applications. To conclude, from the results of the study, it is most suggested that Blockchain with attribute policies can be used for fine-grained access control applications.

References

[1] Electronic health records. https://www.healthit.gov/sites/default/files/factsheets/what-patients-need-to-know-about-ehrs.pdf.

[2] HIPAA basics for providers: Privacy, security, and breach notification rules. https://www.cms.gov/Outreach-and-Education/ Medicare-Learning-Network-MLN/MLNProducts/Downloads/ HIPAAPrivacyandSecurity,.

[3] Miguel Ambrona, Gilles Barthe, Romain Gay, and Hoeteck Wee., 2017. Attribute-based encryption in the generic group model: Auto- mated proofs and new constructions. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, pages 647–664, New York, NY, USA.

[4] Merlinda Andoni, Valentin Robu, David Flynn, Simone Abram, Dale Geach, David Jenkins, Peter McCallum, and Andrew Pea- cock., 2019. Blockchain technology in the energy sector: A systematic review of challenges and opportunities. *Renewable and Sustain- able Energy Reviews*, 100:143 – 174.

[5] L. M. Bach, B. Mihaljevic, and M. Zagar., 2018 Comparative analysis of blockchain consensus algorithms. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1545–1550.

[6] Shaimaa Badr, Ibrahim Gomaa, and Emad AbdElrahman.,2018. Multi- tier blockchain framework for IoT EHRS systems. *Procedia Computer Science*, 141:159 – 166, 2018. The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018).

[7] James Brogan, Immanuel Baskaran, and Navin Ramachandran., 2018. Authenticating health activity data using distributed ledger technologies. *Computational and Structural Biotechnology Journal*, 16:257 – 266.

[8] Roberto CasadoVara, Javier Prieto, Fernando De la Prieta, and Juan M. Corchado. How blockchain improves the supply chain: case study alimentary supply chain.,2018. *Procedia Computer Science*, 134:393 – 398, 2018. The 15th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2018).

[9] N. Chalaemwongwan and W. Kurutach., 2018. State of the art and challenges facing consensus protocols on blockchain. In *2018 International Conference on Information Networking (ICOIN)*, pages 957–962.

[10] J. P. Cruz, Y. Kaji, and N. Yanai., 2018. RBAC-SC: Role-based access control using smart contract. *IEEE Access*, 6:12240–12251.

[11] Damiano Di Francesco Maesa, Paolo Mori, and Laura Ricci., 2017. Blockchain based access control. In Lydia Y. Chen and Hans P. Reiser, editors, *Distributed Applications and Interoperable Systems*, pages 206–220, Cham. Springer International Publishing.

[12] Ali Dorri, Salil S. Kanhere, and Raja Jurdak., 2017. Towards an optimized blockchain for iot. In *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*, IoTDI ’17, pages 173–178, New York, NY, USA.

[13] Ali Dorri, Salil S. Kanhere, Raja Jurdak, and Praveen Gauravaram.,2019. LSB: A lightweight scalable blockchain for iot security and anonymity. *Journal of Parallel and Distributed Computing*.

[14] Dmitry Efanov and Pavel Roschin., 2017. The all-pervasiveness of the blockchain technology. *Procedia Computer Science*, 123:116 – 121, 2018. 8th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2017 (Eighth Annual Meeting of the BICA Society), Russia.

[15] T. M. Ferna´ndez-Carame´s and P. Fraga-Lamas., 2018. A review on the use of blockchain for the internet of things. IEEE Access, 6:32979–33001.

[16] Adem Efe Gencer, Soumya Basu, Ittay Eyal, Robbert van Renesse, and Emin Gu¨n Sirer., 2018. Decentralization in bitcoin and ethereum networks. CoRR, abs/1801.03998.

[17] William J. Gordon and Christian Catalini., 2018. Blockchain technology for healthcare: Facilitating the transition to patient-driven interoperability. *Computational and Structural Biotechnology Journal*, 16:224 – 230.

[18] S. H. Hashemi, F. Faghri, P. Rausch, and R. H. Campbell., 2016. World of empowered iot users. In *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 13–24.

[19] Marko Ho¨lbl, Marko Kompara, Aida Kamisˇalic´, and Lili Ne- mec Zlatolas., 2018. A systematic review of the use of blockchain in healthcare. Symmetry, 10(10).

[20] Zˇ iga Turk and Robert Klinc., 2017. Potentials of blockchain technology for construction management. Procedia Engineering, 196:638 – 645, 2017. Creative Construction Conference 2017, CCC 2017, 19-22, Croatia.

[21] Emanuel Jesus, Vanessa Chicarino, Ce´lio Albuquerque, and Antoˆnio Rocha., 2018. A survey of how to use blockchain to secure internet of things and the stalker attack. Security and Communication Networks, 1–27.

[22] Athina-Styliani Kleinaki, Petros Mytis-Gkometh, Pavlos S. Efraimidis, and Eleni Kaldoudi., 2018. A blockchain-based notarization service for biomedical knowledge retrieval.Computational and Structural Biotechnology Journal, 288 – 297.

[23] Nallapaneni Manoj Kumar and Pradeep Kumar Mallick. Blockchain technology for security issues and challenges in IoT., 2018. Procedia Computer Science, 132:1815 – 1823. International Conference on Computational Intelligence and Data Science.

[24] Zhi Li, Xinlai Liu, W. M. Wang, Ali Vatankhah Barenji, and George Q. Huang., 2019. CKShare: secured cloud-based knowledge- sharing blockchain for injection mold redesign. Enterprise Information Systems, 13(1):1–33.

[25] Chiehyeon Lim, Kwang-Jae Kim, and Paul P. Maglio., 2018. Smart cities with big data: Reference models, challenges, and considerations. Cities, 82:86 – 99.

[26] L. Liu and B. Xu. Research on information security technology based on blockchain., 2018. In 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pages 380–384.

[27] Rob Marvin., 2009. Blockchain: The invisible technology that’s changing the world.

[28] D. Mingxiao, M. Xiaofeng, Z. Zhe, W. Xiangwei, and C. Qijun., 2017. A review on consensus algorithm of blockchain. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 2567–2572.

[29] B. Mo, K. Su, S. Wei, C. Liu, and J. Guo., 2018. A solution for internet of things based on blockchain technology. In 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), pages 112–117.

[30] Satoshi Nakamoto., 2009. Bitcoin: A peer-to-peer electronic cash system.

[31] Truong Nguyen and Kyungbaek Kim., 2018. A survey about consensus algorithms used in blockchain. Journal of Information Processing Systems, 14:101–128.

[32] Aafaf Ouaddah, Anas Abou Elkalam, and Abdellah Ait Ouahman., 2016. Fairaccess: a new blockchain-based access control framework for the internet of things. Security and Communication Networks, 9(18):5943–5964.

[33] Aafaf Ouaddah, Anas Abou Elkalam, and Abdellah Ait Ouah- man., 2017. Towards a novel privacy-preserving access control model based on blockchain technology in IoT. In Advances in Information and Communication Technologies, pages 523–533. Springer International Publishing.

[34] Rafael Pass and Elaine Shi., 2017. Fruitchains: A fair blockchain. In Proceedings of the ACM Symposium on Principles of Distributed Computing, PODC ’17, pages 315–324, New York, NY, USA.

[35] Y. Rahulamathavan, R. C. Phan, M. Rajarajan, S. Misra, and A. Kondoz., 2017. Privacy-preserving blockchain based iot ecosystem using attribute-based encryption. In 2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), pages 1–6.

[36] Ana Reyna, Cristian Mart´ın, Jaime Chen, Enrique Soler, and Manuel D´ıaz. On blockchain and its integration with IoT. challenges and opportunities. Future Generation Computer Systems, 88:173 – 190, 2018.

[37] Alex Roehrs, Cristiano Andre´ da Costa, Rodrigo da Rosa Righi, and Kleinner Silva Farias de Oliveira., 2017. Personal health records: A systematic literature review. J Med Internet Res, 19(1):e13.

[38] T. Salman, M. Zolanvari, A. Erbad, R. Jain, and M. Samaka., 2019. Security services using blockchains: A state of the art survey. IEEE Communications Surveys Tutorials, 21(1):858–880.

[39] Wattana Viriyasitavat and Danupol Hoonsopon., 2019. Blockchain characteristics and consensus in modern business processes. Journal of Industrial Information Integration, 13:32 – 39.

[40] Licheng Wang, Xiaoying Shen, Jing Li, Jun Shao, and Yixian Yang., 2019. Cryptographic primitives in blockchains. Journal of Network and Computer Applications, 127:43 – 58.

[41] Wenbo Wang, Hoang Dinh Thai, Peizhao Hu, Zehui Xiong, Dusit Niyato, Ping Wang, Yonggang Wen, and Dong Kim., 2019. A survey on consensus mechanisms and mining strategy management in blockchain networks. IEEE Access, PP:1–1.

[42] R. Xu, Y. Chen, E. Blasch, and G. Chen., 2018. BlendCAC: A blockchain-enabled decentralized capability-based access control for IoTs. In 2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pages 1027–1034.

[43] Peng Zhang, Jules White, Douglas C. Schmidt, Gunther Lenz, and S. Trent Rosenbloom., 2018. FHIRchain: Applying blockchain to securely and scalably share clinical data. Computational and Structural Biotechnology Journal, 16:267 – 278.

[44] Y. Zhang, S. Kasahara, Y. Shen, X. Jiang, and J. Wan., 2019. Smart contract-based access control for the internet of things. IEEE Internet of Things Journal, 6(2):1594–1605.

[45] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang., 2017. An overview of blockchain technology: Architecture, consensus, and future trends. In 2017 IEEE International Congress on Big Data (BigData Congress), pages 557–564.

[46] S. Zoican, M. Vochin, R. Zoican, and D. Galatchi., 2018. Blockchain and consensus algorithms in internet of things. In *2018 International Symposium on Electronics and Telecommunications (ISETC)*, pages 1–4.

Performance analysis of Load Balanced Ant Colony Optimization in cloud

Richa Pandey *, Aditi Paulb
Dept. of Computer Science, Banasthali Vidyapith, Tonk, 304022,India
Dept. of Computer Science, Banasthali Vidyapith, Tonk, 304022,India

Abstract: Load balancing in cloud environment is a crucial in terms of optimization. The existing load balancing algorithms are mostly centralized in nature which causes inefficient utilization of virtual machines (VMs) by the tasks to be performed. Researchers have been shifted from static load balancing paradigms to dynamic and adaptable load balancing environment in which resource utilization is optimized and performance is enhanced. In the current study a load balanced Ant Colony Optimization (ACO) algorithm is proposed which is able dynamically allocate VMs to tasks depending on the pheromone concentration level. The code in CloudSim simulator is implemented and its performance is compared with static load balancing techniques in terms of cost and response time.

Index Terms: Cloud Computing, load balancing, ACO, CloudSim.

I. INTRODUCTION

In terms of virtualization cloud computing exhibit high performance in processing task efficiently all over the world [1],[2]. However load balancing in datacenter is becoming an issue since VMs have limited capacity as they are built on general purpose machines having very normal configuration. In order to assign tasks to VMs various task scheduling algorithms are proposed so far and cloud architecture is based on assigning these task according to the best algorithm suitable for an environment. In figure1 cloud load balancing model is shown [3]. User requests are divided into tasks and submitted to load balancer through datacenter broker. Load balancer assigns each next task to the VM using load balancing algorithms.

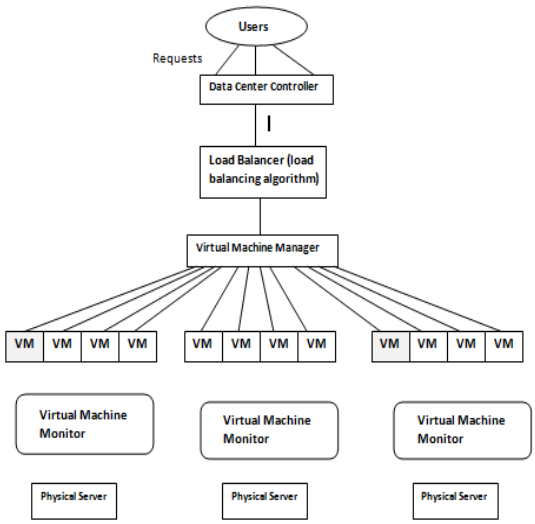


Fig.1. Cloud Load Balancing Architecture

.

At this point of time some observations are crucial:

- User requests come from all over the world.
- VMs should be allocated to the tasks depending on the current load.
- Virtualization should be applied in order to optimize resource utilization during task.

In order to efficiently assign a task to a VM having underutilized load is the basics of dynamic load balancing techniques [4], [5]. With static load balancing tasks are assigned to a VM which is idle or that which is currently available in the schedulers. This creates an uneven assignment of tasks on the VMs as there is no information given about the current load of the VMs. Since resources are expensive and user requests are random there should be a trademark between task allocation and VM load. Dynamic load balancing algorithms play an important role here in order to dynamically allocate resources to the tasks. This results in increase of available resources, power saving, maximum utilization of resources, cost reduction preserving the flexibility, and reduction of carbon emission.

The current study this is analyzed that the performance of dynamic load balancing algorithms over static ones through a modified Ant Colony Optimization load balancing technique with respect to increased response time and decreased cost. Rest of the paper is structured as follows:

Section II describes performance parameters and policies of dynamic load balancing algorithms. Section III implements RR, Throttled, SJF and modified ACO in

CloudSim simulator. Section IV analyzes the performance of each with respect to performance metrics defined in section II. Section V concludes the work.

II. LITERATURE REVIEW

Such as CPU Load (total number currently running and waiting process), memory space used and the network delay load (the time it takes for a bit of data to travel across the network from one node to another). An efficient load balancing strategy is the main challenge in cloud computing as it requires the distribution of workload to the physical servers in an optimized manner during pick hours. In the literature of load balancing various techniques are evolved [6] over time. Each load balancing algorithm should be able to produce higher throughput, minimum response time and maximum resource utilization [7]. Some researches [8, 9] classify load-balancing algorithms into two factors: the state of the system and the sender of the process. Algorithms based on current state of the system are classified as static and dynamic. Some static algorithms are Round Robin, Min-Min and Max-Min Algorithms, and Opportunistic Load Balancing (OLB) [10]. Some of the dynamic algorithms include Ant Colony Optimization (ACO) [11], Honey Bee Foraging [12], Throttled[13], Particle Swarm Optimization (PSO) [14] algorithm, Equally Spread Current Execution (ESCE) [13]etc.

Static algorithms like Round Robin, SJF and FCFS work with a prior knowledge of the system and input load and are not able to achieve optimization. These algorithms work well in systems with low variation in load [15]. The workload is distributed evenly among the nodes. Since these algorithms require a prior knowledge of system resources so the shifting of the load does not depend on the current state of system [16]. Thus static load balancing algorithms have a drawback that they are not flexible to distribute load dynamically [17] based on an instant decision mechanism.

Dynamic load balancing like as mentioned above on the other hand, are flexible to migrate the tasks from local node to remote node by taking the decision on some predefined rules. These rules are called selection rules which are based on some calculated parameters at an instant of time. In this section the principle of each scheduling is discussed in brief for the sake of experimental analysis.

Round Robin scheduling [18]: it is the static scheduling technique based on time scheduling. It maintains a VM allocation table for assigning tasks to the current VM

in a rotation basis. If the current VM is the last one then it reassigns the current VM variable to the first VM to map the task. Otherwise, it directly maps received task with that VM whose id is stored in current_vm variable. This technique performs well for short CPU burst. For larger task short quantum time creates more context switch and hence performance degrades.

Throttled: This is centralized in nature. In this load balancer a single job scheduler is used which keeps track of the current status of the VMs. A virtual machine can have only two states: occupied or idle, denoted by 1 or 0 respectively. As a new task arrives, job scheduler searches the virtual machine which is not busy. If it finds an idle (0) virtual machine it assigns the task to that VM. If no VMs are free task has to wait. No queues are maintained at the virtual machine level. Here only one VM can be allocated to a single task and other task has to wait until the VM is idle i.e. the current task is completed. Hence waiting time is generally large in this technique.

*E**qually Spread Current Execution (ESCE)***: it is a dynamic load balancing algorithm, which works on priority. The task with the least size is given highest priority. This algorithm first checks the size of the process and accordingly transfers the load to the VM which have lesser load. This is why it is called spread spectrum technique.

Particle Swarm Optimization (PSO) Algorithm: The PSO algorithm emphasize on minimizing the total cost of computation of an application workflow. However, this algorithm has larger computation time than others.

Ant colony Optimization (***ACO):*** Ant colony algorithm is inspired from the foraging behavior of ants. Ant colony loop through the entire task nodes according to certain rules rather than random traversal. A task cannot be simultaneously distributed to multiple virtual machines. It should assign task to the idle virtual machine when there has one or more in the system. The algorithm used is as follows:

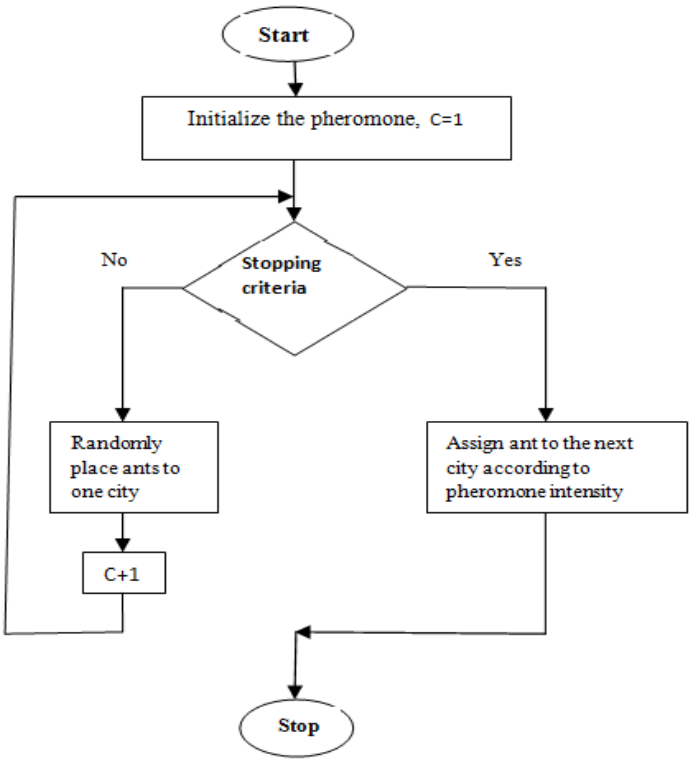


Fig.2. Flow Chart of Ant Colony Optimization

The final convergence path is the optimal solution. Here the probability that task *I* select virtual machine *j* can be calculated below:

$$p_{ij}^{k} = \begin{cases} \dfrac{[\eta_{ij}^{\beta}][\tau_{ij}(t)]^{\alpha}}{\sum_{k \in allowed_k}[\eta_{ij}^{\beta}][\tau_{ij}(t)]^{\alpha}} & \text{if } j \in allowed_k \\ 0 & \text{if } j \notin allowed_k \end{cases}$$

Where

$\tau_{ij}(t)$ = is the pheromone value at time *t* on the path *ij* .

η_{ij}^{β} = which we called visibility, is =1/ d_{ij} where d *is* the length of path *ij*

$allowed_k$ city's number that at time t ant k in city i can goes to.

α= the importance of pheromone; □

β=indicates the importance of distance between cities.

When Ant Colony Optimization compared with other existing load balancing algorithm results in reducing the response time for the given number of tasks and efficiently handles the resources at hand. [21].

When Ant Colony Optimization compared with other existing load balancing algorithm results in reducing the response time for the given number of tasks and efficiently handles the resources at hand. [21].

III. PERFORMANCE METRICS OF STATIC AND DYNAMIC LOAD BALANCING TECHNIQUE

The current study aims at analyzing the strength and weakness of the above discussed load balancing algorithms in complex network scenarios. The load balancing techniques is simulated in CloudSim simulator and compares their performances. This experiment draws some significant analysis for each of the techniques which are uniquely discussed in this study and hence this inference is valuable for further extension of the already existing techniques. Also the modified ACO algorithm discussed in section II is being separately judged on the basis of two important parameters α and . In order to do so we first discuss the load balancing policies based on which cloud computing paradigm is working and along with the performance metrics to understand the efficiency of such techniques. One important aspect to be mentioned here is that in ideal situation any load balancing technique is supposed to achieve the maximum/minimum (optimum) values of these metrics. However, in real scenario this is not the case due to the unpredictable user requests and complex environment. Following steps discuss the general dynamic load balancing policies:

1. ***Allocation /transfer of task***: when a new task enters the allocation is done on the basis of current state of the system i.e. current load of the VMs. The transfer of task from local node to a remote node is done by considering a threshold load units of the nodes.

2. ***Task selection***: during the task scheduling/re-scheduling section of task for migration is done on the basis of task execution time, migration overhead and number of remote system calls required for migration.

3. ***Node selection:*** to transfer a task to a particular node at a remote location depends on node's load factor. If the node is under-utilized and supporting services for migration to the node are available then only migration is possible. In either case the task is processed locally.

The metrics for identifying a load balancing efficiency are evolved over time. Literatures have shown many such performance metrics of cloud load balancing algorithms [19, 20] which are:

Throughput: it is measured as the number of tasks completed per unit time.

Response time: the total time the system takes to serve a submitted task.

Datacenter Processing time: The time taken by data centers to service a user request. It is calculated as total processing time for the entire simulation.

Scalability: It is the ability to uniformly perform the efficiency even with the increased number of nodes (scalable).

Cost: the over cost of operation.

The above parameters are the essential measures of task scheduling techniques in a complex environment where VMs are limited with respect to tasks.

IV. SIMULATION RESULT AND COMPARATIVE ANALYSIS

In this part the static task scheduling algorithm (RR) and dynamic load balancing algorithms (ACO, PSO, Equally ESCE and Throttled) and compare the performance in terms of response time, datacenter processing time and cost. These algorithms are simulated in CloudSim simulator and show the result with some specific input configurations (Table I) to evaluate each algorithm's performance as mentioned.

The average response time for each of the method is shown in Table 1. From Fig.2 it is evident that with the increase of VM for a fixed no. of datacenter, response time becomes larger. This is obvious as increase of VM will cause more number VMs to participate in scheduling which may take more time. If we increase datacenter then the response time logically decrease as the same tasks now have more resources and hence tasks execute faster. From Table 2 and Fig.3 is observed that increasing datacenter the datacenter processing time for RR, ACO and PSO decrease whereas ESCE and Throttled increases. Also in compared to others PSO shows larger processing time as mentioned in section II. Table III shows an important contribution in the current study. As the modified ACO is proposed here for optimizing resource this is reflected in Table III. With increasing no. of datacenter ACO shows lowest cost which shows its best performance. The pheromone concentration and distance between the cities change the performance of ACO. Here we took 10 ants with $\alpha=2$

and □=4 which gives the best result for this simulation. Another important observation we found is that the amount of increase of cost for PSO is almost same as ACO which indicates that with higher resources ACO and PSO perform similar.

TABLE I

Simulation Parameters

Type	Parameters	Values
Datacenter	Number of Data Center	1,2,5,10
Virtual Machine (VM)	Total number of VMs	25,50
	VM memory	512
	Bandwidth	1000
And another entry	Length of task	100

TABLE II. Average Response Time

Data Center containing different virtual Round Robin machines	Equally Spread	Throttled	Ant Colony Optimization	Particle Swarm Intelligent
5 datacenter with 25VMs 51.09	50.86	50.75	51.55	173.53
5 datacenter with 50VMs 51.63	51.26	51.17	52.4	282.83
10 datacenter with 25VMs 50.83	50.75	50.75	51.12	132.11
10 datacenter with 50VMs 51.37	51.29	51.55	51.85	210.1

TABLE III. Average Datacenter Processing Time

DataCenter containing different virtual machines	Round Robin	Equally Spread	Throttled	Ant Colony Optimization	Particle Swarm Intelligent
5 datacenter with 25VMs	1.33	1.09	0.98	1.79	108.34
5 datacenter with 50VMs	1.87	1.5	1.41	2.65	179.72
10 datacenter with 25VMs	1.08	1.01	1	1.38	81.76

10 datacenter with 50VMs	1.63	1.54	1.55	2.12	141.5

TABLE IV. Average Cost

DataCenter containing different virtual machines	Round Robin	Equally Spread	Throttled	Ant Colony Optimization	Particle Swarm Intelligent
5 datacenter with 25VMs	16.06	16.06	16.06	16	16.06
5 datacenter with 50VMs	26.09	26.09	26.09	25.06	26.09
10 datacenter with 25VMs	30.51	30.51	30.51	30.5	30.61
10 datacenter with 50VMs	55.6	55.6	55.6	55.45	55.7

Fig.3. Comparative Analysis of Average Response time

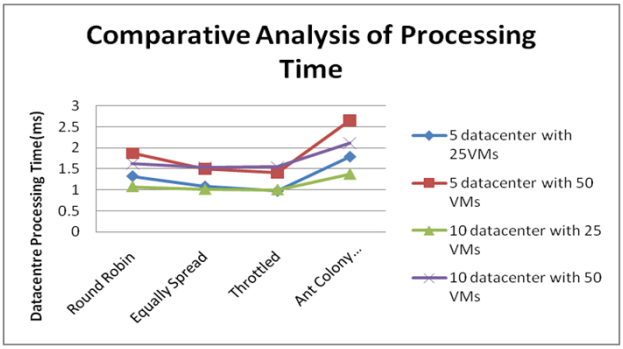


Fig.4. Comparative Analysis of Average DataCenter Processing time

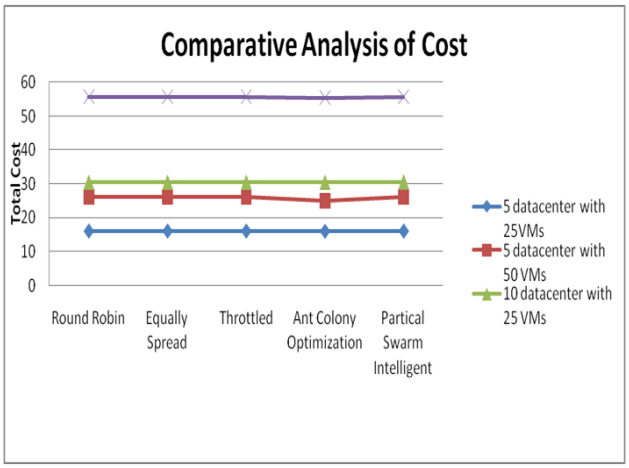


Fig.5. Comparative Analysis of Total Cost

V. CONCLUSION

In the current study analysis of the performance of modified Ant Colony Optimization technique in cloud load balancing paradigm. The aim of this paper is to show how ACO outperforms other load balancing techniques like Throttles, PSO , RR in terms of cost optimization in a complex environment. By varrying the values of pheromone and distance between nodes to find out the the best solution for the current simulation configuration. This indicates a future scope of achieving more variations in performance in a more complex environment. Also during experiment you achieve the performnace of PSO technique which shows an optimized cost but with a drwback of more computation time. This observation can also be evaluated in future experiment by modifying the algorithm according to real cloud environment.

REFERENCE

[1] Kaur, R., Luthra, P., 2014, "Load Balancing in Cloud Computing", International Conference on Recent Trends in Information. Telecommunications and Computing, ITC, pp. 1-8.

[2] Malladi, R. R, ,2015., "An approach to load balancing in cloud computing", Int. J. Abhinav Res. Science. Engineering. Technique., 4 (5), 3769–3777.

[3] Gupta, H., Sahu, K .2014. "Balancing functions in honey bee behavior based cloud computing". International Journal of Science and Research 3 (6).

[4] Goyal, S., Verma, M.K, 2016,"Load balancing techniques in cloud computing environments: a review", Int. J. advice. Race. Computer.

[5] Singh, P., Baga, P., Gupta, S., "Mixed Load Balancing Algorithms in Cloud Computing: A Survey", International Journal of Computer Application 143 (7), 34–40.

[6] E Z Gomi, A M Rahmani, N N Qadar, 2017,"Load-Balancing Algorithms in Cloud Computing: A Survey", J. Network and Computer Applications 88: 50–71 .

[7] Bhatia, J., Patel, T., Trivedi, H., Mazamudar, V.,2012, "HTV Dynamic Load Balancing Algorithm for Virtual Machine Instances in Cloud", International Symposium on Cloud and Services Computing, 15–20.

[8] Mishra, N.K., Misha, N.,2015, "Load Balancing Techniques: Requirements, Objectives and Key Challenges in Cloud Computing — A Systematic Review", International Journal of Computer Applications, 131 (18).

[9] Rastogi, G ,2015, "Analytical literature survey on existing load balancing schemes in cloud computing", International Conference on Green Computing and Internet of Things (ICGCloT).

[10] Aditya, A., Chatterjee, U., Gobata, S., 2015,"Comparative study of various static and dynamic load-balancing algorithms in cloud computing with special emphasis on time factor", Int. J. Chur Engineering. Technique. 3 (5).

[11] Nishant, K. Sharma, P., Krishna, V., Gupta, C., Singh, K.P., Nitin, N., Rastogi, R.,2012," Load balancing of nodes in the cloud using anti- colony optimization", In: Proceedings of the 14th International Conference on Modeling and Simulation, 3–8.

[12] Babu, LDD, Krishna, PV.,2013, " Honey bee behavior induces load balancing of tasks in cloud computing environments", Appl. Cool company 13 (5), 2292–2303.

[13] Domnall, Sridhar and R M Reddy. 2013,"Load Computing in Cloud Computing Modified Throttle Algorithm", Cloud Computing at Emerging Markets (CCEM), IEEE International Conference.

[14] Suraj P, Linlin Wu, Mayura S, and Buyya R, 2010"A particle self-optimizing optimization based for the determination of workflow applications in cloud

computing environments.", Advanced Information Networking and Applications (AINA),, PP. 400–407.

[15] NS Raghav and D. Singh,2014 "Comparative Studies on Load Balancing Techniques in Cloud Computing,", Open Journal of Mobile Computing and Cloud Computing Vol. 1, no. 1, pp. 18–25.

[16] Y. Sahu and R. K. Pateria, 2013,“ Cloud Computing Overview with Load Balancing Technology” , International Journal of Computer Applications (0975 –8887) ,Vol. 75, No. 24, pp. 60-4.

[17] Amandeep, v. Yadav and F. Mohammed, 2014,"Different strategies for load balancing in a cloud computing environment: a critical study," vol. 3, issue. 1, pp. 85–90.

[18] Maltare J, Prajapat B ,2016,"Dynamic Load Balancing in the Cloud Computing Framework for Load Balancing of Web Servers” International Journal of Computer Applications, Volume 148 - Number 5.

[19] Dargami, E.Y., Yuan, S.M.,.2015, “A small world-based overlay network for improving dynamic load-balancing”. J. Cyst. Softw. 107, 187–203.

[20] Milani, AS, Navampour, NJ, ,2014, “Load balancing mechanisms and techniques in cloud environments: systematic literature review and future trends”. J. Network. computer. Appl. 71, 86–98.

[21] SomulaRamasubbareddy, T. AdityaSaiSrinivas, K. Govinda, S.S. Manivannan& E. Swetha, 2019,“Analysis of Load Balancing Algorithms using Cloud Analyst” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6S2.

FPGA based Design and Implementation of modified SIDH Key Exchange Algorithm

Poomagal C Ta1, Sathish kumar* G Ab, Deval Mehtac

a,bDepartment of Electronics and communication Engineering, Sri Venkateswara College of Engineering, Sriperumbudur, 602117, INDIA.

cSAC, ISRO, AHmedabad, INDIA.

Abstract

This paper surveys and implements the Elliptic Curve Supersingular Isogeny Diffie Hellman (EC -SIDH) Key exchange method of post-quantum cryptosystems. The proposed method of key exchanging with Elliptic curve systems provides similar key sizes and computationally productive executions when contrasted with existing schemes. All the while, it is referred with some essential scientific develops that will enhance our fundamental perception of Elliptic Curves. And the investigation has been done between Supersingular elliptic curves and isogenies in the existing models and the proposed model. Finally, this proposed key exchange scheme proves to be a better post-quantum cryptosystems than other similar cryptosystems.

Keywords: Post Quantum Cryptosystem; EC-SIDH key Exchange; Supersingular Elliptic curves; Field Arithmetic; Diffie-Hellman key Exchange.

1. Introduction

It is accessed that the quantum computers will come into a reality in the next decades. In the event that, such systems are acknowledged soon then at that point considerable public key encryption algorithms like RSA, Diffie-Hellman, Elliptic Curve Diffie-Hellman, and Elliptic Curve DSA will end up uncertain security and should be supplanted. It is prudential to start considering such a situation and to investigate potential substitutions.

In the post-quantum cryptography, numerous methodologies will be lacking. The broadly utilized RSA cryptosystem will be superannuated. Any system dependent upon the Discrete Logarithm Problem will be endangered. As shown in [1] the Elliptic curve Discrete Logarithm problem is also not resistant to quantum cryptanalysis. A difficult problem will be required that keeps up the requirement for an exponential

* Corresponding author.
E-mail address: sathish@svce.ac.in

time attack, even with a quantum computer. Hence, the Supersingular elliptic curves and the isogenies between these curves form a system, which has numerous advantages. Primarily, it works on the some of the same primitives as in common with ECC implementations, so the previously available method can be more easily upgraded. Besides, it is not liable under any known patents and thus remains free and open to the research community. The public key of size 330 bytes [2] utilized in SIDH key exchange [1] algorithm is sufficiently stronger enough to safeguard the conveyance in embedded devices. SIDH depends on elliptic curves and offers likenesses with conventional elliptic curve cryptography (ECC); however, the underlying number-theoretic problem is the isogeny-graph problem. An isogeny function $\emptyset$ is an algebraic map relating the two elliptic curves, which are formulated over a finite field of characteristic p. The point multiplication $Q = [\blacksquare\blacksquare]P$ in traditional ECC is treated as a different strategy of an isogeny where $Q = \emptyset(\blacksquare\blacksquare)$ for identical curves. The isogeny of a domain and its co-domain resulting to the isogeny-graph problem and it is known as claw problem [3].

This graph contains nodes and edges, which defines the isomorphism classes of elliptic curves that are representing the isogenies. Alice and Bob traverse begin from the nodal points in the publicly known parameters and navigate this chart through an apparently irregular path. At last, they came up on two curves sharing some values that are utilized as shared secret values. So far, the most famous and accessible algorithm for a quantum computer can sort out this problem with a time complexity of $\blacksquare\blacksquare(p = 6)$. This work investigates on SIDH, which was an algorithm proposed by Jao and De Feo on 2011[1]. Similarly, the Intel Sandy Bridge and Haswell processors utilizing the projective coordinates was implemented and published by Costello et al.

[4]. Recently, Hernandez et al. [5] increases the results with its speed factor. Also in 2016, Koziel et al. [6] it is introduced that advanced execution of affine coordinates on 32-bit Cortex- A8 and Cortex-A15 architecture using the NEON SIMD architecture extension. But, it stays vague how SIDH performs on microcontrollers which have less computational power and need devoted SIMD accelerators.

* Corresponding author. *E-mail address:* ctpoomagall@gmail.com.

Our contributions. It is surveyed that the first hardware implementation of the SIDH isogeny algorithm is quantum-resistance with optimized design for speed and efficient multiplier unit with parallelization of arithmetic in prime field. For the proposed work, it is implemented that the optimized field arithmetic functions i.e. modular multiplication, reduction and other modular arithmetic units in FPGA hardware and implemented the hardware architecture for newly designed EC-SIDH protocol. And the results of the proposed EC-SIDH algorithm is implemented in FPGA hardware is compared to the generic implementation of the isogeny Diffie-Hellman protocol in the previous works, this reduced the certainty of the vulnerability in the isogeny protocol. However, it is shown that the proposed key exchange algorithm applies more security than previous works with the requirement of corresponding increase in hardware for most real-life applications.

Outline. The preliminaries with a brief description of the SIDH protocol and the basics of isogeny are described in section 2. In Section 3, it is presented that how the modified SIDH algorithm is designed and verified in software. In Section 4, it is presented that the implementations of proposed protocol in the virtex-6 FPGA hardware board with the special emphasis on the prime field operations. Section 5 summarizes the performance results of the proposed protocol and previous works with public key validation. Finally, it is concluded in Section 6.

2. Preliminaries

Some of the basis for isogeny-based cryptography is that the isogeny based Diffie - Hellman key exchange was first published by Rostovt-sev and Stolbunov in [7]. This was originally defined over ordinary elliptic curves and was thought to feature quantum resistance. However, Childs, Jao, and Soukharev [8] predicted that a quantum algorithm to compute isogenies between ordinary graphs requires sub-exponential time. Later, David Jao, Luca De Feo, and Jerome Plut captured the isogeny-based key exchange designed over supersingular elliptic curves in [9], which gives no known quantum attack.

2.1 Isogeny Theory. An elliptic curve over a finite field Fp can be written in its Short Weierstrass form as: $\text{\diamond\diamond}(\text{\diamond\diamond}, \text{\diamond\diamond})/Fp : y\ 2 = \text{\diamond\diamond}3 + a\text{\diamond\diamond} + \text{\diamond\diamond}$, where $\text{\diamond\diamond}, b \in Fp$. An elliptic curve is composed of all points (\diamond\diamond, \diamond\diamond) that satisfy the above equation as well as the point at infinity. For a scalar k and a points P, Q ∈ E on the elliptic curve, the scalar point multiplication is Q = kP is

the basis of the elliptic curve Diffie-Hellman, and which forms the abelian group over point addition[10]. A scalar point multiplication is defined by point addition and doubling formulas, and the efficient performance of series of point doublings and additions. However, instead of performing affine point addition and affine point doubling for a scalar point multiplication, we define projective formulas over projective coordinates $(X : Y : ⬛⬛)$ such that $x = ⬛⬛/Z$ and $y = Y/⬛⬛$. With this representation from [11] - [13], only a single inversion is performed at the end of the scalar point multiplication. An isogeny over a finite field Fp is defined as $: E \rightarrow ⬛⬛0$ as a non-constant rational map defined over Fp such that φ satisfies group homomorphism from $⬛⬛(Fp)\ t⬛⬛ \quad ⬛⬛0\ (Fp)$ [14]. An isogeny is defined as a mapping from one elliptic curve class to another that has the same point at infinity then these two curves are isogenous. Specifically, for two elliptic curves to be isogenous over a finite field, they must have the same number of points [15]. Vélu's formulas [3] defines that over a kernel, κ, the isogenies can be computed such that $: E \rightarrow ⬛⬛/h\kappa i$ with the j-invariant is an identifier for an elliptic curve isomorphism class. There are various complex features that are similar to the curves within this isomorphism class.

There is morphism between one isomorphism class and another and hence it is called as an isogeny. Finally, each isomorphism class represented by a node and the isogeny represented by edges are creating a graph of all isogenies.

2.2 Finite Field Arithmetic. For the prime ⬛⬛, a field Fp is used to represent all arithmetic in the supersingular isogeny based cryptosystems. A finite field Fp consists of field elements that are closed under additive and multiplicative axioms, which can be thought of as modular addition and multiplication. We can extend the base finite field Fp to a quadratic extension field by defining the extension field over an irreducible polynomial. Notably, we choose the irreducible polynomial ⬛⬛2 + 1, which is irreducible since $i = \sqrt{-1}$ does not exist in our base field. Thus, we can represent an element $A \in p$ contains the elements ⬛⬛1 , ⬛⬛2 $\in Fp$ in the form $A =$ ⬛⬛0 + ⬛⬛⬛⬛1 . ⬛⬛1 is considered to be the most significant Fp element in the quadratic extension field representation.

SIDH Key Exchange. In the supersingular isogeny key exchange protocol, Alice and Bob compliance on a secret elliptic curve class by separately taking seemingly gone through the respective isogeny graphs. SIDH is made out of a two point

multiplication to form a secret kernel point towards a large-degree isogeny over that kernel point. The difficulty to determine a connection between two distant supersingular isomorphism classes provides security for this protocol. For SIDH implementation, two nodes decide on an isogeny value of a smooth prime p a b of values ►► . ►► . ± 1, where ►►A and ►►B are small primes, a and b are positive integers and f is a small cofactor to make the numbers prime. Over this starting supersingular curve E0 , the two nodes pick the bases {PA , QA } and {PB, QB } which generate the torsion groups E0 [eA , A] and E0 [eB, B], respectively, such that ⟨PA, QA ⟩ = E0 [eA , A] and ⟨PB, QB ⟩ = E0 [eB, B].

The SIDH protocol works as follows. The two participating nodes each perform a double-point multiplication with two selected private keys that span Z/ ►►a Z and Z/ ►►b Z, respectively towards the generation of a secret kernel point that gives a higher degree isogeny. Alice's secret kernel is RA1 = mA PA + nA PA and Bob's secret kernel is RB1 = mBPB + nBPB, where {mA , nA } are Alice's secret keys and {mB, nB } are Bob's secret keys. Next, Alice and Bob use that secret kernel to perform their own secret isogeny walk by computing a large-degree isogeny as in [16] – [20]. Alice computes ►►A : ►►0 → EA = ►►0 /⟨►►A1 ⟩ and Bob computes►►B : ►►0 → ►►B = ►►0 /⟨►►B1 ⟩. For the first round, the opposite party's basis points are also pushed through the isogeny mapping. Alice computes ►►A (PB), ►►A (►►B) with her hidden isogeny and Bob computes ►►B (PA), ►►B (►►A) with his hidden isogeny. At the end of the first round, Alice transmits {E , ►►A (PB), ►►A (►►B)} and Bob transmits {►►B , ►►B (PA), ►►B (►►A)}.

With the exchanged information, Alice and Bob again perform their hidden isogeny walk, but this time with the transmitted public keys as the starting point. Similar to the first round, Alice and Bob calculate a secret kernel with their secret key. Alice's secret kernel for the second round is ►►A2 = ►►A ►►B (PA) + ►►A ►►B (PA) and Bob's secret kernel is ►►B2 = mB ►►A (PB) + ►►B ►►A (PB). With this secret kernel, Alice and Bob compute a large-degree isogeny which is the same isogeny walk as the first round. Alice computes ►►BA : ►►B → ►►BA = ►►B /⟨►►A2 ⟩ and Bob computes ►►AB : EA → EAB = EA /⟨►►B2 ⟩. Essentially, curves ►►BA and EAB were obtained by performing the same two isogeny computations, but in other way. In effect, the two participating nodes serve isomorphic curves that can be used as a shared secret. We illustrate the SIDH key exchange in terms of the isogeny computation in Figure 1.

3. Proposed SIDH Key Exchange Algorithm for Isogeny based Cryptosystem

This work has the major design consideration to implement the Elliptic curve based SIDH key exchange mechanism in hardware with slightly modified SIDH algorithm. To give an initial look into computing large isogenies in hardware, it is chosen to stick with 512-bit primes for the keys, which feature 85-bit quantum security and 128-bit classical security. The computational part of the public key in round 1 computation is slightly modified to increase the security in mathematical hardness of the algorithm. So that it is even more difficult to reverse computation of the shared secret and computation of secret keys by a third party intruder. The proposed method of modified SIDH Key Exchange Algorithm SIDH algorithm presented in this paper is shown below in Fig.1.

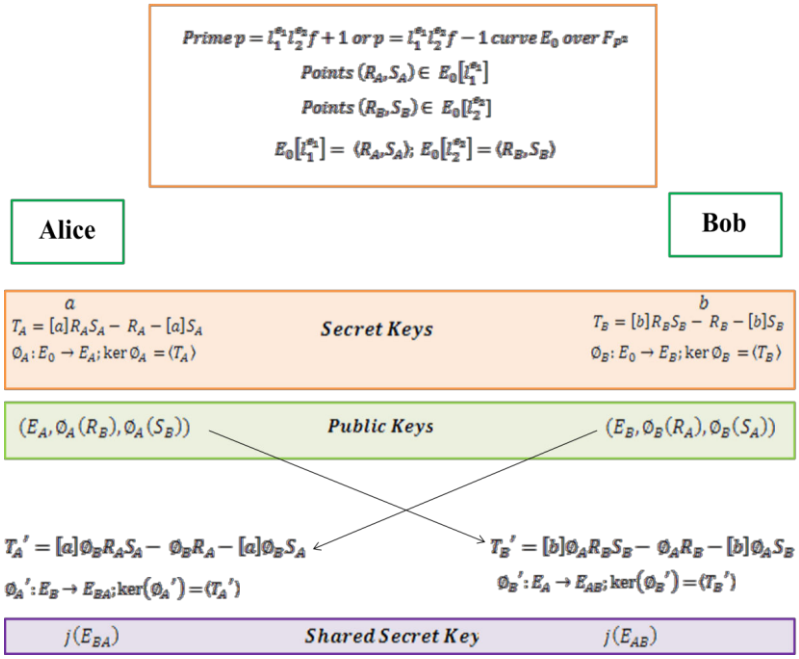


Fig. 1. Illustration of proposed EC-SIDH Algorithm

In this protocol, the Prime ⋯ initial points $\langle \ldots A, SA \rangle$, $\langle \ldots B, \ldots B \rangle$ and the other curve parameters of the corresponding curve ⋯0 [⋯⋯1], ⋯0 [⋯⋯2] are agreed between Alice and Bob. The secret keys of the individuals are selected as and b for Alice and Bob respectively. Similarly, the value of TA, ⋯A and TB, ⋯B are calculated correspondingly to form the public keys as {⋯B, ⋯B (⋯A), ⋯B (SA)} and {EA, ⋯A (⋯B), ⋯A (⋯B)} for Alice and Bob respectively as shown in Fig.1. Then the public keys are shared mutually and shared secret keys are calculated individually. The calculated secret key is then used for future encryption and

decryption processes. Since the proposed protocol is based on the field arithmetic in ECC, it needs to be designed at architecture level for Field arithmetic, Group operation, Isogeny Group operation and completely as Post Quantum Cryptosystem.

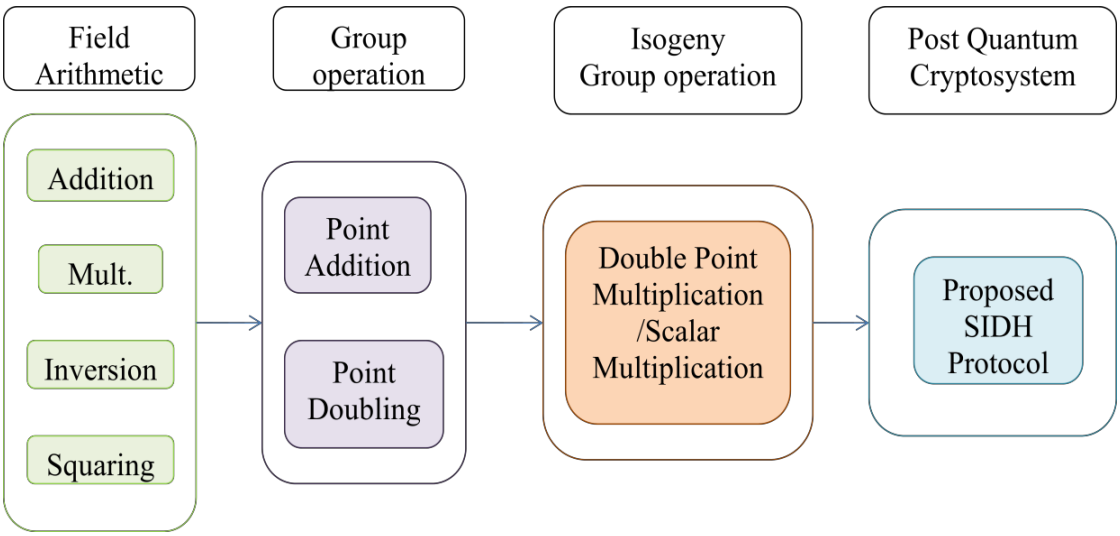


Fig.2. The high level design of basic operations behind the proposed Isogeny protocol.

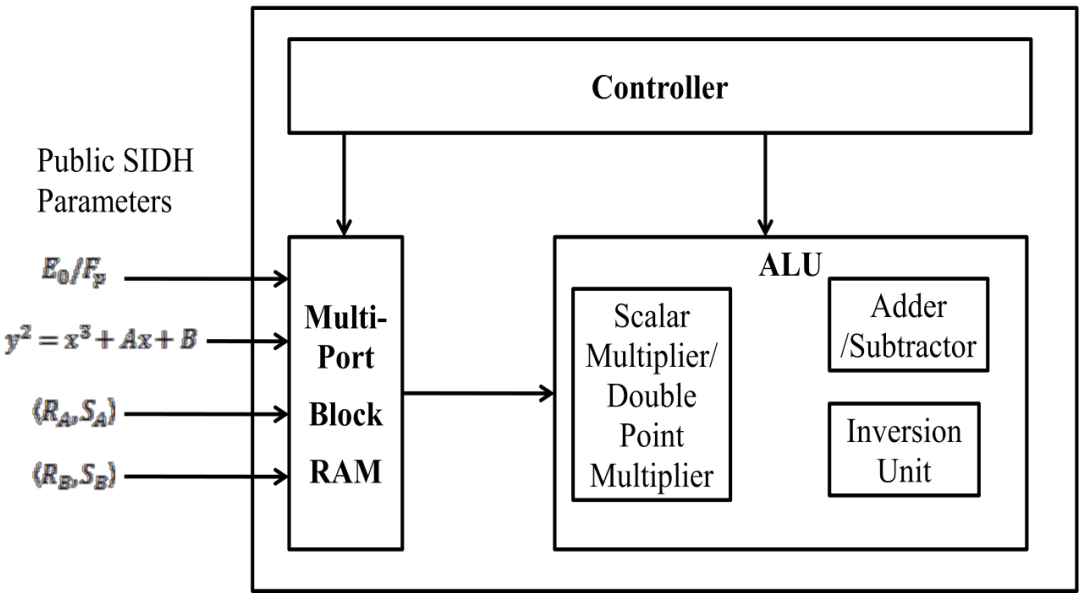


Fig.3. Proposed Architecture for slightly modified EC-SIDH Cryptosystem

The high level design of the proposed isogeny core is depicted in Fig. 2. For the proposed isogeny SIDH protocol, an isogeny group operation of double point multiplication is needed to execute for each round of the calculation in the algorithm. This isogeny group operation in turn needs the Point addition ad point doubling calculations which are supported by the Field arithmetic operation viz, Addition, Multiplication, Inversion, Squaring in any specified and NIST standard Elliptic curve.

And a core has been designed for the implementation of the above depicted high level design of the proposed SIDH algorithm which is shown in Fig.3. This core features an adder/ Subtractor unit, multiplier unit, inversion unit, RAM file for registers, and a controller for controlling every units of the isogeny cryptosystem. The RAM file contains constants for the parameters of the protocol, intermediate values within the protocol, and intermediate values for field arithmetic computations. There are more intermediate values necessary for higher key sizes as the graph traversal through this isogeny is higher, but 256 values is slightly more than enough, which allows more flexibility and optimization with routines. The size of the controller unit mainly depends on the size of the multiplier unit.

4. Implementations in FPGA

In this section, it is presented that the results of proposed supersingular isogeny algorithm on a Virtex-6 FPGA Evaluation Kit for a proof of the concept. Hence the implementation of the proposed isogeny algorithm is depicted as follows. The SIDH core was synthesized with Xilinx ISE to a Xilinx Virtex-6 xc6vlx240t-1ff1156 device. The hardware implementation of the proposed work is shown in Fig 4.

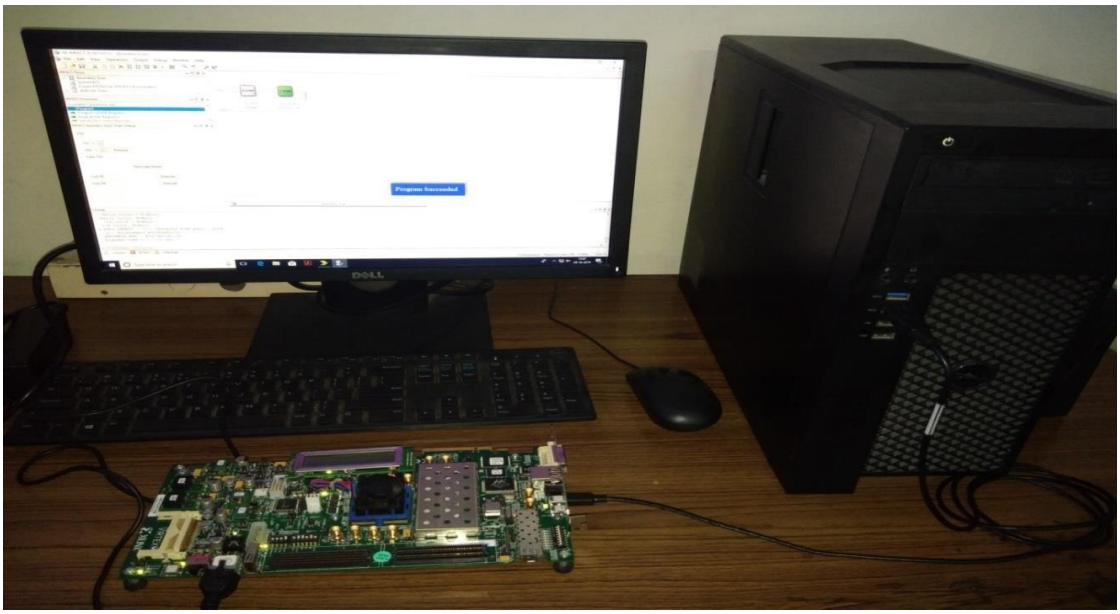


Fig.4. The hardware implementation of the Proposed slightly modified EC-SIDH Cryptosystem

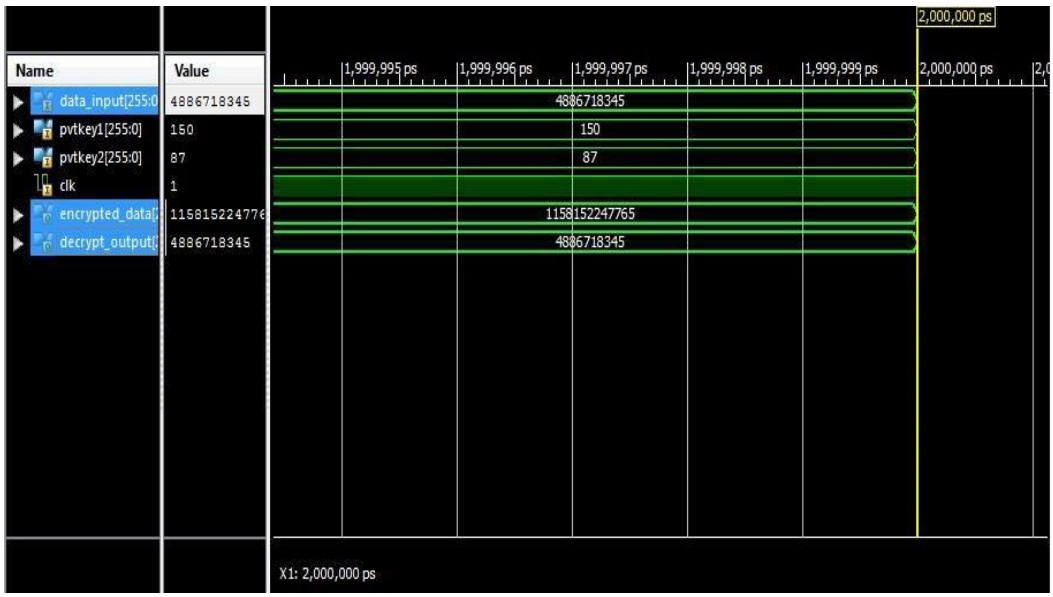


Fig.5. The results after implementation of the Proposed slightly modified EC-SIDH Cryptosystem

All the results are obtained after place-and-route. The output for hardware implementation of the proposed slightly modified EC-SIDH Cryptosystem is shown in Fig 5.

5. Analysis and Comparisons

The area and security results of our SIDH core are shown in the below table 1. The results for different Isogeny algorithm with different devices are presented and it replicate the area requirement for each algorithm with the corresponding devices. And it is compared with our architecture.

Table. 1: Comparison of Proposed work with other security schemes.

Work	**Scheme**	**Platform**	**Quantum security**	**Area**			
				# FFs	**# LUTs**	**# Slices**	**# BRAMs**
This Work	Modified SIDH	Vitrtex – 6	128	34,568	27,640	11,517	77
Wang et al[13]	McBits	Ultrascale +	128	104,684	112,845	40,279	375
Koziel et al[19]	Parallelizing SIDH	Vitrtex – 7	124	48,688	34,742	14,447	59

The proposed scheme in this paper is a modified SIDH protocol implemented in the Xilinx

virtex-6 Evaluation board for the quantum security of 128-bits key size. Likewise, there are some other SIDH protocol proposed by Koziel et al of parallelizing SIDH in 2018 and it was implemented in virtex-7 platform, wang et al proposed McBits protocol designed in Ultrascale+.

The area results of these protocol are compared with our modified SIDH scheme so that it is clearly analyzed that for a 128-bit quantum security, the slices required for computation of the proposed algorithm is merely reduced than other protocols depicted here and showing that the resource requirement is also reduced correspondingly. In the same way, the security of the proposed scheme is also improved than existing scheme since it consists of more mathematical hardness in the calculation of the public key in the round 1 calculation. This shows that the proposed method of modified SIDH algorithm and the corresponding design in FPGA hardware is slightly better than existing methods.

6. Conclusions

There are many different proposals for cryptographic schemes that could one day replace existing schemes that rely upon the Discrete Logarithm or Integer Factorization Problems. Supersingular Isogeny Diffie Hellman (SIDH) is a front runner in this competition. Isogenies using supersingular curves already have achieved in the development and implementation of secure hash functions. Additionally, SIDH can be built on top of existing ECC primitives and when compared to other post-quantum schemes, SIDHs key sizes clearly set it apart (which is of course consistent with the lower key sizes of ECC schemes in general). Lastly, the fact that SIDH does not claims any filed patents makes it very attractive when compared to pre-quantum schemes such as RSA. Several implementations of it already exist for both embedded[6] and non-embedded[7] applications For these reasons, SIDH deserves serious consideration for main usage in a post quantum cryptography.

7. Acknowledgments

The part of the work was supported by Indian Space Research Organization(ISRO) under the RESPOND scheme of Projects with the grants No: ISRO/RES/3/750/18-19.

8. References

[1] Jao, D., Feo, L.D.: Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies. In: Post-Quantum Cryptography - 4th International Workshop, PQCrypto 2011, Taipei, Taiwan, November 29 - December 2, 2011. Proceedings. pp. 19-34 (2011).

[2] Costello, C., Jao, D., Longa, P., Naehrig, M., Renes, J., Urbanik, D.: Efficient compression of SIDH public keys. In: Advances in Cryptology – EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30 - May 4, 2017, Proceedings, Part I. pp. 679-706 (2017).

[3] Costello, C., Longa, P., Naehrig, M.: Efficient algorithms for supersingular isogeny diffie-hellman. In: Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part I. pp. 572-601 (2016).

[4] Faz-Hernandez, A., Lopez, J., Ochoa-Jimenez, E., Rodriguez-Henriquez, F.: A Faster Software Implementation of the Supersingular Isogeny Diffie-Hellman Key Exchange Protocol. Cryptology ePrint Archive, Report 2017/1015 (2017).

[5] Koziel, B., Jalali, A., Azarderakhsh, R., Jao, D., Kermani, M.M.: NEON-SIDH: efficient implementation of supersingular isogeny diffie-hellman key exchange protocol on ARM. In: Cryptology and Network Security - 15th International Conference, CANS 2016, Milan, Italy, November 14-16, 2016, Proceedings. pp. 88-103 (2016).

[6] Rostovtsev, A., Stolbunov, A.: Public-Key Cryptosystem Based on Isogenies. IACR Cryptology ePrint Archive 2006, 145 (2006).

[7] Childs, A., and Jao, D., and Soukharev, V.: Constructing Elliptic Curve Isogenies in Quantum Subexponential Time (2010).

[8] De Feo, L., Jao, D., and Plut, J.: Towards Quantum-Resistant Cryptosystems from Supersingular Elliptic Curve Isogenies. Journal of Mathematical Cryptology 8(3), 209-247 (Sep. 2014).

[9] Silverman, J.H.: The Arithmetic of Elliptic Curves. Volume 106 of GTM. Springer, New York (1992) .

[10] Azarderakhsh, Fishbein, Jao, "Efficient Implementations of A Quantum-Resistant Key-Exchange Protocol on Embedded systems", June. 2014.

[11] L. De Feo, "Quantum-Resistant Cryptosystems from Supersingular Elliptic Curve Isogenies", Nov. 2015. [12] D. Jao and V. Soukharev, "Isogeny-based quantum-resistant undeniable signatures," in Proc. Post-Quantum Cryptography, 2014, pp.160–179.

[13] X. Sun, H. Tian, and Y. Wang, "Toward quantum-resistant strong designated verifier signature from isogenies," in Proc. Int. Conf. Intelligent Networking and Collaborative Systems, 2012, pp. 292–296.

[14] J. H. Silverman, The Arithmetic of Elliptic Curves, New York: Springer, 1992, vol. 106.

[15] J. Tate, "Endomorphisms of abelian varieties over finite fields,"Inventiones Mathematicae, vol. 2, 1966, pp. 134–144.

[16] J. Vélu, "Isogénies entre courbes elliptiques," in Comptes Rendus de l'Académie des Sciences Paris Séries A-B, 1971, pp. A238–A241.

[17] C. Petit, "Faster algorithms for isogeny problems using torsion point images," in Proc. Advances in Cryptology - ASIACRYPT, 2017, pp. 330–353.

[18] P. L. Montgomery, "Speeding the Pollard and elliptic curve methods of factorization," J. Mathematics of computation, pp. 243–264, 1987.

[19] B. Koziel, R. Azarderakhsh, M. Mozaffari Kermani, and D. Jao, "Post-quantum cryptography on FPGA based on isogenies on elliptic curves," IEEE Trans. Circuits and Systems I: Regular Papers, vol. 64, no. 1, pp. 86–99, Jan. 2017.

[20] B. Koziel, R. Azarderakhsh, and M. Mozaffari Kermani, "Fast hardware architectures for supersingular isogeny Diffie-Hellman key exchange on FPGA," in Proc. Indocrypt, 2016, pp. 191–206.

[21] B. Koziel, R. Azarderakhsh, and M. Mozaffari Kermani, "A High performance and Scalable Hardware Architecture for Isogeny-Based Cryptography", in IEEE Transactions on Computers, vol. 674, no. 11, pp. 1594-1609, Jan. 2018.

Wireless Port Design for Sensor and Actuator Networks

Sanskar Biswal, Yepuganti Karuna*

School of Electronics Engineering (SENSE), Vellore Institute of Technology (VIT), Vellore

School of Electronics Engineering (SENSE), Vellore Institute of Technology (VIT), Vellore

Abstract

Internet of Things has heralded the era of connected devices on an unprecedented scale. Traditionally these devices need a custom hardware and firmware to interact with the environment/user. Additionally, there is a large variation in the format of the data. Many of the present designs do a part of the data processing locally, which requires a higher processing capacity at the target device, hereafter referred to as *Node*. The current paper attempts to resolve these issues by suggesting a new hardware design and firmware, so as to reduce the hardware and data format variations. The designed module will have the ability to communicate to the Parent Node wirelessly, using a hexadecimal addressing scheme. Changing the addressing scheme table allows the user to update how the devices are connected (their topologies). This gives the users/system designers to setup custom groups of sensor/actuator networked meshes. Additionally, the dynamic addressing paradigm ensures that the configuration of the established networks can be updated without making changes to the hardware configurations. The fully configurable design in terms of device connectivity and network configurations are the primary novelty suggested by the paper. An Embedded System was designed and tested, which uses an ATtiny85 microcontroller and nRF24L01 module for communication. A simple table of variables stored in the EEPROM of the microcontroller, can fully be modified to suit the type of operation needed by the device. The values of this table may be modified via RF communication from the master-node, thus making the same hardware design feasible and adaptable for both sensor networks, actuator network and even hybrid networks.

Keywords: Node, Parent-Node, Embedded Systems, ATtiny85, RF Communication, nRF24L01, Wireless Sensor Networks, Wireless Actuator Network, Hybrid Wireless Networks, Self-Adjusting Topologies, EEPROM Programming Tables, SPI Protocol, Power Efficient Design

* Corresponding author.
E-mail address: karun@vit.ac.in

1. Introduction

Sensor networks often have a dedicated hardware that are custom built for specific sensors. Thus, any major operational change often involves the overwriting of the device/node firmware and even perhaps substitution of the hardware with another. This process not only adds to the delay in deployment of changes, but also considerably increases the cost of changes. This problem is compounded by several magnitudes with increasing scale of deployment. It is with this problem in mind that this project was conceptualized.

The advent of IoT has made it very easy to perform complex scale processing and decision -making on a server. This is a novel solution for constraints due to local factors, however this also inevitably introduces latency into the system. In many real-time systems, reaction to changing conditions may be a critical factor, not wholly dependent on any decision from a cloud-based master code. In such scenario, rapid inputs from the environment must result in immediate reaction from the system, with minimum latency. Thus, organizing a topology where the sensor sends data directly to the actuator/network of actuators is a potent solution. There is however the chance that the configuration of the sensor or the network needs to be modified. Such scenarios may again lead to excessive delays if there is a need to develop custom hardware or firmware for each use-case.

The proposed solution will aim to mitigate the above-mentioned challenges, by designing a compact unit which has the provision to connect to a variety of sensor and actuator configurations, but also provide a method of wirelessly communication amongst the relevant nodes. A major advantage of wireless networks is that they do not have the need of regular maintenance. To ensure that this holds true for the proposed solution, the design will also need to be as power-efficient as possible.

Thus, the objectives of the project may be broadly listed as the following:

- A compact system, which allows for digital and analog I/O
- Wireless communication module, with configurable addresses to facilitate dynamic topologies
- Power efficient design of the system, to allow the module to function on a battery for the maximum possible duration

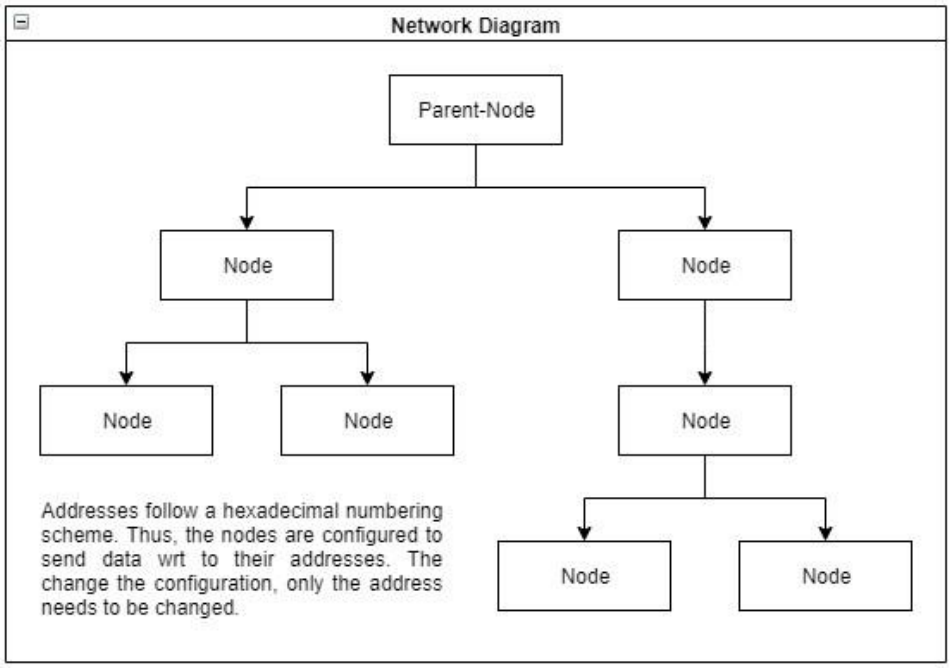


Fig. 1. Networked Node Topology

1.1. Microcontroller and RF Module Selection

The microcontroller for the designated node opted in this project is the ATtiny85. ATtiny85 is a very compact and power-efficient chip which belongs to the Atmel family of microcontrollers. It has very good documentation [3] and a wide base of supporters in the developer community as well. The primary reason for the selection of the ATtiny85 is due to its power-efficiency. In the datasheet [1], it is stated that the ATtiny85 consumes 300µA in active mode and 0.1µA in sleep-mode. This is an ideal feature given the requirements.

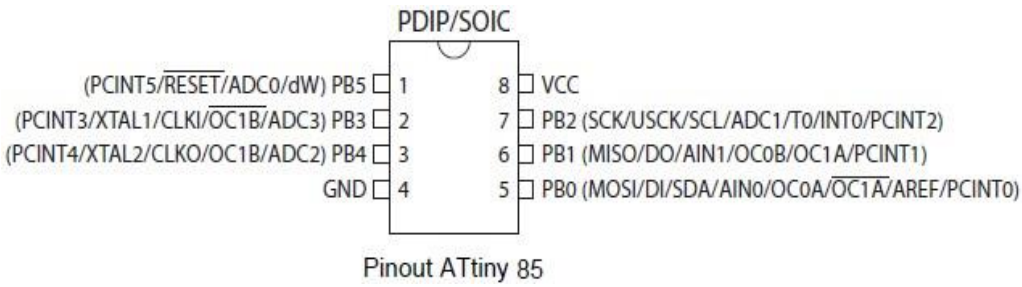


Fig. 2. ATtiny85 Pin Configuration and Purpose

In the figure above, it may be seen that ATtiny85 has a total of 8-pins, out of which, two are power pins (VCC/GND). This leaves us with 6 pins for I/O. Amongst these, pin1 functions as the reset pin and is hence reserved for that purpose. Pin 5,6,7

respectively are the MOSI, MISO and SCK pins. These pins are needed to ensure serial communication with the RF module and hence cannot be used for I/O. Thus, there are only 2 pins remaining, pin3 and pin2 for I/O. A solution to the restricting number of pins maybe the use of a microcontroller with a larger number of I/O pins. However, the aim of the project is to design the most energy and space efficient module possible. Thus, the ATtiny85 module is still the processor of choice. The method of mitigating the small number of I/O pins is discussed in the Design section (Section 2.2) of the paper.

The nRF24L01 module is the chosen for the purpose of RF communication. RF communication is used for a variety of reasons, including but not limited to its range of communication extending up to 100m and low power consumption. nRF24L01 module in specific utilizes the 2.4GHz band for communication and used 1MHz bandwidth for each channel. This setup allows up to 125 simultaneous channels to be utilized in a limited space. The 2.4GHz frequency is very permeable through physical boundaries, and thus do not need line-of-sight setup for operational efficiency. The nRF24L01 module consumes an average power of 13mA in its active mode of operation [2], which is the primary operating mode. Thus, the remainder of power consumption will be either from the actuator or the sensor.

It can thus be assumed that the net power consumption of the fully operational system, including the sensor units will be around 20mA. Actuators will mostly require external power source for operations. The reason for this may be attributed to the working voltages for the microcontroller and the RF module, both of which are 3V. Although they can work at higher voltages, the RF module has a tendency of increased failure. Thus, to prolong the lifetime of the module, 3V DC will be the preferred method to power the module. In all further references to the module will be classified broadly as Node. This is because the Node is in principle, a component of a tree like topology.

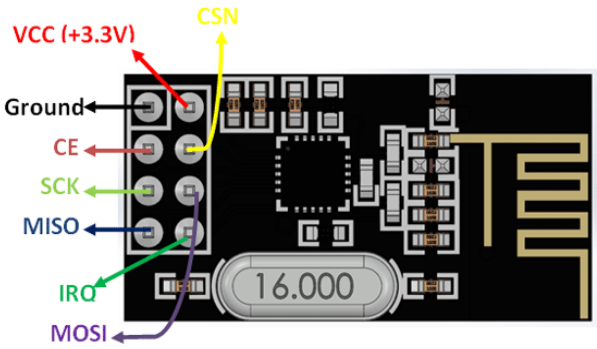


Fig. 3. nRF24L01 pin configuration

The datasheet of the nRF24L01 [3] module specifies that it uses SPI communication protocol for the communication of data to and from the module. The MOSI of RF module connects to the MISO of the microcontroller and vice-versa. The SCK pin connects to the SCK pin of the microcontroller. CE and CSN are Chip Enable and Chip-Select-Not respectively. They can both be set too high to always have the module turned on. The IRQ is the interrupt pin which does not need any form of connection in the present design.

1.2. Wireless Topologies

The present paper aims to design a module that can operate in the following topologies:

- *Sensor Networks*
- *Actuator Networks*
- *Hybrid Networks*

Sensor networks, as the name suggests comprises fully of sensors. The type of sensors may vary, but the nodes only send collected data, with timestamps. The only limitation is that the number of Input ports can be a maximum of two, accounting for the number of I/O ports on the microcontroller. This however, from the study in [4] may not necessarily be limitations as there are very few sensors that have a higher number of input ports. Cameras and combined sensors are the only sensors that may exceed the port capacity. Most other sensors will work well the current design.

Actuator Networks comprise mostly of output modules like motors and speakers. Mechanical and Audio are the two outputs that can be handled by the ATtiny85 without any issues. Display output is not possible due to the port limitation. However, most tasks at home/industrial environments can be easily handled with the two I/O pins. The two pins may also be used to output control signals to the environment, the frequency and the magnitude of which may be set by either the master-node or may be a preset value.

Hybrid networks comprise of both sensor and actuators at the node. These nodes need not be connected to the parent, but instead to a sensor in its operational requirement, thus forming a closed-loop feedback system as proposed in [5].

A hybrid network is more closely related to a real-world operating scenario and is better equipped to perform at more optimal levels when compared to homogeneous element networks. This observation is very commonly seen where the action of an

actuator is closely dependent on the inputs from the direct environment. Examples may include self-balancing bots and automated door opening system. All are essentially a hybrid system. A network of moisture sensors may act as the input for opening of floodgates in a wa ter reservoir, which may again be classified as hybrid network.

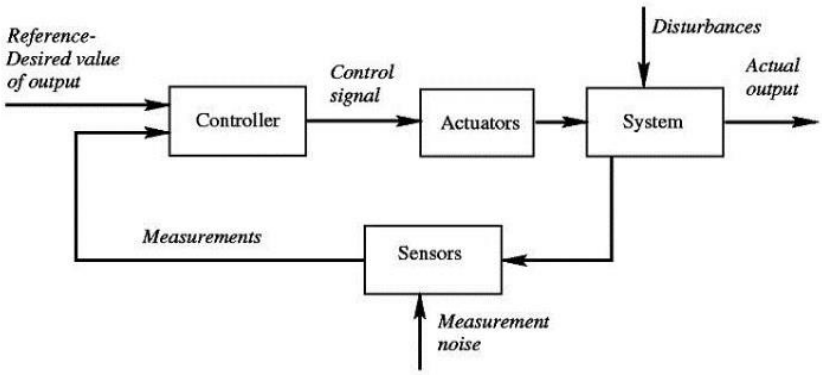


Fig. 4. Sensor Actuator Feedback System Model

2. Embedded System Design Process

The design process depends mainly on the objectives established at the onset of the paper. The designed Node need not have high processing power as its main purpose is to provide an I/O port for sensors or actuators and a communication module to establish a network of connected Nodes.

It has been already established earlier that the ATtiny85 will have only 2 pins available for I/O after accounting for SPI connectivity. The most optimal method for the use of two pins is have them for I/O when needed.

Thus, any sensor/actuator connected to the Node will have connections directly to I/O pins from their O/I pin. Power to the connected module will be provided from the power source of the Node. This power module is limited to 3V, thus motors will normally require external power supply, to avoid a faster drainage of the battery capacity of the Node. In case of motors, the Node may be used as a switch, to control the rate of operation, thus increasing the working life of the node by a considerable margin.

2.1. Battery and Power Systems

The developed power system needs to be capable of providing power to the Node at least one year to be considered a long-term operational module. As stated in Section

1.1, the expected power consumption of the system will be at the maximum threshold of 20 mA/s. Thus, the power consumption of the node is 20mW.

*Battery-life = Battery-capacity / (Load-current * 0.7)*

- *Capacity = Battery-life * Load * 0.7*

When we extend the calculation for one year (8760-hrs), the calculated capacity is

$$8760hrs * 20mA * 0.7 = 122640\ mAh$$

The result indicates that there is a need of a battery back with a capacity of 1.25k-mAh to operate a single node for one year at the transmission interval of 1s. A battery of such large capacity is not just prohibitively expensive, it will be huge, given the current state of battery technology. The only way to mitigate this issue was either to power the Node directly from a power source or to reduce the power consumption of the device. It is a requirement to design the Node to be battery operated. Thus, net power consumption needed to be reduced. To reduce net power consumption, we suggest two potential methods.

The first method is to reduce the frequency at which data is transmitted from the Node. Under the current scenario, where we transmit data once every one second, the annual consumption is 1.25k-mAh. If we reduce the frequency to once every 10s, then the net power consumption for one year drops to 12000mAh. If we further reduce the frequency of transmission to one minute, then the annual power consumption drops to 2000mAh, which is both feasible and cost effective to be used in real-world systems. The AA-cells provide a capacity of 1500mAh. A series connection will yield a total capacity of 3000mAh, which will ensure that the Node can operate independently for well over a year. At this point, the question arises regarding the feasibility of using a system that can send data only once every minute. The case is analyzed from two perspectives, drawing inspiration from results and observations drawn from [6].

The second method is to have a threshold value set in the EEPROM of the Node. The sensor input is compared to the threshold value (can be a limit or a range), and then only if there is a variant scenario, the data is transmitted. Thus, the real power consumption is from the sensor unit, and the transmission rate is drastically reduced. This will further extend the operational lifetime of the Node on a single battery pack of 3000mAh. The prerequisite to operate this method is that the users have a confidence in the limits set for transmission.

In home/domestic scenarios, most of the sensors do not interact in an environment that changes rapidly, thus making almost no impact on the reduced time of access. For

system, that require full up-time, they will need to be connected to an external source. Nodes that connect to actuators can be powered from the power source of the actuators. To account for the variety of input voltages, a safety circuit needs to be built to reduce chances of the Node failing due to power fluctuations.

In industrial scenarios where there is the need for more frequent control, the networked mode of operating may be used. Since industries often operate in a large scale, it reasonable to assume that there will be several sensors/actuators connected to the master. These may operate at a cyclic period in such a fashion that there is always of set of Nodes active at each second. The scale of operation will offset the limited uptime of the Node.

This solution may be unconventional but has the potential to work in several use-cases, some of which will be tested in simulations. Thus, we assume that the required capacity for one-year operation is 2000mAh. Via the comparison of various cell costs and capacities, AA-cells are the best option. They have a potential difference of 1.5V and a rated capacity of 1500mAh. Batteries in Series connection will add-up in capacity, where as in Parallel, they add up in voltage. Thus, a combination of two AA-cell in series in parallel, will yield a voltage of 3V at 3000mAh capacity which sufficiently satisfy the required objective. Thus, each module may be powered either from an external power supply or from 4 AA-cell combination battery pack.

2.2. I/O and RF Module Interfacing

The I/O is supported from two pins on the ATtiny85. A common design in most system is to have sensors ports have either a 4 or 3 pin setup. Two of these pins are the VCC/GND, and they are usually at the extreme ends of the sensor module. The control/data pins are at the centred on the sensor module. This has been utilized to make the connector port accessible to a wide variety of sensors. The GND of the sensor connects directly to the GND of the ATtiny85. It is followed by the 2-pins for I/O and then the Reset-Pin.

If the sensor is a 3-pin module, Pin-2 of the ATtin85 is used as the VCC but having that pin set to HIGH (3V). Thus, Pin-3 will be connected to the I/O of the sensor.

If the sensor is a 4-pin module, Pin-3 and Pin-4 connect to the I/O, whereas the VCC of the sensor connects to the VCC from the power-source. We thus will need two separate ports, one each for 3- and 4-pin sensor modules. Output devices are easier to interface as most have only one pin for control/output.

The nRF24L01 as mentioned in [7] has a much simpler connection to the ATtiny85. The MOSI -> MISO, SCK -> SCK connection scheme is very straight-forward. The main design needs to be worked out in the activation of the module as it is the most power consuming unit on the node. The CE pin of the nRF24L01 module can be set to LOW, to turn-off the module. However as mentioned in [8], the power consumption of the RF module is 900µA in standby mode and only during transmission or reception does it consume 13mA of power. It thus is of almost no consequence if the RF module is always on as it will spend most of its time in standby mode of operation. The following table is taken from [8] indicating the power consumed by the nRF24L01 module in various modes of operation.

Parameter (condition)	Notes	Min.	Typ.	Max.	Units
Idle modes					
Supply current in power down			900		nA
Supply current in standby-I mode	a		26		µA
Supply current in standby-II mode			320		µA
Average current during 1.5ms crystal oscillator startup			400		µA
Transmit					
Supply current @ 0dBm output power	b		11.3		mA
Supply current @ -6dBm output power	b		9.0		mA
Supply current @ -12dBm output power	b		7.5		mA
Supply current @ -18dBm output power	b		7.0		mA
Average Supply current @ -6dBm output power, ShockBurst™	c		0.12		mA
Average current during TX settling	d		8.0		mA
Receive					
Supply current 2Mbps			13.5		mA
Supply current 1Mbps			13.1		mA
Supply current 250kbps			12.6		mA
Average current during RX settling	e		8.9		mA

Table. 1. Power Consumption in Different Modes of Operations in nRF24L01

Thus, the CE and CSN pins are connected to the power source of the Node. The IRQ pin is left unconnected to any unit on the module. A simple voltage control circuit is connected to the CE/CSN pins of the RF module to prevent any excess voltage from affecting the operating of the module. The operating voltage range is up to 3.6V and although the nRF24L01 module supports up to 5V, this circuit will help prevent any unintended module failure due to over-voltage.

3. Schematic and PCB Design

The schematic for the design and development made use of the Autodesk Eagle software. These schematics may be accessed at the link provided in [9].

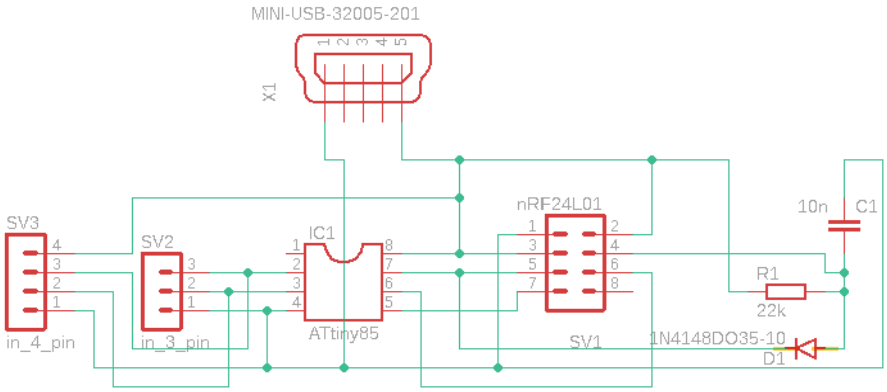


Fig. 5. Layout Schematic

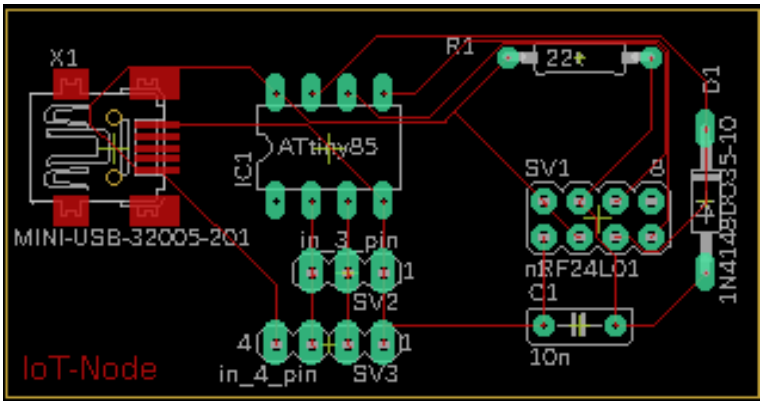


Fig. 6. PCB Layout (Single Layer)

4. Firmware Design

The firmware/code on the module is the most essential and critical component. Since the module has a pre- set hardware, the variety of sensors and actuators will rely on efficient manipulation offered by the firmware for interfacing.

There are 4 major operations that need to be handled in any update:

1. Address Update
2. Pin Configuration Update
3. Transmit Condition
4. Transmit Time Update

Address update is used to store the address of the node. Each Node will have its address on itself, and the parent-hub will store all addresses at a local table. Changing the address of the Node provides the user with the flexibility to modify the topology of the network.

Pin configuration, for this specific use-case needs to set if the pin performs an input or output operation and if the I/O is digital or analogue in nature. There are 2 I/O pins free in the module and thus, 4 variables are used to freely program this configuration set. A set of simple flags (0/1) are used to indicate the configuration.

The transmit condition is used to program the data transmission condition. It can either be periodic or event based, in which case it is to be mentioned which pin I/O is the event pin. Transmit Time Update, is the transmission interval for periodic transmit condition. Fig. 7 indicates the full operational procedure and methods that the firmware has to handle.

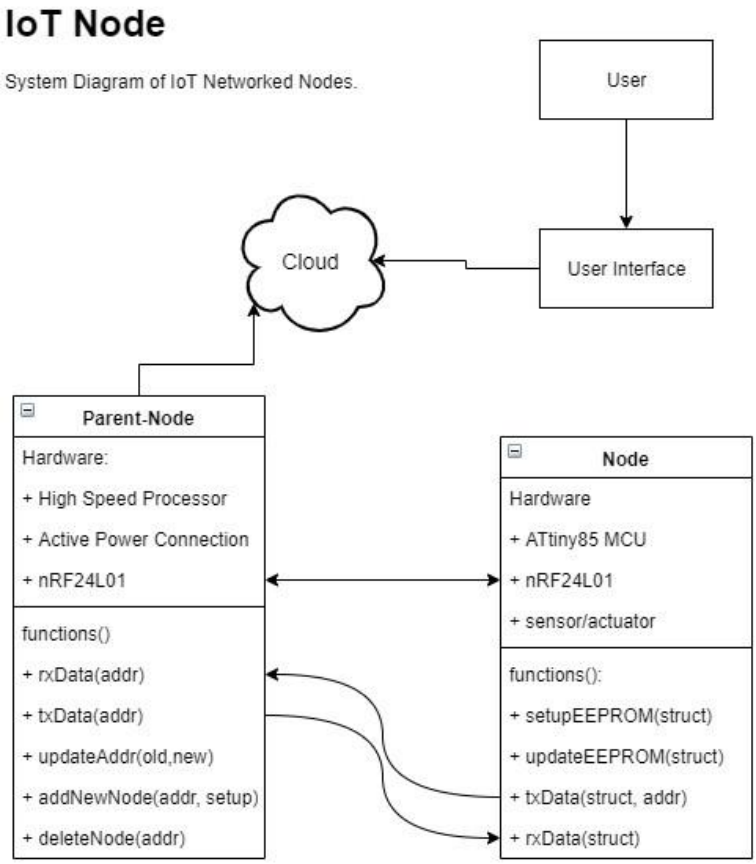


Fig. 7. System Operation Diagram

4.1. System Data Structure

The designed firmware uses concept of ´structures´ to define a standard for data communication between the parent-Node and the child-Nodes. The main structure which is transmitted is called mainConfig which comprises of opCode, which is the integer number of the operation to be carried out. Each operation has its own structure which is a member of the mainConfig structure. The Nodes, will receive the

entire structure and run a switch case in the received opCode. Functions written on the Node then process the configurations and then run the necessary processes to update the EEPROM table. Once this is done, the system is reset to run operations with the updated configuration.

```
struct txData{
unsigned long p3Data;
unsigned long p2Data;
};
// Structure of I/O pin configuration struct ioConfig{
byte p3Mode[1];
byte p2Mode[1];
int p3Dir;
int p2Dir;
};
// Received Data Sturcture
/*
* opCode = 1 => Address Update
* opCode = 2 => Pin Update
* opCode = 3 => Time Config
* opCode = 4 => Time Period Update
*/
struct rxData{ int opCode; ioConfig pinConfig; uint16_t addr;
unsigned long t_period;
};
```

The sensor data is wrapped in txData and delivered to the parent-hub, which then upwraps the structure to finally access the data.

4.2. EEPROM Lookup Table

Table. 2. EEPROM Table Address and Functions

5. Testing of Hardware

Process Loop Structure

1. Initialize PIN from EEPROM
2. Initialize RF Module
3. Get sensor Data

   a. If data is analog -> ADC => Digital Mapping b. If data-dir is Input, read data c. Else Output Value
4. Stop Listening for Signals
5. Send data using data structure via RF Module
6. Start Listening for Signal
7. If data is available from RF radio a. Read data into structure b. Switch on opcode c. Run opcode command sequence function d. Call setup to restart module with new command format

The module was tested with a PIR sensor, the data pin at pin3 of the ATtiny85. The collected data was observed on the Serial Monitor at the parent-hub. The data is at pin3 and pin2 sends an increasing counter as a default setting.

Serial Output at Parent Hub:

Received: 1

Pin 3 = 1

Pin 2 = 100

---------- Received: 1
Pin 3 = 1

Pin 2 = 101

---------- Received: 0
Pin 3 = 0

Pin 2 = 102

Updating the config needs to be done via the Serial Monitor. However, manual input to the Serial Monitor does not confirm to the format of the data structure. Thus, as a potential improvement, an UI can be developed to select the pin configurations and send the data in the right format to the Node.

The output mode was tested using a regular DC motor and all functions were observed to be consistent with the predicted outcomes.

6. Conclusion

The testing on the hardware confirmed the operability and feasibility of the designed Node. The firmware has the functions to write and read from the EEPROM table to update the operation of the Node. Thus, any user can update the EEPROM table and address to make the Node fully customizable and adaptable to the requirements of the user. In the current iteration of the design ,the input structure needed to be put in via the Arduino Serial monitor, which was a tedious and error-prone process. This can easily be rectified with the design of an User Interface and Dashboard. This however is beyond the scope of the project and can be added as a feature in futrue iterations.

The designed Node in its present form is compact, power-effcient, versatile and networking ready. These were the initial requirements from the project and have been achieved to a satisfactory level. It is observed during testing that several new EEPROM control variables may be added to make the ov erall system more versatile and easy to control. By making changes to the EEPROM table, the user can make the Nodes as complex and customizable as necessary.

In conclusion, this paper presents a new paradigm for making IoT Nodes more versatile and network capable. The design allows for the liberty to modify not just the connected device on-the-go, but also to change the overall position/network of the device. It works well with both sensors and actuator and hybrid networks resulting from this may be expected to have better response to real-world input.

Acknowledgements

This work was supported by the faculty and staff at Vellore Institute of Technology, Vellore. We would like to extend our gratitude to the management of VIT, Vellore and Dean of School of Electronics Engineering for their constant support during the process of design and development of this project.

The design paradigm was influenced by various papers [10], [11], [12]. The ideas presented in them may not have directly influenced the direction in which the project was developed, but they provided inspiration in regards to how the system may work better. The result from their analysis helped us develop a functional paradigm. Internet resources are a great medium for reference when designing new hardware design and the following resources have been instrumental in the same regards. [13], [14], [15], [16]

References

[1] ATtiny85 Datasheet: https://ww1.microchip.com/downloads/en/DeviceDoc/Atmel-2586-AVR-8-bit-Microcontroller-ATtiny25-ATtiny45-ATtiny85_Datasheet.pdf, 1

[2] nRF24L01 Datasheet, Section5.1, Power consumption, page-14 [3] nRF24L01 Datasheet: https://www.sparkfun.com/datasheets/Components/SMD/nRF24L01Pluss_Preliminary_Product_Specification_v1_0.pdf [4] https://www.electronicshub.org/different-types-sensors/, Types of common DIY Sensors

[5] Yong Feng, Hai Liu, Jie Yang, Xiaodong Fu, "A localized inter-actuator network topology repair scheme for wireless sensor and actuator networks"

[6] Halah Mohammed Al-Kadhim, Hamed S Al-Raweshidy, "Energy Efficient and Reliable Transport of Data in Cloud-IoT" DOI: https://doi.org/10.1109/ACCESS.2019.2917387

[7] RF module and ATtinty85 connection: nRF24L01+ with ATtiny85 3 Pins - Hackster.io

[8] nRF24L01 Datasheet, Section5.1, Power consumption, page-14

[9] IoT Connect Files and Schematics: https://github.com/sanskarbiswal/IoT_Connect

[10] James Pope, Robert Simon, "Efficient one-to-many broadcasting for resource-constrained wireless networks" DOI: https://doi.org/10.1109/LCN.2015.7366366

[11] Zeeshan Hamid Mir, Young-Bae Ko, "Collaborative Topology Control for Many-to-One Communications in Wireless Sensor Networks" , DOI: https://doi.org/10.1109/ACCESS.2017.2722379

[12] Hium Kim, Abbas Ahmad, Hamza Buqa, Jaeyoung Hwang, "IoT-Taas: Towards a Prospective IoT Testing Framework" DOI: https://doi.org/10.1109/ACCESS.2018.2802489

[13] nRF24L01: In-Depth: How nRF24L01 Wireless Module Works & Interface with Arduino

[14] https://create.arduino.cc/projecthub/arjun/programming-attiny85-with-arduino-uno-afb829, Arduino ISP for ATtiny85

[15] https://arduino.stackexchange.com/questions/25080/can-you-access-the-eeprom-of-an-attiny-with-arduino-code, EEPROM programming

[16] http://www.bajdi.com/sending-structs-with-nrf24l01-modules-and-the-rf24-library/ , Sending struct over Arduino

A Brief Review of Conventional and Deep Learning Approaches in Facial Emotion Recognition

Ajitha Sukumaran[a], Arun M[b] *

[a]*Senior Lecturer,Middle East College,Muscat and P.O Box 79,Oman*

[b]*Head of Embedded Technology,VIT University Vellore,,India*

Abstract

Facial emotion recognition is an important research domain in social signal processing. This paper is a comprehensive review on facial emotion recognition methods using conventional and deep learning approaches adopted for the research in the last few decades. The review is started with the traditional approach for facial emotion recognition with the analysis of some of the novel architectures. Followed by, deep learning method and various algorithms used are discussed. We have also provided a comparison analysis of different deep feature learning architecture using convolutional neural network. A fleeting review of publically available datasets are done from the reviewed papers. In the state of art, analyzed samples of existing novel approaches in both conventional and deep learning method along with the discussion of some use cases in FER. The challenges, when using the deep learning algorithms in real time embedded system are the large memory requirement and increase of computational complexity. Contrasting to this, conventional methods need lower memory and lower processing power. Some of the surveys reflecting that the performance of their conventional method are close to deep learning.

Keywords: Type your keywords here, separated by semicolons ; Deep learning; Conventional; facial Emotion Recognition; convolutional neural network; dataset;embedded system.

5. Introduction

Current generation people occupy the day with the laptop either as a part of their work or as an entertainment. But these computers are emotionally blind and cannot understand the human emotions. Our face reflects majority of human emotions or

* Corresponding author.
E-mail address: arunm@vit.ac.in

otherwise by using the different facial expressions we express our emotions. Hence it is very vital to analyze the facial expression automatically and there by recognize the emotions.

The emotion analysis needs a high level knowledge. Even though the face expressions convey the emotions, but the intention, mental reasoning or the social meaning could be interpreted based on the context, nature of the person, culture, gender or even by body gesture synchronized with facial muscles movement or configuration.

In 1971 Ekman and Friesen [1], showed that particular facial behavior are universally associated with specific emotions irrespective of sex and culture. The emotions considered here are happiness, anger, sadness, disgust, surprise and fear. Recent researches in psychology and social neural science evidenced that the facial expression of images are not culturally universal [2].Different surveys show that the oral communication among humans carry only one third information and the rest conveyed by non-verbal components[3]. In non-verbal components, the facial expressions have a significant role. This paper is an analysis of the different FER algorithms using traditional and deep learning methods.

The paper is organized as follows. The section 2 discuss about an overview of the convention FER approach. Section 3 explicates the deep learning approach and the section 4 provides the review of recent FER novel architectures used by conventional and deep learning approach Included a surveys of some of the use cases in Facial emotion recognition systems, where conventional approaches are more used because memory problem. Section 5 discuss the major challenges and opportunities in the area of FER research

6. The Conventional Facial Emotion Recognition System

The conventional (traditional) approach is based on detecting the human face and extracting the features by geometric or appearance based or a hybrid of both of these methods on the target face.

Figure 1 shows the general block diagram of the conventional FER approach. The traditional FER, is executed in three stages which includes the face detection, feature extraction and emotion classification.

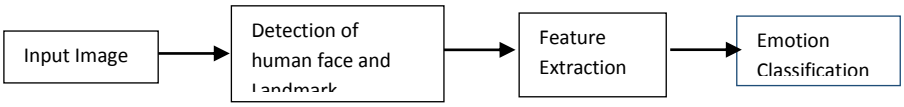


Figure 1: General Block diagram of a traditional FER approach

Face detection is the first process for automatically locating the face region of the input image. In the case of a video sequence, the face is identified in each frame or the first frame and then track the remaining in the sequence. It is important to estimate head pose for performing head alignment. Then the detection of the facial components and the landmarks estimation is done by extracting various spatial and temporal feature information. In the last step the facial expressions are classified by using the pre-trained emotion classifiers such as Support Vector Machines, Local Binary Pattern, Random Forest, Weighed Random Forest, Ada Boost algorithms etc.

Once the face is detected, the next task is to extract the emotions from the facial change. Mainly two methods are used for extracting the features; geometric , appearance based or a mixture of these two. The geometric feature based method centered on confining and tracking the dense set of facial point by Active Appearance based Model(AAM) and its variations. The localized facial points are the landmarks for extracting the shapes of features of face and its movements [4] [5].

Several work have done on geometric feature based algorithms to detect and evaluate FER performance [6][7]. A recent example of an AAM-based technique for facial expression recognition is presented in [8], in which different AAM fitting algorithms are compared and evaluated. Kotisa et al. [6] in their work provided a system based on geometric features for FER . Elastic Bunch Graph Matching (EBGM) and Kanade-Lucas-Tomasi (KLT) tracker were used in some of the research for facial point tracking [7].

In appearance based method the characteristics like intensities of the image, its pixel vales and the histograms are considered. This method has successfully implemented for the FER using Local Binary Pattern[9-12], Gabor Filters[13], Curvelet Feature extraction [14], histogram of orientation gradients (HOG) [15,16] and Principle Component Analysis (PCA) [17]. In [18] introduced a hybrid method, which is the combination of Bilateral 2D Principle Component Analysis (B2DPCA) and extreme learning machine (ELM),with a curvelet-based algorithm.

7. Deep FER Approach

In recent years facial emotion recognition using deep learning approaches become more popular. Deep learning computation can run on GPU, yielding high performance with lower power consumption. Figure 2 shows the general block diagram of an automatic Deep FER approach .The three main steps in the Deep FER approaches are Preprocessing, Deep feature learning and Deep feature classification.

Figure 1: General block diagram of Deep Learning Approach

7.1. Preprocessing

Preprocessing step is required to remove the unwanted areas of the face to line up and regularize the visual semantic information conveyed by the face. In preprocessing stage the alignment of face, data augmentation and normalization of face are done. Various works are dedicatedly done with different algorithms for each of these stages. The Viola-Jones face detector [19] is generally used algorithm for the implementation of face detection. In their paper the object detection is done by Haar feature based cascade classifier with machine learning approach. Here the features of the image are detected and learned by Ada boost algorithm, combined with cascaded complex classifiers. This rejects the background region of the image. In consort with face detection, the alignment of face with the coordinates of localized landmarks will improve the facial emotion recognition performance [18]. Cootes[20] in his paper on "Active Appearance Models" proposed AAM algorithm to extracts the shape and texture of specified object at the same time. Xiangxin Zhu and Deva Ramanan [Zhu] proposed a discriminative combined model for detecting the face, estimation of pose and landmark in real time images, using a tree-structured model. The paper by Zhifeng Li [21] proposed another discriminative model for addressing the face matching for age variation case.

Currently many face alignment methods use deep neural network. Deep Alignment Network (DAN) [22] is a robust method for face alignment, entails multiple stages, in which the locations of the facial landmarks estimated in each stage is improved compared to the previous stage. Miao et al [23] in their paper proposed the direct shape regression network (DSRN) for the face alignment, in which by efficient encoding the correlations of landmarks improved performance. On the whole, the cascaded regression approaches are the most popular state of art methods for the alignment of face because of the accuracy and speed performance. Multiple number of face detectors can be used instead of one for better land mark estimation [24][25].

Next step in preprocessing stage is data augmentation. Since currently available public database won't be having enough images for training, this stem is very essential. It is categorized into on the fly and offline data augmentation. On the fly

data augmentation is used in deep algorithms to overcome overfitting problem. Here the resolution of the image will be resized or downsampled to 32x32, 64x64, 128x128. This is done for checking the performance of Convolutional Neural network (CNN) architecture against the input size. Then the noise like Salt & Pepper, and Speckle are added as data augmentation[26][27][28][29]

The last step in the preprocessing stage is the normalization of face. FER under natural conditions are challenging because of the variation in illumination, head pose and occlusions Histogram Equalization based method (HE) [30] are most commonly used as it increases the global contrast for the images of faces. Block based, Adaptive and Oriented Local Histogram Equalization (BHE, AHE and OLHE) [31] [32] [33] are improved HE algorithms [34]. Illumination Compensation based on Multiple Regression Model (ICR) is another method to overcome variations of illumination in illumination normalization [35]

The facial expressions are affected by head pose variations. A typical solution to this problem is to use a three dimensional or two dimensional face-shape model. In [36] a three dimensional morphable model is used to evaluate the three dimensional facial shape, then the remaining analysis of face is carried out in the frontal view [37]. A two dimensional shape free regression scheme for recognizing the facial-expression in invariant head pose normalization, based on two dimensional geometric features. In recent times, the generative adversarial network (GAN) based deep models were proposed for frontal view synthesis [38][39][40] with encouraging results.

7.2. Deep Feature Learning

Deep learning is very promising, powerful and wildest developing technology in the area of artificial intelligence. It is the sub area in machine learning architecture which combine several multilayer networks with many hidden layers. The deep layer consists of fully connected layers, in such a way that each given neuron is connected with all the nearby neurons around it. The deep learning architecture is formed by pooling the fully connected layer with different machine learning functions. This section provides with a brief overview of the different deep learning architectures used for facial emotion recognition.

Convolutional Neural Networks (CNN)

Convolutional neural network (CNN) architectures has proved good performance rate in FER algorithms. The CNN architecture has four main components. They are

convolutional layers, subsampling/pooling layers, activation functions and fully connected layer [41]. The convolutional layer, will take input images and capture the low level features such as color, edges, and gradient orientation etc. of images. By adding more layers, this architecture acclimates to high level features. After analyzing the image in the convolution layer, the next pooling layer is responsible for the dimensionality reduction of the image. This is required to reduce the computational power while processing data. The main two pooling methods used are Max Pooling or Average Pooling. In max pooling, the maximum value of the Kernel of the image is returned and in average pooling, all values of kernel is returned. Max Pooling, functions as a noise suppressant. The activation function control the data flow between the layers. The high level features of the image, provided by the convolutional layer is learned in the fully connected layer and classification is performed by softmax technique.

The different architectures of CNNs include LeNet, Alexnet, VGGNet, GoogleNet, ResNet, ZFNet etc. They are differ in terms of the number of layers , filter used for convolution and complexity. Table 1 shows a comparison on these architectures.

Table 1. A comparison of various CNN architecture

CNN Architectures	Year	Features	No: of Convolution Layers	No: of Fully connected Layers
LeNet [42]	1998	• First ConvNet, designed by Yann LeCun • Contains all the elements like a modern ConvNet. • Activation function - ReLU *tanh()* • 2 Convolution layers of 5×5 Filters and Stride S=1, and 2 Pooling layers with 2×2 Filters and S=2.	2	-
AlexNet [43]	2012	• Designed by Krizhevsky, Sutskever, and Hinton • Activation function -ReLU • First convolutional layer with 11×11 Filter and S=4 • Database-ImageNet (more than one + million images for	5	Used

CNN Architectures	Year	Features	No: of Convolution Layers	No: of Fully connected Layers
		training, from 1000 categories. • With the help of GPUs, the training time remains same. • Contain two GTX 580 GPUs		
3ZFNet [44]	2013	• Designed by Zeiler and Fergus • First convolutional layer with 7×7 Filter and S=2 • Able to capture the image features with fine resolution because of the small size filters.	5	Used
VGGNet [45]	2014	• Designed by Simonyan and Zisserman • Adopted the classical ConvNet design, but with smaller filter configuration repeating 16 times: • First convolutional layer 3×3 filter, withS=1 and Padding = 1, and 2×2 maxpooling filters with S= 2. • 144 million parameters, (Used in the final 3 Fully Connected layers*)*	13/16	Used
GoogleNet [46]	2014	• All convolution layers use ReLU. • Only using average pooling layer(No fully connected layers) • 5 million parameters • An Inception Module is introduced.	21	Used
ResNet [47]	2015	• Introduced by He et al. • *Ultra-deep* ConvNets era started by the introduction of ResNet • Comparatively large Learning Rate (0.1) because of the extensive use of Batch Normalization at the end of every layer.	151	Used

CNN Architectures	Year	Features	No: of Convolution Layers	No: of Fully connected Layers
		• No Fully Connected layers at the end of the network. • Current popular ConvNets and used by default in practical applicatins.		
DenseNet[48]	2017	• Introduced by Huang et al. • Here concatenation is used. • Each layer gets the combined knowledge from all previous layers,so that the network is compact (number of channels are less) • Batch normalization after each convolution layer and hence no drop out. • Activation function ReLU	-	-

Recurrent Neural Networks(RNN)

Recurrent Neural Network utilize sequential information and the term 'recurrent' shows that the output for a given element of the sequence dependent on the current and previous computations. It is having memory, which stores the information of previous computations.

To alleviate the long term dependency problem in RNN, Hochreiter and Schmidhuber introduced the "Long Short-Term Memory Networks (LSTM)" as a special kind of RNN[49].In case of normal RNN, training is done by back propagation through time, so the problem of vanishing or exploding gradient can come and hence with this it is difficult to learn long sequence. To overcome this problem the RNN cell is replaced by a gated LSTMs cell. Hagry and et al, in their paper [50] using the method of LSTM RNN for emotion recognition from EEG signals. Another extension of RNN is, bidirectional RNNs (BIRNN) in which the output at any time t is depend on both the previous elements and future elements in the sequence. RNN or its extended models are widely used in Speech and video based expression recognition tasks.[51][52].

Deep Belief Networks(DBNs)

DBNs are unsupervised probability based Deep learning model proposed by Hinton et al in 2006[53]. This model is joined by several restricted Boltzmann machines (RBM).Here the top layer have undirected and lower layer have directed connections. The training of DBN is done in two phases,the pre-training and the fine tuning[54].

DBN is pre trained by Greedy learning algorithm**.** It uses layer based approach for identifying features [55]. These features are modified by fine tuning for obtaining better categorization between different classes. Fine tuning can be achieved by using Wake Sleep or by Back Propagation algorithm. In [56] and [57], a hierarchical representation of features and the classification is done by a logistic regression function.. The paper by Liu et al propose a Boosted Deep Belief Network (BDBN) to implement the three stages (feature learning, selection, and classification) iteratively in a unified loopy framework [57].

Digital Auto Encoders

The auto encoder is designed to reduce the data dimensions by ignoring the noise in the data. DAE was first introduced in [58] to learn the dimensionality reduction by efficient coding. It is able to change unsupervised data into a supervised format. It consists of 4 parts encoder, bottleneck, decoder and Reconstruction Loss .The commonly used auto encoder architectures are Feed Forward network, LSTM network or Convolutional Neural Network. Garcia et al [59] in their work showed that in each layer of the model is pre trained as an auto encoder , the performance increased to 92.52% from the existing 91.16% % for a real time empathetic robotic system. Some of the works with DAE for feature extraction outperformed the state-of-art feature selection and dimension reduction techniques [60]

Generative Adversarial Network (GAN)

GAN is introduced by Ian Goodfellow et al at the University of Montreal in 2014[61]. GANs include multiple deep neural networks rather than one network. Here simultaneously two models are trained by an adversarial process. The generator learns to generate images which look real, while a discriminator tries to classify the image as either real (from the domain) or fake (generated). Effective generative modeling provides a more domain-specific approach for data augmentation. To adopt to various applications recently many works with several extensions of GAN network has developed such as conditional GAN[62] Disentangled Expression learning-GAN

[63],Differential GAN[64], Expression Generative Adversarial Network (ExprGAN)[65] etc.

8. FER Database

Facial emotion Recognition of human beings is one of the most challenging tasks in social signal processing. Normally, face expressions are natural and straight means for human emotions and intensions. In FER field several databases have been used for different experiments based on the suitability. In the below table summarizes the publically available databases and its features which are used by some of the reviewed papers

Table 2. Publically available database and their features

Database	Features	Emotions identified	Weblink
CK+	Extended CohnKanade (CK+) database laboratory-controlled database 593 video sequences from 123 Subjects	Happiness ,anger, sadness, disgust, surprise, fear, contempt, neutral	http://www.consortium.ri. cmu.edu/ckagree/
DISFA	Denver Intensity of Spontaneous Facial Action Database Contains stereo videos of 27 adult subjects (12 females + 15 males) with different ethnicities.	Happiness ,anger, sadness, disgust, surprise, fear, contempt	http://www.engr.du.edu/ mmahoor/DISFA.htm
JAFFE	Japanese Female Facial Expression Image database Laboratory controlled Contains 213 samples of posed expressions from 10 Japanese females	Happiness ,anger, sadness, disgust, surprise, fear, contempt and one image with a neutral expression.	http://www.kasrl.org/ jaffe.html
FER2013	Facial Expression Recognition 2013 Dataset Images are collected	Happiness ,anger, sadness, disgust, surprise, fear,	https://www.kaggle.com/c/ challenges-in-representatio n-learning-facial- expression-recognition-

Database	Features	Emotions identified	Weblink
	by Google image search API Contains training images-28,709, validation images-3,589 and; Test images-3,589	contempt	challenge
TFD	Toronto Face Database contains 112,234images, with 4178 annotated with one of the emotion	annotated anger, disgust, fear, happiness, sadness, surprise and neutral.	josh@mplab.ucsd.edu
AFEW	Acted Facial Expressions in the Wild(AFEW) Contains video clips collected from different movies with unprompted expressions, various head poses, occlusions and illuminations. 1,809 videos	Happiness ,anger, sadness, disgust, surprise, fear surprise and neutral.	https://sites.google.com/site/ emotiwchallenge/
BU-3DFE	Binghamton University 3D Facial Expression contains 606 facial expression sequences captured from 100 people.	Happiness ,anger, sadness, Disgust,surprise, fear and neutral.	http://www.cs.binghamton. edu /_lijun/Research/3DFE /3DFE Analysis.html
EmotioNet	Large-scale database One million facial expression images from the Internet.	anger, disgust, fear, happiness, sadness, surprise and 10 compound expressions	http://cbcsl.ece.ohio-state.edu/dbform emotionet.html

9. State of art FER Approaches

This section provide a brief survey on some of the novel research works carried out using traditional and deep learning based FER approach. In addition to this included a brief survey of some of the use cases examples in this field.

9.1. Survey of FER approaches with Conventional Methods

Conventional approaches are based on feature extraction and classification. Feature extraction is done by Gabor filter, Local binary pattern, Histogram of Oriented gradients, Non-negative matrix factorization etc. Most of the conventional approaches use SVM, Adaboost, PCA, Random Forest, Nearest neighbor classifier etc as classification algorithms. Table 3 provides an analysis of some of the state of art methods in conventional FER approach.

Table 3. Analysis of various Conventional FER approaches

Si No:	Author (Year)	Paper Title	Work Carried out	Remarks
1	Olga Krestinskaya and Alex Pappachen James(2018) [66]	"Facial Emotion Recognition using Min-Max Similarity Classifier"	Normalization of pixel is done for removing intensity offsets For feature classification Min-Max metric in a nearest neighbor classifier is used The emotion recognition performance is improved from 92.85% to 98.57% Database -JAFFE	Direct and simple approach. Low computational complexity Less computational time Lower memory space is required. In case of the metric learning it is required to create the templates (each class)chomps additional memory space
2	M. I. N. P. Munasinghe (2018) [67]	"Facial Expression Recognition Using Facial Landmarks and Random Forest Classifier"	Face detection by a histogram of oriented gradients Classification algorithm based on facial landmarks and random forest classifier. By using the model trained with the iBUG	The research proved that it is possible to classify the emotions accurately with the success rate of 90%

Si No:	Author (Year)	Paper Title	Work Carried out	Remarks
			300-W dataset, facial landmarks are identified Database used -s Extended Cohn-Kanade.	
3	Shichuan Du, Yong Tao, and Aleix M. Martinez (2014) [68]	"Compound facial expressions of emotion."	Automatic detection of shape of face is done by configural features Categorization of emotion is done by shape and appearance features. For appearance, the Gabor filters at multiple scales and orientations is used. Author has prepared a new database with a total of 230 subjects with different ethnicities and races Minimized facial occlusions were minimized up to certain extent.	The work proved that the configural features are more good for detecting both basic and compound emotions in the face.
4	Yuanyuan Ding, Qin Zhao (2017) [69]	"Facial Expression Recognition From Image Sequence Based on LBP and Taylor Expansion"	To detect the peak expressions from the video frame, double local binary pattern (DLBP) is used. To handle the illumination variations in DLBP, logarithm-laplace (LL) domain is used. The facial expression feature is extracted by Taylor expansion theorem . Both LBP based Taylor feature pattern	Suited for real-time applications The proposed DLBP method reduce the face detection time by its lower-dimensional size .

Si No:	Author (Year)	Paper Title	Work Carried out	Remarks
			(TFP) and Taylor expansion is used Database - JAFFE and Cohn-Kanade	
5	Yubo WANG, Haizhou AI, Bo WU, Chang HUANG [70]	Real Time Facial Expression Recognition with Adaboost	Haar feature based Look Up Table weak classifiers are used for extracting the facial expression.. In real time able to recognize seven expressions. Database - JAFFE .	This method proved a consistent performance for multi class. Detection speed is improved compared to SVMs method.

9.2. Survey on FER approaches with Deep Learning Algorithms

In recent years deep learning based algorithms are very influential in the area of computer vision based application. The whole approach is based on preprocessing of image, feature learning and classification. Each of these stages were discussed in detail in section 3.Table 4 provides an analysis of various Deep learning approaches employed in FER.

Table 4. Analysis of various Deep Learning approaches in FER.

Si No:	Author (Year)	Title	Work Carried out	Remarks
1	Shervin Minaee1 , Amirali Abdolrashii (2019) [71]	"Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network"	Used a deep learning approach based on attentional convolutional network, By the word "attentional", it is possible to focus important parts of the face For highlighting the salient features of the face, visualization	This achieves significant improvement and consistent performance over previous models on multiple datasets,

Si No:	Author (Year)	Title	Work Carried out	Remarks
			method is used. Database used - FER-2013, CK+, FERG, and JAFFE.	
2	Jiaxing Lia, Dexiang Zhanga, Jingjing Zhanga, Jun Zhanga, (2017) [72]	"Facial Expression Recognition with Faster R-CNN "	Normalization of the face and extraction of features are done with a trainable convolution kernel. In order to reduce the dimension of the extracted features, the maximum pooling is used. High quality region are generated by RPNs (Region Proposal Networks) for Faster R-CNN for detection. Facial expressions are classified using softmax classifier and regression layer. The dataset used is Chinese Linguistic Data Consortium (CLDC),	Good expression recognition accuracy is achieved for 6 basic emotions 100%, recognition rate for disgust and angry. But this algorithm is not giving good result for the identification of emotions urprise and neural By increasing the weight factor of the eyes and mouth from the image by increasing the training data,can expect better performance and accuracy.
3	Xi Ouyang Shigenori Kawaai Ester Gue Hua Goh (2017) [73]	"Audio-visual emotion recognition using deep transfer learning and multiple temporal models."	The method consist of three process, facial emotion recognition, audio emotion recognition and system fusion The face detection is done by a cascaded convolutional network The FER part is the fusion of the stacks of CNN and RNN	The overall accuracy achieved by this system with challenge dataset is 57.2%, against the challenge baseline of 40.47% . The system proved that transfer learning is an effective way to find good image features The combination of geometric and RNNbased video features is more effective than single

Si No:	Author (Year)	Title	Work Carried out	Remarks
			(VGG-LSTM). The emotion recognition from audio signals done by deep neural network and transfer learning. Database - Emotiw2017 But is inadequate for deep learning process, it is trained by Librispeech.	model
4	Xiangjun Wang Yubo Ni (2018) [74]	"Unsupervised Domain Adaptation for Facial Expression Recognition Using Generative Adversarial Networks."	Distributed pseudolabel method (DPL) is used for the unlabelled GAN generated small target dataset. Here a GAN is trained on the target dataset The GAN generated samples are used to fine-tune the pretrained model The dataset used are FER-2013 JAFFE (VGG11 and Alexnet), MMI, and CK+,	This method can apply to any existing CNN architecture. to any existing convolutional neural networks (CNN). The author showed the effectiveness of the system on four facial expression recognition datasets with two CNN structures This show that cross-dataset accuracy of the CNN model is improved here
5	Ping Liu∗ Shizhong Han∗ Zibo Meng Yan Tong (2014) [75]	"Facial Expression Recognition via a Boosted Deep Belief Network"	Model used is a Boosted Deep Belief Network (BDBN) framework.. The set of features which characterize the facial expression was learned and selected to form a boosted strong classifier An objective	The BDBN framework outdid the other state of theart methods using the CK_and JaFFE databases till that time., This network performace need to be checked the various challenging conditions such as spontaneous expressions with face pose variations and in recognizing facial action units.

Si No:	Author (Year)	Title	Work Carried out	Remarks
			function is developed for feature -fine-tuning process As the learning progress, the strong classifier is improved and the discriminative capabilities of the selected features strenghthened. Database - JAFFE and CK+	

9.3. Survey on FER approaches for Use cases

The memory and the computational power required for conventional methods are low compared to deep learning approaches. When an FER model is being implemented for real time application with embedded system, the constraints are memory and computational complexity. Hence the traditional approaches are still being used with hybrid methods or improved ideas Table 5 summarizes the analysis of some of the use cases in FER

Table 5. Analysis of FER approaches in use cases

Si No:	Author (Year)	Title	Work Carried out	Remarks
1	Mira Jeong and Byoung Chul Ko(2018) [76]	"Driver's Facial Expression Recognition in Real-Time for Safe Driving."	The facial region and landmarks are detected in an input image using the face and landmark detector of DLib(open source machine library) The FE is recognized using the hierarchical weighted random fores(WRF) classifier Databases used - Extended Cohn-Kanade (CK+), MMI and the Keimyung University Facial Expression of Drivers (KMU-FED(Accuracy obtained is 92.6% for CK+ and 76.7% for MMI This method is augmented for real time embedded system application	The results of the experiment showed that the performance of the system without using a GPU are similar to DNN based FER approaches Processing cost is much lower than deep learning method. As the tradition method is being used here it entailed a low amount of memory
2	Pranti Dutta N achamai M	Facial Pain "Expression	Used a hybrid model as combination of Constrained Local	This system was able to detect facial emotion

Si No:	Author (Year)	Title	Work Carried out	Remarks
	(2018) [77]	Recognition in Real-Time Videos."	Model , Active Appearance Model and Patch-Based Model cobined with image algebra Used Videos and real time streamed data as input No databases are available for capturing facial pain. .So the database is created by the author Database used is UNBC	correctly for pain from a live stream Work could be extended to include different subjects and a comparison could have been done with preexisting databases for better evaluation of performance. This research could be enhanced by deep learning algorithms and mathematical analysis by partial differential equations
3	Suzan Anwar Mariofanna Milanova (2016) [78]	"Real Time Face Expression Recognition of Children with Autism."	This paper,was on the FER system based on tracking the face triangulation points in real-time. Active shape Model (ASM) tracker, for identifying facial landmarks The facial expression features are extracted using Support Vector Machine (SVM) based classifier Able to recognize seven expressions Database is Child Affective Face Expression CAFE set, and.	The result shows 93% classification accuracy AFEC set and 80% during the real time experiments. Most of the cases the people with Autism Syndrome Disorder exhibit mixed results in behavioral studiers of FER. Hence FER associated with eye-tracking, electrophysiological, and brain imaging studies yield better result.
4	Agieszka Różańska Żaneta Rachwaniec-Szczecińska Aleksandra Dagmara Kawala-Janik Michal Podpora (2013) [79]	"Internet of Things Embedded System for Emotion Recognition."	This paper discuss an IoT-based embedded system for FER. Here a Google Cloud API is installed on the Raspberry Pi platform. The total duration (processing and transmission) time of FER, including sending and saving the file locally on Raspberry Pi, is between 14 to 20 seconds..	The author proved that compared to other APIs Google Cloud API features useful for FER The prototype of the system was working fine,although the processing time is relatively long(14 - 20seconds) The further works could be more oriented on speed concern
5	Moulay Smail Bouzakraoui Abdelalim Sadiq Nourddine Enneya (2017) [80]	"A Customer Emotion Recognition through Facial Expression using POEM descriptor and SVM classifier."	This paper discuss the system developed for customer emotion analysis The face detection is done by using Haar classifier Feature vectors are Patterns of Oriented Edge Magnitudes (POEM), histogram of different block sizes of the face image. Emotion classification by ng Support Vector Machine (SVM). Database - JAFFE	Accuracy can improved by multi modal (eye gaze,eye tracking) methods.

10. Challenges

Database is an essential part of an FER system. Facial expression can change from person to person depending on the gender, age and ethnicity. In order to predict the emotion correctly, it is required to train the system with a dataset which covers all these aspects. The images in the database are subject to head pose variation, illumination variations, occlusions, bias identity and clear differentiation between closely matched expressions. Even though initial section of the paper discuss several algorithms to overcome these problems, there is gap in research for addressing the above problems. The major challenge is the preparation of the database which cover all these characteristics. No specific database is available for performing FER for case to case application. In this context, the researcher need to create an exclusive database.

Available databases were prepared under various conditions and with different subjects. Most of the researches were performed experiments with one or two datasets resulting a dependable accuracy in FER, but the same algorithm may not provide the same performance for a different dataset. FER algorithms should have consistency in detecting emotions correctly across different datasets and research should be focused for improving cross dataset accuracy.

Compound emotion recognition are the area where very few works were dedicated on the detection Most of the researches are restricted to 6 or 7 basic emotions. New research work is focused on the accurate detection of compound emotions, with quality of sensors using are another challenge.

Human emotions are affected by multiple prime factors. Facial expression is only one modality and with this alone it is impossible to come up with accurate judgements. So it is always advisable to have the fusion of different modalities such as audio, physiological signals, body movement etc. and the fusion of modalities will be a promising and developing research area.

Most of the recognized works end up with the algorithm discussion, testing and training in laboratory controlled environment. But when come to the implementation of real world which need to consider many other aspects such as computational complexity and memory requirement.

The memory requirement of deep learning algorithms are more compared to the traditional FER methods. Current research works in real time implementations are

based on conventional approach and these researches are focused on improving accuracy of conventional method or making its performance close to deep learning method. Ongoing researches are in the robust implementation of deep neural network to real time applications, which can be implemented with support of a more suitable processor (highly integrated SoC) for the embedded system application.

11. Conclusion

Extensive research works have been carried out in the field of Facial Emotion Recognition, and many literature reviews are available in the same area. But most of them are either restricted to solely on traditional approach in FER or deep learning approaches. In this paper we have reviewed the basics and algorithms for both convention and deep learning approaches separately. A comparison analysis of features of various available datasets in reviewed papers are discussed. Finally a survey on the analysis of some state of art approaches on conventional, deep learning methods are provided with some use case.

References

[1] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion.” Journal of personality and social psychology, vol. 17, no. 2,pp. 124–129, 1971.

[2] M. Gendron, D. Roberson, J. van der Vyver and L. Barrett, "Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture.", 2019.

[3] F. P. Ede and O. State, “THE USE OF NON – VERBAL COMMUNICATION IN THE TEACHING OF Council for Innovative Research,” vol. 4, no. 3, pp. 428–433.

[4] D. Kim and J. Shin, "A Study on Appearance-Based Facial Expression Recognition Using Active Shape Model", KIPS Transactions on Software and Data Engineering, vol. 5, no. 1, pp. 43-50, 2016. Available: 10.3745/ktsde.2016.5.1.43.

[5] H. Zhou, K. Lam and X. He, "Shape-appearance-correlated active appearance model", Pattern Recognition, vol. 56, pp. 88-99, 2016. Available: 10.1016/j.patcog.2016.03.002.

[6] I. Kotisa and I. Pitas “Facial expression recognition in image sequence using geometric deformation features and support vector machines” IEEE Transaction on Image Processing, vol. 16, pp. 172–187, 2007.

[7] D. Ghimire, J. Lee, “Geometric feature-based facial expression recognition in image sequences using multi-class AdaBoost and support vector machines,” Sensors, vol. 13, pp. 7714–7734, 2013.

[8] F. Extraction, M. Zhang, and F. Xie, “Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM) Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM).”,2018

[9] A. N. Ekweariri and K. Yurtkan, “Facial expression recognition using enhanced local binary patterns,” Proc. - 9th Int. Conf. Comput. Intell. Commun. Networks, CICN 2017, vol. 2018-January, no. September 2017, pp. 43–47, 2018.

[10] G. Prajapat and R. Kumar, "A Hybrid Approach for Facial Expression Recognition Using Extended Local Binary Patterns and Principal Component Analysis", International Journal of Electronics, Communications, and Measurement Engineering, vol. 8, no. 2, pp. 1-25, 2019. Available: 10.4018/ijecme.2019070101.

[11] M. V.Jonnalagedda and D. D. Doye, "Radially Defined Local Binary Patterns for Facial Expression Recognition", International Journal of Computer Applications, vol. 119, no. 21, pp. 17-22, 2015. Available: 10.5120/21360-4369.

[12] M. V.Jonnalagedda and D. D. Doye, "Radially Defined Local Binary Patterns for Facial Expression Recognition", International Journal of Computer Applications, vol. 119, no. 21, pp. 17-22, 2015. Available: 10.5120/21360-4369.

[13] P. Sisodia, A. Verma, and S. Kansal, “Human Facial Expression Recognition using Gabor Filter Bank with Minimum Number of Feature Vectors,” Int. J. Appl. Inf. Syst., vol. 5, no. 9, pp. 9–13, 2013.

[14] Mandal, T.; Majumdar, A.; Wu, Q.J. Face recognition by curvelet based feature extraction. In Proceedings of the International Conference Image

Analysis and Recognition, Montreal, QC, Canada, 22–24 August 2007; Springer: Berlin, Germany, 2007; pp. 806–817.

[15] Ghimire, D.; Lee, J. Histogram of orientation gradient feature-based facial expression classification using bagging with extreme learning machine. Adv. Sci. Lett. 2012, 17, 156–161.

[16] Dhall, A.; Asthana, A.; Goecke, R.; Gedeon, T. Emotion Recognition Using PHOG and LPQ Features. In Proceedings of the IEEE International Conference on Face and Gesture Recognition and Workshop, Santa Barbara, CA, USA, 21–25 March 2011; pp. 878–883.

[17] Lin, D.-T. Facial expression classification using PCA and hierarchical radial basic function network. J. Inf. Sci. Eng. 2006, 22, 1033–1046.

[18] Mohammed, A.A.; Minhas, R.; Wu, Q.J.; Sid-Ahmed, M.A. Human face recognition based on multidimensional PCA and extreme learning machine. Pattern Recognition. 2011, 44, 2588–2597.

[19] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1. IEEE, 2001, pp. I–I.

[20] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,”IEEE Transactions on Pattern Analysis & Machine Intelligence, no. 6, pp. 681–685, 2001.

[21] Z. Li, U. Park and A. K. Jai, "A Discriminative Model for Age Invariant Face Recognition," IEEE Transactions on Information Forensics and Security, pp. 1028 - 1037, 2011

[22] M. Kowalski, J. Naruniec and T. Trzcinski, "Deep Alignment Network: A convolutional neural network for robust face alignment," in IEEE, 2017

[23] X. Miao, X. Zhen, X. Liu and C. Deng, "Direct Shape Regression Networks for End-to-End Face Alignment," in IEEE, 2018.

[24] Z. Yu and C. Zhang, “Image based static facial expression recognition with multiple deep network learning,” in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015,pp. 435–442.

[25] B.-K. Kim, H. Lee, J. Roh, and S.-Y. Lee, “Hierarchical committee of deep cnns with exponentially-weighted decision fusion for staticfacial expression recognition,” in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015,pp. 427–434.

[26] D. A. Pitaloka, A. Wulandari, T. Basaruddin and D. Y. Liliana, "Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition," Procedia Computer Science , pp. 523-529, December 2017

[27] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana,“Enhancing cnn with preprocessing stage in automatic emotion recognition,” Procedia Computer Science, vol. 116, pp. 523–529, 2017.

[28] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos,“Facial expression recognition with convolutional neural networks: coping with few data and the training sample order,” Pattern Recognition, vol. 61, pp. 610–628, 2017.

[29] M. V. Zavarez, R. F. Berriel, and T. Oliveira-Santos, “Cross-database facial expression recognition based on fine-tuned deep convolutional network,” in Graphics, Patterns and Images (SIBGRAPI), 2017 30th SIBGRAPI Conference on. IEEE, 2017, pp. 405–412.

[30] P. Ekman, “Facial action coding system (facs),” A human face, 2002.

[31] H. Gunes and B. Schuller, “Categorical and dimensional affect analysis in continuous input: Current trends and future directions,” Image and Vision Computing, vol. 31, no. 2, pp. 120–136, 2013.

[32] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” Image and Vision Computing, vol. 27, no. 6, pp. 803–816, 2009.

[33] P. Liu, S. Han, Z. Meng, and Y. Tong, “Facial expression recognition via a boosted deep belief network,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1805–1812.

[34] J. Yi, X. Mao and L. Chen, "Illumination Normalization of Face Image Based on Illuminant Direction Estimation and Improved Retinex," PLoS ONE, 23 April 2015

[35] J. Ko, E. Kim and H. Byun, "A Simple Illumination Normalization Algorithm for Face Recognition," in Pacific Rim International Conference on Artificial Intelligence, 2002.

[36] V. Blanz, P. Grother, J. P. Phillips, and T. Vetter. Face recognition based on frontal views generated from non-frontal images. In CVPR, pages 454–461, 2005.

[37] D. Reisfeld and Y. Yeshurun, "Preprocessing of Face Images: Detection of Features and Pose Normalization," Computer Vision and Image Understanding, vol. 71, no. 3, pp. 413-430, September 1998

[38] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, “Towards largepose face frontalization in the wild,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp.3990–3999.

[39] R. Huang, S. Zhang, T. Li, and R. He, “Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2439–2448.

[40] L. Tran, X. Yin, and X. Liu, “Disentangled representation learning gan for pose-invariant face recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1415–1424.

[41] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” ISCAS 2010 - 2010 IEEE Int. Symp. Circuits Syst. Nano-Bio Circuit Fabr. Syst., pp. 253–256, 2010

[42] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition.," in Proceedings of the IEEE, 1998.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.

[44] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," 2013.

[45] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[46] C. Szegedy,W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan,V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[47] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[48] G. Huang, Z. Liu, L. V. d. Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2016

[49] S. Hochreiter and J. Schmidhuber, "Long Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[50] S. Alhagry, A. A. Fahmy and R. A. El-Khoribi, "Emotion Recognition based on EEG using LSTM Recurrent Neural Network," International Journal of Advanced Computer Science and Applications, vol. 8, no. 10, 2017.

[51] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in CML'14 Proceedings of the 31st International Conference on International Conference on Machine Learning, Beijing, China, 2016.

[52] P. Khorrami, T. L. Paine and K. Brady, "How Deep Neural Networks Can Improve Emotion Recognition on Video Data," Computer Vision and Pattern Recognition , 2016.

[53] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets,” Neural computation, vol. 18, no. 7, pp. 1527–1554,2006.

[54] G. E. Hinton, “A practical guide to training restricted boltzmann machines,” in Neural networks: Tricks of the trade. Springer, 2012, pp. 599–619.

[55] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layerwise training of deep networks,” in Advances in neural information processing systems, 2007, pp. 153–160.

[56] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, “Learning active facial patches for expression analysis,” in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 2562–

2569.[45] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., “Challenges in representation learning: A report on three machine learning contests,” in International Conference on Neural Information Processing. Springer, 2013, pp. 117–124.

[57] P. Liu, S. Han, Z. Meng, and Y. Tong, “Facial expression recognition via a boosted deep belief network,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1805–1812.

[58] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” science, vol. 313, no. 5786, pp. 504–507,2006.

[59] A. R. Garcia , M. Elshaw, A. Altahhan and V. Palade, "Stacked deep convolutional auto-encoders for emotion recognition from facial expressions," in IEEE, Anchorage, AK, USA, 2017

[60] M. Usman, S. Latif and J. Qadir, "Using deep autoencoders for facial expression recognition," in 13th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 2017

[61] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., “Challenges in representation learning: A report on three machine learning contests,” in International Conference on Neural Information Processing. Springer, 2013, pp. 117–124.

[62] J. Cai, Z. Meng, A. Khan, Z. Li, J. O'Reilly and Y. Tong, "Identity-Free Facial Expression Recognition using conditional Generative Adversarial Network", arXiv.org, 2019. [Online]. Available: https://arxiv.org/abs/1903.08051. [Accessed: 28- Nov- 2019].

[63] K. Ali and C. Hughes, "Facial Expression Recognition Using Disentangled Adversarial Learning", arXiv.org, 2019. [Online]. Available: https://arxiv.org/abs/1909.13135. [Accessed: 28- Nov- 2019].

[64] W. Baddar, “Differential Generative Adversarial Networks : Synthesizing Non-linear Facial Differential Generative Adversarial Networks : Synthesizing Non-linear Facial Variations with Limited Number of Training Data,” no. November, 2017.

[65] H. Ding, K. Sricharan, and R. Chellappa, “ExprGAN : Facial Expression Editing with Controllable Expression Intensity,” 2014.

[66] K. Olga and A. P. James, "Facial emotion recognition using min-max similarity classifier," in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017.

[67] M. I. N. P. Munasinghe, “Facial Expression Recognition Using Facial Landmarks and Random Forest Classifier,” Proc. - 17th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2018, pp. 423–427, 2018.

[68] S. Du and A. M. Martinez, “Compound facial expressions of emotion: From basic research to clinical applications,” Dialogues Clin. Neurosci., vol. 17, no. 4, pp. 443–455, 2015.

[69] Y. Ding, Q. Zhao, B. Li, and X. Yuan, “Facial Expression Recognition from Image Sequence Based on LBP and Taylor Expansion,” IEEE Access, vol. 5, pp. 19409–19419, 2017

[70] Y. Wang, H. Ai, B. Wu, and C. Huang, “Real time facial expression recognition with adaboost,” Proc. - Int. Conf. Pattern Recognit., vol. 3, pp. 926–929, 2004.

[71] S. Minaee and A. Abdolrashidi, “Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network,” 2019.

[72] J. Li, D. Zhang, J. Zhang, and J. Zhang, “Facial Expression Recognition with Faster R-CNN,” vol. 107, no. Icict, pp. 135–140, 2017.

[73] X. Ouyang et al., “Audio-visual emotion recognition using deep transfer learning and multiple temporal models,” ICMI 2017 - Proc. 19th ACM Int. Conf. Multimodal Interact., vol. 2017–January, no. November, pp. 577–582, 2017.

[74] X. Wang, X. Wang, and Y. Ni, “Unsupervised Domain Adaptation for Facial Expression Recognition Using Generative Adversarial Networks,” Comput. Intell. Neurosci., vol. 2018, 2018.

[75] P. Liu, S. Han, Z. Meng, and Y. Tong, “Facial expression recognition via a boosted deep belief network,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 1805–1812, 2014.

[76] M. Jeong and B. C. Ko, “Driver’s facial expression recognition in real-time for safe driving,” Sensors (Switzerland), vol. 18, no. 12, pp. 1–21, 2018.

[77] P. Dutta and N. M, "Facial Pain Expression Recognition in Real-Time Videos", Journal of Healthcare Engineering, vol. 2018, pp. 1-23, 2018. Available: 10.1155/2018/7961427.

[78] S. Anwar, “Real Time Face Expression Recognition of Children with Autism Real Time Face Expression Recognition of Children with Autism,” no. November, 2016.

[79] A. Rozanska, Z. Rachwaniec-Szczecinska, A. Kawala-Janik, and M. Podpora, “Internet of things embedded system for emotion recognition,” 2018 IEEE 20th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2018, no. November, 2018.

[80] M. S. Bouzakraoui, A. Sadiq, and N. Enneya, “A customer emotion recognition through facial expression using POEM descriptor and SVM claasifier,” ACM Int. Conf. Proceeding Ser., vol. Part F129474, no. March, 2017.

AN AUTOMATIC PARKING ALLOTMENT SYSTEM

Shivnandan Khatri[a], N.Seenu[b], Balaji R*[,b], Harshit Gupta[d], Manju Mohan[c*]

[a]*Department of Automobile Engineering Hindustan Institute of Techonology & Science,Chennai India*

[b]*Department of Mechanical Engineering Hindustan Institute of Techonology & Science,Chennai India.*

[c]*Centre for Automation and Robotics Hindustan Institute of Techonology & Science,Chennai India*

[d]*Department of Computer Engineering Hindustan Institute of Techonology & Science,Chennai India.*

Abstract

In metropolitan regions, the most unavoidable challenge we are facing in our daily lifestyle is vehicle parking. The issue arises from not knowing available parking space; even if it is known, many vehicles may pursue very limited parking spaces to cause serious traffic congestion. Parking control impacts drivers search time and value for parking spaces. In such observation that parking space reservation technique can assist drivers to reduce the search time dramatically; design and implement a Reservation-based Smart Parking System permits drivers to successfully locate and reserve the vacancy in any parking space. Parking price is dynamically updated by time monitoring of parking repute through various sensing technology and android communication, the parking payment is on online mode. However, we have designed a time-driven series technique which solves the hassle of the manual parking system. This paper proposes an android utility with cloud network, which is used to put into effect a prototype of Smart Parking System based on Reservation (SPSR) that allows drivers to locate and reserve the vacant parking lots with the help of the wired connected system, with slot allocation method and automatic billing on parking. A set of parking spaces are connected with the central system to find an optimum parking lot. This paper results in a low cost, upgradable automatic parking system for the manual parking system.

Keywords: Smart Parking System (SPS), Android Application, Smart City, Time-Driven series.

12. Introduction

As the increase in population in the metropolitan towns, the usage of motors was

* Corresponding author.
E-mail address: shivnandankhatri01@gmail.com

increased. It is the reason, problems for parking which ends up in traffic congestion, and air pollutants. When we visit the various public places such as purchasing department stores, multiplex cinema & Restaurants all through the weekends it creates a higher parking hassle. Within the recent studies located that a driver takes almost 4 to 8 minutes to park his vehicle because he spends a greater time in looking at the parking lot [1]. Many smart parking structures are looking to reduce traffic and population congestion and enhance the comfort of drivers. In current operating parking guidance structures that acquire available parking area facts from deployed sensors, provide, and indicate the parking zone to drivers [2]. In locating the unoccupied slots, it reasons not only wastage of money but wastage of fuel also. To clear up this hassle, strategies like buffered parking facts sharing are used [3]. However, figuring out the number of buffered areas is hard. If buffered areas are low, there is an issue of "multiple-cars chase-single- space". If a buffer is big then efficient utilization of parking slots is very low. The system will acquire the supply of parking spaces the use of the GPS enabled community using Android application after you have the customers request. The person may have the utility via which he can send a request to server main utility. After that, he can see the empty spaces for parking his vehicle. If a person wants any area, he can pick that space and reserve it for the confined time [4]. If he cannot make there in that point then that area may be available to others. He can once more request for the same vicinity by sending a message then time will be prolonged. The proposed device uses the GPS community and Google maps for locating the space. We need to develop a system where the vehicle itself will display the close to-via parking to the driver. Display all close by parking areas and internal parking slots, the consumer can e-book parking slot from far away location, the system will maintain "Unique QR-Code" to authenticate person, booked parking slot, and so forth. The system will even display predicted time to get a free car-parking zone in case of full parking. After allocating, the space person gets notification approximately leaving the parking area in some time programmed time if a person says accept then the parking area repute might be set to vacant and then that space can be visible to other users [4-6]. These automatic parking allotment systems enhances the parking zones and time consumption to the users & also comfort the drivers and secure system for users in utilizing this technology surveillance of Parking Allotment System.

Fig.1 The Parking zone at Urban Area

Fig.2 The In-Organized parking system

Sensors are positioned under parking areas and are connected to a portal via a less energy wired network. This portal has a broadband connection to a server allowing the devices to be accessed transparently. These are highly inexpensive, not intrusive detectors. They are therefore appropriate for large-scale applications and small energy consumption [7]. Maintenance, however, costs a lot, which requires that messages are sent in an efficient way to ensure battery life for the sensors. Reservation in a suitable automatic parking Lot the drivers can be notified in advance of the closest parking location. Reserving a specific and appropriate parking lot. This can decrease congestion, but it needs an extensive communications system to communicate user's access to locations [8- 10]. Additional functions that enable the customer to determine where the available places can be parked [11-12]. The benefit of such schemes is that they are regarded to be moderate costs, but not fully automatic parking techniques.

The use of Multi-Agent System in Automated Parking System makes it possible for each user to use pre- determined settings to establish various space allocations [13-15]. In making reservations, two parking information systems must be distinguished by their mechanism processing.

13. Literature Review

Vipin Kumar et al. examined the use of detectors and integrated data processors of tracking for centralized vehicle monitoring and monitoring systems are efficient. This version is referred to as primary functional communications approaches, even if primitive interplay protocols. The version aims to provide a facilitating communication structure to organize multi-agent device designs and constructed for use for a simulation of a street map and to estimate the behaviour (to offer statistics approximately the satisfactory routes). The model assists the drivers to get the desired parking spot considering the current state of affairs of site user's characteristics. It offers the expected arrival time and the corresponding distance among begin and an arrival factor. The essential facts are acquired from current user role the use of sensors [16]. Tsaramirsis et.al proposed a scheme the use of IR sensors, microphones, motion detectors. Those sensing devices were used to detect the supply of parking areas. It acquires real-time parking information via their web programs. However, it generates a huge quantity of information ensuing in excessive strength scavenging and communication bandwidth. The main hindrance of the system is high electricity intake and can suffer from technical components [17]. S. V. Srikanth et.al suggested a smart parking system, with comparable features such as remote parking monitoring, automatic steering and reservation storage. Although a prototype feature, the architecture has been suggested that meets the necessity for vehicle parking monitoring systems. [18]. Mingkai Chen et.al studied on parking mechanism and verifiable information technology relying mainly on sensors. This scheme contains nodes for the parking area, tracking nodes, a drain node, an indication for car parks, and a tracking and control hub. Through cellular sensor networks the sensors transmit information, after reading and processing the information, the information and control centre would distribute the parking data by way of an LED screen and shows for the drivers. The study has also shown that the entire system output can meet the parking management demands [19]. In the parking system "Android primarily based smart car parking using lot allocation and reservations" techniques has been followed. In this paper proposes an android utility which is used to put into effect a prototype of smart automobile parking gadget primarily based reservation that lets in driving force to correctly locate and reserves the vacant parking spaces with the assist of IOT with slot

allocation technique and performs automatic billing manner. The proposed machine courses drivers to find available parking area close to them, less number of drivers searching to park, consequently it reduces traffic congestion, it avoids air pollution and global warming, miles scalable study and reliable, it lessens the drivers stress and improve city vicinity, it presents tools to optimize the parking area management, its far correctly discovering the vehicle occupancy in real time. The main contribution of our proposed machine is to discover the popularity of the parking area and gives comfortable parking. This method is similarly prolonged as a fully computerized device the usage of multi- layer parking method, protection measures such as tracing the automobile quantity face reputation of the drivers to avoid the fraud and automatic billing procedure can also be designed [20]. D.J. Bonde et.al has investigated the' Automated android-operated Car Park System.' The motive of this paper is to properly automate car and automotive parking. It discusses the task of presenting a miniature car park model that can alter and control the number of vehicles which are parked in a certain area, based on the availability of car parks at any time. Computerized storage is a route to park and use sensor systems to present cars. An android relying entirely on utility is used for entering or exiting the traffic crowds. The key distinction between our system and the distinct strategy is that we aim to create our scheme more ethical and viable through automation. [21].

Manuel Quiñones et.al studied the design of a Smart Parking System, the basis was the current System integrado de Estacionamiento (SIMERT). In order to make the project feasible, the proposed solution was regarded a technological platform adaptable to urban circumstances and environmentally friendly, giving us connectivity to a web application for integrated parking and usability of administrators and end consumers [22]. Hongwei wang et.al studied that “A Reservation based automatic vehicle parking device” on this method we model and implement a reservation prototype, a completely smart car parking unit, which enables users to locate and book the empty parking space effectively. By means of manner of periodically learning the parking states from the sensor networks, install in parking lots, the reservation carrier is affected via the exchange of physical parking repute. The users can get directly from their own personal conversations on this cyber-physical machine, as shown in Fig.1. The test results indicate that the proposed automatic parking coverage based on reservation has the potential for simplifying the operation of a car park. [23]. Giuliano Benelli et.al stated that the Automatic payment of parking parks using Near Field Communication technology is suggested in this article. In conjunction with the project SIESTA, a Tuscany-financed research project in Italy to examine innovative facilities for visitors who visit art cities, this scheme has been

planned and established. This scheme allows users to create an electronic ticket with their mobile phones to enter and leave the car park and pay for this automatically as an electronic wallet. There have been four applied protocols designed to enforce this scheme: two protocols handle entry and exit procedures for near-by areas and the other two manage user interaction with the kiosk. It also interacts directly with the NFC phone memory for reading / writing assignments and the database that collects data about users ' attendance and payment [24]. Geng, Y et.al, had stated that in metropolitan areas, parking control influences drivers seek time and value for parking spaces, parking revenue, and traffic and population congestion. In order to address issues associated with parking, numerous clever parking systems aiming to satisfy the drivers and parking vendors need have been developed. based totally on the commentary that parking area reservation can assist drivers to lessen the search time dramatically, we are going to layout and put into effect, a prototype for Reservation-based smart Parking device that let the drivers to successfully discover and reserve the vacant parking areas through producing specific QR-Code. With the actual time monitoring of parking reputation thru numerous sensing technologies and android verbal exchange, a clever parking machine will dynamically replace the parking charge in keeping with the physical parking status, and the parking fee will have an effect on motive force's choice on parking slot selection, therefore, have an effect on the repute of parking. With the motivation to alleviate the parking contention, stability the benefits among parking provider carriers and drivers, coordinate amongst carrier companies, differentiate the wishes of character drivers and decrease the number of site visitors looking for allotment of the vehicle.[25]

A.V. Senthil Kumar et.al experimentally investigated that the Parking allotment and scheduling is an important problem in lots of visitors and civilian programs. With the problems of increasing city traffic congestion and the ever-growing scarcity of area, these vehicles parking plenty are needed to be well prepared with automated parking statistics and steerage systems. Desires of smart parking zone management include counting the wide variety of parked cars and to figuring out the location. This painting proposes a new gadget for supplying parking information and guidance using a CCTV camera. The proposed machine consists of figuring out the vehicle quantity plate and identifying the auto range for parking. The system detects motors via CCTV cameras rather than a manual process. (In real time: A digital camera is established on the entry point of the automobile parking space. It captures photograph sequences. The image sequences are then analysed the usage of digital image processing for automobile detection and according to the popularity of automobile occupancy inner, real-time steerage and statistics is supplied to the incoming motive force [26]. The

scheme called TARIFA has been suggested by Georgios Adam et al. [27]. It estimates traffic and road abnormalities as well as makes accessible data to those with Internet access. This scheme is able to detect potholes and can provide traffic data via the GPS receiver. There are two autonomous components to the architecture of the scheme. An accelerometer and GPS receiver are installed in a vehicle on the smartphone. A local database is also available for temporary information storage. A heurist algorithm is used for detection of potholes and other surface defects and those data in processed to feed the information to temporary database

14. Automated Parking System Technique

Most of the parking system mentioned above in the literature is consisting of wired communication system as they have been taken as a standard method to communicate. Briefly saying wired system are much reliable than wireless system which in-turn makes it better choice. In an automated parking system, the system processes the data in various stages which starts as the user of the system opens the app or site for reserving the parking as the user interact with GUI, the requirements of the user is being sent to central control via internet. Once the central control receives the information the process of confirming the data about optimum slots has to be fast in order to make the user's experience flawless, so here communication method can be through wired or wireless within the parking system. The central control updates the information about the empty slot. In this process of confirming the slot information data from various sensors, cloud database and security are collected and processed. When the data on parking space is available the user is sent to payment gateway which mostly contains a third-party protocol. After the payment user is provided with a QR-code which he uses to get access to the parking slot. During this whole process communication between central control and the sensors should be fast and secure. For this reason, many of the system choses to get a wired system just to be on a safer side because few moments wireless system can lag or hacked which can cause insecurity to the user. These are few features that automated parking systems have:

- *GUI:* Consumer/Client is an Android- based structure and this module can be responsible to take input from the user. The GUI evolved in XML in this system devices enter might be taken via this GUI. The software calls for minimal inputs, maximum features, and many others.

- *Communication technique:* The Communication manager handles the customer server communication; we have used the HTTP standard

communiqué technique for verbal exchange. It is based on a stateless, consumer-server, cacheable communications.

- *Protocol* – SMTP, HTTP, Ethernet, Telnet, Gopher all are protocol instances, in which the HTTP protocol is used. The concept is that, rather than using complex mechanisms consisting of CORBA, RPC to connect between machines, simple HTTP is used to make calls between machines.

- *Network services:* For every get First in- First out (FIFO) of the records internet services is needed .The connection between the App and server the HTTP protocol may be used.

- *Calculation zone*: The GPS is integrated into the app that allows you to calculate the precise vicinity of your tool so that you can upload with the complaint. In addition, the user can manually offer the unique deal

- *Distance calculation set of rules*: Calculate Distance from a modern- day place for close by Parking Slot.

- *Suggestion applicant zone:* This suggestion zone allows the user to feed their feedback via the Android-Based Application App or web page. This can be taken into the account and transfer the feedback messages to the main server applicant and it directs to the cloud database and from there to the customer care server zone.

3.1 Infrared Sensor

An infrared sensor is an electronic device that emits in order to sense some of its surroundings. An IR sensor can measure the heat of an object as well as detects the motion. These types of sensors measure only infrared radiation. Usually, all the objects around us radiate some form of thermal radiations. These types of radiations are not visible to our eyes but can be sensed by an infrared sensor. The emitter is simply an IR LED and the detector is simply an IR photodiode. The photodiode is sensitive to Infrared light of the same wavelength as that emitted by the IR LED [33]. When IR light falls on the photodiode, the output voltages and its resistances, due to the change in proportion to the magnitude of the IR light received. These types of sensors can be used to detect common obstacles. In our case, the IR sensor will be used to detect whether the vehicle is placed in a correct position or it needs to be adjusted and accordingly assist the driver to park the vehicle. The sensor will also be

used as a security asset for the system in case of any unusual fidgeting with the vehicle and inform the parking space operator for taking required actions.

3.2 Pressure Sensor

Pressure sensors are used to monitor the pressure of different matters. This sensor is used for an application where the load is a function. These pressure sensors will be used to monitor the existence of the vehicle and to monitor the current location in parking space. These will be an intricate part of the security system, as they will send the respective data to the on-board diagnostic system; whenever any suspicious activities are captured, the security alert will be deployed. This sensor will also be used to weigh the vehicle to determine the vehicle kind just to crosscheck the tickets with the existing customer details.

3.3 Photoelectric Sensor

A photoelectric sensor is used to discover the distance, absence, or presence of an object by using a light transmitter. These sensors will be integrated with the differently positioned infrared sensors to make the parking more precise and easier for the driver

3.4 Ultra-Sonic Sensor

The ultrasonic sensor works at the concept of SONAR (Sound Navigation and Ranging) i.e. sending a sound pulse of ultrasonic waves of frequency 45 kHz even as a sign of pulse width extra than 10ms is applied at the purpose input of the sensor. On the occurrence of an obstruction, the ultrasonic waves are considered and are sensed which in turn generates a pulse at the Echo pin of the ultrasonic sensor [34]. The width of this pulse at the echo pin is measured in order to discover the distance.

15. System Architecture

After reviewing many designs and methodologies the system has taken a brief structure that satisfies our needs and serves the purpose. The system mentioned below describes how we scan make the design better, simpler and more intricate part of the system. The process starts from logging into the user GUI. The app server validates the user data and compiles it with user database. Once the validation is done the user is connected to the main sever which provides allocation details available at different locations. The allocation details are confirmed by using hardware at parking lots. If

every detail is being confirmed and system allows for allotting a free space then the user is directed to payment gateway. This payment gateway is provided by a third-party application. The payment is chosen according to security provided at the end and it even secures user's private credentials.

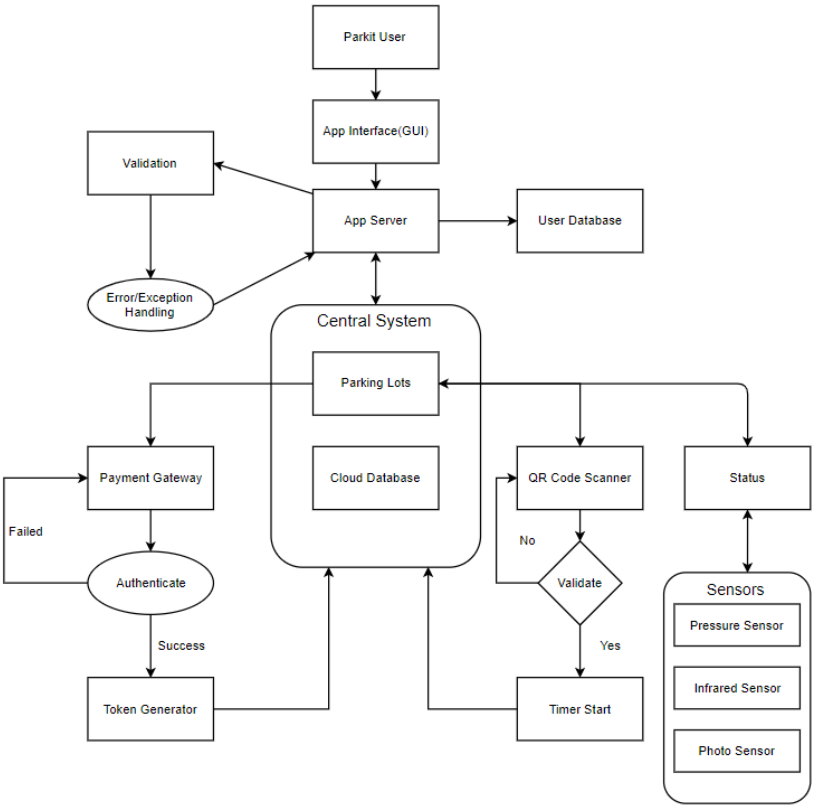


Fig.3 Individual Parking-Lot Architecture

4.1 Guidance in Parking Allotment Information Systems (GPAIS)

The PGI is a parking technological system that offers data and data regarding the accessibility of parking space in significant cities. Vehicle detectors are mounted at entrances, exits and/or parking spaces to collect and calculate the amount of spaces occupied and accessible.

4.2 Data information based on transit system

This scheme specifically offers parking space data and public transport schedules in Park and Ride centres. The primary aim of the scheme is to encourage commuters to park their cars and use buses or trains for their transit. This in turn will decrease traffic congestion, pollution and fuel usage. Vehicle detectors are employed similar to

GPAIS. Messages are then displayed on variable message signs along highways leading to park and ride lots.

4.3 E-parking

E-parking Uses state-of the-art technologies to combine streamline reservation systems and payment systems. This scheme allows a Driver to check accessibility, to reserve parking room for a particular location and to pay when user leaves. Access To the scheme is through cellular, PDA and/or web. To detect approaching cars, standard detectors are still necessary. The System should, therefore, be able to define and provide access to reserved room to clients and/or cars. Confirmation Code access to the client on the cell phone can be used to identify the parking lot.

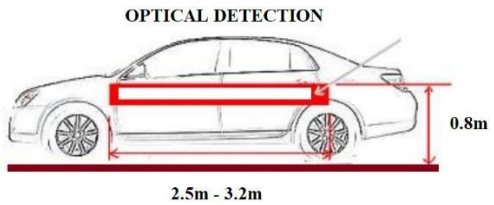


Fig. 4 Vehicle positioning and measuring

Delay & travel time Congestion size method needs to have the "amount of extra a time that is required to make an experience because of congested conditions at the roadway." It examines how fast travel can occur at some stage in the height length by way of specializing in time rather than pace. TRI is a trademark for complete portions of the analysed network based on the respective distance and number of cars supported through each phase [30-32].

4.4 Congestion traffic methodologies and inconsistencies

Congestion is a time that takes for travel or delay in excess of the normally incurred beneath mild or free waft travel situations (Lomax et al., 1997). Unacceptable congestion journey a time or postpone in excess of an agreed-upon norm. The agreed-upon norm may additionally vary with the aid of the type of transportation facility, tour mode, geographic region, and time of the day. The authors of the examiner performed the usage of the U.S. Census records to analyse the unacceptable congestion, concluded that the unacceptable congestion is whilst less than 1/2 of the populace can travel to paintings in much less than 20 minutes or if extra than 10% of the populace can trip to work in greater than 60 minutes or an hour. Mohan Rao A. et al. Measuring city Traffic Congestion – A evaluate in Time & Travelling extra of that

typically incurred beneath mild or unfastened-for journey situations. Traffic congestion is travel time or delays in extra of that typically incurred under light or free- for journey situations. Congestion is the presence of delays along a physical pathway because of the presence of different users [28-29].

Sl no.	Parameters	Wired communication	Wireless communication
1.	Security	Most of the wired communication methods are more secure because all the transmission take place inside the secured wire and hacking the transmission is not easy	Wireless connection provides less security as it can be interrupted and packets can be stolen
2.	Reliability	Data loss is very rare in wired communication	Data loss occurs frequently as a wireless communication depends on many factors such as range, version of wireless card
3.	Power consumption	Wired communication consumes less power because it doesn't use transmitters and directly transfers data only using wires	More power is consumed because of additional devices as transmitters and amplifiers
4.	Response time	Most devices give maximum available data transfer rate in wired connection	Response time depends on how far devices the area. Its response time is always greater than a wired connection
5.	Installation Cost	Professional skill is not needed to install	Cost depends on the area of the system. Larger area requires more devices due to range limits of particular device
6.	Efficiency	Overall when optical fiber is used the life time fatigue for the system will be less it tends to more reliable	Operating cost is high and tends to be less secure

Table 1.1 Comparison between wired and wireless communication

16. Discussion

In a country where most of the automated parking system is costly and reason behind why people are not opting for new gen parking to make things easier. The system described in this paper is a Modular system which means the same system can be used to park different types of vehicle. The development cost and complexity will be much

lesser than any system currently available. The idea is to utilize almost of the parking lots which are not in use for whole time such as movie theatre, railway station parking, government areas, and university parking during season break. Most of these are used only in few peak hours other moment they will be quite free. This system can make use of this free parking hours and this will be beneficial to the associations too. This will improve the land usage efficiency. The space sharing will even help to reduce the traffic congestion as there will be less vehicles searching for the parking space and in addition this will even help to decrease the CO_2 production.

The usage of different sensors will help the drivers to park their vehicles safely and to keep a track of the vehicle. The usage of different types sensors will help to keep vehicle in the safe side as the error with any of the sensor will trigger the system to put the alarm and inform the customer. The bike parking system will be a having the latch auto-lock system which will offer a carrying space for keeping their gears which is always been on demand but never added to any system. This will improve the usability of the space. There are possibilities to make use of environmental lighting to guide the consumer to the allotted space. And the automated barricade system which will be intricate part of the system, will guide the consumer to their space and reduce the human effort.

17. Results

The pictures shown below explains the working of the full working of the parking allotment system. The process starts from collecting the user's basic info and ends with a unique generated code which allows the user to enter the parking lot. The process carried on is similar to shown in the flowchart (fig. 3) above.

When the user is logged into the website, user will be asked some basic info such as Name, location, contact no. , etc. These information will be saved in the website user database. The saved data can also be utilized for keeping a record of user travel log and improve the system.

Fig.5 Information shared by the user while post-registration

Once the general information is filled-up and saved then the user is able to select their current location which will help the system to suggest the nearest parking space available. The user is also provided with option of selecting the desired parking location and space.

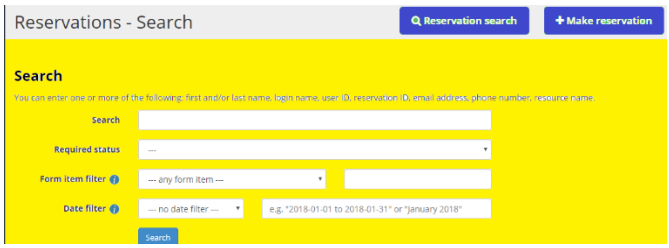


Fig.6 Search bar to check the availability of the parking space

Updated location and other filters applied will open a new window with availability information. The windows will also have the data such as availability time, number of vacancies, and date of availability. The user can select from the given options of dates, places and slot.

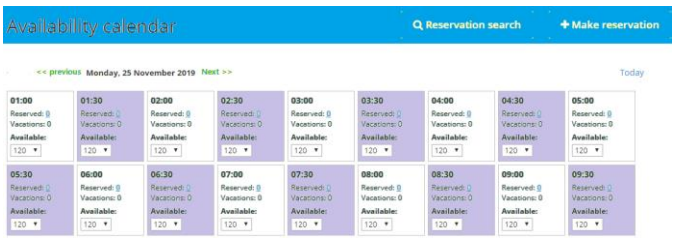


Fig.7 Availability chart showing available parking space

To make the reservation the new window will ask for the user information once again. This option is provided when the user is booking the parking space for another person is not a user. The window will also provide option to autofill the existing user information from the database. The user is kept confidential and can be used in case of any emergency.

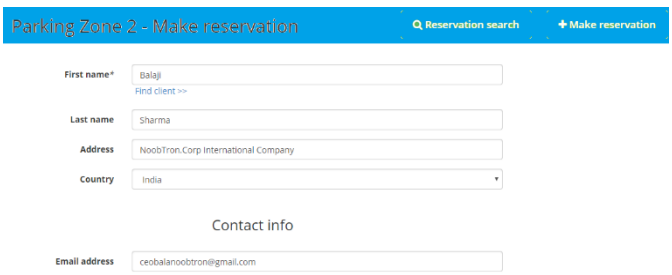


Fig.8 User basic details for making a reservation

For contact and sending the ticket the user has to fill the email-id and contact number which is again given for new or guest users.

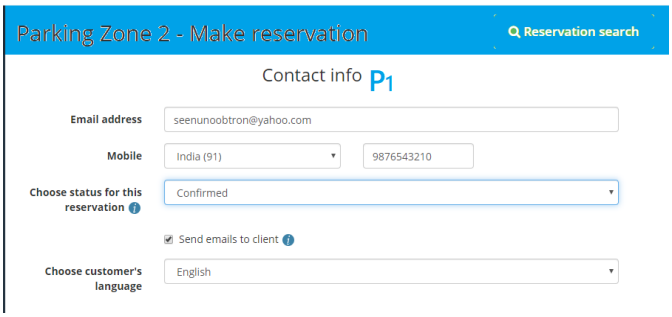


Fig.9 Contact details provided by the user

When the vehicle reaches the destination user have to show the provided ticket and enter the parking space. The existing user can enter the parking lot by entering the reservation credentials

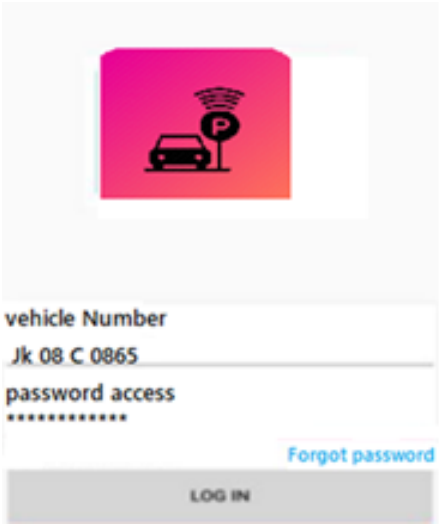


Fig.10 Reservation credentials

The number of user and their basic information related to parking is being saved and tracked by the website. The data is being used to improve the system efficiency and way of working from analysing the delay timing, waiting time, frequency of use etc.

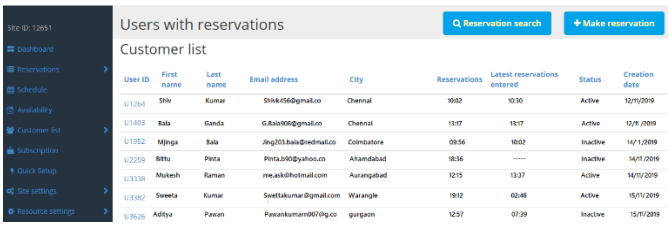


Fig.11 Database of all Allotted Users

Actively the website shows where the space is available and reserved. The parking spaces are differentiated and marked according to the usage.

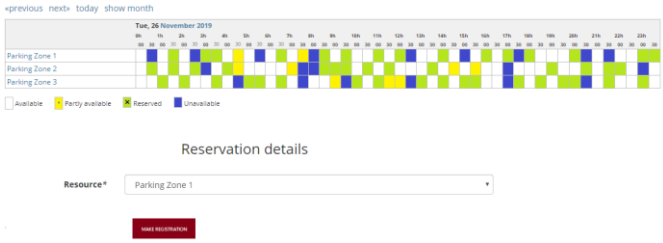


Fig.12 occupancy of the parking spaces

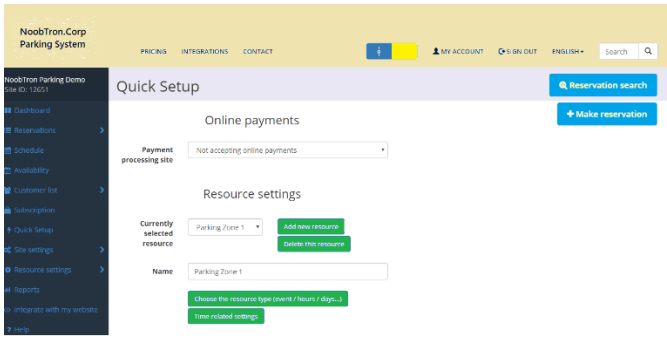


Fig.13 Payment gateway before confirming the lot for individual

Histogram is being prepared from the operating hours collected data. This data can help the system to calculate the operating frequency and also the parking system growth rate estimation.

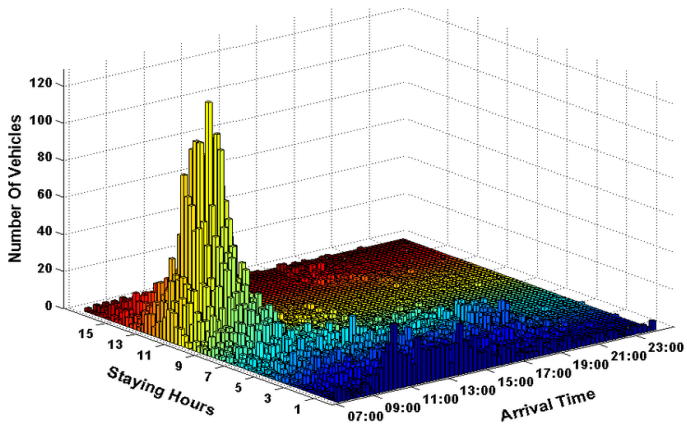


Fig.14 Histogram showing the maximum operating hours allotted for the user

Fig.15 Reservation status of parking space

18. Conclusion

As India is a developing country, we need solutions which will do the work and at the same moment it should be cost efficient. Our proposed system will decrease the cost of parking and the effort customer has to make for manual parking system. When talking about technological advancements the system will be capable of Integrating on its own like Building blocks and will be a part of smart city. Different types of vehicles will be able to use this system such as two-wheelers, 4-wheelers, trucks, buses etc. because of its capability to expand and integrate and easy accessibility. The system will be in approach to customers as it will be easy to install in any type of parking space even in outdoors which make it better in almost every manner for Indian society.

19. Future Work

As the number of user increases, adding more sensors at every park place and assign separate ID for sensor and car is required. It will help us to check whether particular car is parked at particular place properly or not, and it is beneficial to generate proper rate/fare calculation. If the car Id and the sensor id are not matched, it will indicate through alarm. It is also very time effective for the user to avoid the headache of searching for the vacant place in parking place. Safety system to avoid the vehicle theft is to be provided. We can use the finger Print recognition at the out gate to check the particular user and the car ID correctly. A set of Barricade will be couples with the system so that when user enters the parking space. Barricade will guide the user into particular space where the user has to park his vehicle. Moreover, as the user walks away from the park the system will automatically engage the security systems. The lighting system will be made smart so that it can utilize the electricity efficiently and illuminate the place at the same time. An image recognition system can be made into

the system so that at least basic information of the vehicle can be saved for any emergency use.

20. Acknowledgement

We would like to express our sincere gratitude and thanks to Dr. Ashish Selokar, Accendere, C.L. Educate Ltd., New Delhi, India for his valuable suggestions. We express our sincere gratitude and thanks for his helps in the whole duration of executing this research work.

References

[1] Yanfeng Geng, "New Smart Parking System Based on Resource Allocation and Reservations", IEEE Transaction on intelligent transportation systems, Vol. 14, Issue No. 3, Sept 2013, pp. 1129-1138

[2] Jihoon Yang, Jorge Portilla and Teresa Riesgo, "Smart Parking Service based on Wireless Sensor Networks", IEEE Magazine, Feb 2012, pp. 6029-6033.

[3] Benenson, K. Martens and S. Birr. 2008. "Parkagent: An agent-based model of parking in the city," Computer. Environ. Urban Syst. Vol. 32, no. 6, pp. 431–439, November.

[4] K. Cheung and P. Varaiya. 2007. "Traffic surveillance by wireless sensor network: Final report," Univ. California, Berkeley, CA, USA, Tech. Rep.UCB-ITSPRR-2007-4.

[5] S. Chou, S. Lin and C. Li. 2008. "Dynamic parking negotiation and guidance using an agent-based platform," Expert Syst. Appl. Vol. 35, no. 3, pp. 805–817, October.

[6] S. Anandamurugan and C. Venkatesh, "Reducing Power Consumption in Wireless Sensor Networks by using Terrain Based Probabilistic Forwarding Scheme (TBPFS)," Journal of Computer Science, vol. 10, no. 2, pp. 208–217, 2010.

[7] R. Lu, X. Lin, H. Zhu, and X. Shen, SPARK: A New VANETBased Smart Parking Scheme for Large Parking Lots. IEEE, Apr. 2009.

[8] G. Yan, S. Olariu, M. C. Weigle, and M. Abuelela, "Smart Parking: A Secure and Intelligent Parking System Using NOTICE," 2008 11th International

IEEE Conference on Intelligent Transportation Systems, pp. 569–574, Oct. 2008.

[9] S. Srikanth, P. Pramod, K. Dileep, S. Tapas, M. U. Patil, and S. C. Babu N, "Design and Implementation of a Prototype Smart Parking (SPARK) System Using Wireless Sensor Networks," International Conference on Advanced Information Networking and Applications Workshops, pp. 401–406, May 2009.

[10] S.-Y. Chou, S.-W. Lin, and C.-C. Li, "Dynamic parking negotiation and guidance using an agent-based platform," Expert Systems with Applications, vol. 35, no. 3, pp. 805– 817, Oct. 2008.

[11] Z. Pala and N. Inan, "Smart parking applications using RFID technology," International Journal, pp. 21–23, 2007.

[12] G. Ivan, M. O"Droma, and M. Damien, "Intelligent Car Parking Locator Service," Information Technologies and Knowledge, vol. 2, pp. 166–173, 2008.

[13] D. Teodorovit and M. Dell"Orco, "Multi Agent Systems Approach to Parking Facilities Management," in Applied Research in Uncertainty Modeling and Analysis. Springer, 2005, pp. 321– 339.

[14] B. Juan and A. Muñoz, "Developing an Intelligent Parking Management Application Based on Multi-agent Systems and Semantic Web Technologies," Hybrid Artificial Intelligence Systems, pp. 64–72, 2010.

[15] Vipin Kumar Verma, Rahul Chaudhari, Siddharth Kumar Singh, Tapas Mishra, Pankaj Srivastava, "Intelligent Transport Management System using Wireless Sensor Networks", IEEE Intelligent Vehicle Symposium Elndhoven University of technology Elndhoven, Netherlands, June 4-6, 2008

[16] Tsaramirsis, G., Karamitsos, I. and Apostolopoulos, C., 2016, March. Smart parking: An IoT application for smart city. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 1412-1416). IEEE.

[17] K. P., Tapas S., Mahesh U. Patil, Sarat Chandra Babu N, "Design & Implementation of a Prototype Smart Parking (SPARK) System using

Wireless Sensor Networks" International Conference on Advanced Information Networking & Applications workshops, 978-0-7695- 3639-2/09, 2009 IEEE.

[18] Mingkai Chen, Tainhai Chang, "A Parking guidance & Information system based on Wireless Sensor Networks", IEEE International Conference on information & Automation Shenzhen, China, June 2011

[19] Renuka, R. and Dhanalakshmi, S., 2015. Android-Based Smart Parking System Using Slot Allocation and Reservations. ARPN Journal of Engineering and Applied Sciences, 10(7), pp.3116-3120.

[20] Bonde, D.J., Shende, R.S., Kedari, A.S., Gaikwad, K.S. and Bhokre, A.U., 2014, January. Automated car parking system commanded by Android application. In 2014 International Conference on Computer Communication and Informatics (pp. 1-4). IEEE.

[21] Quiñones, Manuel, et al. "Design of a smart parking system using wireless sensor network." 2015 10th Iberian Conference on Information Systems and Technologies (CISTI). IEEE, 2015.

[22] Wang, H., 2011. A reservation-based smart parking system.

[23] Grodi, Robin, Danda B. Rawat, and Fernando Rios-Gutierrez. "Smart parking: Parking occupancy monitoring and visualization system for smart cities." SoutheastCon IEEE, 2016

[24] Geng, Y. and Cassandras, C.G., 2011, September. Dynamic resource allocation in urban settings: A "smart parking" approach. In 2011 IEEE International Symposium on Computer-Aided Control System Design (CACSD) (pp. 1-6). IEEE.Kumar, A.S. and Vineeth, I.V., an Intelligent Car Parking Allotment System Using CCTV Camera.

[25] D. G. Georgios Adam and I. Oikonomidis, "Tarifa: Traffic and abnormalities road instructor for anyone," in 2nd Student Workshop on Wireless Sensor Networks, Athens, Greece, October 2009

[26] Lomax, S.T.T.; Turner, S.; Shunk, G.; Levinson, H.S.; Praž , R.H.; Bay, P.N.; Douglas, G.B. 1997. Quantifying congestion, Volume 1, NCHRP Final Report 398. Washington, D.C.: Transportation Research Board. 108 p.

[27] Kockelman, K. 2004. Traffic congestion. In Monograph Handbook of transportation engineering. 32.

[28] Adler, J. L., & Blue, V. J. (2002). A cooperative multi-agent transportation management and route guidance system. Transportation Research Part C, 10, 433–454.

[29] Luh, P. B., Ni, M., Chen, H., & Thakur, L. S. (2003). Price-based approach for activity coordination in a supply network. IEEE Transactions on Robotic and Automation, 19(2), 335– 346.

Real Time Hand Gesture Recognition Using Template Matching

Clementraj.G[a],Shivakumar.N[b*]

[a]*Thiagarajar collge of engineering, Madurai 625015, India*

bThiagarajar collge of engineering, Madurai 625015, India

Abstract

Hand gesture is a common technique for non-verbal communication human being. For specialized person (especially deaf people) the communication between normal people is complex one. This System provides a prototype system that helps to recognize hand gesture of the specialized people to normal people. The main focus is in the problem of gesture recognition in real time in video streaming. The sign language used by the community of deaf people is mostly ASL (American Sign Language) including the alphabet. For the poor contrast image this model can be applied to get an accurate result. The problem solving will meet by the following module Detection of skin, Segmentation of an image, Filtering of image, and applying some Template Matching techniques. This system recognizes gestures of ASL which also in poor contrast.

Keywords: Gesture recognition; ASL; Template Matching; Gray World; Correlation

21. Introduction

Basically the deaf people communicate through sign language. However, research in sign language recognition is highly influenced by gesture recognition research, as sign language is a form of communicative gesture. Therefore, when reviewing literature in sign language recognition, it is also pertinent to study literature on gesture recognition. Gestures and sign language recognition includes the whole process of tracking and identifying the signs performed and converting into semantically meaningful words and expressions. Ordinary people communicate their thoughts through speech to others, whereas the hearing impaired community the means of communication is the use of sign language and ASL (American Sign Language) is 3rd most used sign language. This language can be recognised, detected and classified by

* Corresponding author.
E-mail address: shiva@tce.edu

hand gesture techniques which can help in to communicate more easily with others and to interact with some computer applications.

1.1 American Sign Language

American Sign Language (ASL) is a visual language. With signing, the brain processes linguistic information through the eyes. The shape, placement, and movement of the hands, as well as facial expressions and body movements, all play important parts in conveying information. Sign language is not a universal language – each country has its own sign language and regions have dialects, much like the many languages spoken all over the world. Like any spoken language, ASL is a language with its own unique rules of grammar and syntax. Like all languages, ASL is a living language that grows and changes over time.ASL is used predominantly in the Unites States and in many parts of Canada. ASL is accepted by many high schools, colleges and universities in fulfilment of modern and foreign language academic degree requirements across the Unites States. It is the most well documented and most widely used language in the world. American Sign Language (ASL) is a complex visual-spatial language that is used by the Deaf community. It is a linguistically complete, natural language. It is the native language of many Deaf men and women, as well as some hearing children born into Deaf families.

1.2 Computer Vision

Computer vision is described as the eyes of the computer. It is a scientific field which relates with how a device can gain high level of understanding of information or messages from the digitalised images and videos. It includes acquiring, analysing and processing of digital images and extraction of the features from the digital videos.

1.2.1 Application of Computer Vision

The following are some of the field where computer vision applied.

- Military Sector.
- Medicinal Field.
- Non-Verbal Communication.

1.3 Gesture Recognition

Gesture is a human activity to express the thought in the mind. It also used to interact with the virtual world. In virtual reality, the interaction between the object in virtual

world and the user in the real world is done by the gesture. Gesture of human is of two types,

- Gesture obtained from human face.
- Gesture obtained from human hand.

1.3.1 Facial Gesture

Facial gesture is the gesture obtained from the face of the human. Based on some controversial theories, it expresses the emotional state of each and every one. In this, the eyes are played a vital role in facial gesture. It's done by blinking rate of the eyes and skin tone of the face. Even though everyone has unique face the recognition gets little complex.

1.3.2 Hand Gesture

Hand gesture is the form of the communication in Non-Verbal between the people who want to express their thoughts. The Hand Gesture can be expressed in two ways,

- Single Handed.
- Double Handed.

22. Literature Survey

[1] In this paper the authors Khetarpal, Poras, Neeraj Kumar, and Nikita Rai presents a model that uses segmentation technique to recognise the static hand gesture of the human. To read the hand gesture from the human, Sensor with gloves has been attached to measure the position of the finger. The gesture image is taken and pre-processed first. That image is known gone under process of segmentation by thresholding methods. Background Subtraction method is used to subtract the background image from the hand to get a clear image of hand gesture. After that the image is to be filter to reduce noise of the hand gesture image by Mean and median filter. Now the image is compared with the database and gives the result.

[2]In this paper the authors Frey, Gabriel, Arno Jurkschat, Safak Korkut, Jonas Lutz, and Rolf Dornberger describes the development of virtual reality played a well place in computer era. The hand gesture is not only for the sign language it also for the interaction of the objects in virtual reality. The intuitive hand gesture is detected. A set of intuitive gestures when interacting with information visualization were developed. It improves a touch less interaction with the virtual reality environment.

Based on the VISM (Visual Information Seeking Mantra) framework the model was developed. Only few gestures were developed users cannot remember the developed gestures. Gestures may be some familiar one.

[3] In this paper the authors Li, Feng, and Jun Fei describes about the gesture recognition model for hand gesture. Multi-sensor information fusion model of the virtual environment for gesture recognition were developed. Multi sensor used to sense the gesture input. By the fusion of multi sensor with the information gives a great result. Based on image fusion a gesture recognition algorithm was proposed. By this algorithm hand gesture is detected. Multiple gestures have same starting or ending point. In segmentation method the algorithm fails.

[4] In this paper the authors Jang, Youngkyoon, Ikbeom Jeon, Tae-Kyun Kim, and Woontack Woo describes that it deals with the recognition of metaphoric hand gestures. These metaphoric hand gestures are used for the interaction of virtual objects presented in virtual environment. An interaction model within virtual environment is developed. For detection of these special kind of hand gesture a 3D palm post estimator and a static-dynamic (SD) gesture estimator were used. These estimators find the gesture and recognized accurately. The transition of static and dynamic metaphoric hand gesture estimations from function-equipped AR/VR objects to objects not in real time objects. In visual recognition process, the method fails when occlusion happens during process.

[5] In this paper the authors Ercan, M. Fikret, and Allen Qiankun Liu describes that it deals with Hand gesture recognition to improve user experience for VR applications. The model has the ability to see the environment in the virtual environment. There is no need of removing VR headset to see the environment. By controlling the source of image stream using hand gestures to interact with virtual environment. It is computationally costly to implement and FPS drops radically on process which makes it difficult to apply for a VR application.

[6] In this paper the authors Wu and Xiao Yan describes that it deals with a supervised hand gesture model. By using DC-CNN (Double Channel Convolutional Neural Network).Two CNN used both are independent. For first CNN input was gesture image after preprocess and second input was edge gesture. For edge detection of the hand gesture canny edge detection algorithm is used. The rate of dynamic gesture recognition still has to improve. Use of supervised learning data, dependency is high on large number of tag data. In future, this model can develop as semi-supervised and unsupervised.

[7] In this paper the authors Sethi, Rajni, and Sreedevi Indu describes that it deals with the recognition of underwater images which posses poor contrast and color. To recognize such type of images in underground gray world based algorithm was used. SLIC (Simple Linear Iterative Clustering) Algorithm was used for segmentation process. The drawback of this model is the SLIC segmentation needs to be done for each images. In lot of water on background false colors are recognize.

[8] In this paper the authors Ma, Chunbo, Xuewei Lv, and Jun Ao describes that it deals with the biggest challenge in an image processing is to detect the noise and remove. To remove such type of noise median filter was used. It removes random value impulse noise in images. The drawback of this model is the median value calculate in median filter is altered even though it is not corrupted.

[9] In this paper the authors McBride, Timothy James, Nabeel Vandayar, and Kenneth John Nixon describes that it deals about the comparison of three algorithms used for skin detection .YCbCr, RGB, KNN algorithms are compared among themselves. All the three algorithms have implemented in skin detection technique to compare them with some parameters. Even though RGB gives high accuracy but it is inconsistent. YCbCr gives moderate accuracy but it is consistent.

[10] In this paper the authors Wang, YongHong, XiZuo Dan, JunRui Li, SiJin Wu, and LianXiang Yang describes that it deals with a multi perspective digital image is evaluated using correlation method. The images were taken in 3d camera. The 3 CCD (Charged Coupled Device) cameras were used. The main drawback is the model requires high mono-chromaticity of the band-pass filters to separate the images into three sub-images perfectly.

23. Motivation

In this approach, Segmentation technique is used to recognise the static hand gesture of the human. To read the hand gesture from the human, Sensor with gloves has been attached to measure the position of the finger. The gesture image is taken and pre-processed first. That image is known gone under process of segmentation by thresholding methods. Background Subtraction method is used to subtract the background image from the hand to get a clear image of hand gesture. After that the image is to be filter to reduce noise of the hand gesture image by Mean and median filter. Now the image is compared with the database and gives the result. For the normal image this approach was a good one but for the poor contrast image, the image

of the finger cannot pre-process accurately. So that, during Background Subtraction technique provides a worse result on recognition of hand gesture.

24. Proposed Approach

The proposed approach uses Template Matching technique to recognise the static hand gesture of the human. A real time webcam or an inbuilt camera is used to read the hand gesture from the human. The gesture image is taken from the real time streaming video and pre-processed by YCbCr method. And by using Gray world Algorithm the image is converted to gray, there is no subtraction of image so that no pixel can lose. After that the image is to be filter to reduce noise of the hand gesture image by median filter. Median filter filters the noise by taking nearest pixel value and find median. Now the image is compared with each and every pre-processed hand gesture database. The template matching method of correlation provides a result of correlation coefficient value as the result. The correlation coefficient value is calculated by comparing similar images in database with the current image. For the normal image and the poor contrast image, the image of the finger can pre-process accurately. So it gives better result on recognition of hand gesture.

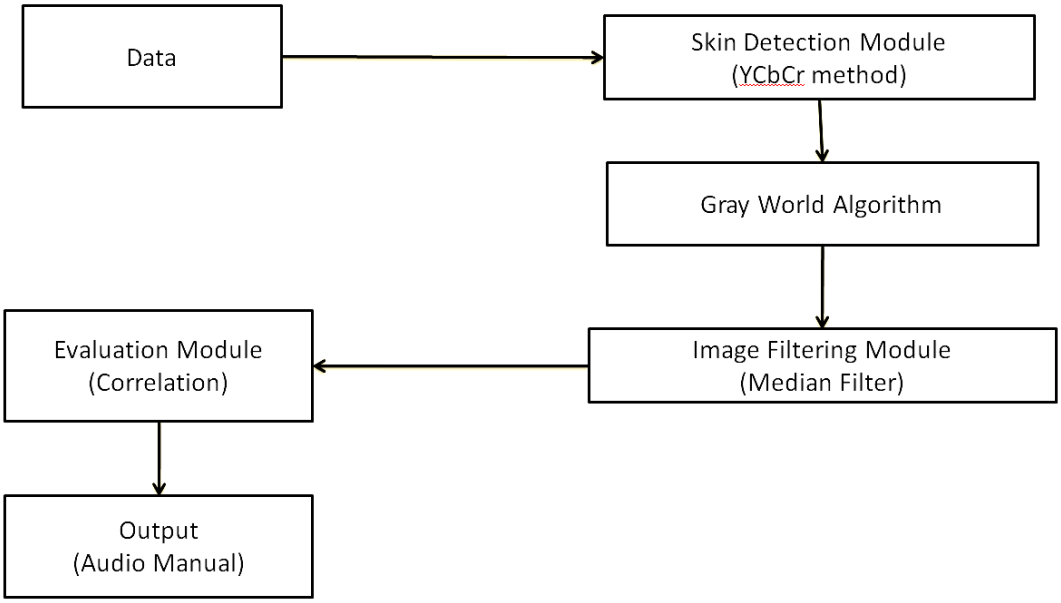


Fig. 1. Architectural Design

4.1 Dataset

The data set has been collected from the online database called kaggle. This kaggle dataset is just like a github where there are many types of datasets will be available.

The dataset taken consists of a two folders, one is of gesture image and another is of the pre-processed gesture image. The gesture image available is representing the alphabets ‘A’ to ‘Z’ and the numeric values ‘0’ to ‘9’ totally a 36 unique gestures. For reference the image of gesture is taken 1500 images for the recognition process and each image is pre-processed also. Therefore the total number of images in dataset is 56K in gesture image folder and another 56K in pre-processed folder. Totally nearly a 1K dataset is taken for the recognition process.

4.2 YCbCr

YCbCr is one of the colour spaces just like RGB. When digital media like cameras and laptops need to be use colours, it needs to be presented in numbers. Therefore the colour space is a set of rules that allows numbers to describe about colours. Y represents the luma component of the colour. Luma component is nothing but the brightness of the colour. That means the light intensity of the colour. The human eye is more sensitive to this component. Cb and Cr are the blue component and red component related to the chroma component. That means Cb is the blue component relative to the green component. Cr is the red component relative to the green component. These components are less sensitive to the human eyes. Since the Y component is more sensitive to the human eye, it needs to be more correct and Cb and Cr are less sensitive to the human eye. Therefore it needs not to be more accurate. When in JPEG compression, it uses these sensitivities of the human eye and eliminates the unnecessary details of the image.

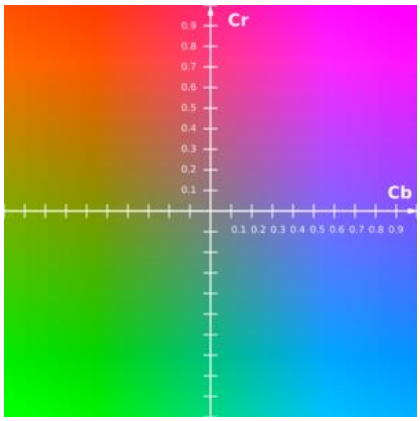


Fig. 2. Cb and Cr color composition

4.3 Gray world Algorithm

A white balance method that assumes an image as a neutral grey is referred as the Gray World Algorithm. The assumption of gray world will hold if a perfect

distribution of colors found in the image. Let us assume that the image is at a perfect distribution of colors in the image, the color which is averagely reflected is assumed to be the light color which is reflected. Therefore, by looking at the average color and comparing it to gray. The estimation of the illumination color will be calculated. By computing the each channel mean value of the image, the Gray World algorithm produces an illumination estimation results. There are many methods for the normalization. One of the methods of normalization is that the mean of the components of three was used as an illumination estimate of the image. The image is normalized by using the components channel i and the pixel value. It is scaled by dividing the average value of the channel mean to the average value of the illumination estimated value.

$$r_i = \max(avg_R, avg_G, avg_B) / avg_i \qquad (1)$$

4.3 Correlation

Correlation is one of the basic operation by which the features or the information can be extracted from the given image. It is a simple operation but the information gathered by this is extremely useful for image processing. This method correlates the information of the image or the features extracted from the image with the other images and find the similar items. After the images found, it estimates a correlation coefficient value among the images. The correlation coefficient value gives the accuracy of the recognized image.

$$F \circ I(x) = \sum_{i=-N}^{N} F(i) I(x+i) \qquad (2)$$

where the circle denotes correlation. With this notation, we can define a simple box filter as:

$$F(i) = 1/3 \quad \text{for } i = -1,0,1$$

$$0 \quad \text{for } i \neq -1,0,1$$

25. Experiment And Results

The results of each phases of the process is described and depicted below in the sequential order of their implementation.

Fig. 3. Sample Dataset of alphabet

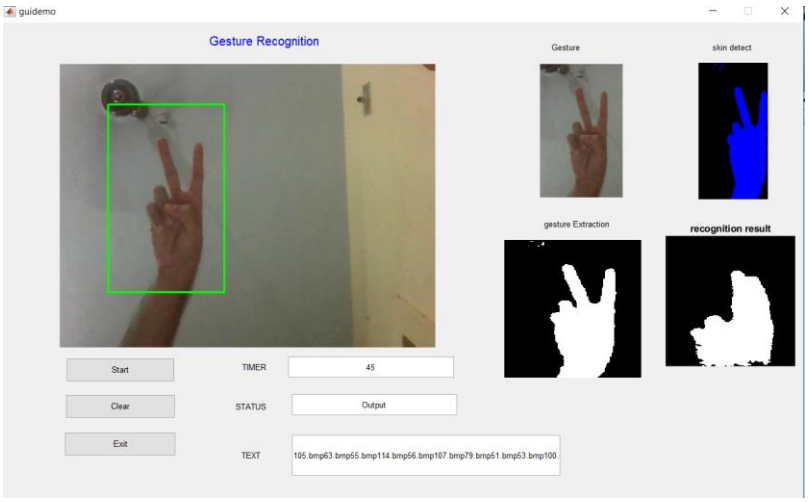


Fig. 4. Capture of image using digital camera

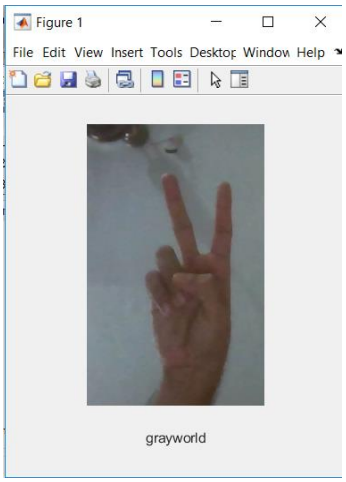


Fig. 5. Original gesture image at poor contrast

26. Conclusion And Future Work

In this paper, the recognition of the static hand gesture from a real time streaming video. This model is designed in way that a poor contrast image can also recognized accurately by implementing the gray world algorithm. The median filter is reducing the noise as much as possible which helpful in the noiseless image. To take the frames from live video the timer is set as 45 seconds. The user must wait up to 45 seconds it will reduce in future. The hand gesture produces a accuracy of 75% by introducing it in the neural network like CNN, ANN and KNN gives more accuracy and also reduces the programming part.

References

[1] Khetarpal, Poras, Neeraj Kumar, and Nikita Rai. "Real-Time Static Hand Gesture Recognition Using MATLAB." In Applications of Computing, Automation and Wireless Systems in Electrical Engineering, pp. 1149-1158. Springer, Singapore, 2019.

[2] Frey, Gabriel, Arno Jurkschat, Safak Korkut, Jonas Lutz, and Rolf Dornberger. "Intuitive hand gestures for the interaction with information visualizations in virtual reality." In Augmented Reality and Virtual Reality, pp. 261-273. Springer, Cham, 2019.

[3] Li, Feng, and Jun Fei. "Gesture recognition algorithm based on image information fusion in virtual reality." Personal and Ubiquitous Computing (2019),Springer: 1-11.

[4] Jang, Youngkyoon, Ikbeom Jeon, Tae-Kyun Kim, and Woontack Woo. "Metaphoric hand gestures for orientation-aware VR object manipulation with an egocentric viewpoint." IEEE Transactions on Human-Machine Systems 47, no. 1 (2016): 113-127.

[5] Ercan, M. Fikret, and Allen Qiankun Liu. "Hand Gesture Detection and Its Application to Virtual Reality Systems." In 10th International Conference on Robotics, Vision, Signal Processing and Power Applications, pp. 499-504. Springer, Singapore, 2019.

[6] Wu, Xiao Yan. "A hand gesture recognition algorithm based on DC-CNN." Multimedia Tools and Applications (2019),Springer: 1-13.

[7] Sethi, Rajni, and Sreedevi Indu. "Local Enhancement of SLIC Segmented Underwater Images using Gray World based Algorithm." In 2017 Ninth International Conference on Advances in Pattern Recognition (ICAPR), pp. 1-6. IEEE, 2017.

[8] Ma, Chunbo, Xuewei Lv, and Jun Ao. "Difference based median filter for removal of random value impulse noise in images." Multimedia Tools and Applications 78, no. 1 (2019): 1131-1148.

[9] McBride, Timothy James, Nabeel Vandayar, and Kenneth John Nixon. "A Comparison of Skin Detection Algorithms for Hand Gesture Recognition." In 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), pp. 211-216. IEEE, 2019.

[10] Wang, YongHong, XiZuo Dan, JunRui Li, SiJin Wu, and LianXiang Yang. "Multi-perspective digital image correlation method using a single color camera." Science China Technological Sciences 61, no. 1 (2018): 61-67.

A Survey on Automatic Plant Identification Using Convolutional Neural Network and Transfer Learning

Rini Euchrist.J[a],Sivakumar.M[b*]
[a]*Thiagarajar College of Engineering,madurai 625015,India*
[a]*Thiagarajar College of Engineering,madurai 625015,India*

Abstract

Deep learning is an effective field derived from the area of artificial intelligence that deals with machine learning concepts. It is used to process heavy data along various neural networks like artificial, convolutional, deep etc., Deep learning approaches are used widely across various language processing, voice recognition, image processing, medical image analysis etc. It is used to extract high level features across multiple layers from the given input. Along the computational power of the system, there is an increase in dataset exponentially where deep learning plays a major role along processing increased dataset.

Keywords: NN-Neural Network;CNN-Convolutional Neural network;BOW-bag of words; RELU-Rectified Linear Unit;

1. Introduction

Image processing is a state of art technique on converting an image into a digital form to extract some useful information. The purpose of image processing includes various stages like visualizing, image sharpening, retrieval of image, detecting the patterns and finally recognition to identify different objects. A neural network is a connection of neurons, a network composed of artificial nodes. It is an information processing system similar to the biological neural networks that processes data in form of neurons. A simple neural network consists of a set of neutrons where one node is connected to the other and so on in form of different layers. Deep learning uses deep neural network architecture along multiple data in form of big data and so deep learning models are referred to as deep neural networks.

* Corresponding author.
E-mail address: shiva@tce.edu

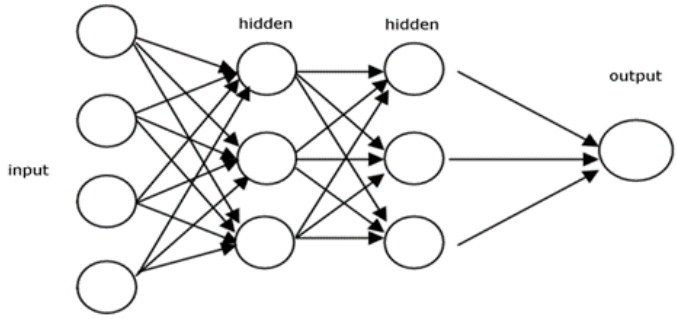


Fig. 1. neural networks

Plant Identification is essential in the field of natural products .Not all plants we see in our day to day life are healthy rather poisonous and produces harm. According to World Health Organization(WHO),more than half percent of the total population depend upon primary health care. plants Play a major role in curing diseases with less side effects and hence the importance of plants must be known by all people before they become extinct. Plants also forms the basis of food chain. It is useful for large parts of areas and people ranging from professionals like architects, foresters, farmers, conservationists, and biologists to the general. But the identification of plants manually is tedious and time consuming for novices as it contains different botanical names. Hence an automated classification system is needed for identifying and classifying plant species. In recent years, computer science research in the field of image processing and pattern recognition techniques, has made development for efficient classification and identification of different plant species.

2. State of the art

2.1. Convolutional Neural Network

CNN is one of the neural networks used in visual imagery along image classification, recognition etc.., It is a Feed Forward network and its basic layers are an input layer, many hidden layers and an outer layer. There are five layers of CNN namely convolutional layer, pooling layer, activation, max pool and fully-connected layer. The inputs are considered as weights and biases of each image. Both actual and target values are compared where when it has too low error rate is performed as iterating for deviation.

2.1.1 Layers of CNN

2.1.1.1 Convolutional Layer

This is the first layer of CNN consisting of linear filters. All input images are considered along the filters and later added and divided along the pixels. The first level of features are extracted along the convolutional layer.

2.1.1.2 Activation Layer

Activation Function determines non-linearity of the data obtained and transfers the output into definite values in form of non-negative terms (1).

$$F(x) = \max(x, 0) \quad (1)$$

2.1.1.3 Pooling Layer

This layer is generally to reduce or shrink the size of image using a window depending upon the data size.

2.1.1.4 Fully Connected Layer

This layer is to transform the image into a single vector. Later the input image is compared along the single image and predicts the accuracy.

2.1.1.5 Overfitting

Overfitting is a condition where those inputs being trained are not considered for testing which leads to this factor. Models gets more trained with the known data that makes overfitted to a model.

2.2. Transfer Learning

Transfer learning is the idea of overcoming the isolated learning procedure and uses the knowledge acquired for one task and in turn solve the related task. Getting a huge dataset like ImageNet for every domain is tough. Besides, most deep learning models are very specialized to a particular domain or even a specific task. While these might be state-of-the-art models, with really high accuracy and beating all benchmarks, it would be only on very specific datasets and end up suffering a significant loss in performance when used in a new task which might still be similar to the one it was trained on. This forms the motivation for transfer learning, which goes beyond specific tasks and domains. Transfer learning should enable us to utilize knowledge

from previously learned tasks and apply them to newer, related ones. In the case of problems in the computer vision domain, certain low-level features, such as edges, shapes, corners and intensity, can be shared across tasks, and thus enable knowledge transfer among tasks.

Transfer learning includes various approaches:

- Parameter transfer
- Relational-Knowledge transfer
- Instance transfer
- Feature representation transfer

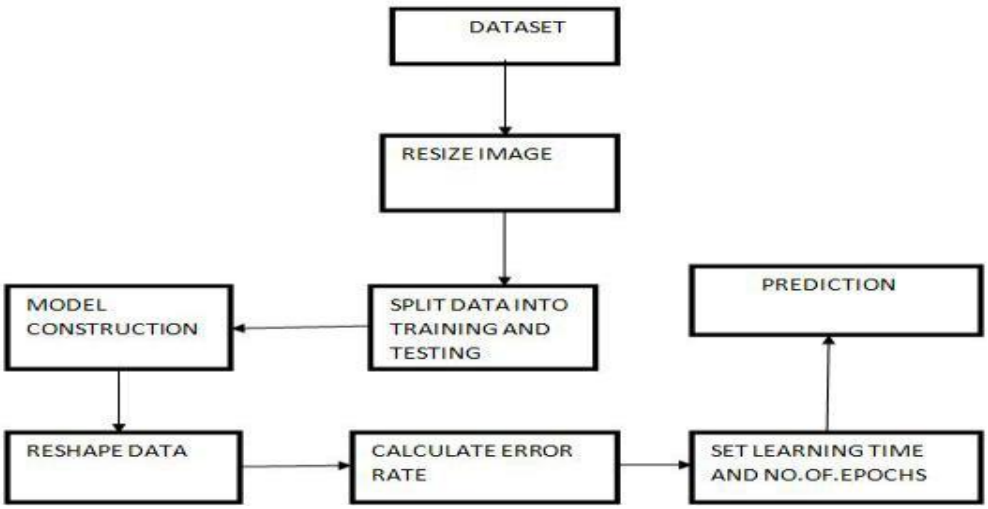


Fig. 2. Architecture Overview

2.3. Related Works

[1] This paper investigates about identifying a neural network model for diagnosing plant disease using images of healthy and diseased plants. Hence it encounters convolutional neural network algorithm along comparison of various architectures over an open dataset in identification of 26 different plant diseases in both lab and field test. Different architectures like AlexNet, AlexnetOWTBn, GoogleNet, Over Feat and VGG along Torch7 using the LUaJIT programming knowledge and training was performed using CUDA. Along comparison of all other architectures using parameters like percentage of successful classification of classes, corresponding average error of models, training epochs, along its average time VGG net provided a high success rate achieving a correct identification of 17466 images out of 17548 images. Hence proves that it is highly advisable for automated detection

and diagnosis for operating in real cultivated conditions along its less computation which makes it feasible for integrating into mobile application to be used in mobile devices.

[2] This paper suggests the use of deep convolutional neural network on identifying plants based on their leaf vein patterns along three different legume species like white, red and soya bean along the leaf shape, texture and color. Leaf fingerprint is considered as a main feature along classification. This method overcomes the usage of hand crafted extraction methods and also improves the accuracy by increasing the particular model's depth. Steps include vein segmentation, central patch extraction, vein measure and their classification along machine learning supervised algorithms like SVM, PDA, and RF along two different setups. The CNN models were trained on an increasing depth from 2 to 6 layers that includes five convolutional layers along one soft max layer. Optimizations of parameters were done using stochastic gradient decent over a set using mini batches along 20 samples with a 50% dropout rate. The experiments were done using Pylearn2.Later CNN algorithm is applied along legume species of different classes for more accuracy. To gain an insight, occlusions of different parts of image was done with 10*10 black patch and the output probability was obtained for the class.

[3] This paper implies the use of deep learning technique for image classification to identify the plant disease by only considering the lesions and spots instead of the whole leaf that improves more accuracy and helps in identifying multiple diseases along a single leaf. Also, data augmentation is performed that reduces the usage of different image along identification of multiple diseases along the same leaf. Most disorders were related to fungi, bacteria, phytotoxicity, algae, nutritional deficiency, and senescence. The leaf was identified based on the classification of symptoms like small numerous small lesions, large numerous with large lesions, single lesions or spots etc., The accuracy is increased along 12% in Google net architecture which is a pre-trained CNN. The following experiment is performed along matlab environment under unmodified, original and expanded conditions and accuracy is tabulated below as Healthy, mildly diseased, moderately diseased and severely disease. The solution not only increase the size of the data but also increases the diversity of data as the natural variability within each image is indirectly considered by the subdivision into smaller regions. This approach also has some shortcomings, but it clearly leads to more reliable results in a context of limited data availability.

[4] This paper is to develop a deep learning system to learn discriminate features from leaf images along with a classifier for species identification by comparing results with customized systems like leaf snap that proves CNN is better to hand crafted approach. Leaf net system contains pre-trained datasets like leaf snap, foliage and flavia where leaf snap is taken for training the data and data augmentation is performed along the datasets where different transformations along each images are performed that increases the data size. Later leaf net system also trained other two open datasets like foliage and flavia where top 1 and top 5 accuracies are provided below for leaf net, foliage and flavia respectively. Further flavia dataset was processed for classification. Later approaches like SVM, was applied across leafsnap but not a better performance was achieved. Hence a fine grained standard CNN is required and overall learning features through CNN can provide better results along leaf net was discovered.

[5] This paper deals with the identification of pests by developing an agricultural pest identification system based on computer vision using deep residual learning where it achieves more accuracy along complex farmland background than SVM and BP network. The accuracy was highly predicted along one of the pre-trained CNN architecture known as AlexNet Along deep residual approach higher accuracy. Automatic feature extraction is evolved by visualizing the level along convolution, pooling, softmax layers where deep residual approach is applied that fuses a residual map by stacking multilayer non linear network and then expressing the actual map relation. Compared to plain conventional network the accuracy is further improved along more using deep residual learning.

[6] This paper investigates about convolutional neural network to learn unsupervised feature representation along the convolutional network and propose a deep learning technique along bottom up and top down approach. After performing plant identification along CNN using ILSVRC2012 dataset deconvolutional network(DN) is used for visualization to identify the characteristics learned by CNN and to avoid the hand crafted features. DN enables us to observe the transformations of the features by projecting the feature map back to input pixel space. Specifically the feature maps from layer n are alternately de convolved and un pooled continuously down to input pixel size. A comparative performance evaluation was made along datasets as D1 and D2 where D2 was obtained by manually cropping images into patches from D1 and more accuracy was obtained in D2.This result that CNN approach is quite better among others and venation part of plant provides better accuracy and DN quite helps in identifying internal behaviour of CNN.

[7] This paper proposed an automatic disease recognition system on recognizing plant viral diseases as they are very hard to clear which can damage the entire agricultural field. Hence earlier diagnosis is done based on two different classes along healthy plants. Convolutional neural networks are employed along the diseased plant images for better classification including image rotation. Image rotation of various angles increase the dataset size and also has more accurate prediction on virus detection on plants. Compared to state of art techniques, CNN on automatic feature extraction provides its best performance on identification of viruses and also image rotation provides good feature extraction. But as the images are collected from a single source, unknown images with sunlight, drooping may be difficult to identify and hence the system could further be improved with more pre processing and can identify various other diseases.

[8] This paper investigates about the plant species identification along evaluation of the factors that affects the performance .Various neural nets are performed along the datasets including Google net, VGG net and Alex nat. Transfer learning is performed along the datasets for pretraining, and finetuning along LIFECLEF .Data augmentation is also performed along plant identification that involves process like rotation, transformation, alteration along utilizing pretrained and fine tuning of neural networks. Alex net trained on roughly 1.2 million image sets over 1000 different categories along five convolutional layers, two normalization layers, three max-pooling layers, three fully-connected layers, and a linear layer with Soft max activation in the output. Google net roughly possesses 6.8 million parameters along nine inception modules, two convolutional layers, one convolutional layer for dimension reduction, two normalization layers, four max-pooling layers, one average pooling, one fully-connected layer, and a linear layer with Soft max activation in the output. VGG net possesses 144 million parameters along 16 convolutional layers with very small receptive fields (3×3), five max-pooling layers of size 2 × 2, three fully-connected layers, and a linear layer with Soft max activation in the output. All the above methods use drop-out regularization to reduce overfitting. Along average accuracy, later using 80 patches with 300000 iterations along 60 batches size both google and VGG net achieved a good accuracy.

[9] This paper explains about the classification of Indian herbal plants based on powdered microscopic images .Different techniques like CNN, transfer learning, deep learning are applied along the powdered images. The chemical images were collected in a Charge coupled camera .Datasets were employed across SVM and CNN along ImageNet that uses 14 million images along 20 categories.CNN used both MNIST

and NORB dataset that provided a very higher accuracy. Different models were performed along CNN that include VGG, Inception, MobileNet and Xception . Transfer learning was enhanced along pretrained models with fine tuning along flat, dense, dropout and softmax layer. Two different datasets were taken consisting of herbals liquorice and rhubarb. Liquorice has iconic medical property that treats eczema, dyspepsia, prostrate cancer, kidney problems et., Rhubarb is to cure digestive problems.CNN models were employed along non-linear kernel with pooling and fully connected layer. Softmax function was employed for edge and blur detection. Further training, testing and validation accuracy are performed along the four models. Finally VGG16 provides a better accuracy but consumes more time whereas Mobile net takes less time hence to prefer them based on accuracy or execution time.

[10] This paper investigates about the classification of plant seedling based on deep learning. Here plant identification was performed along a set of 12 species using various kinds of transforms that includes resize, rotation, flip, scaling etc., CNN model was employed across the datasets that provided a very high accuracy along testing accuracy of medicinal plants. A forward deep learning approach along convolutional network was applied along plant classification.CNN involves the automatic feature extraction from the normalized images. AlexNet considered input images along 227x227x3 size with colour space conversion and image enhancement was performed along each image to obtain increased chromaticity and luminosity. Alex net models 25 layers,4 is input 1 output , Convolutional layer along ReLU ,2 cross channel layer along 3 max pool layer with 2 fully connected 4096 layer and 1 fully connected 2 layer along 1 dropout and soft max layer. This approach uses SGDM with 0.001 rate along 64 size mini batches along 100 epochs. All the procedure were carried out under the matlab environment. Data augmentation is useful along reducing the noise with fin e tuning .The total images along data augmentation obtained was 118750 out of which 83126 were considered along training 23750 along validation and 1174 along testing. Overall 99.98% were achieved along species like Charlock, Common wheat, Fat Hen, Maize that got the highest sensitivity rate and Common Wheat, Maize, Scentless Mayweed, Shepherds Purse, got the highest specificity rate along 12 different species.

[11] This paper investigates about plant disease detection and its classification along two different plant species including apple and tomato plants. They are classified as either diseased or healthy .A dataset of around 3600 images are collected for classification. A CNN model is developed including 4 convolutional layers followed by activation and pooling layers,2 fully connected layers and softmax layer.

Images are fed along 64*64 in pixel size along image binarization that helps in extracting high level features along CNN. Datasets were split in the ratio 80:20 respectively. An average accuracy is obtained along the CNN developed and overfitting is reduced by adjusting the drop out rate from 0.25 to 0.20.The concept of transfer learning can be included in future work for achieving more accuracy along unknown data

[12] This paper provides a comparison of various plant classification models along different publicly available datasets .The concept of transfer learning is defined that increases efficiency and reduces the time taken by the model to learn from the scratch. Datasets like flavia, Swedish leaf, UCI Leaf, Plant Village are taken along different models that includes CNN from scratch , different transfer learning models with fine tuning-cross dataset fine tuning, fine tuning of CNN models, pre-trained CNN models and finally combination of RNN and CNN. All 4 different datasets are individually trained and tested along these five different classification models that obtained a good accuracy on applying transfer learning .Also models that are applied deep features provide better performance compared to others. This can be implied for various applications in future with an increase in dataset.

[13] This paper investigates about the method of using fine grained approach and using wide and deep model that uses both deep network and discrete features along linear model at the same time. The experiment is done along meta data obtained which is Exchangeable Image File Format that contains various useful information including exposure, lens size, date resolution etc., Inception V4 model is also applied along the Korean flower species dataset with a comparison along experimentation of different wide and deep model along date, colour, petals along the meta data obtained that combines both neural networks and discrete features along linear model. Including the external features like date of flowering, location of flowering provides good accuracy when compared to baseline method. Also by improving the angles and lightings of images the accuracy can be further improved in future work.

[14] This paper proves a plant recognition system on identifying plants using both traditional method and deep learning feature. Traditional method includes using filters like Haralick features, local binary patterns, that helps in extracting features and then further employing data along various machine learning approaches like k-nearest neighbour, Naive bayes, random forest and bagger classifier. This paper suggests the use of CNN which is more efficient when compared to other traditional approaches. Standard publicly available datasets like flavia, swedish leaf dataset, folio CNN

models were employed like VGG16,VGG 19,inception-resnet.Among all models compared, VGG19 along logistic regression provided a very good accuracy for effective plant recognition. Plant recognition can further be revised and improved by using data augmentation that shows diverse data types for more efficiency.

3. Conclusion

The papers listed above provide an efficient automated system on identifying plant related species and classification using several CNN models. The models achieve good accuracy only with an increase in learning time. Transfer learning provides better performance, but some pre-trained models are very huge along training that requires more training time for improvement in accuracy is challenging.

References

1. Ferentinos, Konstantinos P. "Deep learning models for plant disease detection and diagnosis." Computers and Electronics in Agriculture 145 (2018):311-318.

2. Grinblat, Guillermo L., et al. "Deep learning for plant identification using vein morphological patterns." Computers and Electronics in Agriculture 127 (2016): 418-424.

3. Barbedo, Jayme Garcia Arnal. "Plant disease identification from individual lesions and spots using deep learning." *Biosystems Engineering* 180 (2019): 96-107

4. Barr´e, Pierre, et al. ”LeafNet: A computer vision system for automatic plant species identification.” Ecological Informatics (2017): 50-56.

5. Cheng, Xi, et al. ”Pest identification via deep residual learning in complex background.” Computers and Electronics in Agriculture 141 (2017): 351-356.

6. Lee, Sue Han, et al. "Deep-plant: Plant identification. Lee, Sue Han, et al. "Deep-plant: Plant identification with convolutional neural networks." *2015*

7. Kawasaki, Yusuke, et al. "Basic study of automated diagnosis of viral plant diseases using convolutional neural networks." *International Symposium on Visual Computing*. Springer, Cham, 2015.

8. Lee, Sue Han, et al. ”How deep learning extracts and learns leaf features for plant classification.” Pattern Recognition 71 (2017): 1-13

9. Marwaha, Rohan, and Bhupendra Fataniya. "Classification of Indian Herbal Plants based on powder microscopic images using Transfer Learning." *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. IEEE, 2018.

10. ., Ashqar Belal AM, Bassem S. Abu-Nasser, and Samy S. Abu-Naser. "Plant Seedlings Classification Using Deep Learning." (2019).

11. .Francis, Mercelin, and C. Deisy. "Disease Detection and Classification in Agricultural Plants Using Convolutional Neural Networks—A Visual Understanding." *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 2019.

12. Kaya, Aydin, et al. "Analysis of transfer learning for deep neural network based plant classification models." *Computers and electronics in agriculture* 158 (2019): 20-29.

13. . Lee, Jun Woo, and Yeo Chan Yoon. "Fine-Grained Plant Identification using wide and deep learning model." *2019 International Conference on Platform Technology and Service (PlatCon)*. IEEE, 2019.

14. . Pearline, S. Anubha, V.Sathiesh Kumar, and S. Harini. "A study on plant recognition using conventional image processing and deep learning approaches." *Journal of Intelligent & Fuzzy Systems* Preprint (2019): 1-8.

An Optimized Privacy Preserving Mechanism in Cloud

P.Jhansi Rani *, Dr.M.Akkalakshmi*
Research Scholar, GITAM, Hyderabad
Professor, Dept of IT, GITAM., Hyderabad

Abstract

cloud computing is the range of computing services that gives the organizations to remotely save their data into the cloud and access on-demand applications. In the present methodology explained an Extendable Access Control System strategy speculated that the authority is the trusted in party, yet a significant part of the time, they may play out an unlawful action which causes the data setback. The proposed work encoded the data through Uncrackable Cipher Dynamic Double Encryption Standard (UCDDES). The UCDDES contains the key length of 32, 40 and 48. To subjectively pick the key length lessened the data security issues. After continuously picking the key length the data agent sent the key requesting to the power. By then reliant on the got key length the data delegate delivered the midway riddle key. It is moreover used to unscramble the data and set away in the circulated stockpiling. The eventual outcome of the examination overhauls cloud security and access control. It reduces the issue of unapproved customer/software engineers getting to data. It assembles the cloud security and keeps from word reference ambushes, creature control attacks, sway attacks, SQL implantation ambushes, and so on. It is an augmentation work of our past work on UCDDES.

keywords: Optimazation; GF Multiplication; UCDDES based data encryption; cloudnetwork security.

4. Introduction

cloud computing as the rising innovation has become prevalent nowadays and built up significant enthusiasm from both scholarly community and business. Despite the fact that there is no remarkable portrayal for cloud computing, in any case, one average importance given by numerous specialists starts from the National Institute of

* Corresponding author.
E-mail address: jhansi.infoarea@gmail.com

Standards (NIST): a model for organize access for sharing assets, for example, application, stockpiling, system, administrations and servers that can be discharged with less exertion [1]. The clients store their data at remote stockpiling servers in the cloud computing. These remote stockpiling servers are adapted by a Cloud Service Provider (CSP) as often as possible as an outsider [2, 3]. In addition, PC equipment like memory, plate space, and processor are virtualized and given to the end clients as an office through the open Internet [4, 5]. A various virtual machine are conveyed over a lot of servers with various geographical focuses that fills in as a cloud office, which is interrelated utilizing the media transmission joins. Besides, the cloud clients need to pay dependent on the genuine measure of administration they have used as relating to water or power bill [6, 7]. For an end user, the rapid elasticity, less maintenance cost, minimal upfront investment, measured service and ubiquitous access to cloud services are the advantages [8, 9]. In cloud computing, the virtualization technology utilization outcomes in high resource utilization and therefore imposes fewer electricity costs to service providers. Although clouds are more dependable and have more powerful infrastructure compared to personal computers, there are still security worries that prevent users to deploy their businesses in the cloud and therefore decreases the growth of cloud computing. The reason why the people are not trusting the cloud provider to manage the data is that they lose their control over the data[10, 11]. Apart from that, the sensitive information in the cloud storage needs to be secured from illegal access. As a result, the people who are generating data are wished to know about the privacy of the data by using cryptographic access control systems (ACS). In recent times, investigators have suggested numerous data ACS safeguard the stored data in the cloud computing. Such schemes authorize the data governor to handle authorized users securely and repeal their authorization rights. Attribute-based encryption (ABE) is a significant technique interpreting the dissimilar attributes of the user, data governor, or cloud environment to implement the control of data access (Zhang et al, (2017) recommended a cyber-physical system for patient-centric healthcare applications and services. This was recognized to be as Health-CPS. It constructed by using the technologies of cloud and big data analytics. This system encompassed a data collection layer which has a unified standard, a data management layer for distributed storage and parallel computing, and a data-oriented service layer. The outcomes of this study presented that the cloud as well as big data technologies could be utilized to improve the healthcare system's performance. Then only humans can enjoy various smart healthcare applications and services.

Users are expecting to secure data transmission and storage on cloud , which require efficient cryptographic algorithms. Hence, there is increasing need for efficient software implementation cryptographic algorithm. Many of the IEEE standards, such as 802.11, 802.15 and 802.16 use AES-CCM [2][1] as the basis for their security; and this is a good choice because it provides both encryption and authentication in a single efficient solution. Security and performance considerations are therefore both imperative during the design phase of encryption algorithms. With the advent of AES in IEEE 802.11i, as considered in this work, and the high prospects of wireless systems, this research holds a prominent position in the evaluation and analysis of the structure of the Rijndael algorithm (AES-CBC) from the resource constraint wireless systems point of view. Many cryptographic algorithms, such as AES Rijndael, which are compact and efficient to implement on high-performance microprocessors, may not be implementable efficiently on smaller and less powerful microprocessors found in low-power mobile devices. A resource-constrained system's efficiency is invariably related to the size of the code [3]. The efficiency of a program increases, as the code size decreases and the execution speed increases. Therefore, implementation of cryptosystems on with very tight memory constraints mobile devices introduces new challenges. In [4], the authors investigate the speed measurement of several cryptographic system libraries to determine if they are feasible for Palm devices or if they are too complex. The main finding is that it can be valuable to encrypt real-time data if pre-computation of certain steps is allowed but it may cause a problem for devices with very limited space. Hence, optimization is often possible through a closer inspection of the encryption software algorithms. The security suites can be more broadly classified by their properties: encryption only (AES-CTR), followed by authentication only (AES-CBC MAC), and finally encryption and authentication (AES-CCM) [1]. AES-CTR (counter mode of cryptographic operation with AES) means that the CTR mode uses AES as the block cipher; and provides access control, data encryption and optional sequential freshness. Authentication is done using the cipher block chaining with message authentication code (CBC- MAC), which creates a message integrity code using a block cipher in CBC mode, and computes a MAC over the packet and includes the length of the authenticated data. The code can be computed upon packet reception and can be compared with the one received. AES-CCM is a combination of the encryption and authentication suites detailed above. It has three inputs; the data payload to be encrypted and authenticated, the associated data (header etc.) to be authenticated only, and the nonce to be assigned to the payload and the associated data [1]. There are varying MAC lengths to choose from for AES-CBCMAC and AES-CCM modes of operation (4, 8 or 16 bytes), allowing

for some scalability of security depending on application requirements. In this paper, we study the performance of UCDDES software execution and its operation complexity. An overview of UCDDES algorithm and complexity is presented in section 3.. Simulation results are analyzed, and discussions are provided in section 4 and 5 respectively. Finally, we conclude and summarize the main findings of the paper.

5. Literature survey

Qiu et al., (2018) expressed a framework that ensures the monetary clients' protection information utilizing Attributed-Based Access Control (ABAC) conspire. The proposed framework planned to guarantee the information protection in which a semantic way to deal with client get to control is made. Next, a client driven methodology that keeps the information from dangerous procedures on the cloud side was proposed. Taking everything into account, the prescribed plan has more elevated level secure maintainability as it could contract with dynamic dangers, together with the creating and future dangers.

Zuo et al., (2018) proposed a technique that ensures the data in the capacity of cloud. At first, te cryptographic key was ensured by thinking about the two angles. These perspectives are used to keep up the key discharge. Next, the mystery key can be renounced effectively by joining the elective re-encryption alongside key partition techniques. Taking everything into account, the information was verified in a fine-grained path by executing the ABE system. Also, the exhibition valuation and security examination displayed that the proposition was secure and compelling, individually.

Yuan et al., (2018) expressed a technique called Oblivious Random Access Memory (ORAM) utilized for high security and information sharing. The information square can be kept away from change through rearranging. The IND-CPA security is furnished for the framework with an ID-Based mark and the Path-ORAM security properties. The framework demonstrated the best calculation intricacy.

Ning et al., (2018) managed the two classes of access abuse they are, semi-confided in power's unlawful access, and cloud client's unapproved get to. These kinds of abuse can be decreased by a distributed storage conspire called CP-ABE with reviewing and white-box recognizability known as CryptCloud+. The paper demonstrated the security and utility of the plan.

Srinivasan et al., (2018) clarified the insurance worry that incorporates the cloud assaults, respectability, security, powerlessness in asset sharing and spillages. The

framework associated with the data insurance from misfortune and security from programmers and aggressors of a continuous domain. The framework gave the validation and approval. The administrations on nature of administration, information transmission and exclude the critical data are guaranteed. This capable system safeguards the cloud with improved execution assessment.

Divya et al., (2018) concentrated on an information sending strategy named "Secure Cloud Data Forwarding Framework (SCDF)" which accommodatingly achieved security and execution issues. Presently, the document was isolated into various parts and put away in a cloud server and transmitted to the clients utilizing DADR (Distributed Autonomous Depth First Routing) convention when they mentioned. The multi-gushing element of DADR makes the sections to choose an elective way if any disappointment occurred. Utilizing elliptic bend cryptography and T-shading procedures, the system achieved uncommon information security.

Cui et al., (2018) introduced a property based distributed storage framework with secure provenance. At first, the capacity framework was without denial, and later it reaches out to the repudiation system to frustrate the information clients from recovering the recently encoded information. At that point, the calculation was actualized in the two both the methodologies. The exploratory results exhibited that the proposed frameworks are satisfactory to be applied by and by.

Iyapparaja et al., (2017) recommended an alternate encryption innovation and mark key on cloudlets likewise and those mark key send to a register email id. Each cloud was isolated in cloudlet, for explicit cloudlet expected to get to client must enroll on them. In this client can utilize other cloud data and evacuate other information with the proper confirmation in client side and principle cloud server side. The client may store any data, for example, pdf, picture, content, and so on. When the client can store information on cloudlet, another client can ready to utilize that information.

Sookhak et al., (2017) gave a total examination on get to control frameworks which depended on properties and related each plan's usefulness notwithstanding the trademark. The quality put together approaches are set up with respect to the structure; get to control mode, invalidation technique, dissolution mode, cancellation issue, and controller are exhibited. Brought together, Decentralized, and Hierarchal based groupings are made. This was performed to know the advantages and disadvantages of the ABE system.

Tang et al., (2018) proposed a powerful three-layer encryption framework made on DES notwithstanding system coding. Set up in the hypothetical examination, the novel plan was presented to have the advantage to achieve a powerful progress among productivity and security. It expands its versatility to various digital conditions. The recreation results additionally uncovered that the running proportion of the creative plan was to some degree lower or proportionate to significantly increase DES.

Jhansi Rani et al., (2019) proposed an entrance control framework made on AES. It has improved the encryption execution with dynamic twofold encryption standard technique which performs encryption in two phses to improve the encryption execution.

6. Proposed methodology

The encryption system for the cloud storage is an important one. In the existing technique described the Extendable Access Control System (EACSIP) algorithm which is built on top of an ideal cryptographic primeval, namely Functional Key Encapsulation with Equality Testing (FKEET). It assumes that the authority is a trusted party and get the security parameters from the authority. Then the data from the data owners are encrypted through the authority. Further, they could send to the cloud storage. Then the user can download the data from the cloud through private keys and access policies. The existing system does not provide the efficient encryption technique; it used the classical symmetric encryption technique which is not efficient for the real-time scenario. So, the proposed work encrypted the data through UCDDES. Generally, the UCDDES contains the key length of 32, 40 and 48. Because of the three-key length, the intruder cannot obtain the combinations for the entire three key which is the huge task. Further, in the proposed system, even the security user cannot crack the codes once it generated. To make the plain text into more secured ciphertext, the proposed framework is divided into two phases, they are

1) Phase I –Cipher Text1: The initial phase which gets the plain text as input. The input text is then converted into the first form of ciphertext with Advanced Substitution method with 128 bit and a key length of 16

2) . Phase II – Cipher Text2: The ciphertext which is generated in phase I is given as input to the phase II. The ciphertext is further encrypted with the symmetric key of the length 16 again(Private key). This key is like a one-time pad key which can be used once for encryption and decryption.)

The main advantages of UCDDES are • Protecting from the collision attack, the SQL injection, the dictionary attack and from the brute force attack. • Data integrity, data confidentiality, and data privacy are achieved through UCDDE.

Algorithm UCDDES for Encryption phase 1

input: Plain Text (t1), R: Random, Nr: Total_Rounds, Nb: Constant worth

output: Cipher Text (C1)

1 Compute: Block Size R[128, 192, 256-piece Plain_Text]

2 For Cipher (t1, Secret_Key) at that point/UCDDES contains the Secret_key length of 32, 40 and

48

Byte_State [4, Nb]

4 for (State = t1)

5 Add_Round_Key (State, w[0, Nb-1])/Key Expansion

6 for round ==1 : Nr-1 at that point

7 Sub_Byte(State)

8 Shift_Rows(State)

9 Mix_Columns(State)

10 Add_Round_Key(State, w[round *Nb, (Round+1)*Nb-1])

11 end for

12 Sub_Byte(State)

13 Shift_Bytes(State)

14 Add_Round_Key(State, w[Nr*Nb, (Nr+1) * Nb-1)

15 C1 = State

16 call(Cipher2)

Algorithm UCDDES for Encryption phase 2

Input: Cipher Text (C1), Nr: Total_Rounds, Nb: Constant worth

output: Cipher Text (C2)

1 in the event that (C1 != NULL) at that point

2 For Cipher2 (C1, Secret_Key) at that point/UCDDES contains the Secret_key length of 32, 40

also, 48

3 Byte_State [4, Nb]

4 for (State = C1)

5 Add_Round_Key (State, w[0, Nb-1])/Key Expansion

6 for round ==1 : Nr-1 at that point

7 Sub_Byte(State)

8 Shift_Rows(State)

9 Mix_Columns(State)

10 Add_Round_Key(State, w[round *Nb, (Round+1)*Nb-1])

11 end for

12 Sub_Byte(State)

13 Shift_Bytes(State)

14 Add_Round_Key(State, w[Nr*Nb, (Nr+1) * Nb-1)

15 C2 = State

16 end

Algorithm UCDDES for Decryption phase 1

Input: Cipher Text (C2), R: Random, Nr: Total_Rounds, Nb: Constant worth

Output: Cipher Text (C1)

1 Compute: Block Size R[128, 192, 256-piece Plain_Text]

2 For DeCipher1 (C2, Secret_Key) at that point/UCDDES contains the Secret_key length of 32,

40 and 48

3 Byte_State [4, Nb]

4 for (State = C2)

5 Add_Round_Key (State, w[0, Nb-1])/Key Expansion

6 for round ==1 : Nr-1 at that point

7 Inv_Sub_Byte(State)

8 Inv_Shift_Rows(State)

9 Inv_Mix_Columns(State)

10 Add_Round_Key(State, w[round *Nb, (Round+1)*Nb-1])

11 end for

12 Inv_Sub_Byte(State)

13 Inv_Shift_Bytes(State)

14 Add_Round_Key(State, w[Nr*Nb, (Nr+1) * Nb-1)

15 C1 = State

16 call(Decipher2)

Algorithm UCDDES for Decryption phase 2

Input: Cipher Text (C1), R: Random, Nr: Total_Rounds, Nb: Constant worth

Output: Plain Text (t1)

1 Compute: Block Size R[128, 192, 256-piece Plain_Text]

2 For DeCipher2 (C1, Secret_Key) at that point/UCDDES contains the Secret_key length of 32,

40 and 48

3 Byte_State [4, Nb]

4 for (State = C1)

5 Add_Round_Key (State, w[0, Nb-1])/Key Expansion

6 for round ==1 : Nr-1 at that point

7 Inv_Sub_Byte(State)

8 Inv_Shift_Rows(State)

9 Inv_Mix_Columns(State)

10 Add_Round_Key(State, w[round *Nb, (Round+1)*Nb-1])

11 end for

12 Inv_Sub_Byte(State)

13 Inv_Shift_Bytes(State)

14 Add_Round_Key(State, w[Nr*Nb, (Nr+1) * Nb-1)

15 t1 = State

16 end

Optimization:

We have utilized pre-processed look up tables to streamline program tasks and improve execution. In this methodology MixColumn activity with the S-box, and includes 4 tables with 256 4-byte sections, i.e., 4 KByte of memory. Each call to the MixColumn or InvMixColumn tasks brings about sixteen field increases. Recursive assignments should be tried when the guidance arrangement should bounce unaware of what's going on. For few repeats, the additional calculation could be expelled by substituting the loop with the code modules for that fixed number of times. This strategy is called loop unrolling. At the point when two loops are being executed with comparable undertakings that can be consecutively balanced, it is smarter to join the two circles into a solitary circle. This procedure is called loop merging. This lessens the complete additional calculation time of executing various loops to the overhead of a single loop. After software optimization techniques were applied: • SubBytes() gained 20% execution time performance. • ShiftRows() gained 30% execution speed performance.. • AddRoundKey() gained 21 % in execution speed. Section headings should be left justified, with the first letter capitalized and numbered consecutively, starting with the Introduction. Sub-section headings should be in capital and lower-case italic letters, numbered 1.1, 1.2, etc, and left justified, with second and subsequent lines indented. You may need to insert a page break to keep a heading with its text.

7. Results & Conclusion

UCDDES encryption and decryption code size for optimized and un optimized versions are given in below table.

UCDDES Encryption Operations	Unoptimized Code / Bytes	Optimized Code / Bytes
KeyExpansion	1016	1703
AddRoundKey	139	443
SubBytes	142	476
ShiftRows	479	583
MixColumns	277	2533

UCDDES Decryption Operations	Unoptimized Code / Bytes	Optimized Code / Bytes
KeyExpansion	1016	1703
AddRoundKey	139	443
InvSubBytes	142	476
InvShiftRows	479	572
InvMixColumns	317	3374

Table 1: Optimized vs Unoptimized UCDDES Code

Each encryption methods will have upsides and downsides, so we need to choose the well-suited cryptography calculation for an application. The parameters like execution, shortcoming, and quality of the techniques are to be scholarly [26]. The encryption strategy must be breaking down dependent on a few measurements, and looked at are portrayed underneath:

4.1. Encryption time – It is perceived as the change time from plain content to the ciphertext that is based on the key length, input square length, and mode. This time is taken in the unit of milliseconds. This time is significant for execution, and in the event that the time is less, at that point the framework works quick.

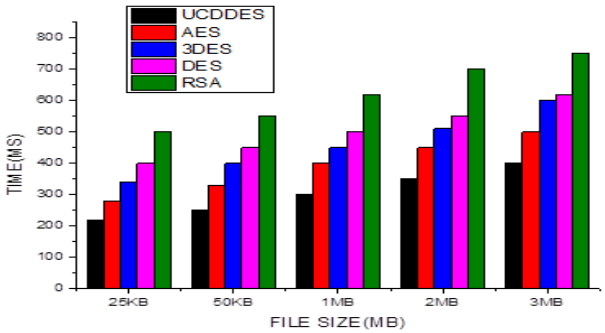


Figure 2: Evaluation of time of Encryption among various algorithms

4.2. Decoding time - It is known as the change time from ciphertext to the plaintext that depends on the key length, input square length, and mode. This time is taken in the unit of milliseconds. This time is significant for execution, and in the event that the time is less, at that point the framework works quick.

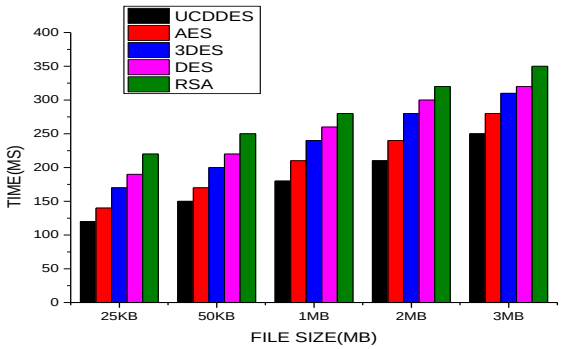


Figure 3: Comparison of Decryption time among various algorithms

4.3. Transfer time – It is perceived as the time taken to store the information into the cloud framework. This time is taken in the unit of milliseconds. This time is significant for execution, and on the off chance that the time is less, at that point the framework works quick.

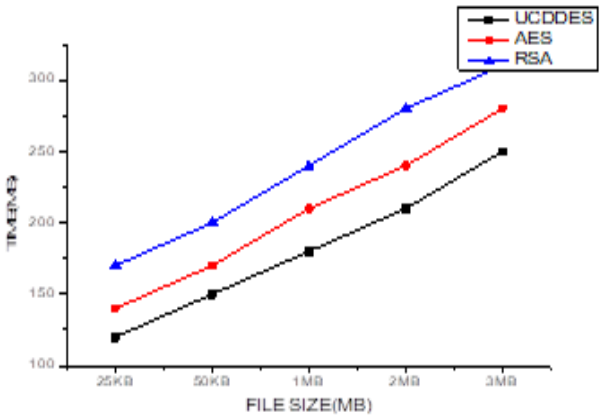


Figure 3: Comparison of Upload time among various algorithms

4.4. Download time – It is recognized as the time taken to recover the data from the cloud framework. This time is taken in the unit of milliseconds. This time is significant for execution, and on the off chance that the time is less, at that point the framework works quick.

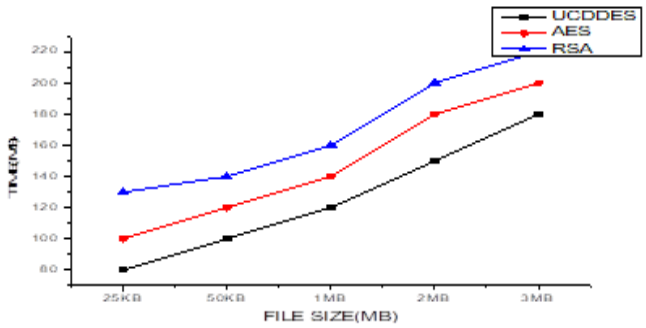


Figure 4: Comparison of Download time among various algorithms

We presented an optimized UCDDES – Uncrackable Cipher Dynamic Double Encryption Standard, which encodes the information. The encoding is finished in two phases the principal organize scrambles the plaintext with the key, and the yield is given as the contribution to the second stage where again the twofold encoding happens with another key. At that point the content is spared in the cloud. At that point the content can be downloaded from the cloud and unscrambles the content with the keys given. Through this encoding procedure the information trustworthiness, information classification, and information protection are improved. This tale method is used to anticipate SQL infusion, Brute power, Collision assault and lexicon assault. The encoding time and decoding time will be additionally decreased later on work to accomplish better execution

References

1) P. Jhansi Rani, Dr. M. Akkalakshmi “An Enhanced Data Access Control and Privacy Preserving Mechanism using UCDDES” (2019)

2) [1] Mell, Peter, and Tim Grance. "The NIST definition of cloud computing." (2011).

3) Zhou, Z., & Huang, D. (2012, October). Efficient and secure data storage operations for mobile cloud computing. In *Proceedings of the 8th International Conference on Network and Service Management* (pp. 37-45). International Federation for Information Processing.

4) Sookhak, M., Talebian, H., Ahmed, E., Gani, A., & Khan, M. K. (2014). A review on remote data auditing in single cloud server: Taxonomy and open issues. *Journal of Network and Computer Applications*, *43*, 121-141.

5) Kumar, K., & Lu, Y. H. (2010). Cloud computing for mobile users: Can offloading computation save energy, *Computer*, *43*(4), 51-56.

6) Sookhak, M., Akhunzada, A., Gani, A., Khurram Khan, M., & Anuar, N. B. (2014). Towards dynamic remote data auditing in computational clouds. *The Scientific World Journal*, *2014*.

7) Whaiduzzaman, M., Sookhak, M., Gani, A., & Buyya, R. (2014). A survey on vehicular cloud computing. *Journal of Network and Computer Applications*, *40*, 325-344.

8) Sookhak, M., Gani, A., Khan, M. K., & Buyya, R. (2017). Dynamic remote data auditing for securing big data storage in cloud computing. *Information Sciences*, *380*, 101-116.

9) Sookhak, M., Yu, F. R., Khan, M. K., Xiang, Y., & Buyya, R. (2017). Attribute-based data access control in mobile cloud computing: Taxonomy and open issues. *Future Generation Computer Systems*, *72*, 273-287.

10) Sookhak, M., Gani, A., Talebian, H., Akhunzada, A., Khan, S. U., Buyya, R., & Zomaya, A. Y. (2015). Remote data auditing in cloud computing environments: a survey, taxonomy, and open issues. *ACM Computing Surveys (CSUR)*, *47*(4), 65.

11) Wang, C., Ren, K., Lou, W., & Li, J. (2010). Toward publicly auditable secure cloud data storage services. *IEEE Network*, *24*(4).

12) Shiraz, M., Sookhak, M., Gani, A., & Shah, S. A. A. (2015). A study on the critical analysis of computational offloading frameworks for mobile cloud computing. *Journal of Network and Computer Applications*, *47*, 47-60.

13) Di Vimercati, S. D. C., Foresti, S., Jajodia, S., Paraboschi, S., & Samarati, P. (2007, November). A data outsourcing architecture combining cryptography and access control. In *Proceedings of the 2007 ACM workshop on Computer security architecture* (pp. 63-69). ACM.

14) Sahai, A., & Waters, B. (2005, May). Fuzzy identity-based encryption. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques* (pp. 457-473). Springer, Berlin, Heidelberg.

15) Goyal, V., Pandey, O., Sahai, A., & Waters, B. (2006, October). Attribute-based encryption for fine-grained access control of encrypted data. In *Proceedings of the 13th ACM conference on Computer and communications security* (pp. 89-98). Acm.

16) Hur, J. (2013). Improving security and efficiency in attribute-based data sharing. *IEEE transactions on knowledge and data engineering*, *25*(10), 2271-2282.

17) Qiu, M., Gai, K., Thuraisingham, B., Tao, L., & Zhao, H. (2018). Proactive user-centric secure data scheme using attribute-based semantic access controls for mobile clouds in financial industry. *Future Generation Computer Systems*, *80*, 421-429.

18) Zuo, C., Shao, J., Liu, J. K., Wei, G., & Ling, Y. (2018). Fine-Grained Two-Factor Protection Mechanism for Data Sharing in Cloud Storage. *IEEE Transactions on Information Forensics and Security*, *13*(1), 186-196.

19) Yuan, D., Song, X., Xu, Q., Zhao, M., Wei, X., Wang, H., & Jiang, H. (2018). An ORAM-based privacy-preserving data sharing scheme for cloud storage. *Journal of information security and applications*, *39*, 1-9.

20) Ning, J., Cao, Z., Dong, X., Liang, K., Wei, L., & Choo, K. K. R. (2018). CryptCloud+: Secure and Expressive Data Access Control for Cloud Storage. *IEEE Transactions on Services Computing*.

21) Srinivasan, S., & Raja, K. (2018). An Advanced Dynamic Authentic Security Method for Cloud Computing. In *Cyber Security: Proceedings of CSI 2015* (pp. 143-152). Springer Singapore.

22) Divya, S. V., Shaji, R. S., & Venkadesh, P. (2018). An Efficient Data Storage and Forwarding Mechanism Using Fragmentation-Replication and DADR Protocol for Enhancing the Security in Cloud. *Journal of Computational and Theoretical Nanoscience*, *15*(1), 111-120.

23) Cui, H., Deng, R. H., & Li, Y. (2018). Attribute-based cloud storage with secure provenance over encrypted data. *Future Generation Computer Systems*, *79*, 461-472.

24) Iyapparaja, M., Navaneethan, C., Meenatchi, S., Kumar, P. J., & Suganya, P. (2017). A Privacy-Preserving Secure Access Control Mechanism in Cloud.

25) Sookhak, M., Yu, F. R., Khan, M. K., Xiang, Y., & Buyya, R. (2017). Attribute-based data access control in mobile cloud computing: Taxonomy and open issues. *Future Generation Computer Systems*, *72*, 273-287.

26) Tang, H., Sun, Q. T., Yang, X., & Long, K. (2018). A Network Coding and DES Based Dynamic Encryption Scheme for Moving Target Defense. *IEEE Access*, *6*, 26059-26068.

27) Patil, P., Narayankar, P., Narayan, D. G., & Meena, S. M. (2016). A comprehensive evaluation of cryptographic algorithms: DES, 3DES, AES, RSA, and Blowfish. *Procedia Computer Science*, *78*, 617-624.

.

Textile Industry Waste Water Monitoring System Using LoRa Technology

P.Visalakshi, C.Aarthi*
PSG College of Technology, Coimbatore 641041, India
PSG College of Technology, Coimbatore 641041, India

Abstract

Textile Industry waste water is one of the major sources of water pollution. The waste water must be continuously monitored and treated before letting it in to water bodies. The Long Range (LoRa) technology is used for monitoring the waste water and alert owner in times of need is given in this paper. The turbidity, pH, and temperature sensors are used to monitor the quality of the water. The sensor values are monitored using web server and IoT platforms by the owner. The owner receives alerts, if the threshold values are exceeded.

Keywords: LoRa Technology, Monitoring, IoT, turbidity

1. Introduction

The waste water from the textile industry contains many harmful chemicals, dyes and toxins. This water must not be let in to water bodies without water treatment. A continuous monitoring system at low power and long range is required, to perform water treatment. One of the main reasons for the lack of waste water treatment is, the condition of the water has to be checked each time manually. The samples of the water must be collected from the waste water plant and tested in laboratory. The important parameters to determine the quality of water are, Temperature, pH and Turbidity [1]. Based on this, the type of the treatment it requires in the effluent plant is decided and treated. If continuous monitoring is done, the treatment can be done regularly in the required time. This reduces the cost and improves the efficiency of treatment [5].

This process can be made simple and real time using the sensors and microcontrollers. The data can be stored in the cloud and can be made available to the owner using any cloud server and IoT platforms. The LoRa technology can be used to make the

* Corresponding author.
E-mail address: c.aarthi94@gmail.com

process power efficient and long range [6]. In the proposed method, the temperature, humidity, pH and turbidity level of the waste water of the textile industry is monitored continuously using the sensors and when the threshold level increases, an Email alert or SMS alert is sent to the owner or person in charge for monitoring it. The LoRa technology is used as the number of nodes can be increased based on the requirement in the industry.

The testing of the water is done continuously and the results are updated to the cloud periodically. The data can be viewed from the cloud server.

2. Literature review

The LoRa technology is implemented for various applications globally. Among them, Monitoring systems using LoRa are used widely. In the Bristol Floating Harbor, the monitoring system is implemented using multiple sensors and camera. The solar panel is kept for power supply. The Aeration caused due to the waves generated by the ship movement deposited bubbles on the sensors. The camera recorded only the reflection of the sky in water [1]. The velocity of the cooling air and temperature are to be monitored during the export of apples. The circuit boards with antenna are placed in between the boxes of apples inside the container. The temperature of the container is monitored and it is varied based on the monitored level to preserve the Apples [2].

The wastes from the Tea factory are used for the cultivation of the Earthworms to produce vermicompost. The Temperature and Humidity level of the tea factory waste must be monitored for the growth of the Earthworms [3]. The key feature of LoRa is the Low power consumption and Long range. This is one of the important features for IoT. The Low power and Long range of data transmission is possible in LoRa by the Adaptive data rate. The Adaptive data rate is varying of the data rate based on the Payload size .This is achieved by the chirp spread spectrum modulation type used in LoRa technology. In the chirp spread spectrum, the frequency increases or decreases linearly with the data rate [4]. The major issue in the Internet of Things is the Data Management and Power consumption. The power consumption can be optimized by using the Renewable Energy resources. The data are sent to the cloud for every minute. The storage of data are difficult in real time. Hence the data are averaged at an interval of time and stored in the memory. By this the memory utilization is reduced and data are managed [5]. The utilization of the Renewable energy helps in power management unit cost reduction. The renewable energy resources like solar energy, Wind energy and Tidal energy can be used. The Real time Web services like Email Alert and SMS alert can be used to alert the owner or person in charge during

the time of emergency. The alerts can be given using the IoT platforms [6]. The different applications of LoRa [1-6] help in understanding of the working of LoRa and how effective it can be utilized.

The three major sensors which are used for monitoring the quality of the water in any industry are the pH sensor, temperature and humidity sensor and turbidity sensor [7].The LoRa is important for IoT as it provides low power consumption and long range, which is important for an IoT device [8]. This makes LoRa suitable for long distance communication. LoRa is better than GPRS and the next technology which can overcome them, are the NB-IoT boards [9]. The LoRa provides high security as each device has to be registered in the LoRa network. The gateway used must also be registered [10]. The LoRa has 128 bit AES algorithm for security. It uses the App key and Network session key of the application for the secured transmission of data [11]. The Network session key is used to provide security between the node and the network server. The Apps key is used to encrypt and decrypt the data. The Extended Unique Identifier (EUI) is generated by the LoRa Alliance, uniquely for the device and application [12].

3. Methodology

3.1. Hardware design

The hardware design includes the temperature and humidity sensor, turbidity sensor, pH sensor, LoRa shield and gateway. The LoRa shield has the Arduino Uno Microcontroller. The hardware design is shown in the Fig. 1.

The function of each component is as follows:

- Temperature and humidity sensor is used to measure the heat and humidity of waste water plant.
- Turbidity sensor is to measure the waste water turbidity level.
- pH sensor is to measure the level of acidity and alkalinity of the waste water.
- Microcontroller (Arduino Uno) is used to process sensor reading values and determine the waste water status.
- Shell (Dragino) is to forward the data read from a microcontroller to the LoRa gateway.

- The LoRa gateway is to connect the server via the internet. The server can be interfaced with IoT platform and display the sensor values.

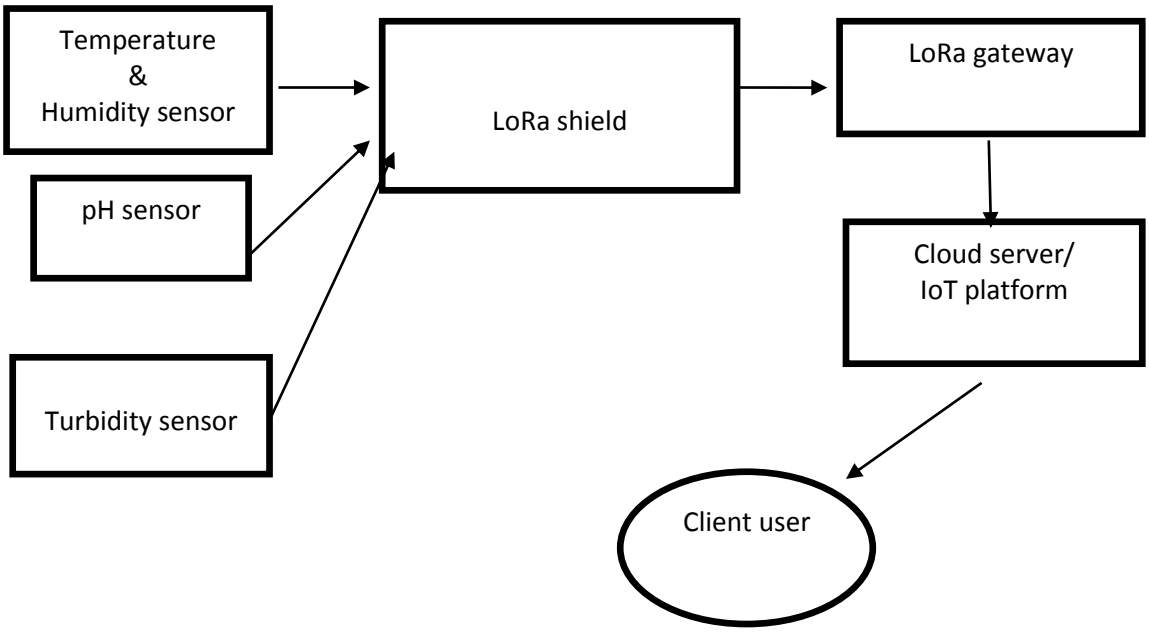


Fig.1. Block diagram

3.2. Software design

The software design is used, to program, display data from hardware. The user name and password has to be fed in the server and IoT platform to display the values.

4. Result

The system is created and tested. The results are obtained using the Arduino 1.8.10. The sensor values are displayed in the web server The Things Network and IoT platform Ubidots. The user can understand the data easily from them. The accuracy of the pH and turbidity sensor is checked using different samples. The turbidity and pH range of different water samples are given in the Table 1.

The unit used for measuring turbidity is Nephelometric Turbidity Unit (NTU). The pH range is determined using the pH scale. NTU= 1milligram of Silica per liter of water. The Temperature is measured in Celsius and Humidity is given in percentage.

Table 1 pH and Turbidity for different samples

Type of sample	pH (in pH scale)	Turbidity (in NTU)
Tap water	8	0.98
Muddy water	9	2.3
Drinking water	7.5	0.12
Silk dyed waste water	12	4

The sensor values displaying in The Things Network is given in the Fig. 2

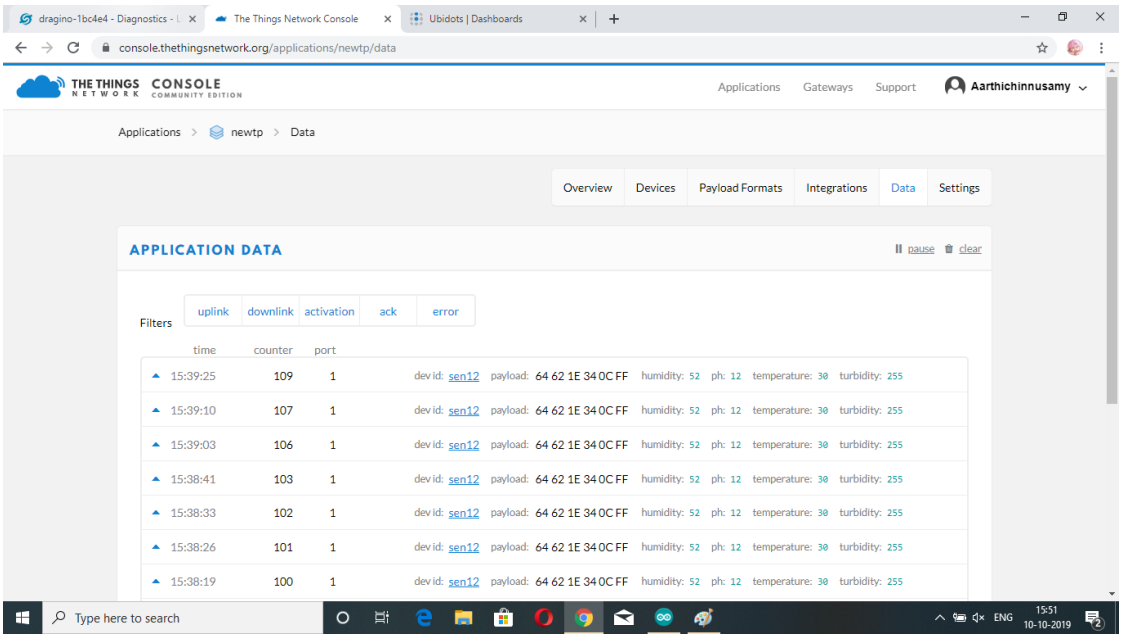


Fig. 2 Sensor values in TTN

The sensor values in Ubidots IoT platform is given in the Fig. 3

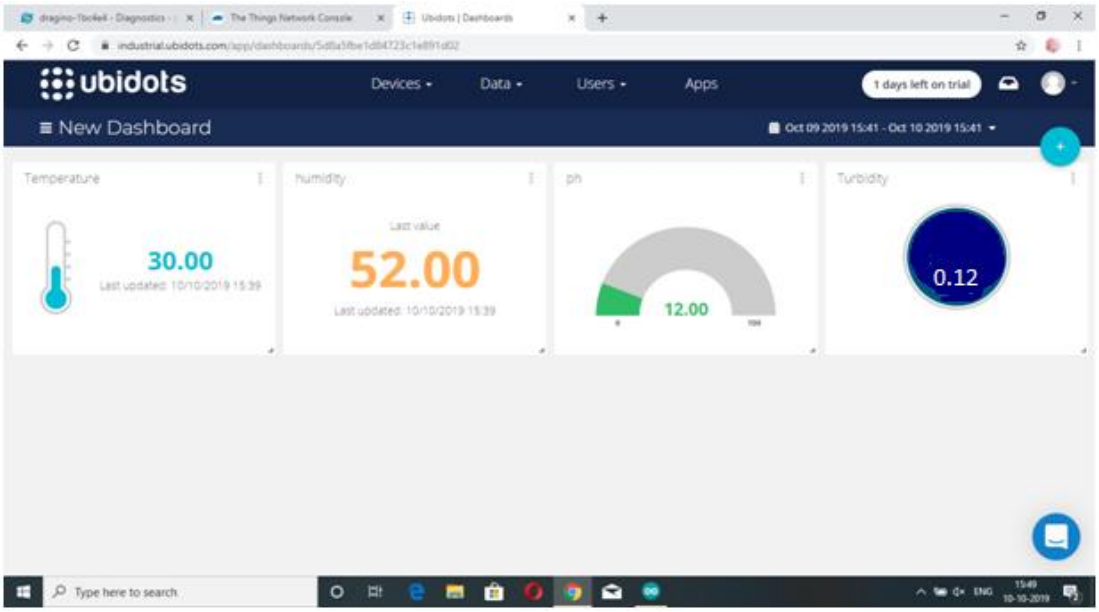


Fig. 3 Sensor values in Ubidots

4.1 SMS and Email Alert

The SMS Alert or Email Alert is sent to the registered phone number and Email ID of the factory owner or person in charge when the threshold level exceeds. The threshold level setting and the registration of the phone numbers are done in Ubidots. There are also other options in Ubidots such as call alert, alerting the workers in the factory by led.

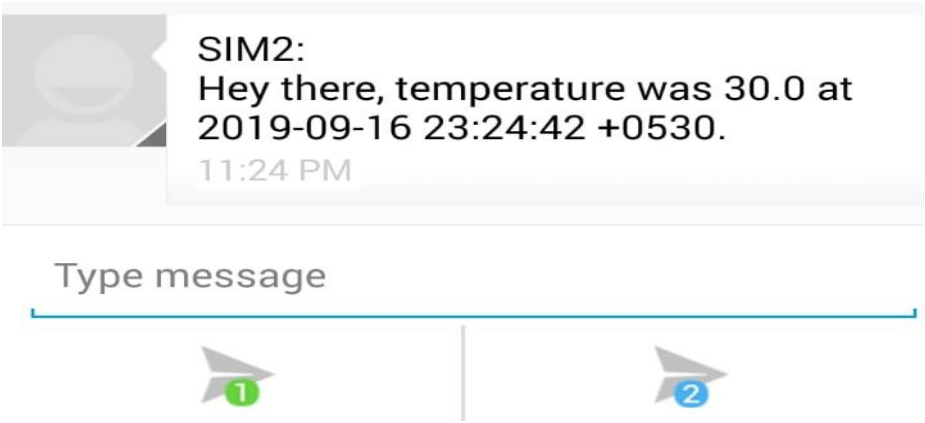


Fig. 4 SMS Alert

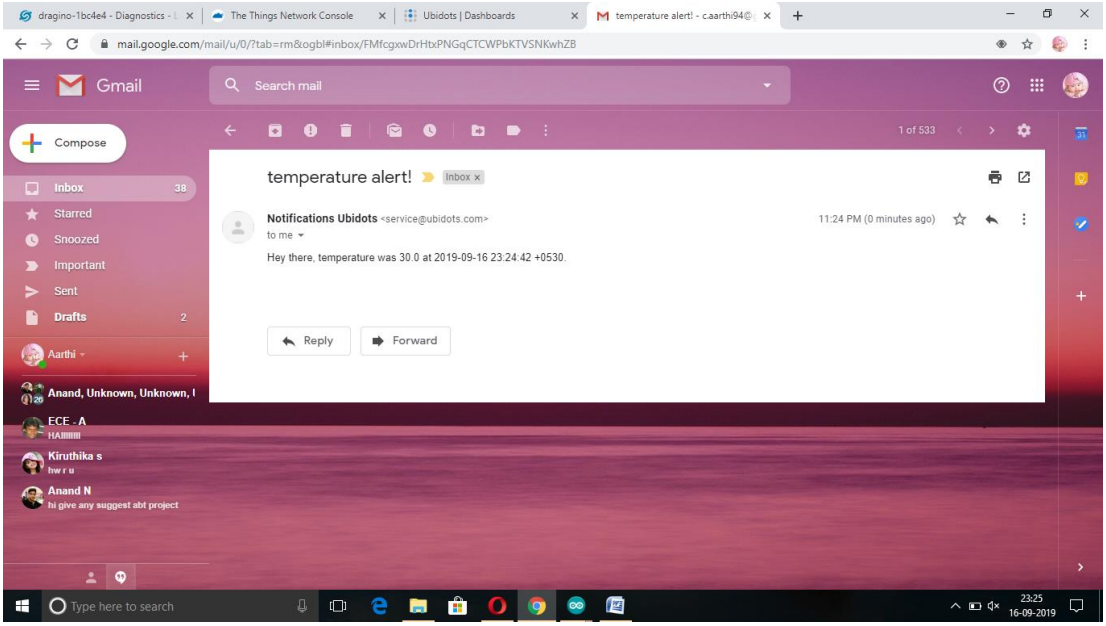


Fig.5 Email Alert

5. Conclusion

The system displays the waste water data from factories. The alerts are sent to the factory owner when the waste water thresholds level exceeds. The user can view and understand the data easily. The GPS shield can be used to view the location of the waste water plant and reverse link can be established to make the user take some actions during emergency. Chemical sensors like sodium; nitrate can be used to know the amount of chemicals in the waste water. This can help in chemical specific treatment of the waste water.

Acknowledgements

We like to thank the PSG College of Technology for providing the necessary equipments. We thank the Enthu technologies, Coimbatore for providing hands on training in LoRa. We like to thank the teaching and non teaching staff members, family and friends for their immense support.

References

[1] Chen, Y., Han, D., 2018.Water Quality Monitoring in Smart City: A Pilot Project, *Elsevier*, *Automation in Construction*, vol 89, pp.307316.

[2] Jedermanna, R., Borysova, M., Hartgenbuscha, N., Jaegerb, S., Sellwige, M., Lang, W., 2018.Testing Lora for food applications- Example application for

airflow measurements inside cooled warehouses with apples, *Elsevier, Procedia Manufacturing*, vol.24, pp. 284-289.

[3] Gosha, S., Goswamia, A.J., Ghoshb, G.K., Pramanika, P.,2018. Quantifying the relative role of phytase and phosphatase enzymes in phosphorous mineralization during vermicomposting of fibrous tea factory waste, *Elsevier, Ecological Engineering*, vol 116.pp. 97-103.

[4] Song, T., Lin, J., Tang, M., Dong, S.,2017. An Internet of Energy Things Based on Wireless LPWAN, *Elsevier, Engineering*, vol 3,pp.460-466.

[5] Reza Nourjou, Mehdi Hashemipour,2017.Smart Energy Utilities based on Real-Time GIS Web Services and Internet of Things, *Elsevier, Procedia computer science*, vol.110,pp.8-15.

[6] Wixted, A.J., Kinnard, P., Larijani, H., Tait, A., Ahmadinia, A., Strachan, N.,2016. Evaluation of LoRa and LoRa WAN for Wireless Sensor Networks, *IEEE SENSORS*,pp.1-3.

[7] Yudi, E., Windarto, Agung, B., Prasetijo, Galang F., Damara. 2018.A GIS-based Waste Water Monitoring System Using LoRa Technology, 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE).

[8] Somayya Madakam, R.Ramaswamy, Siddharth Tripathi,2015.Internet of Things(IoT):A Literature Review, *Journal of Computer and Communications*, Vol 3.pp.164-173.

[9] Petric, T., Goessens, M., Nuaymi, L., Pelov, A., Toutain, L.,2016.Measurements, Performance and analysis of LoRa FABIAN, a real-world implementation of LPWAN, *IEEE SENSORS*, pp.9-11.

[10] Mohamed Elde frawya, IsmailButuna ,Nuno Pereirab,Mikael Gidlun,Formal Security Analysis Of LoRaWAN , *Elsevier, Computer Networks*, 2018.

[11] A White Paper prepared for the LoRa ALLIANCE by Gemalto, Actility and SEMTECH, feb 2017.

[12] Emekcan Aras, Gowri Sankar Ramachandran, Piers Lawrence and Danny Hughes, Exploring the Security Vulnerabilities of LoRa, *IEEE Conference*, 2017.

Multi Document Abstractive Summarization using Graph-Based Technique for Big Data

Anish Mathew Kuriakose,DrV.Umadevi*

Ph.D Research Scholar, Department of Computer Science,

Jairams Arts and Science College Karur affiliated to Bharathidasan University Tiruchirappalli,India

Director, Department of Computer Science,

Jairams Arts and Science College Karur affiliated to Bharathidasan University Tiruchirappalli,India

Abstract

The advancement of internet technologies impressed the people and keep every day of people to be with its digital power. This development in technologies enforced public users to make and share the contents through online media such as Facebook, LinkedIn, blogs, forums and news channels etc. These data generation due to the over usage leads to data deluge and nowadays it is known as 'Big Data'. The information collected from these news, blogs, forums are precious and unpredictable. Finding the insights from the variety of unstructured data is the most challengeable for the industrialists and researchers. This paper mainly focusing on collecting the voluminous text information and providing them a valuable insights. These unstructured text are presented in hundreds and thousands of web pages. This research work aims to provide an abstractive summarization technique to summarize the big data. A novel framework has been proposed in this paper using the hybridization of the topic modelling and graph based text ranking to summarize the huge amount of text data into a meaningful short summary. The core idea of this research work is to summarize the huge text from multiple documents using the graph theory technique.

Keywords: Big data, Text summarization, Abstractive text summarization, Topic modelling.

* Corresponding author.
E-mail address: anishmathewkuriakose1@gmail.com

1. Introduction

The data on World Wide Web is creating at an exponential pace. Nowadays, people use the web to find information through information recuperation instruments, for instance, Google, Yahoo, and Bing and so forth [1]. Regardless, with the exponential improvement of information on the web, information reflection or abstract of the recouped results has ended up being indispensable for customers. In the present time of information over-trouble, text summarization has transformed into an indispensable and promising instrument for customer to quickly grasp the sweeping volume of information.

The goal of modified substance layout is to merge the chronicles into a shorter frame and spare fundamental substance. A chronicle diagram keeps its rule content, encourages customer to fathom and decipher huge volume of substance in the record and in this way decline customer's the perfect open door for finding the key information in the record. Overview, as done by individuals, incorporates examining and understanding an article, site, or document to find the key core interests. The key centres are then used to deliver new sentences, which outline the abstract. For individuals, making a diagram is an immediate technique yet the opportunity has arrived eating up. In like manner, the necessity for motorized frameworks is twisting up evidently increasingly clear to normally deliver the once-over and get the general idea of since quite a while ago printed information. One of the basic things to ask in abstract is "What are the compositions that should be addressed or kept in a summary?" The layout must be delivered by picking the fundamental substance or, on the other hand ends in the principal content. Finding crucial information transforms into an extremely troublesome endeavour. At present, the prerequisite for customized content summary has appeared in various districts, for instance, news stories abstract, email layout, short message news on versatile, and information plot for agent, government specialists, investigate, online web search apparatuses to get the structure of pages discovered accordingly on [1].

Rundown is the assignment of making a compact and exact synopsis of an enormous report or a lot of archives. Manual content outline is consistently utilized in the field of news coverage where journalists need to make features, rework meets, and gather a lot of data into a couple of sentences [2]. In any case, outline isn't confined to news coverage and books. Indeed, even in the scholarly world, we frequently make outlines of diary and meeting papers, propositions, and expositions to feature the key commitments of research. To help and substitute for manual rundown, PC researchers

and language specialists have chipped away at programmed content synopsis since late 1950s [3] with shifting degrees of progress.

In text mining, text summarization is a one among the major issues. [4]. It gives various advantages to clients and various productive genuine applications can be created utilizing content rundown. In content synopsis a huge assortments of content records are changed to a diminished and reduced book report, which speaks to the review of the first content assortments. A condensed record assistances in understanding the substance of the enormous content assortments rapidly and furthermore spare a ton of time by abstaining from perusing of every individual archive in a huge book assortment. Numerically [5], content synopsis is an element of changing over enormous content data to little content data in such a way, that the little content data conveys the general image of the huge content assortment as given in condition (1), where T expresses to the huge content assortment and t expresses to the outlined content record and the size of huge content assortment T is bigger than the size of condensed archive t.

$f: T \rightarrow t$, where $T > t$ …… (1)

The calculation plays out the assignment of content synopsis is called as content summarizer. The content summarizers are extensively classified in two classifications which are single-report summarizer and multi-archive summarizers. In single-archive summarizers, a solitary enormous book report is condensed to another single record outline, though in multi-archive synopsis, a lot of content records (multi archives) are abridged to a solitary archive rundown which speaks to the general look at the different records.

Multi-archive outline is a strategy used to condense numerous content records and is utilized for seeing enormous content report assortments [6]. Multi-record rundown produces a minimal synopsis by extricating the applicable sentences from an assortment of reports based on archive subjects. In the ongoing years scientists have specified a lot of consideration towards creating report synopsis systems. Various synopsis procedures are proposed to create outlines by take away the significant sentences from the given assortment of records.

Multi-record outline is utilized for comprehension and examination of huge record assortments, the significant wellspring of these assortments are news documents, online journals, tweets, pages, investigate papers, web indexed lists and specialized reports accessible over the web and different spots [7]. A few instances of the

utilizations of the multi-report rundown are breaking down the web indexed lists for helping clients in further perusing [8], and creating outlines for news stories [3]. Report handling and outline age in an enormous book archive assortment is computationally mind boggling task and in the period of Big Data investigation where size of information assortments is high there is need of calculations for abridging the huge content assortments quickly. In this paper, a MapReduce structure based rundown strategy is proposed to create the outlines from enormous content assortments. Test results on UCI AI storehouse informational collections uncover that the computational time for abridging huge content assortments is radically decreased utilizing the MapReduce system and MapReduce gives adaptability to pleasing enormous content assortments for condensing. Execution estimation metric of outline ROUGE and Pyramid scores are additionally gives satisfactory qualities in abridging the enormous content assortments.

Single-record synopsis is anything but difficult to deal with since just a single content report should be broke down for outline, though taking care of multi-archive rundown is a mind boggling and troublesome task. It requires some of (various) content records to be dissected for producing a smaller and educational (important) synopsis. As the quantity of reports increments in multi-record synopsis, the summarizer gets more challenges in playing out the rundown. A summarizer is said to be great, in the event that it contains progressively productive and important reduced portrayal of enormous content assortments. Considering semantic comparative terms give benefits regarding producing increasingly pertinent outline yet it is more register concentrated, since semantic terms will be created and considered for making rundown from an enormous book assortment. In this work the issues with multi-archive content synopsis are tended to with the assistance of most recent advancements in content investigation.

The main goal of this research work is to provide a novel architecture to summarize the text documents from multiple webpages using graph based methodology. The detailed state-of-art is discussed in the background study. The proposed architecture is described in the section three. The research work is concluded in the section four.

2. Background Study

The two primary ways to deal with programmed content rundown are extractive and abstractive [9]. Extractive outline is a method that concentrates or chooses notable sentences from the report. This for the most part includes positioning each sentence as indicated by how significant it is in the archive [10]. Abstractive synopsis, then again,

summarizes sentences from the record to make an outline as a human would. Cutting edge usage of abstractive rundown are neural arrange based strategies [11].

The author [1] had proposed a framework using the popular parallelized algorithm Map Reduce for Density-based spatial clustering of applications with noise **(**DBSCAN) clustering algorithm along with the Hidden Markov Model for big data text summarization. The framework provided an integrated application of enormous amount of web pages (text documents) for summarization. The author had used machine learning techniques for pre-processing and topic based model for identify the topic of the documents.

One of the principle testing assignments in various archive synopsis is sentence requesting and positioning. There are a few techniques has been proposed for sentence positioning/requesting up until now. Record synopsis has tremendous applications in regular language preparing. In light of the comparability between sentences, the last synopsis will be made. So regardless the sentences that don't have a similar surface structure, the copy judgment ought to be taken. As of late created methodology for sentence requesting is by [15] introduced an inclination learning system to sentence requesting for multi document synopsis. This work utilized five inclination estimates, for example, order, probabilistic, topical closeness, priority and progression to discover likeness between the various sentences in the rundown. Eager calculation is utilized to locate the absolute requesting among the sentences and diagram model is utilized to arrange the sentences. List wise way to deal with sentence positioning has been created by Fen Xia et al. The properties of the misfortune capacities including consistency, coherence, sufficiency, convexity, effectiveness and differentiability are examined to discover rank. An adequate condition on consistency for positioning is given first time in this work. Utilizing bolster vector machine, [16] built up a sentence requesting technique for different information archive outline. The rundown sentences are placed into various classifications bunch as indicated by the source archives. Arrangement of each gathering is chosen dependent on the directional relativity of contiguous sentences. In [3], a component explicit sentence positioning technique has been created. It utilizes reliance parser instead of normal vector space model, so that the reliance between each pair of words is unequivocally determined to make viable outline. In [12], a methodology is created which customizing the vector space model and using Latent Semantic Analysis (LSA) on it to deliver update synopsis from different archives. A two way printed entailment framework was created by pakray et al.

The lexical and syntactic highlights are utilized to decide the entailment among content and speculation. The AI calculation is prepared utilizing these two highlights. The classifier return either yes or no dependent on the entailment choice. All the above existing way to deal with sentence requesting in numerous information record synopsis doesn't tended to the semantic relationship and sensible surmising between sentences in the rundown which is imperative to produce important outline. The proposed framework primarily centers on finding the semantic and coherent connection between sentences by building compelling sentence requesting methodologies.

MapReduce [13] is a software design for usage of dispersed calculations of high volume or enormous information and an execution structure for huge scale information preparing on groups of servers. One of the MapReduce open-source execution is Hadoop [4] by Apache Foundation Projects. The mapper is applied to each information key-esteem pair to create a self-assertive number of middle of the road key-esteem sets. The reducer is applied to all qualities related with a similar transitional key to create yield key-esteem sets. Mappers and reducers are objects that actualize the Map and Reduce strategies, individually.

The conventional sequence-to-sequence centred prototypes are the causes for the producing the incorrect aspect of the summarization when it is compared to the original content. The authors [17] have proposed a new and novel method named as Reader-Aware Summary Generator with the help of adversarial learning. The proposed method is not same as the usual summarization techniques. It has more advantages to reduce the noises in the causal comments that are posted by the public users in the forums, blogs, etc. and also it joints the documents published by the news media along with the user’s blogs.

In this big data era, providing scalable summarization technique is the most challengeable task. To propose a machine learning based model, it is needed that, the training dataset must be huge in size. The authors [18] have handled this problem by proposed a scalable machine learning model for evaluating the abstractive summarization technique. They also discussed the state-of-art models in the summarization research field.

3. Proposed Architecture

The previous section discussed the related work that had been proposed for the text summarization. To solve issues mentioned in the related work, this research work has been structured with a novel framework named as Big Data Abstractive Summarization (BDAS). Figure 1 depicts the high level architecture of the novel framework. The following are the objectives of the proposed framework.

- Collecting the web documents based on the topics using topic modelling approach
- Linking the contents between same topic in various documents using the graph-based technique
- Combining topic modeling and ranking the contents and to provide the abstractive summary

To achieve the above said objectives, the framework has been divided into three phases. These phases are playing the major role at its levels. Each phases are important and have its own characteristics to define their vital presence in the framework. The three phases are as follows:

Phase – 1: Pre-processing
Phase – 2: Data preparation
Phase – 3: Summarization

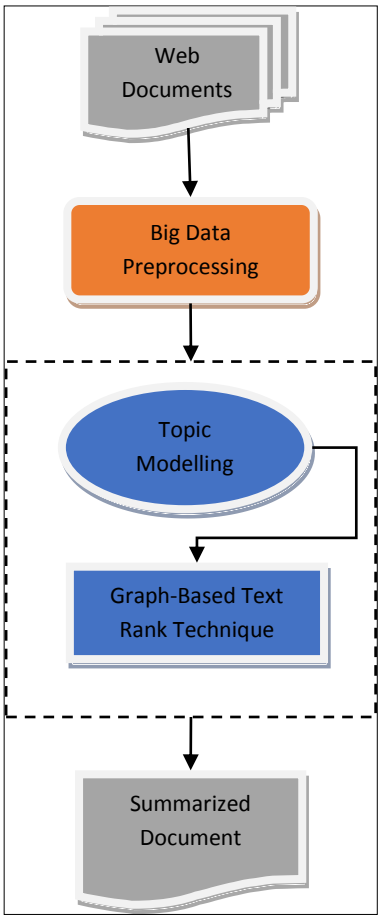


Fig 1. Big Data Summarization Framework

3.1 Preprocessing:

Preprocessing is the initial and most common stage for any kind of data analytics work. It is the key phase to determine the quality of the data. Here in this research work, the complexity of preprocessing is very high, since, the data are huge in size with more varieties and unstructured in format. Cleaning the unstructured data is the most challengeable task for preparing the data. The web documents contain huge amount of valuable information as well as the high amount of the noisy data. It contains URL links, video tags, audio tags, images, etc. For summarizing the text documents these noisy data are not needed and to be eliminated during the data collection.

For collecting the data, web crawling is the most common method that has been used. The noisy data are removed during the data collection and stores the raw text. The raw text is not enough to measure the data quality. It contains lot of stop words and jargons. The people used their colloquial language to express their views and thoughts. This leads to change the meaning of the sentence during summarization and provides the low quality of summary.

To avoid those mistakes, the common NLP (Natural Language Processing) tasks has been used. The tasks are as follows:

i. Keeping the key terms

ii. Tokenizing the important phrases

iii. Lemmatize the words and keep the root words

In this research work, the basic step of NLP removing stop words is not being used. The stop words are acting like the sentence determination. Hence it is not removed during the preprocessing phase.

Data collection and preprocessing has been done in the initial phase. The preprocessing is done during collection of data. Here it is proposed to use the 'In-Memory analytics' to preprocess the data during data collection. This avoids the data storage and increase the computational time.

3.2 Data Preparation

For any type of data analytics related work, data preparation is another important phase. This phase is applicable to prepare the data for the analytics environment. After cleaning the raw text, the cleaned data are transformed to the required format.

The proposed work has been implemented on the various news articles. The data has been collected on the events news. The different types of media publish the news in their own way of content. After collecting data from multiple web documents, the collected data are preprocessed during the data collection and stored into a dump of text file. Each document contains its own id which is referred in the parent document. The following is the mathematical function which represents the data preparation of big data text summarization.

Definition – 1: Let's assume the keyword as K, event as E, parent document (parent website where the information has been crawled) as d, where d contains the rich informative text data and represents as T, the informative text contains the most commonly used key phrases which is denoted as t and the data for summarization as D.

$$D \rightarrow \{K, E, d\} \quad (1)$$

$$K \rightarrow \{E: e\} \quad (2)$$

$$d \rightarrow \{T: t\} \quad (3)$$

From the above equations, 'D' denotes the set of functions containing K, E and d. Where 'K' is the key term which contains the event. The event such as news, sports, politics, disasters, natural events, science, technologies, etc. 'E' is the collection of events. 'd' denotes the preprocessed textual data.

3.3 Summarizer

This is the final and valuable phase of the proposed framework. The prepared data are processed through the summarizer phase and provides the meaningful summary. This phase contains two techniques that are as follows:

i. Topic modelling

ii. Graph based text ranking

The novelty of the proposed framework is hybridization of topic modelling along with the graph based text rank algorithm.

Topic modelling is the most common technique that has been used across several applications. In the proposed research work, the topic modelling has been used to identify the various websites based on the topic (key term) given during the web crawling. This identifies the webpages across the internet under the same topics and

dumped the data into a single document. For topic modelling the popular technique Latent Dirichlet Allocation (LDA) has been used. It is a unsupervised learning model. LDA performs the following tasks at macro level for topic modelling on each every document d:

1. Consider there are 't' topics through all over the webpages under the same event 'e'.
2. Allocate these 't' topics through the web document *d* by allocating topic to each word.
3. Consider the topic 't' is incorrect to the other word in the document which is allocated as correct topic for each word in the web documents.
4. Find the probability of correctness across the topics
5. The above steps has to be repeated until the document reaches it end.

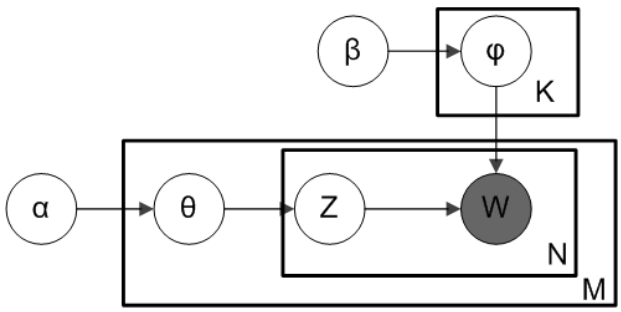


Fig 2. LDA model

Figure 2 shows the work flow of the LDA model [14], where the notations are described as follows.

α is the single document taken for the topic distributions across the multiple documents,
β, the single topic word distribution,
θ, topic distribution for document *d*,
φ, word distribution for identified topic *t*,
z, topic for the nth word in web document *d*, and
w, exact word

It is observed from the diagram that w is turned grey out. This is on the grounds that it is the main detectable variable in the framework while the others are dormant. Along these lines, to change the model there are a couple of things that can disturb and beneath the centre around two. α is where each line is a record and every section

speaks to a subject. An incentive in push x and segment y speaks to how likely report x contains point y. A symmetric dissemination would imply that every point is equitably circulated all through the archive while a lopsided appropriation supports certain themes over others. This influences the beginning stage of the model and can be utilized when it have an unpleasant thought of how the themes are appropriated to improve results. β is where each line speaks to a theme and every segment speaks to a word. An incentive in push x and segment y speaks to how likely that point x contains word y. Generally each word is appropriated equitably all through the point with the end goal that no theme is one-sided towards specific words. This can be misused however so as to predisposition certain points to support certain words.

For instance on the off chance to realize the theme about political events it tends to be useful to predisposition words like "Donald" and "Modi" for one of the points so as to push the model towards finding that specific subject. Because these two names are most popular and used in the political events across the internet.

For summarizing the large text, the data obtained from the topic modelling are linked to each other using the text rank algorithm. The text rank algorithm ranks the key phrases. Graph based approach is used to rank the key phrases by creating relationships among the phrases.

Definition – 2: Graph G is a set of functions contains phrases p as vertices and edges based on the rank R.

$$G \rightarrow (P, V, E, R) \qquad (3)$$

Where P denotes the phrases, V denotes the vertices, E as edges and R as ranker id of the phrases.

The ranked phrases are parsed into the huge set of English corpus. The corpus contains the synonyms of words. The matched words are replaced by the synonyms and it creates the new sentences. Grammar of the new sentences is verified by the predefined grammar rules which are stored in the grammar corpus. The abstractive summarization from the huge set of unstructured documents are generated by hybrid the topic model and graph based text ranking algorithm.

4. Conclusion

The proposed framework has being testing with the real world data. The challenges faced during the testing are, lack of dedicated corpus for the words related to the

specified events or topics. Creating those corpus will reduce the complexity of classifying of topics and will provide more accuracy on the extracted document. The issues founded from this work will be solved in the future and it will be provided to the social media for reducing the complexity of the multi-document big text data summarization.

References

[1] Krushnadeo Tanaji Belerao, Dr. S. B. Chaudhari, "Summarization using Mapreduce Framework based Big Data and Hybrid Algorithm (HMM and DBSCAN)", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017), 2017.

[2] Li HG, Wu GQ, Hu XG, Zhang J, Li L, Wu X (2011), "K-means clustering with bagging and Mapreduce". Proceedings of 44th Hawaii International Conference on IEEE System Sciences (HICSS). Kauai/Hawaii, US, 2011

[3] Kogilavani, P. Balasubramanie, "Semantic summary generation from multiple documents using feature specific sentence ranking strategy", Elixir Comp. Sci. Engg. 2011, Vol 40, pp. 5372-5375.

[4] Borthakur, D. "The hadoop distributed file system: Architecture and design" 2011, https://hadoop.apache.org/docs/r1.2.1/hdfs_design.pdf.

[5] Russom P, "Big Data Analytics. TDWI Research Report", US, 2011, pp 1–38

[6] Chen S-Y, Li J-H, Lin K-C, Chen H-M, Chen T-S, "Using MapReduce Framework for Mining Association Rules", In Information Technology Convergence Springer, Netherlands, 2013, pp 723–731

[7] Steve L, "The Age of Big Data. Big Data's Impact in the World, New York, USA, 2012, pp 1–5

[8] Chananuparp P., Hu X. and Shen X. "The evaluation of sentence similarity measures", 10th International Conference on DW and Knowledge Discovery 2018.

[9] Li F, Ooi BC, Özsu MT, Wu S, "Distributed Data Management Using MapReduce", ACM Computing Surveys, 2013, Vol 46:1 pp:41-46.

[10] Verma A, Llora X, Goldberg DE, Campbell RH, "Scaling Genetic algorithms using MapReduce", Intelligent Systems Design and Application(ISDA). Ninth International Conference, Pisa, Italy, 2017, pp 13–18

[11] Cambria E, Rajagopal D, Olsher D, Das D, "Big Social Data Analysis", Big Data Computing, 2013, Chapter 13:401–414

[12] Dean J, Ghemawat S, "MapReduce: Simplified data processing on large clusters", Proceedings of the 6th Symposium on Operating System Design and Implementation, San Francisco, California, 2004, pp 137–150

[13] Dean J, Ghemawat S, "MapReduce: A flexible data processing tool", ACM, 2010, vol 53:1, pp:72–77

[14] LDA model, internet source https://towardsdatascience.com/lda-topic-modeling-an-explanation-e184c90aadcd as on dated 15-Nov-2019.

[15] Bollegala, D, Khan, M.A.H, Liu, G, Sezaki, K, "Multi-tweet summarization of real-time events", Socialcom, 2013.

[16] Peng, H., Poria, S., Hussain, A., Howard, N., Cambria, E. "Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis", Neuro Computing, 2017.

[17] Shen Gao, Xiuying Chen, Piji Li, Zhaochun Ren, Lidong Bing, Dongyan Zhao and Rui Yan, "Abstractive Text Summarization by Incorporating Reader Comments", The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19), 2019, pp 6399-6406.

[18] Wojciech Kryscinski, Bryan McCann, Caiming Xiong and Richard Socher, "Evaluating the Factual Consistency of Abstractive Text Summarization", arXiv:1910.12840v1, 2019.

Detection of Pulsars using Machine Learning Algorithms – A Study

Partha Sarathi Chakraborty*

**Assistant Professor, Dept. of Information Technology, University Institute of Technology, The University of Burdwan, Burdwan,713104, India*

Abstract

Pulsars, a type of neutron stars, are of massive interest for the space scientists due to several reasons. The light beams produced by the pulsars are detectable from the earth. In this paper, after a brief introduction of pulsar stars and surveys conducted on pulsar stars, a study of works done by different researchers for detecting pulsars using machine learning algorithms has been made by discussing the approaches under two categories-offline and online. Three machine learning algorithms namely Logistic Regression, Random Forest and k-Nearest neighbor have been tested on HTRU2 dataset to separate radio emissions coming from true pulsars and their performance have been compared**.**

Keywords: Neutron stars; Pulser Detection; Machine Learning; Random Forest; Logistic Regression; K-Nearest Neighbors;

1. Introduction to Pulsar

The pulsar stars are a type of neutron stars which have been named after combining two words-"Pulse" and "Star".

When a massive star runs out of fuel it crashes and at the core of the star, every proton and electron turns into a neutron. If the resulting core which consists of neutrons has weight between one to three solar masses, a neutron star is formed. Neutron stars can be of type Pulsars or Magneters. Pulsar stars are most common types of neutron stars which periodically emits pulses of radiation. Pulsar stars produces light beams which are detectable from earth while crossing our line-of-sight.

Pulsars have been described as excellent natural laboratories [1] due to several reasons. Pulsars help to study stellar evolution, the nature of gravitation, the

* Corresponding author.
E-mail address: psc755@gmail.com

composition of the Interstellar medium (ISM) and much more. [1], [2], [3], [4] elaborates the importance of research on pulsars. The paper is organized as follows. Section II provides a brief introduction to radio telescopes and pulsar surveys. Section III discusses different categories of approaches proposed by researchers for detecting pulsars. Section IV concentrates on machine learning based strategies. Section V describes HTRU2 dataset first and then compares performance of three machine learning algorithms in detecting pulsars using HTRU2 dataset and finally, the paper is concluded in section VI.

2. Radio Telescopes and Pulsar Surveys

Unlike optical telescopes, radio telescopes receive radio waves instead of light beams. They are used to receive radio waves on earth from pulsars, galaxies and other astronomical objects. A radio telescope is consisting of at least two parts-an antenna and a receiver. A large antenna or an array of large antennas is used to receive very weak signals emitted from astronomical objects which are million miles apart from earth. The weak signals are amplified for making further analysis possible.

Radio Telescope at the Arecibo Observatory (305-metre)located in Puerto Rico, Five-hundred-metre Aperture Spherical Telescope (FAST) located in China, Robert C. Byrd Green Bank Telescope (GBT) located in West Varginia are examples of some single large radio telescopes. On the other hand, Very Large Array (VLA) consisting of 27 parabolic antennas of 25 metres each in diameter located in New Mexico, Westerbork Synthesis Radio Telescope antennas consisting of 14 antennas, each of 25 metres in diameter, Giant Metrewave Radio Telescope (GMRT) consisting of 30 antennas located in India are examples of some radio telescope arrays.

The era of pulsar surveys have been started from the end part of sixty's decade [5]. The process followed to detect pulsars was totally manual in the initial surveys and the number of pulsars detected was very small. Gradually, use of computer and automated searching tools takes the place of searching "by eye". Table 1[6] shows a listing of some major surveys along with their year of conduction and Name of the Telescope.

Table 1. Major pulsar surveys along with year of conduction and name of Telescope

Name of the Survey	Name of the Telescope	Year
1st Molonglo Survey [5]	Molonglo	1968
Jodrell Survey C (610 MHz, 925 MHz, 928 MHz, 1420 MHz) [7]	Jodrell Bank	1985-87
Parkes 20-cm Survey (System I, System II) [8]	Parkes	1988
High Galactic Latitude Pulsar Survey of the Arecibo Sky (H1,H2,H2,H4,H5) [9]	Arecibo	Between 1990-1995
Arecibo Survey 4 Phase I & Phase II [10]	Arecibo	1991 &1992
Parkes Southern Pulsar Survey [11]	Parkes	1991-93
Parkes Multibeam Pulsar Survey [12]	Parkes	1997
Parkes high-lat multibeam pulsar survey [13]	Parkes	2000-03
Survey of the Magellanic Clouds [14]	Parkes	2000-01
1.4 GHz Arecibo Survey (DM < 100 & DM > 100) [15]	Arecibo	2001-02
Large Area Survey for Radio Pulsars [16]	Parkes	2001-02
EGRET 56 Pulsar survey [17]	Parkes	2002-03
Parkes deep northern Galactic Plane survey [18]	Parkes	2004-05
P-ALFA Survey (WAPP) [19], [20]	Arecibo	2004-10
6.5 Ghz Multibeam Pulsar Survey [21]	Parkes	2006-07
Green Bank 350 MHz (Spigot, Drift Scan) [21], [22]	Green Bank	2007
P-ALFA Survey (MOCK) [23], [20], [24]	Arecibo	2009-present
GBNCC (GUPPI) [25], [26]	Green Bank	2009-14
Southern HTRU (Low, Medium, High) [27]	Parkes	2010-12
LOTAS [28]	LOFAR	2010-11
Northern HTRU (Low, Medium, High) [29], [30]	Effelsberg	2010-14
LOTAAS [31], [32]	LOFAR	2013
GMRT High Resolution Southern Sky Survey (MID & HIGH) [33]	GMRT	2014
SKA (Configuration A 3000 Dishes & Configuration B Aperture Array) [34]	SKA	2020-22

3. Categories of Pulsar Detection

The pulsar detection procedures can be broadly divided into three categories [6]- manual detection, semi-automatic detection and automatic detections. The manual detection involves identifying probable pulsar candidates by manual inspection of digitally produced diagnostic tools. Some graphical tools like RUNVIEW [35], REAPER [36], JREAPER [37] had been developed between 2001 and 2010 to facilitate the process of manual detection. Semi-automated selection approaches like PEACE [38] help manual detection of pulsars by ranking probable pulsar candidates. Automated approaches of pulsar detection highly rely on machine learning, specifically sophisticated classification algorithms for labelling candidates as pulsars and non-pulsars. The next section provides a detailed study of the detection algorithms using machine learning algorithms. Table 2 shows a listing of some major software tools along with their brief description.

Table 2. Major software tools for pulsar detection along with their brief description

Software Tool	Description
RUNVIEW [35]	• Interactive display and plotting tool.
REAPER [36]	• A collection of programs that shows each pulsar candidate as a single point on a phase-space diagram. • User interface provides of an X–Y plot of a pair of chosen parameters (typically period and S/N) from a list of 10 parameters. • The Galactic position of a possible candidate and the exact position, period and DM of known pulsars can be shown on the plot.
JREAPER [37]	• Supports wide range of data formats. • Easy pulsar searching by tracking viewed candidates and filtering displayed candidates by user selectable parameters. • JREAPER features a scoring algorithm that attempts to do this by assigning a numerical score to each candidate pulsar signal based on the data that are produced by the search algorithms. The scoring routine uses a number of tests that are then combined with weights to give a final score. • JREAPER was designed to automatically detect and remove known pulsars from the results.
PEACE [38]	• In Pulsar Evaluation Algorithm for Candidate Extraction (PEACE), a score function is used to rank pulsar candidates. • Validated using human-based ranking results.

All the above mentioned categories of detection are common in one way-they work on stored data obtained from different surveys and may be called as offline detection. Radio telescopes which have been installed in last few years (and are going to be installed within next few years) are producing so large amount of data that cannot be stored and then processed offline. Inevitably, online processing of stream data is the only choice which has been discussed in the second part of the next section.

4. Pulsar Detection using Machine Learning

1.1. Offline Detection Strategies

Detection of pulsar stars using machine learning algorithms is relatively a new area of research. Until recently this process was done manually. [39] summarizes the works on searching pulsars using machine learning algorithms.

Artificial neural network and deep learning have been extensively used in detecting pulsrs. [40], [41] and [42] used artificial neural network on Parkes, HTRU and HTRU-S survey data respectively. In their work, [40] investigated two neural network architectures (8:8:2 and 12:12:2) for processing nearly sixteen million pulsar candidates obtained from the Parkes multi-beam survey (after reprocessing data). Using their search software, twelve scores, as they name, were fed to neural network as input vectors. The problem of unbalanced training sets was not addressed in their implementation and they advocated in favour of more sophisticated neural network architecture. The paper [41] applied neural network on High Time Resolution Universe (HTRU) mid-latitude survey data where each candidate was described by twenty two feature values. The Straightforward Pulsar Identification using Neural Networks (SPINN) system was proposed by [42] which uses neural network with six input features namely (1) Signal-to-Noise ratio of the folded profile (log-scale), (2) Intrinsic equivalent duty cycle of the pulse profile, (3) Ratio between barycentric period and dispersion measure (log-scale), (4) Validity of optimized dispersion measure, (5) Persistence of signal through the time domain and (6) Root-mean-square distance between the folded profile and the sub-integrations. Table 3 shows a listing of some major research works which uses machine learning algorithms.

Table 3. Major research works on pulsar detection using machine learning algorithms

Papers	No of Features Used	Machine learning Algorithm
[40]	twelve features	Artificial neural network
[41]	Twenty two features	Artificial neural network
[42]	Six features	SPINN (artificial neural network)
[38]	Six features	PEACE
[6]	Eight features	Gaussian Hellinger Very Fast Decision Tree
[43]	Eight features	Fuzzy *k*-nn classifier

In paper [44], artificial neural network has been used in combination with Support Vector Machine and Convolution Neural Networks to develop Pulsar Image-based Classification System (PICS) for working on Pulsar Arecibo L-band Feed Array (PALFA) survey data for training and Green Bank North Celestial Cap survey (GBNCC) data for validation.

Other than artificial neural network and Convolution Neural Networks, Random Forests have been used by different researchers [45], [46]. [45] uses a random forest based classifier for improving their previously developed V-FASTR system for detecting rare fast radio bursts from Very Long Baseline Array (VLBA) data. [46] applied six machine learning algorithms- ANN , SVM, direct rule learner, standard tree rule learner, hybrid rule and tree learner and Random Forests. Author of paper [43] developed a fuzzy k-nearest neighbour based classifier which has been tested on HTRU2 dataset [6] . The author Ford J. M. made a detailed study on the performance of different classifiers (including ensemble classifiers) in the detection of pulsars in his PhD thesis [47]. The author also addressed the problem of class imbalance of training and test datasets and devised methods to reduce the effect. In their recent work, the authors of the paper [48] have used eleven deep learning models for pulsar detection.

1.2 *Online Detection Strategies*

The recent surveys for pulsar detection produce a huge volume of raw data and consequently candidate data which becomes impossible to store permanently. Table 4 [6] shows a listing of some recent surveys along with volume of data they produced.

Table 4. List of some recent pulsar surveys along with year of conduction and volume of data produced

Name of the Survey and Year Conducted	Volume of Data
Southern HTRU (LOW) (2010-12)	16gb per beam. 1PB data total (Ng 2014).
Southern HTRU (MED) (2010-12)	2gb per beam.
Southern HTRU (HIGH) (2010-12)	1gb per beam.
Northern HTRU (LOW) (2010-14)	13.4 gb per beam, ~38MBs-1. 5PB data total (Ng 2014).
Northern HTRU (MED) (2010-14	1.6 gb per beam.
Northern HTRU (HIGH) (2010-14)	0.8 gb per beam.
LOTAAS (2013)	Each 1-hour pointing produces 16 TB of raw data.
GMRT High Resolution Southern Sky Survey (MID) (2014)	76GB per beam.
GMRT High Resolution Southern Sky Survey (HIGH) (2014)	45GB per beam.

Among the surveys which are running currently, Square Kilometer Array (SKA) [34] is in the centre of attention to the space scientists. The main aim of Square Kilometer Array (SKA) is to find pulsars not only in our galaxy but also in other parts of the universe. In other words, SKA will make a “cosmic census” [49] of the pulsars of the universe. SKA1-MID and SKA1-LOW will be used to find pulsars in the Milky Way whereas potentially all the Galactic pulsars which emits radio signals towards the earth will be discovered in SKA2 survey. To meet this goal, SKA surveys will have to investigate a significantly large number of potential pulsar candidates which are multiple times greater than the numbers investigated by the previous surveys.

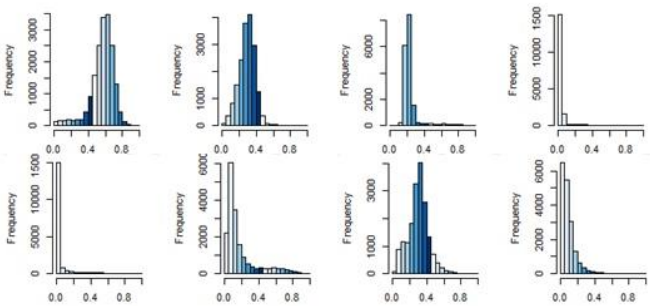
This leads the requirement of processing data online and developing stream classifier in this problem area. In recent times, a significant contribution has been made in this research area by [50], [51] , [6] studied HTRU-S survey data by applying some stream classifiers. In their work, [51] devised a novel classifier using Hellinger distance measure for imbalanced data streams. The authors of paper [6] developed Gaussian Hellinger Very Fast Decision Tree for detecting pulsars from the HTRU-1 and LOTAAS surveys. Twenty new pulsars have been discovered from the LOTAAS [31], [32] survey data.

5. Automatic Pulsar Detection using HTRU2 Dataset

HTRU2 [6] is a data set which describes a sample of pulsar candidates collected during the High Time Resolution Universe Survey (South) [6]. The data set contains 16,259 spurious examples caused by RFI/noise, and 1,639 real pulsar examples. Each pulsar candidate in the dataset is described by eight attributes namely Mean of the integrated profile, Standard deviation of the integrated profile, Excess kurtosis of the integrated profile, Skewness of the integrated profile, Mean of the DM-SNR curve, Standard deviation of the DM-SNR curve, Excess kurtosis of the DM-SNR curve and Skewness of the DM-SNR curve.

When analysing the HTRU2 [6] dataset, high correlation is observed between Excess Kurtosis Integrated Profile and Mean Integrated Profile, between Excess Kurtosis Integrated Profile and Skewness Integrated Profile, between Excess Kurtosis DM SNR Curve and Standard Deviation DM SNR Curve and between Skewness DM SNR Curve and Excess Kurtosis DM SNR Curve. Histogram of the eight individual attributes of the candidate pulsars are shown in Figure 1.

Figure 1. Histogram of the eight individual attributes of the candidate pulsars in HTRU2 dataset



Three machine learning algorithms namely Logistic Regression, Random Forest and k-Nearest neighbour have been considered for detecting pulsars from HTRU2 dataset [6]. The results have been shown in the table1. Four evaluation metrics (definition of each is given in table 2.) have been considered for measuring the efficiency of the models-accuracy, recall, precision and F-score.

Accuracy of a machine learning (ML) algorithm is defined as the ratio of number of correct classifications to the total number of samples examined. This metric is more suitable in cases where dataset having no class imbalances. Precision indicates the relevance of the results obtained from the ML algorithm. The metric, Recall, on the other hand, highlights the percentage of detection of correctly identified positive

samples. The Precision or Recall value which will be given more priority in analyzing the result of the ML algorithm depends on the problem domain. The metric, F-Score-harmonic mean of precision and Recall values indicates the performance of the algorithm in a more simple way. The definitions of the above-mentioned metrics have been shown in Table 5.

Table 5. Performance metrics

Accuracy Measure	Definition
Accuracy	(TP+TN)/(P+N)
Recall	TP/P
Precision	TP / (TP+FP)
F-score	(2 X precision X recall) / (precision + recall)
TP, TN, P and N are True Positive, True Negative, Total number of positives and Total number of negatives respectively.	

Table 6 shows the performance of three machine learning algorithms namely Logistic Regression, Random Forest and K-Nearest Neighbours in detecting pulsars from the HTRU2 dataset. It can be seen that Logistic Regression and Random Forest shows better performance than K-Nearest Neighbours algorithm.

Table 6. Performance of Logistic Regression, Random Forest and K-Nearest Neighbours algorithm

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	97.93	93.81	82.97	88.06
Random Forest	97.93	93.81	84.5	88.64
K-Nearest Neighbors	97.62	79.63	93.57	86.04

6. Conclusion

In this paper, researches on pulsar detection that have been done by researchers over the last few decades have been briefly studied. The necessity of shifting from offline detection strategies to online detection strategies has been mentioned and some recent research works in this direction has been mentioned. HTRU2 dataset which describes a sample of pulsar candidates collected during the High Time Resolution Universe Survey (South) has been used to show the effectiveness of machine learning

algorithms in detecting pulsars. The new generation pulsar surveys like SKA generate very large volume of data at a very high rate. This scenario demands more and more research on designing efficient online classification algorithms which can detect pulsars from data streams coming from software pipelines.

References

[1] Keane E. F. et al., E., 2014, Proceedings of Science, PoS(AASKA14)040.

[2] Cordes J. M., Kramer M., Lazio T. J. W., Stappers B. W., BackerD. C., Johnston S., 2004, New Astronomy Reviews, 4, 11-12,p.1413.

[3] Lorimer, D.R., Bailes, M., McLaughlin, M.A., Narkevic, D.J., Crawford, F.,2007. A Bright Millisecond Radio Burst of Extragalactic Origin. Science 318, 777. 0709.4301.

[4] Lyon R. J., 2015, PhD thesis, Univ. Manchester.

[5] Large M. I., Vaughan A. E., Wielebinski R., 1968, Nature, 220, 5169, p.753

[6] Lyon, R.J., Stappers, B.W., Cooper, S., Brooke, J.M., Knowles, J.D., 2016. Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach. MNRAS 459. 1603.05166.

[7] Biggs J. D., Lyne, A. G., 1992, MNRAS, 254, 257-263

[8] Johnston S., Lyne A. G., Manchester R. N., Kniffen D. A., D'Amico N., Lim J., Ashworth M., 1992, MNRAS, 255, 401

[9] Foster R. S., Cadwell B. J., Wolszczan A., Anderson S. B., 1995, ApJ, 454, 826

[10] Camilo F., Nice D. J., Taylor J. H., 1993, ApJ, 412, p.L37

[11] Manchester R. N. et al., 1996, MNRAS, 279, 1235

[12] Manchester R. N. et al., 2001, MNRAS, 328, 17

[13] Burgay M. et al.,2006, MNRAS, 368, 283

[14] Manchester R. N., Fan G., Lyne A. G., Kaspi V. M., Crawford F., 2006, ApJ, 649, 235

[15] Hessels J. W. T., Ransom S. M., Stairs I. H., Kaspi V. M., Freire P. C. C., 2007, ApJ, 670, 363

[16] Jacoby B. A., Bailes M., Ord S. M., Edwards R. T., Kulkarni S. R., 2009, ApJ, 699, 2, p.2009

[17] Crawford F. et al., 2006, ApJ, 652, 2, 1499

[18] Lorimer D. R., Camilo F., McLaughlin M. A., 2013, MNRAS, 434, 347

[19] Cordes J. M. et al., 2006, ApJ, 637, 446

[20] Deneva J. S. et al., 2009, ApJ, 703, 2, 2259

[21] Bates S. D. et al., 2011a, MNRAS, 411, 1575

[22] Boyles, Jason *et al.* Astrophys.J. 763 (2013) 80 arXiv:1209.4293 [astro-ph.HE].

[23] Spitler L. G. et al., 2014, ApJ, 790, 2

[24] Lazarus P., 2012, The PALFA Survey, IAU Symposium 291, http://www.pulsarastronomy.net/IAUS291/download/ Oral/IAUS291 LazarusP.pdf (accessed January 6th, 2016)

[25] Deneva J. S., Stovall K., McLaughlin M. A., Bates S. D., Freire P. C. C., Martinez J. G., Jenet F., Bagchi M., 2013, ApJ, 775, 1

[26] Stovall K. et al., 2014, ApJ, 791

[27] Keith M. J. et al., 2010, MNRAS, 409, 619

[28] Coenen T. et al., 2014, A&A, 570, A60

[29] Barr E. D. et al., 2013, MNRAS, 435, 2234

[30] Ng C., 2012, IAU Symposium, S291, 8, p.53-56

[31] LOFAR Pulsar Working Group, 2013, presentation at LOFAR Status Meeting, Dwingeloo, The Netherlands, March 6th, 2013 (http://www.lofar.org/wiki/lib/exe/fetch.php?media= public:lsm new:2013 03 06 hesself.pdf (accessed January 6th, 2016)

[32] Cooper S., 2014, presentation at LOFAR Science 2014, Amsterdam, The Netherlands, 7-11 April 2014. http://www.astron.nl/ lofarscience2014/Documents/ Tuesday/Session%20III/Cooper.pdf (accessed January 6th, 2016)

[33] Bhattachatyya B., 2014, presentation at ”Transient Key science project meeting 2014”, Manchester, UK, 9-10 September 2014. http://www.jb.man.ac.uk/meetings/transients2014/pdfs/Bhaswati.pdf (accessed January 6th, 2016) Bhattachatyya B. et al., 2015, astro-ph/1509.07177

[34] Smits R., Kramer M., Stappers B., Lorimer D. R., Cordes J., Faulkner A., 2009a, A&A, 493, 3, p.1161

[35] Burgay M. et al.,2006, MNRAS, 368, 283.

[36] Faulkner A. J. et al., 2004, MNRAS, 355, 147.

[37] Keith M. J, Eatough R. P., Lyne A. G., Kramer M., Possenti A., Camilo F., Manchester R. N., 2009, MNRAS, 395, 837.

[38] Lee KJ, Stovall K, Jenet FA, Martinez J, Dartez LP, Mata A, et al. PEACE: pulsar evaluation algorithm for candidate extractionea software package for post-analysis processing of pulsar survey candidates. MonNot R Astron Soc 2013;433(1):688e94.

[39] Bethapudi S., Desai S., Separation of pulsar signals from noise with supervised machine learning algorithms, Astronomy and Computing, Volume 23, April 2018, Pages 15-26.

[40] Eatough RP, Molkenthin N, Kramer M, Noutsos A, Keith MJ,Stappers BW, et al. Selection of radio pulsar candidates using artificialneural networks. Mon Not R Astron Soc 2010;407(4):2443e50.

[41] Bates SD, Bailes M, Barsdell BR, Bhat NDR, Burgay M, Burke-Spolaor S, et al. The high time resolution universe pulsar surveyeVI. Anartificial neural network and timing of 75 pulsars. Mon Not R Astron Soc2012; 427(2):1052e65.

[42] Morello V, Barr ED, Bailes M, Flynn CM, Keane EF, van Straten W.SPINN: a straightforward machine learning solution to the pulsarcandidate selection problem. Mon Not R Astron Soc 2014;443(2):1651e62.

[43] Mohamed, Taha. (2017). Pulsar Selection Using Fuzzy knn Classifier. Future Computing and Informatics Journal. 3. 10.1016/j.fcij.2017.11.001.

[44] Zhu, W.W., Berndsen, A., Madsen, E.C., Tan, M., Stairs, I.H., Brazier, A.,Lazarus, P., Lynch, R., Scholz, P., Stovall, K., Ransom, S.M., Banaszak,S., Biwer, C.M., Cohen, S., Dartez, L.P., Flanigan, J., Lunsford, G., Martinez,J.G., Mata, A., Rohr, M., Walker, A., Allen, B., Bhat, N.D.R., Bogdanov,S., Camilo, F., Chatterjee, S., Cordes, J.M., Crawford, F., Deneva,J.S., Desvignes, G., Ferdman, R.D., Freire, P.C.C., Hessels, J.W.T., Jenet,F.A., Kaplan, D.L., Kaspi, V.M., Knispel, B., Lee, K.J., van Leeuwen, J.,Lyne, A.G., McLaughlin, M.A., Siemens, X., Spitler, L.G., Venkataraman,A., 2014. Searching for Pulsars Using Image Pattern Recognition.ApJ 781,117. 1309.0776.

[45] Wagstaff, K.L., Tang, B., Thompson, D.R., Khudikyan, S., Wyngaard, J.,Deller, A.T., Palaniswamy, D., Tingay, S.J., Wayth, R.B., 2016. A MachineLearning Classifier for Fast Radio Burst Detection at the VLBA.PASP 128,084503. 1606.08605.

[46] Devine, T., Goseva-Popstojanova, K., & McLaughlin, M., 2016, MNRAS, 459, 1519.

[47] John M. Ford. 2017. Pulsar Search Using Supervised Machine Learning. Doctoral dissertation. Nova Southeastern University. Retrieved from NSUWorks, College of Engineering and Computing. (1001) https://nsuworks.nova.edu/gscis_etd/1001.

[48] Agarwal D, Aggarwal K, Burke-Spolaor S, Lorimer D R,Garver-Daniels N, Towards deeper neural networks for Fast Radio Burst detection, MNRAS 000, 1-14 (2019).

[49] Keane, E.F. et al. PoS AASKA14 (2015) 040 arXiv:1501.00056 [astro-ph.IM].

[50] Lyon, R., Brooke, J., Knowles, J., Stappers, B., 2013. A study on classification in imbalanced and partially-labelled data streams, in: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, IEEE. pp. 1506–1511.

[51] Lyon, R.J., Brooke, J.M., Knowles, J.D., Stappers, B.W., 2014. Hellinger Distance Trees for Imbalanced Streams. ArXiv e-prints 1405.2278.

AIIoT-2019

International Virtual Conference on Artificial Intelligence for IoT 2019

Reviewers Response Sheet

Paper ID: 41

Title:AN IOT BASED APPROACH TO REDUCE TRAFFIC ON ROADS IN METROPOLITAN CITIES

Name of the Author(s): a)chethana.s.patil (corresponding author) *

b)Dr.sujatha Terdal

Is the title of the paper is modified as per reviewers suggestion: YES/No/Not Applicable

Reviewer No# 1

Query No.	Response	Paragraph No.	Line No.	Page No.
1	Reviewer suggestion:*Literature review part is inadequate in the paper. More latest citations on the selected topic shall be included in the paper. Response: latest citation papers are added(old papers citations are removed) Done with new 2019,2018,2015,2016 Grammar & sentence correction (done with new survey)	Section 3	Section 3	3
2	Reviewer suggestion:*Motivation to select specific work and type shall be elaborated bit more in introduction section. Response: as per suggestion in introduction part is updated with specific work (section 1. Starting with line elaborated Till end of Para(Grammar & sentence correction)	Section 1 Paragraph 2	Section2	1-2
3	Reviewer suggestion:* Language: English representation and Grammar can be carefully copy-edited before preparation of final paper. Response: grammar mistakes and sentence	Section 1-5		1-8

* Corresponding author.
E-mail address: chethana.patil@gmail.com

Query No.	Response	Paragraph No.	Line No.	Page No.
	engish correct representation with proper sentence and word. edited .section 4(line.no-5 –line.no 6, Line.no-28-line no.34*start mqtt to end thing speak line) Section 5 (5.1 in snapshot 4 line.no-4 to line.no.5 * this collection-thing speak) ,5.2.4(line.no4-line 5),5.2.5(line.no-3 to 5for the purpose-accomplished) Abstract:(second line-due to this –is replaced by because of this on road pollution and Increasing-growing, road-on-road, smart solution –intuitive way, vehicle-automobiles traffic-transportation bottleneck Help-assist) (as per new format line.number)			

An IoT Based Approach to Reduce Traffic on Roads in Metropolitan Cities

Chethana.S.Patil[a], Dr.Sujatha P Terdal[b*]

[a]*P.D.A college of Engineering ,Kalaburagi-585102,India*
[b]*P.D.A college of Engineering ,Kalaburagi-585102,India*

Abstract: The world around is suffering with many types of pollution such as road pollution which leads to noise pollution, air pollution and many more. Because of this on-road pollution, the amount of pollutants is growing in the atmosphere at a very high pace causing asthma and other serious diseases to humans and other living organisms. In order to solve the issue of on-road pollution, intelligent conveyance systems are a must to be implemented in the country in order to reduce the amount of contamination and also provide an easy access of living as traffic bottlenecks can be a nail in the coffin of a person. Intuitive way out are necessary to reduce the amount of pollutants such as PM 2.5, PM 10 and other harmful gases being released by the combustion of fuel in the engine of automobiles. We are trying to find the solution by improving the ITS by collecting various data through the medium of an SensorDataAPP Android application such as acceleration, latitude, longitude and many more which helps to find user traffic movements on-road and Further this data can be used over the algorithm to assist people find better paths to ignore the traffic and also prevents the transportation bottlenecks on the road.

Keywords: Internet of Things, Cloud Computing, MQTT channel, ThingSpeak Channel

1. Introduction

The world is engulfed with various types of pollution around that is working as an anti-human task to wipe the existence of living race from the face of earth. These are the results of our actions that are being performed for centuries. One of the pollution out of all is road pollution which is caused by high fuel consumption by vehicles on roads. One of the reasons of high fuel consumption is traffic jams caused by a very large number of on-road automobiles. Due to these large numbers of vehicles, the roads are jammed with a lot of traffic.

* Corresponding author. Tel.876287611;
E-mail address:chethana.patil@gmail.com.

As a consequence, a various difficulties can be faced by us such as we may be late at our work place or educational institutes, patients' condition can be deterioted waiting in the traffic, valuable productive time can be wasted ,unnecessary burning of fuels leading to wastage, emergency service automobiles may be prohibited, increasing the possibilities of injuries and death due to accidents, thinning of economic health and more .The solution of this problem exists in preparing and following a traffic arrangement for purpose of Minimizing on-road traffic. These systems are known as intelligent transportation system (ITS)[1].

The intelligent transportation system (ITS) have evolved significantly in recent years which in turn have influenced the demand of mobility as towns a well as cities are being expanded day by day. As consequence more number of automobiles has started occupying the road network. The problem is that the road infrastructure has not been enhanced in a direct variation of number of increasing automobiles on road. As a result we are facing tiring traffic jams along with related sound and huge air pollution [3].

For the objective of managing this huge traffic a static signalling system is implemented at the road crossings which may prohibit emergency service automobiles like ambulances, police jeeps or fire engines resulting to causalities. A common traffic control system utilizes static signalling timer at intersections and does not provide priority to emergency vehicles such as ambulances ,fire fighters and police cars ,possibly causing a loss of lives, loss of public or government properties, rising expense due to unnecessary fuel burning as well as related air pollution[4,5] .so the real time data analysis & management system is in need for efficient controlling huge volume of traffic .As the automobile on-road increases the air pollution diseases also increases[6], so there is necessity of implementing artificially intelligent traffic controlling systems is the demand of the situation day by day[2].

Internet of things or IoT will build the internet in the time a head .the term iot refers to collection of wired or wirelessly interconnected devices attached with sensors. The sensors attached with devices accumulate information from the environment and sent to a place where analysis of this huge amount of data and appropriate decisions can be made making these devices smart. There are lots of application domains of IoT including health services, economics analysis and environment monitoring[7].

2. Problem Statement

The City traffic in major cities and metropolis around the globe is becoming extremely severe problem for city authorities. Google has introduced Real Time city traffic monitoring system this along with most of the other vehicular sensors or virtual sensors based systems are aimed for providing real time visualization of the traffic conditions around the streets. However, traffic depends upon several parameters, for example:

- Direction of the vehicles
- Probable office and working places and working hours
- Width of the road and traffic density
- Peak hours
- Type of vehicles.

Without considering these parameters, it is impossible to get a realistic estimation of the traffic problems. The existing systems are more concerned with obtaining the traffic map rather than optimizing the flow of the traffic to reduce city traffic congestion. So Existing work do not seamlessly integrate positional data gathering with traffic flow estimation, traffic density prediction and traffic optimization. Hence we propose to bridge this gap to solve traffic problem of city.The present system only considers current vehicle density on the road their direction and speed whereas the traffic jam depends upon the various other aspects including but not limitation to.

- Historical data
- Scheduled future events that can offend the traffic
- Sudden events accidents.
- Other mutually exclusive but intertwined

Therefore the present system fails to predict the best route/timing that can lead to minimum traffic jams.

3. Literature survey

An economically , greatly extensible ,adaptable ,knowledgeable, and IoT aided transportation management framework has been proposed by Hasan Omar Aisakran 2015[8]the said research work has integrated the concepts of RFID technology, wireless sensor network, cloud platform and GPS navigation along with other and other leading techniques for the purpose of integrating intelligence to the

management of transportation. the advantages are minimization of traffic volume at a time ,efficient controlling of traffic and avoiding casualties.

Al-Sayed Ahmed Al-Sobky , Ragab M. Mousa 2015[9] Proposes a new technique spatial data analyasis technique which utilizes capabilities of smart phones and measures the original traffic thickness on the portion of the road .The data compilation analysis is done on the basis of spatial road segment length and it is measured macroscopically. The data analysis technique is flexible to survive with leading Smartphone operation and it is constrained by scenarios containing less number of weighty transports.

Pallavi A.Targe ,Dr. M. P. Satone [10] discusses a summary of existing approaches in the field of vehicular ad-hoc networks and presents a real-time intelligent traffic framework supported by vanet addressing the problems of previous approaches .the author proposes a technique which is supported by RFID as well as ARM controller aided for reducing the transportation bottleneck with the help of accurate controlling of signals as well as planning of the least path with respect to distance or time and implemented as Smartphone application.

The author Pampa Sadhukhan, Firoj Gazi 2018[11] Presents arrangement for controlling city overcrowding traffic it is a new system for reducing transportation bottleneck and provides efficient minimization and control of blocking automobiles on-road traffic with the help of accurate controlling of signals the traffic situations are accommodated in dynamic manner and avoids the problems predetermined static process signal and signal function works on time period based bulk, weight(density) of automobiles at crossings of road.

The author Asma Farheen, et,al 2019[12]proposes a model for metropolitan city transport administration purpose as populace is huge in growth day by day .The IR sensor, IP (webcam), GSM module and image processing based techniques is Used for the detecting jams of traffic on-roads ,The primary goal of the system is intuitive tracking ,controlling of city traffic, for communication purpose the “GSM module” is used were the information can be passed to police force whenever the operator of the automobile avoids signal and penality can be issued.

In metropolitan cities as there are huge number automobile users and the Traffic growth increases in every second continuously and The existing systems as many challenges so there is need of advanced traffic management systems with new technologies for controlling in efficient manner.

4. Methodology

One of the main goals is to reduce the traffic load in specific area of the city. The research proposal methodology uses the technologies like vehicular IOT, cloud sensor network and other advanced technologies for real time urban city traffic data collection analysis and optimization techniques for management. Integration of IOT with the cloud can get benefits such as unlimited capacity and resources for technical constraints. The concept of IoT can contribute to the cloud technology for furnishing innovative services in real world situations following a dynamic manner. So we assume that every user who is driving in city areas as their mobile with GPS (global positioning system) enabled and as they are driving the GPS co-ordinates of this device has been logged and send to cloud.

The cloud runs the machine learning services which tells us the following:

- The probable flow of traffic.
- Estimated traffic over next hour.
- Estimated direction of traffic.
- Estimated density of traffic.
- The Paths or roads that assume to have over traffic.

The machine learning technique is based on the road network gives the estimation of the probable flow of traffic that can be used to change the traffic routes in order to have balanced traffic over the entire city. In short the mobile devices of the users driving log their GPS co-ordinates. These GPS co-ordinates are logged in cloud network and this data is passed to machine learning technique located in cloud. The machine learning technique analysis and tells the possible flow of traffic that is the possible flow of traffic that could be used to redirect the traffic or monitor the traffic. The machine learning technique which is heart of it could be the probability graph theory that runs or Bayesian belief networks that learns from the existing traffic flows the road networks and predicts the possible behaviour of traffic. By simple terms the technique runs the cloud services that gathers GPS data from the mobiles of the users who are driving and analysis this data in service oriented architecture .The real time traffic data is collected by developing real time traffic android app and which is distributed among different users. The process would be done in three steps-

- The sensor data such as latitude, longitude, acceleration would be collected from the mobile phone with the help of an SensorData App android which would be installed on the mobile phone. These data is then used to predict the route for the vehicle.

- We have to make a channel on the cloud (we used myMQTT app[16] and Thing Speak channel for this purpose downloaded from play store &Open source IOT platform respectively.) and will use the channel to display all the data for every vehicle. MQTT (Message Queuing Telemetry Transport) is a Communication protocol transmitting information from device to device. It has less and thin data load that are shared between clients &servers following publish/subscribe actions[13,14].

- ThingSpeak refers to an IoT application and application programming interface which saves and fetches data from IoT devices. It also helps to analyze and view live data in cloud platform[15].

- The connection will then be established between the channel and the android app with the help of API in order to store all the information. These data would be further used for calculating the result.

5. EXPERIMENTAL SETTINGS

The real time traffic data is collected by developing real time traffic android app and which is distributed among different users

- To collect the real time traffic SensorDataApp android app is developed.

- The sensor data apk is installed

- Then we have to open SensorData app in mobile phone and get connected to it

- Then this app starts sensing the information about latitude, longitude(GPS data),accelerometer, gyroscope, Magnetic field, light, noise level through the help of different sensors in Smartphone[17]

- This data is collected are sent to cloud by my MQTT android Application

- It is a lightweight publish and subscribe system. This data is transferred to Thing Speak cloud platform for Analysis.

- The data of accelerometer, gyroscope, magnetometer, and GPS can be used to find the perfect vehicle navigation system in our Smartphone and for real time traffic analysis.

5.1 App demonstration

Snapshot 1: To collect the real time traffic SensorDataApp android app is developed and the corresponding apk file is installed on Smartphone. MyMQTT App from play store is installed[16].

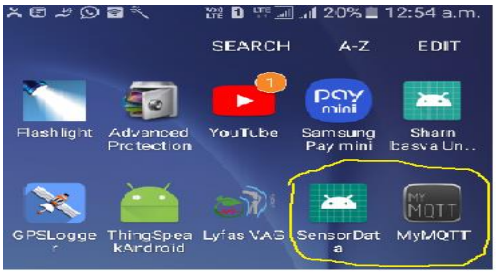


Figure 1: SensorData Android App and my MQTT protocol App at the desktop of the smart phone

Snapshot 2

The SensorData App is opened. We are getting connected to the app. The SensorData App make use of sensors for collecting information like latitude, longitude, gyroscope, acceleration, magnetic field, light etc from physical environment

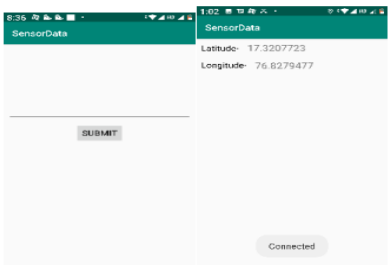
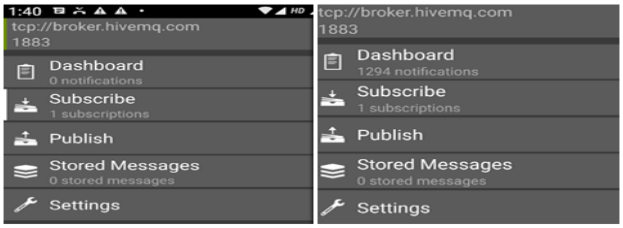


Figure 2: SensorDataApp interface

Snapshot 3

Real time data collected by the SensorData App is transferred by the aid of MQTT App to Thing speak IOT cloud Platform for Analysis. In the figure below the MQTT App dashboard interface has been illustrated. It demonstrates collected data notifications on MQTT App dashboard. At the initial state, when data collection has not been started, then there are no MQTT dashboard notifications. But as soon as SensorData App starts collecting data in real time, dashboard notification section becomes non-zero. These collected sensor data items like acceleration, latitude,

longitude, gyroscope, light, magnetic field, etc. will be pushed into the cloud storage for analysis.



a) **b)**

Figure 3: MQTT App interface a) Initial state of MQTT app, when no notifications have been received b) State when SensorDataApp collects sensor data showing dashboard notifications.

Snapshot 4

The MQTT App dashboard has been demonstrated below. The dashboard of MQTT App displays sensor information collected by SensorDataApp such as acceleration, latitude, longitude, gyroscope, accelerometer, light, magnetic field, noise level etc. This collection of data is transmitted to ThingSpeak data cloud platform to be analyzed in internet ThingSpeak channel.

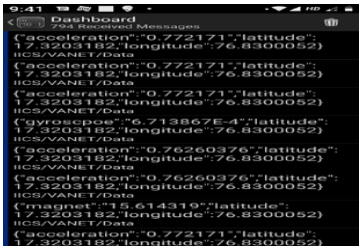


Figure 4: The MQTT App dashboard

5.2 *Results and Discussion*

The following graphs demonstrate monitoring of different users' latitude, longitude, and accelerometer and gyroscopes data collected from their smart phone using SensorDataApp at different time intervals. These results have been obtained from ThingSpeak.

5.2.1 Latitude monitoring

The Graphical Record of latitude monitoring has been displayed in the graph below. Figure 5 shows the record of current latitude of user using a SensorData App over a period of time. It can be observed from the graph of latitude vs. time that whenever the user latitude changes in traffic, those changes are tracked properly and updating of users' movements is performed after every 15 seconds. As the SensorData App is distributed between different users, real time traffic data is obtained in form of latitude which indicates location of the user.

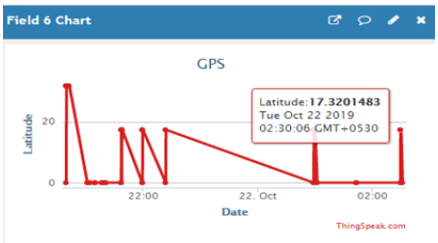


Figure 5: Location monitoring of the user by latitude, latitude vs. time

5.2.2 Longitude monitoring

The graphical record of longitude monitoring has been illustrated in the figure below. The figure 6 below shows that the records of current longitude of user over a period of time. It is clear from the graph whenever the user longitude changes in traffic, those changes are reflected by the App and movement data is updated after an interval of 15 seconds. Here, the longitude field is used to denote the location of a user.

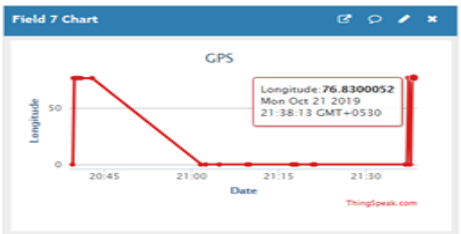


Figure 6: Location monitoring of the user by longitude, longitude vs. time

5.2.3 Acceleration monitoring

The graphical record of acceleration monitoring is given in the figure below. The figure7 below shows the record of current accelerometer reading of user over a period of time. The graph plots accelerometer vs. time. Whenever the user acceleration changes in traffic those changes are updated after an interval of 15 seconds. In the

graph below different users acceleration is plotted and the data from accelerometer is used to follow the movement of the user.

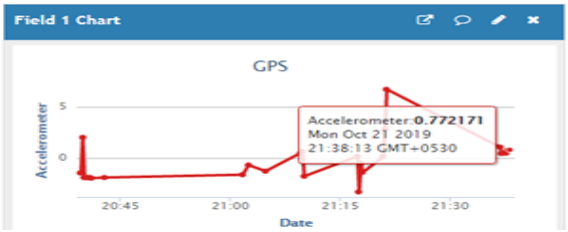


Figure 7: Location monitoring of the user by accelerometer data, acceleration vs. time

5.2.4 Gyroscope monitoring

The Figure 8 below shows the records of current gyroscope field of user over a period of time. The graph plotted is gyroscope vs. time where the orientation and angular velocity changes of the user are tracked. Whenever the user gyroscope changes in traffic those changes are updated after an interval of 15 seconds. For the purpose of accumulating transport data in real time manner data the *gyroscope field data* is used to monitor the velocity & orientation (direction) in all the angles of the user which is used to follow the user location.

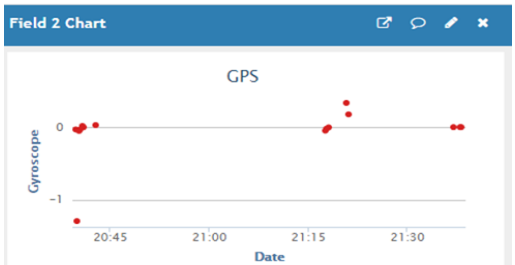


Figure 8: Location monitoring of the user by gyroscope data

5.2.5 Magnetic field monitoring

The figure below shows the record of current magnetic field of over a period of time. The graph of magnetic field vs. time indicates the change in the magnetic field. Magnetic changes in traffic are updated after an interval of 15 seconds. In graph we can see different users' magnetic field data are plotted .For the purpose of accumulating transport data in real time manner the Magnetic field data is used to monitor the angular speed of user[18,19]. The functions like tracking the magnetic field, detecting and identifying passing automobiles are accomplished.

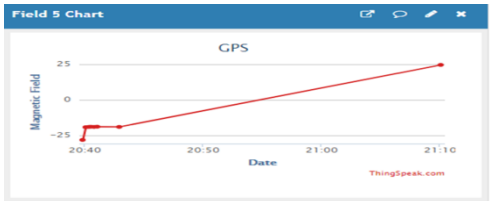


Figure 9: Magnetic field monitoring

6. CONCLUSION

This research focuses towards bettering each of the current state of traffic control system with support of robust, high performance, scalability, virtualization and, elasticity. The technique must address current issues challenges in real time urban traffic data collection analysis for city traffic optimization management using vehicular IOT and cloud sensor network.

The current research work has developed an Android Application called SensorDataApp. Various sensor data such as latitude, longitude, accelerometer data and gyroscope data are collected from users' smart phone. This collection of data is sent with the help of by MQTT protocol to the ThingSpeak cloud platform for analysis. It can be observed from the results that user and traffic movements can be easily tracked by analyzing sensor data collected from a smart phone.

In future this work can be optimized by obtaining possible prediction of probable traffic over next hour, density and direction of traffic flow and must perform the optimal traffic management and should also minimize travel time and stoppage. To obtain possible routes for the vehicles in order to reduce the traffic load in specific area over an entire city can be another future improvement.

Acknowledgement :

I woul like to thank Dr.sujatha.Terdal for guiding me and I thank one and all who assisted me in paper.

References:

[1] Nan Ding, Guozhn Tan, Wei Zhang, And Hongwei Ge "Distributed Algorithm For Traffic Data Collection And Data Quality Analysis Based On Wireless Sensor Networks" , International Journal Of Distributed Sensor Networks Volume 2011

[2] Sanjay Kumar Singh "Urban Transport In India: Issues, Challenges, And The Way Forward" European Transport Issue 52, Paper 5, ISSN 1825-3997 2012

[3] Kai Zhanga and Stuart Battermanb ". Air pollution and health risks due to vehicle traffic" Sci Total Environ. 2013 April 15; 0: 307–316. doi:10.1016/j.scitotenv.2013.01.074.

[4] Nellore, Kapileswar; Hancke, Gerhard P "Traffic management for emergency vehicle priority based on visual sensing" Sensors (Switzerland) MDPI, 16(11), [1892]. Nov 2016.

[5] Kapileswar Nellore and Gerhard P. Hancke "ASurvey on Urban Traffic Management System Using Wireless Sensor Networks" ,Sensors (Basel). Jan,2016

[6] https://www.aaaai.org/conditions-and treatments/library/asthma-library/air-pollution-asthma

[7] Jayavardhana Gubbi, Rajkumar Buyya,Slaven Marusic, Marimuthu Palaniswami, "Internet Of Things (IoT): A Vision, Architectural Elements, And Future Directions", Elsevier, Volume 29, Issue 7, September 2013.

[8] Hasan Omar Al-Sakran "Intelligent Traffic Information System Based On Integration Of Internet Of Things And Agent Technology", International Journal Of Advanced Computer Science And Applications (IJACSA), Vol. 6, No. 2, 2015

[9] Al-Sayed Ahmed Al-Sobky , Ragab M. Mousa " Traffic density determination and its applications using smartphone". Alexandria Engineering Journal, 55(1), 513–523. doi:10.1016/j.aej.2015.12.010

[10] Pallavi A. Targe, Prof. Dr. M. P. Satone. Real–Time Intelligent Transportation System Based On VANET International Research Journal Of Engineering And Technology (IRJET) E-ISSN: 2395 -0056, Volume: 03 Issue: 01 | Jan-2016

[11] Pampa Sadhukhan, Firoj Gazi "An IoT based Intelligent Traffic Congestion Control System for Road Crossings" *IC3IoT*, 978-1-5386-2459-3/18, 2018 IEEE

[12] Asma Farheen, S. Ashwini, K. G. Priyanka, P. R. Divya,and R. Bhagya "IoT Based Traffic Management in Smart Cities" international Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI), Springer Nature Switzerland AG 2019, 10.1007/978-3-030-03146-6_121.

[13] https://1sheeld.com/mqtt-protocol/

[14] https://randomnerdtutorials.com › what-is-mqtt-and-how-it-works

[15] https://thingspeak.com

[16] https://play.google.com/store/apps/details?id=at.tripwire.mqtt.client&hl=en

[17] https://developer.android.com/guide/topics/sensors/sensors_motion

[18] Xiao Chen , Xiaoying Kong , Min Xu, Kumbesan Sandrasegaran1 , Jiangbin Zheng "Road Vehicle Detection and classification Using Magnetic Field Measurement"IEEE Access, date of current version May 1, 2019.

[19] Vytautas Markevicius, Dangirutis Navikas, Mindaugas, Zilys, Darius Andriukaitis, Algimantas Valinevicius and Mindaugas Cepenas "Dynamic Vehicle Detection via the Use of Magnetic Field Sensors" MDPI,JOURNAL, SensorsPublished: 19 January 2016.

Advances in science, engineering and technology have paved way to solve many challenging problems targeting societal needs. Artificial Intelligence (AI) and Internet of things (IoT) are some of the areas which have a vast potential to provide tangible benefits in leaps and bounds. The role of the scientific community is to reap the maximum benefits of the existing technology and grow parallel with it. The Book, Emerging Trends in Artificial Intelligence for Internet of Things presents the chapters of researchers and academicians who contribute their work in the areas of Artificial Intelligent and IoT.

Book chapters of this book are presented in the Virtual Conference of Artificial Intelligence for Internet of Things which are research outcomes and technological advances at the crossroads of artificial intelligence, industrial automation and network security, all of them tailored for IoT. This book would be a resource for engineers and scientists to refer theoretical, experimental, and simulation results targeting diverse applications of AI for IoT.

Editors have organized the publication of the original work, simulation, hardware implementation and review material in this book, pertaining to AI and IoT. The editors would like to thank each and every one for their time and effort towards the successful publication of the work.

Arun Manoharan
Hemprasad Patil
Sujatha Rajkumar
Usha Rani Seshasayee

Emerging Trends in Artificial Intelligence for Internet of Things

ISBN 9789389640083 Vol 1, Dec 2019

9 789389 640083