

**PROCEEDINGS OF THE
INTERNATIONAL CONFERENCE ON RECENT TRENDS IN
ENGINEERING AND TECHNOLOGY**

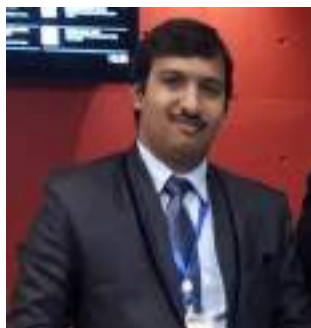
(ICRTET – 2K21)

11th – JUNE – 2021

**Organized By the Department of Information Technology,
QIS College of Engineering and Technology, Ongole-523272, Andhra
Pradesh, India**

Prof. Dr. S. Jafar Ali Ibrahim, Convener

Prof. Dr. R. Suneetha Rani, Co-Convener




Dr. S. Jafar Ali Ibrahim
B. Tech, M. Tech, Ph.D.
Convener
M +91 98426-88952



Dr. R. Suneetha Rani
B. Tech, M. Tech, Ph.D.
Co Convener
M +91 91825-98402

ROYAL BOOK PUBLISHING – INTERNATIONAL

Book Title	Proceedings of the International Conference on Recent Trends in Engineering and Technology (ICRTET – 2K21)
Book Size	6.69 x 9.84 inch
Paper	Natural Shade
Publisher	 Royal book publishing – International KM Nagar, Ayodhiyapatinam, Salem, Tamilnadu - 636103
Website	www.royalbookpublishing.com
Email id	contact@royalbookpublishing.com

ISBN Assigned by Raja Ram Mohun Roy National Agency for ISBN, New Delhi – 110066 (India)

ISBN: 9789391131289



Sri N. Nageswara Rao
B.E., MIE., President

President – SNES

President Message

As we face the end of the first decade of the new millennium, we must come together to share ideas and resources so we can plan and create a better future for the next generation of students, faculty, and researchers. In the tight, competitive global markets of the 21st century, the leading companies in various industries have embarked on massive reorganizations, mergers, partnerships, and cutting-edge collaborative projects with their like-minded peers—including their rivals—primarily to survive but ultimately to grow.

I hope this (ICRTET – 2K21) conference will reshape and create a better future for the next generation of students, faculty, and researchers in academia and industry.

My best wishes for this conference's grand success.



Dr. N.S.Kalyan Chakravarthy
M.Tech, Ph.D.

Secretary & Correspondent
QIS Educational Institutions

Secretary & Correspondent Message

The theme of the conference is “Recent Trends in Engineering and Technology”. This theme has been chosen after careful deliberation, to reflect the fact that there have been fundamental changes to both the character and the dynamics of the global industrial need since the turn of the century, and that this may, in turn, affect the way the industry addresses the challenges that lie before it.

Today, technologies like AI, IoT, big data, 5G, autonomous robots, and blockchain are stand-alone solutions. It is not a small task to ensure a variety of IoT sensors can speak with a manufacturing execution system, which is, in turn, able to talk with a cloud-based data analytics package. That leaves producers with two choices: They can either find a vendor who packages all these capabilities together, though this may lock them into a single and often expensive proprietary system. Or, if they want to mix and match best-of-class applications, they must pay programmers to integrate devices and software, so data formats are compatible up and down the system.

I am very proud to invite global researchers to join hands with our QIS Educational Institutions through this conference to carry out research and make new discoveries that are beneficial to mankind in the years to come.

I am delighted to declare open this conference on Recent Trends in Engineering and Technology, and I hope this conference will share more knowledge with all. Also, I greet this conference will be a successful one.



Dr. C.V.Subba Rao
M.Tech, Ph.D
Principal

Principal Message

The sector of Mechanical Engineering is the primary consumer of Artificial Intelligence as a technology. It is more than any other industry; it is consumed the most in Mechanical designs or engineering works. Sections of Mechanical Engineering like Robotics, Automation, or sensor technology, use Artificial Intelligence as a technology. So, it is easy to say that Mechanical Engineering disseminates the application and use of AI in the eco-system.

Artificial Intelligence isn't a long-listed dream anymore. More and more industries are taking advantage of it and providing some fantastic results that can help the human ecosystem. Yes, AI might affect an industry's HR by eradicating the least essential employee. Yes, it might kill the human element attached to the process. Yet, against all of it, it is helping our generation to imagine an entirely new world by impacting the different fields.

I hope this conference will be a good platform for all of you for advancing your knowledge.

Key Note Speaker – 1



Prof. Dr, Reynaldo G. Alvez,

**Dean – College of Information and Communication Technology,
Taguig City University, Philippines**



Prof. Ramesh Chandra Panda

**Dean - Research & Development
Synergy Institute of Engineering & Technology
Dhenkanal, Odisha, India**

Table of Content

Sl. No	Title	Page number
1	Analyzing transactional data of credit cards in the frequency domain Dr. R. Suneetha Rani, K. Bhuvana Chandana, D. Amrutha, D. Naga Teja, M. Sree Yaraswini, B. Supraja	1-11
2	Sentiment Classification Using Machine Learning Techniques N. Suresh, N. Naga Mounika, T. Navya Haritha , N. Sahithi, G. Vineela, V. Rakesh	12-21
3	Smart Assistance For Vehicle Breakdown K. Sreenath, O. Vinay Kumar, G. Sudhakar, P. Srikanth, B. Ajarath babu	22-33
4	Securing A File Using Cloud Computing Adopting Framework Dr. G. L. VaraPrasad, CH. Mahesh, P. Nikhil, R. Vincent Babu, S. SivaSai, G. Swami reddy	34-49
5	Relationship Identification & Predication Of Diseases Association Using Micro-Rna Of Genomic Data N. SaiKiran1, B. DeepthiPriya, V. Srilekha, P. Lakshmi Pavani, P. Manasa	50-64
6	Message Passing Over Cloud Using Cascade Ciphering With Randomized Algorithm G.L.V. Prasad, CH. V. Sowmya, N. Divya, K. Susmitha, T. Manasa	65-79
7	Online Voting Process For Government With Securitys K. Srinath, CH. Akhil, N. Satish, G. Kalyani, S. Bhanu Prasanthi	80-90
8	Integrating Secure Network Coding Techniques to store the data securely in Cloud with Data Dynamics Dr. R. Suneetha Rani, D. Goutham, D. Gopi babu, K.	91-104

	Anil kumar, R.BV. Saikumar, V.Aravindh	
9	CDA Generation and Integration for Health Information Exchange Based on Cloud Computing System- under cyber security. Dr. G.Lakshmi Vara Prasad, K.Ganesh ,V.Amarnatht, N.Manideep, K.Sandeep	105-114
10	Utilizing Hidden Search Patterns and Access Patterns in Searchable Encryption Scheme R.Suneetha Rani, P.Haritha, J.Mounika, M.Haritha, M.JayaSankar, K.Adithya	115-127
11	IOT based digital filtering system for COVID - 19 N.Suresh, B.Avinash, B.V.Ahalya, M.Niharika, M.Mounika, B.Sruthilaya	128-143
12	Automatic E-Timetable Generating System D. Ashok, CH. Sowmya, K. Susmitha, N. Divya, T. Manasa.	144-148
13	Android Malware Detection Using Machine Learning K. Sreenath, D. Akhila, G. Jahnvi, N. Preethi, T. Udaya Bhanu, V. Srikanth reddy	149-159
14	A Comparison Approach In Image Optimization Techniques Using MNIST Handwritten Digit Dataset N. SaiKiran, N. Naga Mounika, N.V. Sahithi, G. Vineela, T. Navya Haritha, V. Rakesh	160-174

ANALYZING TRANSACTIONAL DATA OF CREDIT CARDS IN THE FREQUENCY DOMAIN

**Dr. R. Suneetha Rani, K. Bhuvana Chandana, D. Amrutha, D. Naga Teja,
M. Sree Yasaswini, B. Supraja,**

Department of Information Technology
QIS College of Engineering and Technology
Ongole.

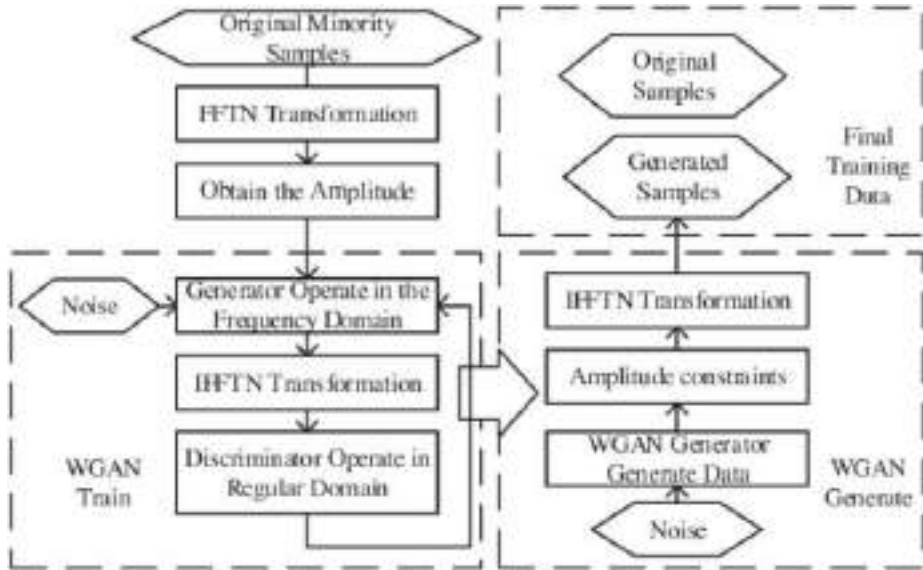
Abstract:

The massive increase in financial exchanges made in the e-commerce field has led to an equally massive increase in the dangers related to fraudulent activities. It is a problem directly correlated with the use of credit cards, considering that practically every one of the operators that offer merchandise or services in the e-commerce space permit their customers to use them for making payments. The primary disadvantage of these powerful methods of payment concerns the way that they can be used by the legitimate users (cardholders) yet additionally by fraudsters. Literature reports a considerable number of techniques designed to face this problem, in spite of the fact that their effectiveness is jeopardized by a series of basic problems, like the imbalanced conveyance and the heterogeneity of the involved information. The methodology presented in this paper takes advantage of a novel evaluation criterion based on the investigation, in the frequency space, of the spectral pattern of the information. Such strategy permits us to acquire a more stable model for representing data, with respect to the standard ones, reducing both the problems of imbalance and heterogeneity of information. Experiments show that the performance of the proposed approach is comparable to that of its state-of-the-craftsmanship competitor, albeit the model definition does not use any fraudulent previous case, embracing a proactive strategy able to differentiate the coolbeginning issue

1. Introduction:

The credit card frauds (i.e., when someone makes purchases without approval or counterfeits a credit card) represent one of the significant issues

that affect the e-commerce environment [1], especially in this age of exponential development of it. Authoritative studies performed by American Association of Fraud Examiners report that this sort of extortion involves the 10-15% of all extortion cases, for a complete financial value close to 75-80%. Just in the United States of America, this scenario leads towardan estimated average



Schematic Diagram of Flow 1

misfortune per misrepresentation case of 2 million of dollars. Such circumstance has given rise to a great interest by there search local area to develop increasingly effective techniques able-to detect thefraudulent exchanges. This is an errand that can be performed by exploiting several techniques yet there are some regular issues that the researchers should face. The generally significant between them are the imbalanced dissemination and the heterogeneity of the involved information.The main issue is given by the way that the fraudulent exchanges are typically less than the legitimate ones [3-6], designing an unbalanced conveyance of information that reduces the effectiveness of the machine learning strategies (Japkowicz and Stephen, 2002). The regular criterion used in practically all the stateof-the-craftsmanship approaches of extortion

detection is significantly based on the examination between the set of previous legitimate exchanges of a user and the new exchanges under evaluation. This is a rather paltry criterion that by and large, due to the high heterogeneity of information, leads toward misclassifications [2]. In order to overcome this problem, an extortion detection approach ought to be able to use however much as could be expected data about the exchanges during the evaluation process, however this isn't generally possible due to the powerlessness of some approaches to managesome data (e.g., Random Forests, one of the most performing approaches, can't manage types of information that involve a large number of categories).

2. Background Work

The fundamental errand of an extortion detection system is the evaluation of a new financial exchange with the point to characterize it as legitimate or fraudulent, by utilizing the data gathered before (i.e. value of the features that compose each exchange and on the off chance that it was an extortion or not) [9]. This section provides a general overview of the context taken into account in this paper, beginning with the presentation of the most used strategies and approaches, proceeding with the description of the exceptional problems, and finishing up with the presentation of core concepts on which the proposed approach is based, giving some details about the state-of-the-craftsmanship approach chosen to evaluate its performance [11].

2.1. Strategies and Approaches

Operative Strategies. The extortion detection approaches operate by following a supervised or unsupervised strategy (Phua et al., 2010)[13]. A supervised strategy works by exploiting the previous fraudulent and non-fraudulent exchanges gathered by the system, utilizing them to define a model able to group the new exchanges in a specific class (i.e., legitimate or fraudulent). It is evident how such strategy needs a series of examples concerning the two classes, and how its effectiveness is limited to the recognition of known patterns. An unsupervised strategy operates instead

by investigating the new exchanges to evaluate when they present anomalies in their values, compared to the ordinary range of values that characterizes the context taken into account. It is an inefficient strategy because a fraudster can operate to stay away from that the exchange presents anomalies in its values, and therefore the definition of effective unsupervised strategies represents a hard challenge. Operative Approaches. The most well-known approach to operate to detect fraudulent events in a financial information stream related to a credit card action is the appropriation of a static methodology (Pozzolo et al., 2014). It operates by separating the information stream into squares of equal size, preparing its model by utilizing just a limited number of beginning and touching squares. A different methodology is instead adopted by the refreshing methodology (Wang et al., 2003), where at each new square, the model is updated via preparing it by utilizing a certain number of latest and bordering blocks [16]. The forgetting approach (Gao et al., 2007) represents another possible operative way. By following this methodology, the user model is updated when a new square appears, by utilizing just the legitimate exchanges present in the last two squares, yet by utilizing every one of the previous fraudulent exchanges. The models generated by these approaches can be directly exploited to evaluate the future squares, or they can be used to define a bigger model of evaluation [14].

2.2. Open Problems

Information Scarcity Issue. The undertaking of the researchers working in this area is complicated by the shortage of public real- world datasets. This predominantly happens due to the restrictive policies adopted by those working in this field, which don't permit them to release data about their business activities for protection, competition, or legal issues. Not even a release in mysterious type of the information is typically taken into account by numerous financial operators, since likewise in unknown structure such information can provide precious data about their customers, and hence they could reveal potential vulnerabilities of the related e-commerce infrastructure.

Non-versatility Issue. This problem concerns the powerlessness of the misrepresentation detection models to correctly characterize the new exchanges, when their features give rise to different patterns (wrt the patterns used to define the evaluation model). Both the supervised and unsupervised extortion detection approaches are affected by this problem (Sorournejad et al., 2016) [15], which leads toward misclassifications, due to their powerlessness to detect new legitimate or fraudulent patterns.

Information Heterogeneity Issue. Pattern recognition represents a very significant part of the machine learning, since it very well may be used to solve a large number of real-world problems. The effectiveness of these processes is nevertheless jeopardized by the heterogeneity of the involved information. Such problem happens due to contradiction between comparative features resulting in the same information being represented differently in different datasets (Chatterjee and Segev, 1991) [8].

Information Unbalance Issue. Another significant issue that the approaches of extortion detection have to face is the unbalanced dissemination of information during the preparation of their evaluation models. This means that the data available to prepare an evaluation model are ordinarily composed by a large number of legitimate cases also, few fraudulent ones, an information design that reduces the effectiveness of the arrangement approaches (Japkowicz and Stephen, 2002; Earthy colored and Mues, 2012). A typical strategy adopted to face this problem is the fake balance of information (Vinciotti and Hand, 2003), made by performing an over-testing or under-inspecting process: in the first case the balance is obtained by copying some of the exchanges that are less in number (typically, the fraudulent ones), while in the second case it is obtained by removing some of the exchanges that are in greater number (normally, the legitimate ones).

Cold-start Issue. The cool beginning problem arises when the set of information used to prepare an evaluation model does not contain enough data about the area taken into account, making it impossible to define a

reliable model (Donmez et al., 2007). In other words, it happens when the preparation information are not representative of the multitude of involved classes of data (Attenberg and Provost, 2010) (i.e., in our case, legitimate what's more, fraudulent)[12].

3. Problem Definition

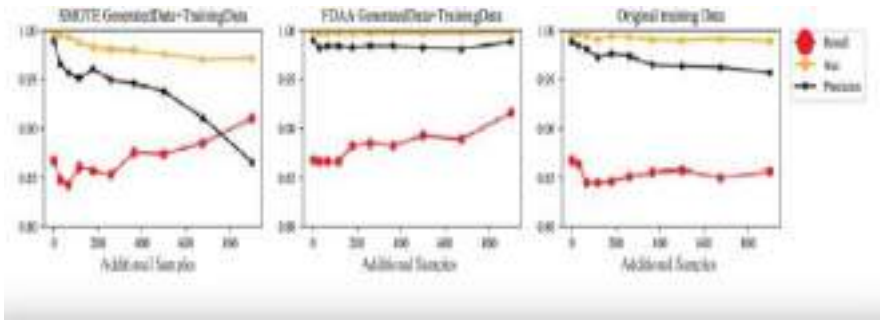
A definitive objective of our work is to create learning tests with genuine data qualities, add these produced data R to the AI classifier preparing sets S , get the last preparing set T to prepare the classifier, and afterward improve the acknowledgment impact of the classifier. Instructions to create these excellent preparing data R is the primary piece of our work. For the classifier C , with the first preparing set and classifier preparing set CT after extension dataset, the anticipated outcomes are communicated as $CS(V)$ and $CT(V)$. Data unit V methods one test of S , its genuine class is communicated as $R(V)$.

4. Proposed Approach

The execution of the proposed approach was performed through the accompanying three stages, which will be clarified later:

Data Definition: definition of the time arrangement in terms of grouping of qualities expected by the exchange highlights;

Data Processing: change of the time recurrence space (Smith et al., 1997). It implies that the portrayal in the recurrence area permits us to distinguish a particular example, notwithstanding the situation of the arrangement in the recurrence range by repeating to the DFT measure; In this progression, we move the time arrangement of the exchanges to the recurrence space by a DFT interaction performed through the FFT approach presented. In a starter study we thought about the exchange portrayals in the time area (time arrangement) to those in the recurrence area (recurrence range). Without diving deep into the benefits of the proper attributes of a Fourier change be that as it may, by restricting our examination at the setting taken into account, in our approach we need to misuse the following two properties:



Dataset 1

1. Stage Invariance: the main property shows that there are not varieties in the ghostly example in the event of significant worth translation to more officially, it is one of the stage properties of the Fourier change, i.e., a shift of a period arrangement in the timespace leaves the greatness unaltered in the qualities expected by the exchange includes that begin it;
2. Adequacy Correlation: the subsequent property rather demonstrates the presence of an immediate connection between the qualities accepted by the highlights in the time area and the relating extents expected by the ghostly segments in the recurrence space. All the more officially, it is the homogeneity property of the Fourier change (Smith et al., 1997), i.e., when the adequacy is changed in one space, it is changed by a similar element in the other domain

Data Evaluation: formalization of the calculation ready to arrange another exchange as genuine or fake based on a range examination measure.

5. Experiments

This segment reports data about the trial climate, the utilized datasets and measurements, the embraced methodology, and the consequences of the performed experiments.

Environment

The proposed approach was created in Java, where we utilize the JTransforms library to work the Fourier changes. The best in class approach (i.e., Irregular Forests) and the measurements to assess it have been executed in, by utilizing randomForest, DMwR, and ROCR bundles.

The RF boundaries have been tuned via looking through those that augment the execution. For reasons of reproducibility of the RF experiments, the R work set.seed() has been utilized in the code to fix the seed of the irregular number generator.

DataSet

This present reality dataset utilized for the evaluation of the proposed approach is identified with a progression of Visa exchanges made by European cardholders . In additional detail, this dataset contains the exchanges conveyed out in two days of September 2013, for an aggregate of 492 fakes out of 284,807 exchanges. It ought to be seen how it addresses an exceptionally unequal dataset (Pozzolo et al., 2015), taking into account that the false cases are just the 0.0017% of the multitude of exchanges.

Metrics

Cosine Similarity. The cosine similitude (Cosim) between two non-zero vectors v_1 and v_2 is determined in terms of cosine point between them, as demonstrated in the Condition (6). It permits us to assess the closeness between two ghostly examples by looking at the vectors given by the greatness of their recurrence parts.

6. Conclusions and Future Work:

Mastercard misrepresentation discovery frame works assume a vital part in our online business age, where an expanding number of exchanges happens through this amazing instrument of installment, with every one of the dangers that it implies. More than needing to supplant the current state-of-the-workmanship arrangements, the approach introduced in this paper needs to present a novel recurrence area based model that permits an extortion discovery framework to work proactively. The outcomes acquired are fascinating, since it is important to think about that the best in class contender considered (i.e., Random Forests), as well as utilizing the two classes of exchanges to train its model additionally preprocesses the dataset by utilizing an adjusting procedure (i.e., SMOTE). It ought to be noticed that the Mastercard setting considered is just one of the potential situations,

since the proposed approach can be utilized in any setting described by monetary electronic exchanges.

References

- [1] Y. Xia, C. Liu, Y. Li, and N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," *Expert Systems with Applications*, vol. 78, pp. 225–241, Jul. 2017.
- [2] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, "Credit Card Fraud Detection Using AdaBoost and Majority Voting," *IEEE Access*, vol. 6, pp. 14 277–14 284, 2018.
- [3] I. Dutta, S. Dutta, and B. Raahemi, "Detecting financial restatements using data mining techniques," *Expert Systems with Applications*, vol. 90, pp. 374–393, Dec. 2017.
- [4] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, Jan. 2015.
- [5] J. West and M. Bhattacharya, "Some Experimental Issues in Financial Fraud Mining," *Procedia Computer Science*, vol. 80, pp. 1734–1744, 2016.
- [6] A. M. Rather, V. N. Sastry, and A. Agarwal, "Stock market prediction and Portfolio selection models: A survey," *OPSEARCH*, vol. 54, no. 3, pp. 558–579, Sep. 2017.
- [7] M. D. Godfrey, "An Exploratory Study of the Bi-Spectrum of Economic Time Series," *Applied Statistics*, vol. 14, no. 1, p. 48, 1965.
- [8] R. Bradley, "Adaptive data cleaning," US Patent US20 060 238 919A1, Oct., 2006.
- [9] R. Saia and S. Carta, "Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach," in *Proceedings of the 14th International Joint Conference on E-Business and Telecommunications*. Madrid, Spain: SCITEPRESS - Science and Technology Publications, 2017, pp. 335–342.
- [10] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," *IEEE*

Signal Processing Magazine, vol. 35, no. 1, pp. 53–65, Jan.2018.

[11] I. Brown and C. Mues, “An experimental comparison of classification algorithms for imbalanced credit scoring data sets,” *Expert Systems with Applications*, vol.39, no. 3, pp. 3446–3453, Feb. 2012.

[12] C.-L. Huang, M.-C. Chen, and C.-J.Wang, “Credit scoring with a data mining approach based on support vector machines,” *Expert Systems with Applications*, vol. 33, no. 4, pp. 847–856, Nov. 2007.

[13] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*. New York: Assoc Computing Machinery, 2016, wOS:000485529800092.




[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002.

[15] D. M. McNeish and L. M. Stapleton, “The Effect of Small Sample Size on Two- Level Model Estimates: A Review and Illustration,” *Educational Psychology Review*, vol. 28, no. 2, pp. 295–314, Jun.2016.

[16] H. He, Y. Bai, E. A. Garcia, and S. Li, “ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning,” in *2008Ieee International Joint Conference on Neural Networks, Vols 1-8*. New York: Ieee,2008, pp. 1322–1328,wOS:000263827200212.

Author details:

	<p>Dr.R.Suneetha Rani (tanneeru.suneetha@qiscet.edu.in) Associate professor & HOD Department of Information Technology QIS college of Engineering andTechnology</p>
	<p>K.Bhuvana Chandana (Korrapatibhuvana2000@gmail.com) Student Department of Information Technology QIS college of Engineering andTechnology</p>

	<p>D. Amrutha (amurthadondapati3@gmail.com) Student Department of Information Technology QIS college of Engineering and Technology</p>
	<p>D. Naga Teja (Teja.darsi.19@gmail.com) Student Department of Information Technology QIS college of Engineering and Technology</p>
	<p>M. Sree Ysaswini (yasaswinimarella8999@gmail.com) Student Department of Information Technology QIS college of Engineering and Technology</p>
	<p>B. Supraja (suprajachowdary347@gmail.com) Student Department of Information Technology QIS college of Engineering and Technology</p>

SENTIMENT CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

^[1]N.Suresh, ^[2] N.Naga Mounika, ^[3] T.Navya Haritha , ^[4] N.Sahithi , ^[5] G.Vineela , ^[6] V.Rakesh

^[1] Associate Professor, Department of Information Technology, QISCET, Ongole.

^{[2], [3], [4], [5], [6]} Students, Department of Information Technology, QISCET, Ongole.

Abstract: Large amounts of information are available online on the web. The discussion forum, review sites, blogs are some of the opinion rich resources where review or posted articles is their sentiment, or overall opinion towards the subject matter. The opinions obtained from those can be classified into positive or negative which can be used by customers to make product choices and by businessmen for finding customer satisfaction. This paper studies online movie reviews using sentiment analysis approaches. In this study, sentiment classification techniques were applied to movie reviews. Specifically, we compared two supervised machines learning approaches SVM, Naive Bayes for Sentiment Classification of Reviews. Results state that Naïve Bayes approach outperformed the sum. If the training dataset had a large number of reviews, Naive bayes approach reached high accuracies as compared to others.

Key words: Sentimental Analysis, supervised Algorithm, Naive bayes, Support vector machine

1. INTRODUCTION

Opinions are important to almost all human activities because they are key influencers of our behaviours. At whatever point we have to settle on a choice, we need to know other's opinion. In this present reality, organizations and associations dependably need to discover buyer or general feelings about their Items and administrations. Singular buyers additionally need to know the sentiments of existing clients of an item

before buying it, and others' feelings about political competitors before settling on a voting choice in a political decision. Before, when an individual required assessment, he asked loved ones. At the point when an association or a business required shopper opinion, it led studies, assessment surveys, and centre gatherings. Securing open and buyer suppositions have for some time been an immense business itself for promoting, advertising, and political crusade organizations. With the hazardous development of online networking (e.g., audits, gathering dialogs, sites, smaller scale websites, Twitter, remarks, and postings in interpersonal organization destinations) on the Web, people and associations are progressively utilizing the substance as a part of these media for choice making. These days, if one needs to purchase a customer item, one is no more restricted to approaching one's loved ones for conclusions in light of the fact that there are numerous client audits and examinations in broad daylight gatherings on the Web about the item. For an association, it might never again be important to direct studies, conclusion surveys, and centre gatherings with a specific end goal to accumulate popular assessments on the grounds that there is a wealth of such data openly accessible. Our goal is to calculate the polarity of sentences that we extract from the text of reviews. We will find the sentiment of this review and find whether the movie is successful or not. So that we can find whether movie is positive or negative. We examine the effectiveness of applying machine learning techniques to the sentiment classification problem. Our analysis helps concerned organizations to find opinions of people about movies from their reviews, if it is positive or negative. One can in turn formulate a public opinion about a movie. The challenging aspect in sentiment analysis is an opinion word which is considered as a positive in one situation may be considered as negative in another situation. The traditional text processing considers that a little change in two bits of content has no change in the significance or meaning [1]. But in sentiment analysis a little change in two bits of content has change in the significance or meaning, consider Example "story is good" is different from "the story is not good". The system processes it by analyzing one by one sentence at a time [3]. However, blogs and twitter contain more informal sentences which user can

understand and but system cannot understand it. Consider example, “that movie story was as good as its previous movie” is dependent on previous movie whose details is not available. Another challenging aspect of this problem that seems to distinguish it from traditional topic- based classification is that while topics are often identified by keywords alone, sentiment can be expressed in a more subtle manner For example, the sentence “How could anyone watch this Drama?” contains no single word that is obviously negative. Thus topic-based classification can easily understandable then sentiment. So, apart from presenting our results obtained via machine learning techniques, we also understand the problem to gain a better understanding of how difficult it is. Consider another example visual effect of movie were good but storyline was terrible this convey both positive and negative meaning respectively. Thus, review can be helpful by providing useful information to customer as well as businessmen. For customer it provides useful information that which product is good by examining the rating that come with it. Opinions or sentiment, can also provide researchers, businessmen, and policy-maker with valuable information ranging from rates of customer satisfaction to public opinion trends.

I. Literacy Survey

A set of questions with a choice of answers, devised for healthcare surveys helps a lot to gain knowledge on effectiveness of the treatment. There are various surveys such as patient experience survey which captures every patient’s voice using email, phone and mail to provide a deeper data and a clearer picture of patient perception for the entire care.

Preliminary Investigation

Watching interviews of doctors, patients and drug developers helps in being updated with the latest technology and effectiveness of the treatments available.

Feedback

Now-a- days many pharmaceutical companies are asking for

consumer feedback. These pharmaceutical companies make record of both make record of both negative and positive feedback and use these records to come up with a better optimized drug technology and effectiveness of treatments available.

II. Existing System

There are many traditional methods which provide the benefit of having knowledge and updates on the latest technology in the medical field.

- Surveys and Questionaries
- Interviews
- Feedback

III. Proposed System

In the proposed system, searching the information based on category and keywords from the twitter database is performed. Searching keywords in twitter is one of the hardest tasks because of the diversity of the language and the slangs used on the internet.

The logics that have been used in the proposed system has the following major steps:

- Collecting tweets
- Pre-processing tweets
- Sentiment analysis
- Suggestion of other treatment

Levels of sentiment

Due to scarcity of opinion text available in digital form, very less research interest on computational linguistics in the last decade of twentieth century was witnessed. The escalation of social media text on internet attracts young researchers to define the level of granularities of text. The web text is classified into three levels viz. document level, sentence level and word level. In the fourth level granularity is defined by using deep convolution neural network. This fourth level is character level feature extraction approach used

for extracting features of each character window from given word.

IV. SYSTEM REQUIREMENTS

BFTREE algorithm used for sentiment prediction

HARDWARE REQUIREMENTS:

- Operating System : Windows10
- RAM:8 GB
- Processor: Intel(R) Core I5
- Hard Disk:50GB

SOFTWARE REQUIREMENTS:

- Platform : Python(3.9)
 - Libraries : Pandas NumPy, html, core, components,matplotlib_lib
- Plots: Scatter plot bar plot 5 5
- Languages: Machine Learning

V. Methodology

Naive Bayes used for sentiment classification:

The dichotomy of sentiment is generally decided by the mindset of an author of text whether he is positively or negatively oriented towards his saying. Naïve Bayes. The classifier is a popular supervised classifier, furnishes a way to express positive, negative and neutral feelings in the web text. Naïve Bayes classifier utilizes conditional probability to classify words into their respective categories. The benefit of using Naïve Bayes on text classification is that it needs small dataset for training. The raw data from web undergoes preprocessing, removal of numeric, foreign words, html tags and special symbols yielding the set of words. The tagging of words with labels of positive, negative and neutral tags is manually performed by human experts. This preprocessing produces word-category pairs for training set. Consider a word y from test set (unlabeled word set) and a window of n -words (x_1, x_2, \dots, x_n)

from a document. The conditional probability of given data point 'y' to be in the category of n-words from training set.

J48 algorithm used for sentiment prediction

The hierarchical mechanism divides feature space into distinct regions followed by the categorization of sample into category labels. J48 is a decision tree based classifier used to generate rules for the prediction of target terms. It has an ability to deal with larger training datasets than other classifiers. The word features for sentences of corpus taken from labeled raff file of training set are represented in the leaf nodes of decision tree. In the test set every time when a near feature qualifies the label condition of internal feature node, its level is lifted up in the same branch of decision tree. The assignment of labels to the word features of test set gradually generates different two branches of decision tree. J48 algorithm uses entropy function for testing the classification of terms from the test set. Another classification approach outperforms J48, C4.5 and CART by expanding only b e s t node in the depth first order. BF Tree algorithm excavates the training file for locating bestsupporting matches of positive and negative terms in the test file. BF Tree algorithm keeps heuristic information gai

One R algorithm used for sentiment prediction

One R algorithm is a classification approach which restricts decision tree to level one thereby generating one rule. One rule makes prediction on word feature terms with minimal error rate due to repetitive assessment of word occurrences. The classification of most frequent terms of a particular sentence is made on the basis of class of featured terms from training set. The demonstration of One R algorithm for sentiment prediction with smallest error of classification is given below:

Step 1: Select a featured term from training set.

Step 2: Train a model using step3 and step4.

Step 3: For each prediction term. For each value of that predictor. Count frequency of each value of target term. Find most frequent class. Make a rule and assign that class to predictor.

Step 4: Calculate total error of rules of each predictor.

Step 5 Choose predictor with smallest error. For each value of that predictor. Count frequency of each value of target term. Find most frequent class. Make a rule and assign that class to predictor.

Step 4 Calculate total error of rules of each predictor.

Step 5: Choose predictor with smallest error.

Proposed methodology for optimization of sentiment prediction using weka "

The preprocessing of raw text from web is done in

$P(w)$: Probability that a document d contains term w .

$P(c)$: Probability that document d does not belongs to category c .

$P(w, c)$: Joint probability that document d contains word term w of category c .

$P(c/w)$: Conditional probability that a document d belongs to category c under the condition that d contains word term w .

Similarly other notations like $P(w')$, $P(w/c)$, $P(w/c')$, $P(c/w')$ and $P(c'/w)$ are taken and $\{c\}$ is the set of categories.

N_1 : Number of documents that exhibit category c and contain term w .

N_2 : Number of documents that do not belong to category c but

contains term w . N_3 : Number of documents that belong to category c and do not contain term w . N_4 : Number of

documents that neither belong to category c nor contain term w .

N : Total number of document reviews.

DF method qualifies only those documents in which a higher frequency terms are considered.

$$DF = \sum_{i=1}^m N1i$$

The MI method measures features of text by computing similarity of word terms w and

$$SimInfo(w, c) = \log \frac{P(w/c)}{P(w)}$$

$$MI = \log \frac{N1 \times N}{(N1 + N3)(N1 + N2)}$$

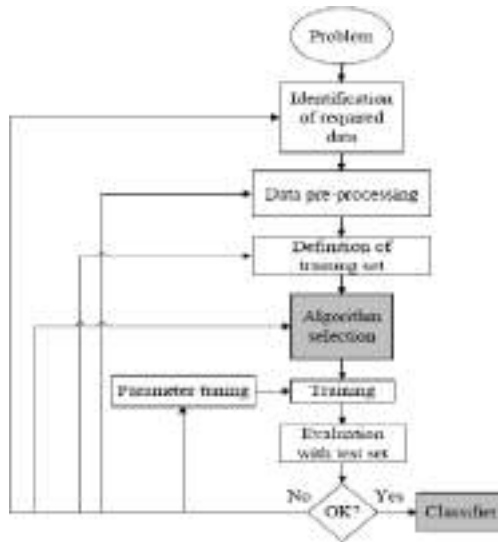
category c . The IG-construct measures similarity information for category by exploiting probabilities of absence or presence of terms in a document review.

$$IG(w) = - \sum P(c) \cdot \log P(c) + P(w) \sum \sum P(c/w) \cdot \log P(c/w) + P(w) \sum \sum P(c/w) \cdot \log P(c/w)$$

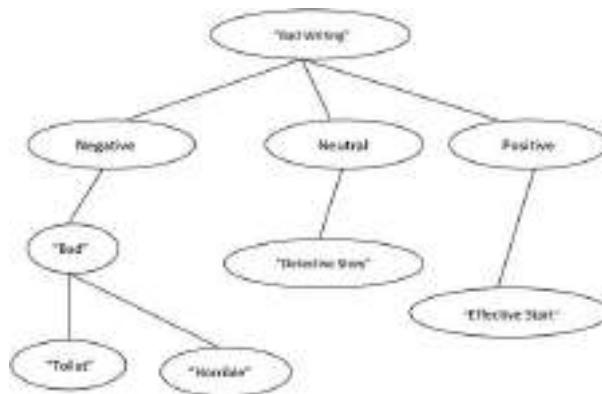
The normalization module converts all letters into lowercase, removal of punctuation marks and special symbols, conversion of numbers into words, expansion of abbreviation and limiting the average length of twenty words in a sentence. Each sentence is delimited by a newline character. The Python's NLTK and bs4 libraries are used for this purpose. Data splitter take the ratio of (80:20) of (Train: Test) subsets. We have used manual splitting of dataset at the time of retrieval of data from web. The four classifiers are trained with training subsets followed by performance evaluation. The evaluation metrics taken in the experiment are precision, recall, accuracy and F-measure.

MODULES

The process of supervised learning machine learning



Decision Tree



Conclusion

Sentiment analysis of data collected from social media is beneficial for mankind in providing them better healthcare. The workflow of analysing healthcare content in the social media helps to overcome the limitations of large-scale data analysis and manual analysis of user generated textual content in social media.

VI. References

- www.medicalnewstoday.com/articles/44967.php
- <http://www.idf.org/latest-diabetes-figures-paint>

grim-

globalpicture.

- Mohammad Ali Abbasi, Huan Liu and Reza Zafarani, "Social Media Mining - An introduction, Arizona state university", April 2014.
- [http://www.socialmediaexaminer.com/5-new-studies-showfacebook-a-marketing- powerhouse/](http://www.socialmediaexaminer.com/5-new-studies-show-facebook-a-marketing-powerhouse/)

SMART ASSISTANCE FOR VEHICLE BREAKDOWN

K. Sreenath¹, O. Vinay Kumar², G. Sudhakar³, P. Srikanth⁴, B. Ajarath babu⁵

¹Associate Professor, Department of Information Technology, QISCET, Ongole

^{2,3,4,5} IV BTech Students, Department of Information Technology, QISCET, Ongole

Abstract:

There are numerous aspects of mechanics that must be kept up to date using an internet technique and a software application. Activities, mechanical details, and a variety of other issues can be better controlled by programmes that are capable of doing so. The main goal is to give better service, make the procedure easier, and then immediately assign a mechanic. The mechanics database project application will assist the admin management authorities to save time and money. The mechanic details, users, and feedback system performance of users that may be supplied in certain problems that need to be improved may all be readily dealt with utilising the car breakdown database project application.

Keywords:- Android, users, Car ,Bike , vehicle, breakdown

I. INTRODUCTION

1.1 Project Background

The majority of individuals nowadays travel in their own vehicles. Most drivers confront difficulties on the road as a result of car breakdowns. Vehicle breakdown cause to waste the user valuable time. That is a worst experience they have to face. It also makes you fatigued during the ride. When a car breaks down on the road, the driver must look for a mechanic and check if there are any spare-part shops nearby. At that time if driver unable to search a good mechanic they have to ask for help someone, but that may be not good technological help for driver. However, if a motorist has an Android phone and uses this car

breakdown help, he or she may find a qualified repair in a matter of minutes. The most significant benefit is that the user may locate a mechanic based on their location. If user need vehicle spear-part shop, there are the facility search any shops. If user have technical problem related to vehicle they can ask it via chat.

1.2 Project Aim and Objectives

Aim

- To offer a framework that will enable technicians and transporters operate more efficiently.

Objectives

- To solve the mystery of a system malfunction.
- Build a web taxonomy that can detect a proper mechanic.
- To offer a framework that integrates professionals and riders

1.3 Description About Artifacts

After completing User Registration, the user logs in to the App. The user's current location is then tracked using GPS. The user's location is then sent to the database, where it is matched with the mechanic who signed up for the app.

The mechanic that is closest to the user's location is displayed. There is a feature that allows users to search for spare part businesses based on their location. An administrator can add a mechanic to the application. Admins have access to user and mechanic information.

I. Literature Survey

Following the analysis of the task's requirements, the next stage is to evaluate the problem and comprehend its context. The analysis of the old system is the initial activity in this phase, followed by an understanding of the new system's needs and domain. Both processes are equally vital, but the first one acts as a foundation for providing functional specifications and subsequently designing the proposed system successfully. Knowledge the qualities and needs of a new

system is more challenging and demands creative thinking. Understanding the characteristics and needs of an existing running system is similarly challenging, and an incorrect understanding of the current system might lead to a diversion from the solution.

2.1. STUDY OF THE SYSTEM

The interface has been built with the flexibility of the users in mind, with a graphical idea in mind, and linked through a web interface. The top-level graphical user interfaces have been classified as

1. Administrative user interface
2. The operational or generic user interface.

The administrative user interface focuses on consistent data that is practically part of organizational activities and that requires adequate authorization for data collecting. The interfaces assist administrators with all transactional states, such as data input, deletion, and updating, as well as comprehensive data search capabilities.

The operational or general user interface aids users in completing transactions on the system by providing access to current data and services. The operational user interface also assists ordinary people in handling their own data in a customized manner according to the aided flexibilities

MODULES

Following a thorough examination, the system was determined to have the following modules:

1. User
2. Admin
3. Worker

1. **User:** In customer module consists of various locations of a mechanics can get nearest services center address and contact any type of mechanic servicing.

2 **Admin:** In admin module admin has privileges to view all the data regarding the mechanics, information about workers works and test drive booking.

3 **Worker:** In worker module worker heshould give information about his work i.e. car details, name of mechanic center, send feedback to admin etc.

2.2.PROPOSED SYSTEM

To debug the existing system, remove procedures those cause data redundancy, make navigational sequence proper to offer information about users at various levels, as well as to represent current work status based on the organization. To create a robust password system.

NEED FOR COMPUTERIZATION

- We are all aware of the significance of computerization. The world is moving at breakneck pace, and everyone is pressed for time. One's goal is to obtain knowledge and complete a task in a short amount of time, with the highest level of efficiency and accuracy possible. The computerization application areas have been chosen based on the following criteria:

- Reducing the number of manual records held in various places.
- Data integrity will be improved.
- Facilitating the presentation of needed information in a timely manner by obtaining data from users.
- Facilitating a variety of statistical data that aids decision-making?
- To lessen the amount of physical labour required in operations that are repetitive in nature.

It will be easy to update and delete such a large volume of data.

FUNCTIONAL FEATURES OF THE MODEL

The functionality of the project is basic as far as it has been built, but the goal of the proposal is to improve and reinforce the working of Audit Status Monitoring. Coordinator Level, Management Level,

Auditor Level, User Level, and State Web Coordinator Level are the five streams that make up the overall scope. The suggested software will meet the information demands of each user group's request, such as accepting the request, delivering a vulnerability document report, and the audit's current progress.

ACCESS CONTROL FOR DATA WHICH REQUIRE USER AUTHENTICATION

The access control IDs specified in the following instructions are normally used to approve and authenticate the user (command codes are shown in parentheses).

USER NAME (USER)

The server requires user identity before granting access to its file system. This is usually the first command sent by the user when the control connections are established (some servers may require this).

PASSWORD (PASS)

This command must be followed by the username command, and it completes the user's identity for access control on some sites. Because password information is so sensitive, it's best to "mask" it or prevent it from being typed out.

II. METHODOLOGY

3.1. INTRODUCTION

Software design is the technical core of the software engineering process, and it applies to all development paradigms and application areas. For any engineered product or system, design is the initial stage in the development process. The purpose of the designer is to create a model or representation of an object that will be created later. Following the specification and analysis of system requirements, system design is the first of three technical processes necessary to construct and validate software: design, coding, and testing.

The significance may be summed up in a single word: "Quality." In software development, design is where quality is nurtured. Design offers us with software representations that we can evaluate for quality. We can only correctly transform a customer's vision into a finalized software product or system through design. All of the software engineering phases that follow are built on top of software design. We risk developing an unstable system without a robust design - one that will be difficult to test and whose quality will not be judged until the very end. During the design phase, the data structure, programmed structure, and procedural elements are refined, evaluated, and documented in stages. From a technical or project management standpoint, system design might be seen. Architectural design, data structure design, interface design, and procedural design are the four activities that make up design from a technical standpoint.

3.2 NORMALIZATION

It's the process of transforming a relationship into a standard format. The procedure is used to deal with issues that might develop as a result of data redundancy, such as data duplication in the database, as well as issues that might develop as a result of insertion, updating, and deletion anomalies.

Decomposing is the process of breaking down relationships into different connections in order to eliminate anomalies and maintain data integrity. To accomplish so, we employ standard forms or standards for relation structure.

Insertion anomaly: Due to a lack of additional data, it is impossible to contribute data to the database.

Deletion anomaly: Data loss that was unintended as a result of the deletion of other data.

Update anomaly: Due to data redundancy and incomplete updates, there is data inconsistency.

Normal Forms: These are the guidelines for organizing relationships in such a way that anomalies are avoided.

FIRST NORMAL FORM

If the values in a relation are atomic for each attribute in the relation, it is said to be in first normal form. This simply means that no attribute value can be a collection of values, or a repeating group, as it is frequently called.

SECOND NORMAL FORM:

When a relationship is in second place, it is said to be in second place. It is in first normal form and must meet one of the following requirements in order to be considered normal.

- 1) Primary key is not a composite primary key
- 2) No non key attributes are present
- 3) Every non key attribute is fully functionally dependent on full set of primary key.

THIRD NORMAL FORM:

If a relationship has no transitive dependencies, it is said to be in third normal form.

Transitive Dependency:

Two non-key qualities are said to be transitively dependent if they rely on each other as well as the main key.

The data was decomposed into many tables using the after mentioned normalization methods, allowing the data to be preserved in a consistent condition.

System security refers to the safeguarding of computer-based resources such as hardware, software, data, procedures, and people from unwanted access or natural disasters.

System security may be broken down into four categories:

- Security
- Integrity

- Privacy
- Confidentiality

SYSTEM SECURITY refers to the technical innovations and processes used in hardware and operating systems to safeguard against intentional or unintentional harm from a specific danger.

DATA SECURITY is the safeguarding of data against loss, disclosure, alteration, and destruction.

SYSTEM INTEGRITY refers to the efficient operation of hardware and software, as well as adequate physical security and protection against external dangers like as eavesdropping and wiretapping.

PRIVACY specifies the user's or organization's rights to decide what information they are willing to share with or accept from others, as well as how they can be safeguarded from undesired, unjust, or excessive distribution of information about them.

CONFIDENTIALITY is a designation given to sensitive data in a database to reduce the risk of privacy breach. It's a characteristic of information that defines its necessity for security.

System security refers to a variety of data validations in the form of checks and controls that are used to keep the system from failing. Only legitimate data should be entered into the system, and only valid actions should be done on it.

CLIENT-SIDE VALIDATION

Various client-side validations are employed to verify that only legitimate data is entered on the client side. Client-side validation reduces server load and saves time when dealing with incorrect data.

The following are some of the controls in place:

- VB Script is used to guarantee that only the appropriate data is entered into the relevant fields. The maximum lengths of the fields on the forms have been set correctly.
- Forms cannot be submitted without the required data being filled in, so that human errors such as entering empty necessary fields may be handled out at the client side, saving the server time and load.

Tab-indexes are created depending on the demands of the user and the ease with which they can use the system.

Some checks aren't possible to do on the client side. Server-side checks are required to prevent the system from failing and to notify the user that an invalid operation has been performed or that the operation done is limited. The following are some of the server-side checks:

- A server-side constraint has been implemented to ensure that the main key and foreign key are valid. It is impossible to replicate the value of a primary key. Any attempt to duplicate the main value generates a notice informing the user of those values. Forms employing foreign keys may only be modified with the current foreign key values.
- The user is notified of successful activities or server-side exceptions through suitable messages.

A variety of access control mechanisms have been used to ensure that one user does not irritate another. According to the organizational structure, access permissions for various sorts of users are managed. Only authorized users are allowed to log in and have access to the system, which is determined by their category. On the server side, usernames, passwords, and permissions are managed.

Server-side validation is used to place limits on a number of restricted operations.

III. CONCLUSION

Working on this intriguing and hard project has been a genuine pleasure for me. This project was beneficial to me since it taught me

not only how to write in Java, but also how to handle all aspects of the "College Management System." It also gives information on the most up-to-date technologies for creating web-enabled applications and client-server technologies, both of which will be in high demand in the future. This will give better opportunities and advice in developing initiatives on your own in the future. By utilising this application, you will have access to data from the College. It contains information about each student and faculty member, organised by student and faculty member. Students may examine all of the facts and file complaints using this system. The user will be able to make rapid decisions and save time and money by using this programmed.

REFERENCES

[1] Android Developer Guide:

<http://developer.android.com/guide/index.html>

[2] AndroidAPI: <http://developer.android.com/reference/packages.html>

[3] Java 6 API: <http://downloadlnw.oracle.com/javase/6/docs/api/>

[4] Google Maps API: <http://code.google.com/android/add-ons/googleapis/reference/com/google/android/maps/packagessummary.html>

[5] Android Fundamentals:

<http://developer.android.com/guide/topics/fundamentals.html>

[6] The Java Tutorials: <http://downloadlnw.oracle.com/javase/tutorial/index.html>

[7] Android Native Development Kit:

<http://developer.android.com/sdk/ndk/index.html>

[8] Android User Interfaces:

<http://developer.android.com/guide/topics/ui/index.html>

[9] Declaring Layout: <http://developer.android.com/guide/topics/ui/declaring-layout.html>

[10] Common Tasks: <http://developer.android.com/guide/appendix/faq/common tasks.html>

[11] Maps External Library: <http://code.google.com/android/add-ons/googleapis/maps-overview.html>

[12] Maps API Key:

<http://code.google.com/android/addons/google-apis/mapkey.html>

[14] Icons: http://developer.android.com/guide/practices/ui_guidelines/icon_design.html

[15] Sample Source Code: <http://developer.android.com/resources/samples/get.html>

[16] List of Sample Apps: <http://developer.android.com/resources/samples/index.html>

[17] apps-for-android Sample Apps: <http://code.google.com/p/apps-for-android/>

[18] Android Developer's Blog: <http://androiddevelopers.blogspot.com/>

[19] Developer FAQ: <http://developer.android.com/resources/faq/>


[20] Developer Forums: http://developer.android.com/resources/community_groups.html

[21] Android Developer's Group: <http://groups.google.com/group/androiddevelopers?lnk=>

[22] XDA-Developers Forums: <http://forum.xdadevelopers.com/>

[23] A book by O'reilly called "android application development".

Authors Profile

	K. Sreenath, (srinathits@gmail.com) Associate Professor, Department of Information Technology, QIS college of Engineering and technology
	O. Vinay Kumar (vinavoguri1125@gmail.com) Student Department of Information Technology, QIS college of Engineering and technology

	<p>G. Sudhakar (sudhagorantla9@gmail.com) Student Department of Information Technology, QIS college of Engineering and technology</p>
	<p>P. Srikanth (srikanthpaladugu3@gmail.com) Student Department of Information Technology, QIS college of Engineering and technology</p>
	<p>B. Ajarath Babu (ajarath1232@gmail.com) Student Department of Information Technology, QIS college of Engineering and technology</p>

SECURING A FILE USING CLOUD COMPUTING ADOPTING FRAMEWORK

Dr.G.L.VaraPrasad¹ CH.Mahesh² P.Nikhil³ R.VincentBabu⁴ S.SivaSai⁵
G.Swamireddy⁶

¹Associate Professor, Department of IT, QISCET, Ongole.

^{2,3,4,5,6} IV B.Tech Students , Department of IT, QISCET, Ongole.

ABSTRACT

In the modern world distributed computing is the most well-known arising figuring innovation that has altered data innovation through adaptable provisioning of registering assets. The overall public cloud administrations market is gauge to become 17% in 2020 to add up to \$266.4 billion, up from \$227.8 billion out of 2019 as per Gartner. In any case, on the off chance that it neglects to guarantee appropriate security insurance, cloud administrations could eventually bring about greater expense and possible loss of business in this manner disposing of the relative multitude of likely advantages of cloud innovation. The numerous encryption is a cycle to scramble the information on various occasions utilizing similar calculation or various calculations. This method is broadly utilized because of its component of improved security for information correspondence over the weak remote organization as the web. The idea of various encryption can be portrayed as a procedure to give multi-layer and staggered security over questionable remote organization. In this undertaking, we propose a computationally proficient staggered encryption structure that joins the strength of symmetric, the encryption calculation AES, SHA 128, SHA 256, MD5, WHIRLPOOL, BLOWFISH. It gives the examination of information security issues and protection assurance undertakings identified with distributed computing by forestalling information access from unapproved clients, overseeing delicate information, giving exactness and consistency of information put away. In this system, user can upload their files in cloud by encrypting that file three times by using different algorithms. During the encryption process user will get secret key, if anyone wants to decrypt the uploaded file, they need to enter that secret key. Without the key values no

one can decrypt the file. By using this method file can be uploaded in cloud with more secured and in efficient manner.

Key Words: Symmetric, Multiplealgorithm, AES, SHA, MD5, Whirlpool, Blowfish.

I INTRODUCTION

Cloud Computing is the liberation of computing services such as servers, networking, analytics, storage, intelligence, database, software and more, over the other cloud. When an industry buys in third-party cloud services as either a public or hybrid cloud offering, it is likely they will not be provided with a full services description, giving details how the platform works, and the security processes the seller operates. This lack of service precision makes it difficult for customers to wisely evaluate whether their data is being stored and processed securely at all times. Surveys have shown that around 85% of IT managers are slightly confident that company data is being stored securely by their cloud storage. Using public or hybrid cloud assistance/discounts can expose a business to security vulnerabilities caused by other users of the same cloud infrastructure.

The liability is upon the cloud vendor to see that this does not occur, yet no vendor is perfect. It is almost possible that a security issue caused by another client system in the same cloud will affect every other client. Cloud encryption is the development of encoding or converting data before it's transferred to cloud storage. Encryption uses mathematical algorithms to transform data (plaintext), file, may it be a text, code or image, to an unreadable form (ciphertext) that can mask it from unauthorized and malicious users. It is the easiest and most vital way to make sure that cloud data can't be stolen, breached, and read by someone with a bizarre motive. Cloud storage providers store an encrypted data and give encryption keys to the users. These keys are used to securely decipher data when required. Decryption modifies the concealed data back into readable data.

There are two methods used to code and decode data, and these methods are universally evolved as the field of information technology changing its ways of data protection and privacy security. These are also called encryption algorithms. These methods are as follows:

1. Symmetric algorithm

In this system, encryption and decryption keys are the same, which makes it best for closed systems and individual users. These keys are used for reliable communication. This is also known as the secret key algorithm and is usually used for major data encryption. This is doubtlessly and rapidly performed by hardware and faster than the asymmetric method.

2. Asymmetric algorithm

In this method, two keys are used (private and public) and they are mathematically associated together. It is called asymmetric as they keys/keys are coupled with each other, but they are not the same. The private key must be kept concealed and secret, yet the public key can be shared with anyone.

In symmetric key algorithms, the same key/key may be used to encipher /decipher a plain text. Subsequently, these algorithms are simple and computationally efficient. Examples of such algorithms: AES, SHA 128, SHA 256, MD5, Whirlpool, Blowfish. On the other hand, asymmetric encryption algorithms need two keys for encryption and decryption, thereby making the generation of these keys computationally high and counterproductive for encrypting large data. For instance, RSA is an Asymmetric encryption algorithm. With the advancement of technology and computational power, hackers have developed different models to ambush these primary encryption algorithms. One of the outcomes to deal with this problem is to build a synthesis model that consists of multi-level encryption algorithms.

The ideology here is to combine the strength of multiple basic encryption algorithms together to build a complex, sophisticated

encryption technique. Two challenges are facing these techniques: computation efficiency and security. The initial one is needed when sending massive data and the latter ensures that the combination between these steps will lead into a complete system that is secured. Many multi-level encryption approaches were put forward in literature, but they were lagging from computational or security problems or being specified for a certain type of networks.

In this paper, we propose a multi-level encryption framework to encrypt the file based on its size that is secure and computationally efficient. The framework comprises using a symmetric encryption algorithm AES, SHA etc. for encrypting and decrypting data.

Most cloud storage suppliers grip the encryption key on behalf of the user, which requires him or her to impetuously trust the company and will not misuse the access to your files, release the key to hackers, or give it to prying government authorities. Furthermore, anyone with physical access to your phone or laptop could easily get to files on the cloud, because most of these facilities leave you registered in by default. A handful of companies with cloud backup, such as iDrive, authorize the user to create their own private key, encrypting data on their personal computer prior uploading it to the cloud. However, this is reserved for highly valuable business-tier cloud backup, but not the average individual user.

Cloud encryption is needed because its main aim is to secure and protect confidential information or file as it is transmitted through the Internet and other computer systems. Best way to analyze an organization's security and privacy status is through the CIA triad. This embraces for Confidentiality, Integrity, and Availability. By conventional method, the platform of information technology the availability of the data and its integrity is well-defined. IT does not provide enough impression on classified data. That is why, companies/institutions/organizations should prefer cloud encryption. Besides, encryption is not just about protection, securing data and its privacy: rather its core, digital data is significantly

transferring, and encryption is needed to perform the transmission in a safe way. Users make sure that their particulars and data is secured when transmitted to another user and that the other user is who they intend to provide the data to and not any malicious attackers.

1. Accomplished statistics security of all times

when data or the information is being transferred or secured, that's when encryption works. This is a flawless solution no matter what is being done with the data. Even if data is weak and liable to vulnerabilities while being transferred from elsewhere to somewhere else. Encryption verifies the security during this process.

2. Protection of privacy

Encryption defends liable data like personal information of every individual user. This enables privacy and inconspicuousness, alleviating chances of surveillance by government agencies, criminals, and cyber attackers.

3. Part of compliance

Encryption is one of the most reliable tools to share and save data as it follows with the restrictions proposed by an organization.

4. Collective devices protection

Several types of communication devices have become essential of our lives nowadays. Transfer of data from one device to another constitute high threat and vulnerability, hence encryption aid in protecting data across multiple devices.

5. Maintains integrity

Hackers benefit from modifying the information by swindling, just by stealing data. It is possible for these hackers to swap and amend ciphered data. However, the recipient of the information has the potential to identify if it is corrupted, allowing for a prompt response and solution to the attack.

Cloud Encryption Challenges Even though we agree that encryption is the best tool for data protection in the cloud, it's better to also acknowledge that there is no absolute approach when it comes to privacy and security. Besides any other tool to conflict vulnerabilities, threats in the information superhighway, there are challenges that an organization or user may face by using encryption. These are as follows:

1. Loss of data

The limitations about encryption: if a user loses track of his decryption keys and has no backup of the data, it's similar to closely losing data and wrecking it.

2. Encryption functions like a password

When operation on encryption to protect devices, files, and disks, the cue is a password that's chosen by the user. Human passwords are effortlessly easy to seize and hack unlike other solutions like AES-256 (Advance Encryption Standard) which involves long random keys.

3. Complexity of encryption

For a daily-basis user, some encryption programs are complex and they may end up using it inappropriately. This could end up as a failure in encrypting the data which they wanted to secure and encrypting data but that they did not want to encode. The synthesis of deciphering extracts the processor time in the computer. The more muddled up of the encoding, the longer it takes to process.

II TYPES OF ALGORITHMS:

1. AES:

The Advanced coding customary (AES) is AN agreement chosen by the U.S. government to fortify classified data. AES is run in package and hardware throughout the planet to cypher protected information. it's crucial for presidency network security, cyber security and electronic information protection. AES could be a kind of bilateral algorithmic program. It applies the same key for each coding and secret writing. AES

work with block cipher. AES includes 3 block ciphers: AES-128, AES-192 and AES-256. AES-128 operates on a 128-bit key length to cipher and decode a group of messages, whereas AES-192 operates on a 192-bit key length and AES-256 a 256-bit key length to cipher and decode messages. Each cipher codes and decodes data in blocks of 128 bits victimization cryptologic keys of 128, 192 and 256 bits, accordingly.

There are a unit ten rounds for 128-bit keys, twelve rounds for 192-bit keys and fourteen rounds for 256-bit keys. A spherical incorporates of many process steps that embrace substitution/replacement, transposition and mixture of the input plain text to rework it into the ultimate yield of cipher text.

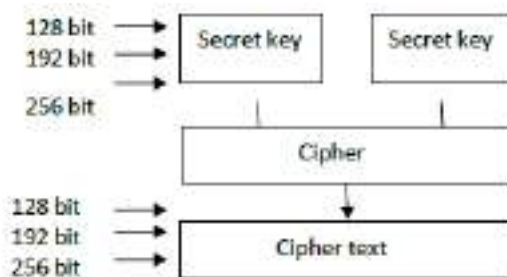


Fig 1 AES algorithm

Each spherical has for transformation.

- Sub bytes
- Shift rows
- Mix columns
- Add spherical keys

2. SHA

In cryptography, SHA-1 (Secure Hash algorithmic program 1) could be a cryptologic hash perform that takes the information And fabricates a 160-bit (20-byte) hash price referred to as a messagedigest – undoubtedly used as an hex variety, forty digits long. Secure Hash Algorithms, conjointly referred to as SHA, is a member of cryptologic functions designed to stay information safe. It works by transfiguring the information employing a hash function: AN algorithmic program includes of bitwise operations, standard additions, and compression functions. The hash perform creates a

fixed-size string that appears obscurely different from the original. These algorithms are not made public to be unidirectional functions, therefore reworking into their several hash values, that is much not possible to rework them back to the initial information. A couple of algorithms are SHA-1, SHA-2, and SHA-3, every of that were linearly designed with sturdy coding in response to hacker attacks. SHA-0, as an example, is currently noncurrent thanks to the manifested vulnerabilities. SHA-256 is one in all the most effective hash functions to SHA-1, and it's one in all the secured and it has well-defined hash functions on the market. SHA-256 isn't a lot of complicated to code than SHA-128. The 256-bit key makes it a best partner-function for AES.

3. MD5:

MD5 is far faster than different class of message digest that extracts the plain text of 512 bit blocks that is moreover classified into sixteen blocks, every of thirty two bit and produces the 128 bit message digest that could be a set of 4 blocks, every of thirty two bits. MD5 produces the message digest through 5 steps i.e. padding, append length, divide input into 512 bit blocks, initialize chaining variables a method blocks and four rounds, uses contrastive constant in every iteration.

Step 1: Append Artifact Bits

Padding suggests that computing additional bits to the initial message. In MD5 original message is cushioned in order that its length in bits is congruent to 448 modulo 512. Artifact is completed specified the full bits area unit sixty-four less being a multiple of 512 bits length.

Padding is completed though the length of the initial message is already congruent to 448 modulo 512. In Artifact bits, the sole initial bit is one and also the remainder of the bits area unit zero.

Step 2: Append Length

After Artifact, sixty-four bits area unit compact at the top that is employed to document the length of the particular input. Modulo 2^{64} , at this locus, the ultimate message contains a length multiple of 512 bits.

Step 3: Initialize MD buffer

A four-word buffer (A, B, C, D) is operated to enumerate the values for the message digest. Here A, B, C, D are unit 32-bit registers and are unit initialized.

Step 4: process message in 16-word block

MD5 uses the auxiliary functions that confiscates the input as 3 32-bit variety and produces a 32-bit output. These functions employ logical operators like OR, XOR, NOR.

4. Whirlpool

Whirlpool could be a cryptologic hash perform that was derived from sq. and Advanced coding customary. it's a block cipher hash perform and invented once block cipher. It considers but 2^{256} bits length input and converts it into 512 bit hash. {the initial the primary} model of whirlpool is termed Whirlpool-0 and switched to Whirlpool-T once it's first revision in 2001. Consistent with this style the S-box is modified and become easy to use in hardware.

Every block cipher in whirlpool could be a 8×8 matrix. The state of the perform changes in ever spherical by victimization four operations:

- 1) Mix Rows(MR)
- 2) Substitute Bytes(SB)
- 3) Add Spherical Key(AK)
- 4) Shift Columns(sc)

Hash price is calculated by victimization the formula:

$$\text{State} = \text{MR} * \text{AK} * \text{SC} * \text{SB}(\text{State})$$

5. Blowfish

Blowfish is a coding technique designed by Bruce Schneier in 1993 as another to DES coding Technique. it's unco quicker than DES and provides an honest coding rate with no constructive scientific discipline technique found until date. it's one in all the primary, secure block ciphers

not subject to any patents and thence honestly on the market for anyone to use.

1. blockSize: 64-bits
2. keySize: 32-bits to 448-bits variable size number of subkeys: eighteen [P-array]
3. number of rounds: sixteen
4. number of substitution boxes: four

Based on the file dimensions any three algorithmic program are determined and cipher it. Once coding, the encrypted file are hoarded within the cloud.

III PROPOSED IMPLEMENTATION

Step 1: Before upload the file, the user have to register for login by entering basic information about the user.



Fig 2 Registration page

Step 2: After register, the user can login and then upload their files which they want to encrypt and store in the Cloud.



Fig 3 Login page



Fig 4 User page

Step 3: After uploading the file, based on the file size randomly three algorithm will be selected by the system and encrypted it by three times by using different algorithms.



Fig 5 Encrypted file

Step 4: During the encryption process, the user will get some secret key values or private key values. If the user wants to decrypt the file they have to enter that secret key. If the user doesn't have the secret key they can't access that file. If the user enters that correct key values that file will be decrypted.



Fig 6 Decrypted file

In this project we developed an algorithm to secure the user data in more efficient way. We have proposed high secured techniques to protect the user data which they want to be upload in the cloud without any third party access. Encryption is regarded as one of the most effective approaches for data security. In this system, we are using more than 3 algorithms which are more efficient for data encryption.

IV PERFORMANCE

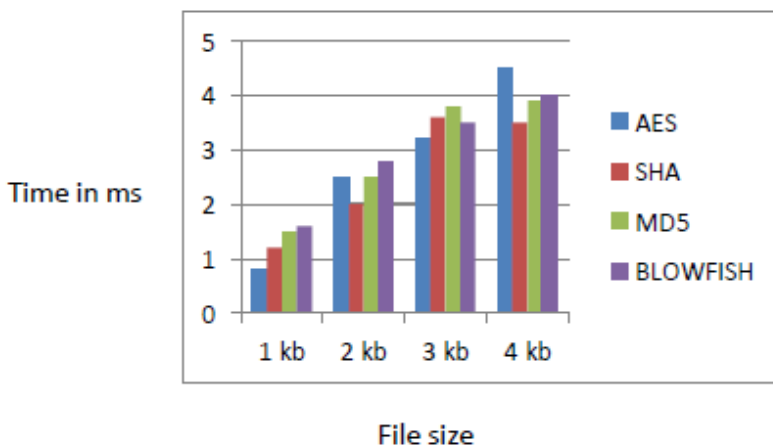


Fig 7 Comparison of Encryption Algorithm

Success of the execution test or analysis is required to evaluate the efficiency and security. Every encryption algorithm has some strength and weaknesses. In this performance analysis it describes the encryption time in milliseconds based on file size.

V CONCLUSION

Cloud computing is the arrangement of assets or administrations given through the web to the clients on their interest by cloud specialist co-ops. Since every single association is moving its information to the cloud, implies it utilizes the capacity administration given by the cloud supplier. Consequently it is compulsory to ensure that information against unapproved access, alteration or disavowal of administrations and so forth Distributed computing can turn out to be safer utilizing cryptographic calculations. Cryptography is the workmanship or study of keeping messages secure by changing over the information into non intelligible structures. In any case, the current cryptographic calculations are single level encryption calculations. Digital crooks can undoubtedly break single level encryption. Consequently we propose a framework which utilizes staggered encryption and decoding to give greater security to Cloud Storage. As our proposed calculation is a Multi-level cryptographic calculation. Regardless of whether some interloper (unapproved client) gets the information unintentionally or purposefully, he/she should need to unscramble the information at each level which is a troublesome assignment without a secret key. It is normal that utilizing staggered encryption will give more security to Cloud Storage than utilizing single level encryption.

VI. REFERENCES

[1] S. Ahmad, K. M. R. Alam, H. Rahman, and S. Tamura, "A comparison between symmetric and asymmetric key encryption algorithm based decryption mixnets," in Proceedings of the IEEE International Conference on Networking Systems and Security, 2015.

- [2] S.A.M. Daa, M.A.K. Hatem, and M.H. Mohiy(2010). "Evaluating The Performance of Symmetric Encryption Algorithms" International Journal of Network Security, 2010, 10(3), pp.213-219
- [3] Priti V. Bhagat, Kaustubh S. Satpute and Vikas R. Palekar "Reverse Encryption Algorithm: A Technique for Encryption & Decryption" International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 2 Issue 1 January 2013, pp 90-95.
- [4] Z. Hercigonja, D. Gimnazija, and C. Varazdin, "Comparative analysis of cryptographic algorithms and advanced cryptographic algorithms," International Journal of Digital Technology & Economy, vol. 1, no. 2, pp. 1-8, 2016.
- [5] Gagandeepshahi, Charanjitsingh "Cryptography and its two Implementation Approaches" International Journal of Innovative Research in Computer and Communication Engineering, Vol.1, Issue 3, May 2013, PP 668-672.
- [6] Kuldeep Singh, Rajesh Verma, Ritika Chehal "Modified Prime Number Factorization Algorithm (MPFA) For RSA Public Key Encryption" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012, pp 204-206
- [7] Rostyslav Barabanov, Stewart Kowalski and Louise Yngström, "Information Security Metrics", DSV Report series No 11 -007, Mar 25, 2011
- [8] Pallavi Vaidya and S. K. Shinde, "Application for Network Security Situation Awareness", in International Conference in Recent Trends in Information Technology and Computer Science (ICRTITCS - 2012), IJCA, ISSN: 0975 - 8887, 2012.
- [9] SunJun Liu, Le Yu and Jin Yang, "Research on Network Security Situation Awareness Technology based on AIS", in International Journal of Knowledge and Language Processing, ISSN: 2191-2734, Volume 2, Number 2, April 2011.

[10] Disina, A. H., Pindar, Z. A., & Jamel, S., "Enhanced caesar cipher to exclude repetition and withstand frequency cryptanalysis," Journal of Network and Information Security, 2015.

Authors Profile:

	<p>Dr. G. Lakshmi Vara Prasad, Associate Professor, Department of Information Technology, QIS College of Engineering and Technology Email: glv.prasad19@gmail.com</p>
	<p>R. Vincent Babu Student Department of Information Technology QIS College of Engineering and Technology Email: vincentrayapati8@gmail.com</p>
	<p>CH. Mahesh Student Department of Information Technology QIS College of Engineering and Technology Email: maheshch610@gmail.com</p>

	<p>P. Nikhil Student Department of Information Technology QIS College of Engineering and Technology Email: pamidinikhilchowdary9999@gmail.com</p>
	<p>S. Siva Sai Student Department of Information Technology QIS College of Engineering and Technology Email: sivasai.sigirisetty143@gmail.com</p>
	<p>G. Swami Reddy Student Department of Information Technology QIS College of Engineering and Technology Email: guthikonda.swamireddy@gmail.com</p>

RELATIONSHIP IDENTIFICATION & PREDICATION OF DISEASES ASSOCIATION USING MICRO-RNA OF GENOMIC DATA

N. SaiKiran¹ B. DeepthiPriya² V. Srilekha³ P. Lakshmi Pavani⁴ P. Manasa⁵

¹Associate Professor, Department of IT, QISCET, Ongole

^{2,3,4,5} IV BTech Students, Department of IT, QISCET, Ongole

ABSTRACT

The current process of finding the relationship between the father and the son and also predicting the diseases that is yet to occur is quite inaccurate because it includes only the gene-id of the respected person. In order to handle or to make this system more accurate, we propose this system by using the chromosome structure of the person. This system takes the input of the chromosome structure of the son that has been partitioned from the father's chromosome structure. It initially preprocesses the image of the son using the collaborative filtering for making it look different from the input image to show the similarity between the father and the son. It then detects the edge of the structure after preprocessing it using the SOBEL edge detection algorithm. The SOBEL edge detection algorithm is that the gradient of the image is calculated for each pixel position in the image. After detecting the edges of those input images, matching process starts between the input image and the list of father chromosome images. Then the matched output appears. In order to predict the diseases which is yet to come in future for the son is represented graphically by dividing it into three colors, firstly green represents there is less possibility of the son getting the disease, secondly yellow represents there may be any chance of son getting the disease and finally red represents there is high possibility of son getting the disease.

1. INTRODUCTION

Information mining is the path toward discovering plans in broad enlightening accumulations including methods at the intersection purpose of AI, experiences, and database systems. It is an essential technique where keen systems are associated with evacuate data plans. It is an interdisciplinary subfield of programming building. The general target of the data mining process is to remove information from an enlightening file and change it into a sensible structure for further use. Other than the unpleasant examination step, it fuses database and information the

authorities' points of view, information pre-taking care of, model and induction contemplations, enrapturing quality estimations, multifaceted nature contemplations, post-arranging of found structures, acknowledgment, and web empowering Information mining is the examination adventure of the "learning disclosure in databases" system or KDD. Image processing is planning of pictures using numerical exercises by a flag getting ready for which the data is an image, a movement of pictures or a video, for example, a photo or a video design; the yield of picture pre-preparing may be either an image or a great deal of characteristics or parameters related to the image. Most picture planning techniques incorporate seeing the image as a two-dimensional banner and applying standard banner taking care of strategies to it. Pictures are also taken care of as three-dimensional signs with the third estimation being time or the z-center point.

An Overview Of Relationship Identification & Prediction of Diseases Association Using Micro-Rna Of Genomic Data. In recent years, the technology is being increased to greater level in several fields. This system focuses on the medical field. Our system helps the doctors or the admins to predict the diseases easily with help of some tools and technologies.

In our system, we are going to find the relation of father and son using two of the factors. One is the gene id and the other is the chromosome structure. Next the diseases that the father has, can also occur for the son are being predicted. When dealing with gene id it was bought from the real time hospitals. Using the father's gene id, the son's gene id is bought by the way of molecular weight and the possibilities of the diseases for the son are displayed by using some of the algorithms. When dealing with chromosome structure, son's structure is given as input and the fathers structure is retrieved using the cM(Centimorgan) and SNP (Single-nucleotide Polymorphisms) values and the diseases are predicted and the disease severity is displayed in a graph. This project not only deals with human chromosome but little process on rat chromosome structure are also being included.

The performance is more accurate when dealing with chromosome structure than with gene id. It also becomes more realistic when dealing the chromosome. The understandability is also very easy with chromosomes. When dealing with gene id it speaks of molecular weight which is the back-end calculation. When with chromosome structure it gives the visual similarity which the user or the admin can easily view the

relation between the father and the son. The gene id format differs from organization to organization but the structure is standard format for any type of chromosome.

1.1.1. CM

Centimorgan (contracted CM) is a proportion of hereditary linkage. Consider it a proportion of DNA data inside a chromosome. Every chromosome contains diverse measures of data. Chromosome 1 contains 281.5cM of data. Chromosome 2 has 263.7cM. Chromosome 21 has just 70.2cM.

1.1.2. SNP

SNPs, or single-nucleotide polymorphisms, are modest bits of a chromosome that contain particular squares of data. There are a great many of them for each chromosome. SNPs are contrasted between two individuals with check whether they coordinate. The measure of data in coordinating SNPs is estimated in CM.

The CM esteems for SNP matches are here and there alluded to as "chromosome length" or "match length". Be that as it may, data is all the more thickly pressed in specific regions or SNPs inside chromosomes, so there's not an immediate connection between's number of SNPs and CM sum. When you see GED match's graphical portrayal of chromosome coordinates, a greater coordinating square does not constantly mean a higher CM esteem.

1.1.3. Start and End Location

Singular markers (called base sets - the things that SNPs are made of) inside a chromosome are numbered. There are a huge number of these markers per chromosome. A section of a chromosome can be recognized by these area numbers.

1.2. MICRO-RNA OF GENOMIC DATA

Quality articulation in cells and tissues of each staggering animal is unquestionably controlled and, all things considered, subject to different conditions, (for instance, improvement, changes in nature, ailments or meds). Distinctive cells and organ structures inside such living thing (checking individuals) contain differing quality verbalization profiles, along these lines genuine understanding of managerial frameworks

connected with such enunciation addresses one of the key issues in genomic medicine.

Separating miRNAs from different classes of little RNAs that are available in the phone is frequently unwieldy – especially the qualification from endogenous little meddling RNAs (siRNAs). The most critical refinement among miRNAs and siRNAs is whether they quietness their very own appearance. Practically all siRNAs (paying little notice to their viral or other beginning stage) calm a comparable locus from which they were resolved. On the other hand, most miRNAs don't tranquil their own one of a kind locus, yet extraordinary characteristics. miRNAs control grouped pieces of progress and physiology, consequently understanding its natural employment is showing progressively basic. Examination of miRNA enunciation may give gainful information, as deregulation of its ability can incite human ailments, for instance, dangerous development, cardiovascular and metabolic diseases, liver conditions and resistant brokenness.

SYSTEM ANALYSIS

Existing System

In this system, we have used the molecular weight of the patients which is based on the chromosome structure. Using the chromosome structure, we cut it into several pieces which is then allowed to undergo the test of chromatography where the pieces of the chromosome structure gets diffused to some extents and based on the heights where the diffusion gets stopped is used to identify the molecular weight. In this system, when the admin login to it, it will ask for the gene ID as the input, then it shows the related details like disease name, symptoms, precautions and then in further it gives the protein values of that patient which then displays the molecular weight that matches another gene ID which is related to that patient gene id. We use co-regulatory modules between Transcription Factor, gene and MiRNA on functional level with genomic data. The integration technique is implemented between miRNA, Transcription Factor (TF) and gene. After integration, Iterative Multiplicating update algorithm is used to check the optimization function between the regulatory modules. We get the expression or some value from this algorithm then compare to protein values. The protein values are from Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) with the help of cross ontology technique. After getting the protein values the diseases

associated for each value is found. Then the molecular weight is found by chromatography test for each person's chromosome. By using the molecular weight, the father and the son relationship is identified.

SNCONMF ALGORITHM

In order to identify miRNA-TF-gene co-regulatory modules(miRNA-TF-gene), we develop an approach called SNCoNMF. The final optimization function has described above. Similar to standard NMF, we adapt the "iterative multiplicative update" method to minimize the Euclidean error function. The following are the updating rules for the factor matrices W and H_i ($i = 1, 2, 3$). Specifically, the SNCoNMF algorithm starts by randomly initializing matrices W , H_1 , H_2 and H_3 and iteratively

DRAWBACKS

- The major drawback of this method is the visual representation of finding the relationship and predicting the diseases based on the relationship is not possible.
- The relationship between the father and the son is shown using the molecular weight of the person, where the accuracy is less.
- This system has less reliability as molecular weight is the only feature predicting the relationship.

PROPOSED SYSTEM

From the existing system, it is seen that the current process of finding the relationship between the father and the son and also predicting the diseases that is yet to occur is quite inaccurate because it includes only the gene id of the respected person. In order to handle or to make this system more accurate, we propose this system by using the chromosome structure of the person.

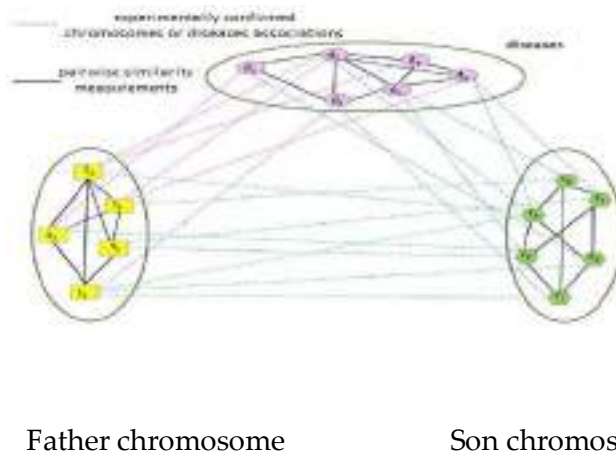
This system takes the input of the chromosome structure of the son that has been partitioned from the father's chromosome structure. It initially preprocesses the image of the son using the collaborative filtering for making it look different from the input image to show the similarity between the father and the son. It then detects the edge of the structure after preprocessing it using the SOBEL edge detection algorithm. The SOBEL edge detection algorithm is that the gradient of the image is calculated for each pixel position in the image. After detecting the edges of those input images, matching process starts between the input image and the list of father chromosome images. Then the matched output appears. In

order to predict the diseases which is yet to come in future for the son is represented graphically by dividing it into three colors, firstly green represents there is less possibility of the son getting the disease, secondly yellow represents there may be any chance of son getting the disease and finally red represents there is high possibility of son getting the disease.

ADVANTAGES

1. User can easily predict the severity of the disease that is given in the bar graph. The red color indicates the more chances, yellow moderate and the green for less chances.
2. The relation between the father and the son is more accurate and reliable.

SYSTEM ARCHITECTURE

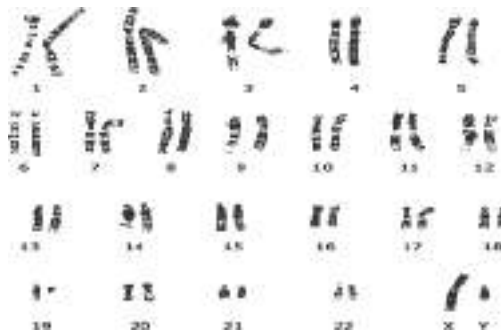


MODULE DESCRIPTION DATASET

The picture chromosomes of dataset have been made for examination of ailment. This dataset contains twenty-two cases and five characteristics are utilized in this similar examination. The chromosome legacy and how-to examining chromosomes by utilizing GED match databases to discover One-to- numerous Matches, One-to-one Compare, and People who coordinate either of two kits, the ideas are relevant to comparative databases given by the DNA real testing administrations. ⁷ This implies your DNA information is in the GED match database so it very well may be utilized to contrast with others. Here the backend of the coding utilizing by java stage. Java is an extensively important PC programming language that is synchronous, class-based, object-

masterminded, and unequivocally wanted to have as few execution conditions as could reasonably be typical. It is intended to give application originators "an opportunity to compose once, run wherever ", suggesting that collected Java code can continue running on all phases that assistance Java without the necessity for recompilation.

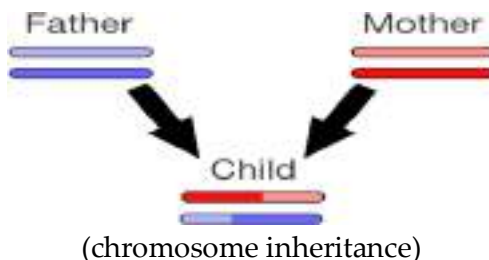
Chromosome Basics



Chromosomes are small structures found inside your cells.¹³ They contain the DNA data and guidelines that characterize your identity - what you resemble, how your body works, and even what hereditary ailments you may have. People have 46 chromosomes. Be that as it may, chromosomes come in sets, so we ordinarily consider them 23 sets of chromosomes. The initial 22 chromosome sets (called autosomes) are numbered 1 through 22. We'll basically concentrate on these autosomal chromosomes.

Chromosome Inheritance

One autosomal chromosome from each pair originates from your mom and alternate originates from your dad.⁴ This implies you get half of your DNA from your mom and half from your dad. Every chromosome they pass on to you is their very own blend pair of chromosomes which they got from their folks (your grandparents).



The picture above delineates how one sets of chromosomes might be passed from your folks to you. The hues don't mean anything unique - they basically portray the individual chromosomes and chromosome areas

ALGORITHMS SPECIFICATION

Collaborative Filtering

For every client, recommender frameworks prescribe things dependent on how comparable clients loved the thing. Suppose Alice and Bob have comparable interests in computer games. Alice as of late played and delighted in the amusement Legend of Zelda: Breathe of the Wild. Sway has not played this amusement, but since the framework has discovered that Alice and Bob have comparative tastes, it prescribes this diversion to Bob. Notwithstanding client likeness, recommender frameworks can likewise perform synergistic sifting utilizing thing comparability ("Users who preferred this thing additionally loved X").

In the more general sense, collaborative filtering is the process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. Applications of collaborative filtering typically involve very large data sets. Collaborative filtering methods have been applied to many different kinds of data including: sensing and monitoring data, such as in mineral exploration, environmental sensing over large areas or multiple sensors; financial data, such as financial service institutions that integrate many financial sources; or in electronic commerce and web applications where the focus is on user data, etc. The remainder of this discussion focuses on collaborative filtering for user data, although some of the methods and approaches may apply to the other major applications as well.

Sobel Edge Detection Algorithm

Edge detection is in the forefront of image processing for object detection, it is crucial to have a good understanding of edge detection algorithms. Sobel which is a popular edge detection algorithm is considered in this work. There exists a function, edge's which is in the image toolbox. In the edge function, the Sobel method uses the derivative approximation to find edges. Therefore, it returns edges at those points where the gradient of the considered image is maximum. The Sobel operator performs a 2-D spatial gradient measurement on images. It uses a pair of horizontal and vertical gradient matrices whose dimensions are 3x3 for edge detection operations. It will also demonstrate how to build a Sobel

detector function of 5×5 dimension in mat lab to find edges.

The Sobel operator, sometimes called the Sobel–Feldman operator or Sobel filter, is used in image processing and computer vision, particularly within edge detection algorithms where it creates an image emphasizing edges. It is named after Irwin Sobel and Gary Feldman, colleagues at the Stanford Artificial Intelligence Laboratory (SAIL). Sobel and Feldman presented the idea of an "Isotropic 3×3 Image Gradient Operator" at a talk at SAIL in 1968. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel–Feldman operator is either the corresponding gradient vector or the norm of this vector. The Sobel–Feldman operator is based on convolving the image with a small, separable, and integer-valued filter in the horizontal and vertical directions and is therefore relatively inexpensive in terms of computations. On the other hand, the gradient approximation that it produces is relatively crude, in particular for high-frequency variations in the image.

Modified Greedy Algorithm

A modified greedy algorithm called Multi-Tree-based Orthogonal Matching Pursuit (MTOMP). It is an algorithmic paradigm that follows the problem-solving heuristic of making the locally optimal choice at each stage with the intent of finding a global optimum. In many problems, a greedy strategy does not usually produce an optimal solution, but nonetheless a greedy heuristic may yield locally optimal solutions that approximate a globally optimal solution in a reasonable amount of time.

For example, a greedy strategy for the traveling salesman problem (which is of a high computational complexity) is the following heuristic: "At each step of the journey, visit the nearest unvisited city." This heuristic does not intend to find a best solution, but it terminates in a reasonable number of steps; finding an optimal solution to such a complex problem typically requires unreasonably many steps. In mathematical optimization, greedy algorithms optimally solve combinatorial problems having the properties of matroids, and give constant-factor approximations to optimization problems with sub modular structure.

RESULTS AND DISCUSSION

In this system a collaborative filtering is used for preprocessing method. In the preprocessing the required information that are related to

the chromosome and diseases are processed. We use edge detection method to detect the edges of the chromosome structure and so that it gets the parent chromosome. For the edge detection we use the algorithm called Sobel Edge detection. This algorithm is being used because it gives the exact outer structure when compared to that of other algorithms. Then for matching purpose the Modified Greedy algorithm is being used to match the related chromosome. Then the bar graph is being generated. From the bar graph the possibility or the chances of the diseases for the son are displayed with the percentage of diseases. The chances are also being displayed with the color differences. Then finally the histogram can be displayed for each of the input structure that is being given. Thus, by using these predictions and the bar graph, the admins and the doctors can be benefited for explaining or show casing the patients or the users to understanding an easy way.

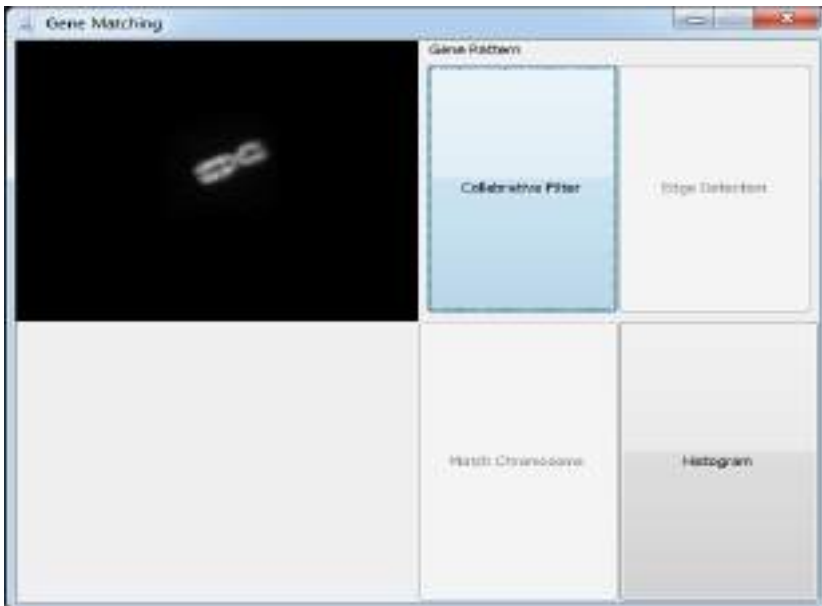


Figure. Image preprocessing

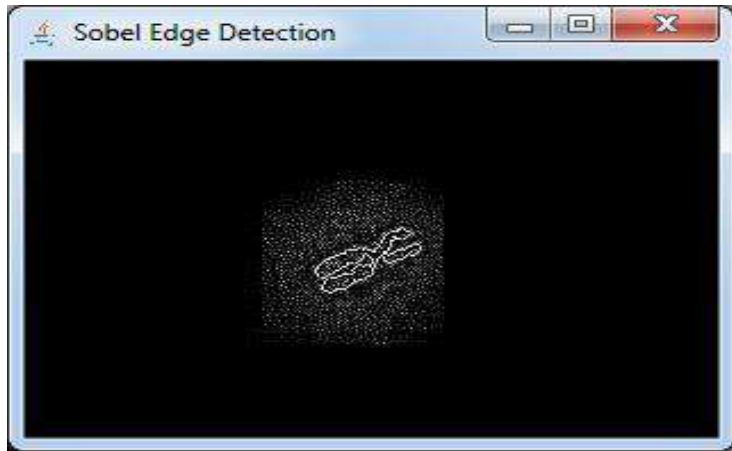


Figure: Edge Detection



Figure: Image Matching

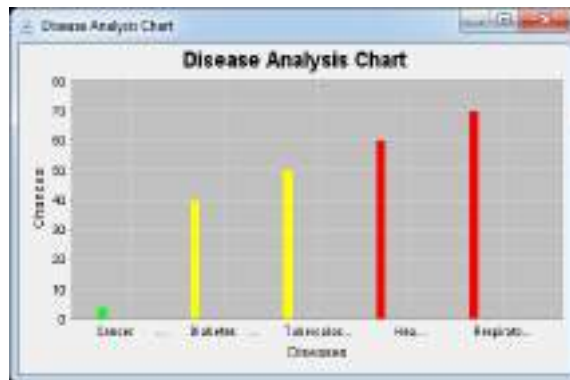


Figure : Graphical Representation

CONCLUSION & FUTURE ENHANCEMENTS

In this system a collaborative filtering is used for preprocessing method. In the preprocessing the required information that are related to the chromosome and diseases are processed. We use edge detection method to detect the edges of the chromosome structure and so that it gets the parent chromosome. For the edge detection we use the algorithm called Sobel Edge detection. This algorithm is being used because it gives the exact outer structure when compared to that of other algorithms. Then for matching purpose the Modified Greedy algorithm is being used to match the related chromosome. Then the bar graph is being generated. From the bar graph the possibility or the chances of the diseases for the son are displayed with the percentage of diseases. The chances are also being displayed with the color differences. Then finally the histogram can be displayed for each of the input structure that is being given. Thus, by using these predictions and the bar graph, the admins and the doctors can be benefited for explaining or showing cases to the patients or the users to understand an easy way.

REFERENCES

- [1] Hailin Chen And Zuping Zhang(2018) prediction of Drug-Disease Associations for Drug Repositioning Through Drug-miRNA Disease Heterogeneous Network Hailin ChenAnd Zuping Zhang Received June 23, 2018, accepted July 24, 2018, date of publication July 27, 2018, date of current version September 5, 2018. Digital Object Identifier 10.1109/ACCESS.2018.2860632.
- [2] Z. Sun, Z. Wu, F. Zhang, Q. Guo, H. Chen, J. Zhao, D. Song, Q. Huang, L. Li, and J. Xiao(2016), "Prone is critical for breast cancer growth and

metastasis," *Gene*, vol. 594, no. 1, pp. 160–164.

[3]. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, (2015) "limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic acids research*, p. gkv007.

[4]S. Zhang, C.-C. Liu, W. Li, H. Shen, P. W. Laird, and X. J. Zhou,(2012) "Discovery of multi-dimensional modules by integrative analysis of cancer genomic data," *Nucleic Acids Research*, vol. 40, no. 19, pp. 9379– 9391.

[5] Y. Lu, Y. Zhou, W. Qu, M. Deng, and C. Zhang, (2011) "A lasso regression model for the construction of microrna-target regulatory networks," *Bioinformatics*, vol. 27, no. 17, pp. 2406–2413.

[6] D. H. Tran, K. Satou, T. B. Ho, and T. H. Pham, (2010) "Computational discovery of mir-tf regulatory modules in human genome," *Bioinformation*, vol. 4, no. 8, pp. 371–377, [2010].

[7] X. Peng, Y. Li, K.-A. Walters, E. R. Rosenzweig, S. L. Lederer, L. D. Aicher, S. Proll, and M. G. Katze,(2009) "Prediction of regulatory modules comprising micrnas and target genes," *BMC Genomics*, vol. 10, p. 373.

[8] C.-Y. Chen, S.-T. Chen, C.-S. Fuh, H.-F. Juan, and H.-C. Huang,(2011) "Coregulation of transcription factors and micrnas in human transcriptional regulatory network," *BMC bioinformatics*, vol. 12, no. 1, p. 1.

[9] F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons,(2006) "Document clustering using nonnegative matrix factorization," *Information Processing and Management*, vol. 42, no. 2, pp. 373–38.

[10] Carmona-Saez P, Chagoyen M, Rodriguez A, Trelles O, Carazo JM, et al. (2006) Integrated analysis of gene expression by Association Rules Discovery. *BMC Bioinformatics* 7: 54.

[11] M. Kim and B. Tidor, (2003) "Subsystem identification through dimensionality reduction of large-scale gene expression data," *Genome Research*, vol. 13,pp. 1706–1718,

[12]S. V. Gayathiri Devi, C. Nalini, N. Kumar, "An efficient software verification using multi-layered software verification tool "International Journal of Engineering & Technology,7(2.21)2018 454-457

[13] J.-P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, (2004) "Metagenes and molecular pattern discovery using matrix factorization," *Proceedings of the national academy of sciences*, vol. 101, no. 12, pp. 4164–4169.



[14] Hia Fang, Julian Gough,(2013) A domain centric solution to functional genomics via dcGO predictor *BMC Bioinformatics*. 2013; 14(Supply 3): S9. Published online 2013 Feb 28.

[15] Samta Gupta, Susmita Ghosh Mazumdar, (2013) Sobel Edge Detection Algorithm International Journal of Computer Science and Management Research, Vol 2 Issue 2 February 2013 ISSN 2278-733X

[16] Kavitha, R., Nedunchelian, R., "Domain-specific Search engine optimization using healthcare ontology and a neural network backpropagation approach", 2017, Research Journal of Biotechnology, Special Issue 2:157-166

[17] Dr. C Nalini Sruthy KG2," Data Mining Techniques To Predict Diabetes Disease", International Journal of Pure and Applied Mathematics, Volume 119 No. 12 2018, 10891-10900 ISSN: 1314-3315 (on-line version)

Authors Profile

	<p>N. Sai Kiran Associate Professor Department of Information and Technology QIS college of Engineering and Technology Email: neelam.saikiran534@gmail.com</p>
	<p>B. Deepthi Priya Student Department of Information and Technology QIS college of Engineering and Technology Email: deepthipriyabattula26@gmail.com</p>

	<p>V. Srilekha Student Department of Information and Technology QIS college of Engineering and Technology Email: srilekha1917@gmail.com</p>
	<p>P. Lakshmi Pavani Student Department of Information and Technology QIS college of Engineering and Technology Email: lakshmipavani99494@gmail.com</p>
	<p>P. Manasa Student Department of Information and Technology QIS college of Engineering and Technology Email: manasaparitala14@gmail.com</p>

MESSAGE PASSING OVER CLOUD USING CASCADECIPHERING WITH RANDOMIZED ALGORITHM

G.L.V.Prasad¹, CH.V.Sowmya², N.Divya³, K.Susmitha⁴, T.Manasa⁵

¹Associate Professor, Department of Information
Technology , QIS College of Engineering and Technology,
Ongole

^{2,3,4,5} IV B.TechStudents , Department of Information
Technology , QIS College of Engineering and Technology,
Ongole

ABSTRACT

The multiple encryption is a process to encrypt the data multiple times using the same algorithm or different algorithms. This technique is widely used due to its feature of enhanced security for data communication over the vulnerable wireless network as the internet. The concept of multiple encryption can be described as a technique to provide multilayer and multi-level security over unreliable wireless network during communication. In this project, we propose a computationally efficient multi-level encryption framework that combines the strength of symmetric, the encryption algorithm AES (Advance Encryption Standard), MD5, SHA129, SHA 256, Whirlpool, Blowfish. The framework was evaluated and compared to a set of benchmark symmetric encryption algorithms, such as RC5, DES, and 3-DES. It also provides the analysis of data security issues and privacy protection affairs related to cloud computing by preventing data access from unauthorized users during data transmission, managing sensitive data, providing accuracy and consistency of data transmission.

Keywords: Symmetric, Multiple algorithm, AES, SHA, MD5, Whirlpool, Blowfish.

INTRODUCTION

Cloud Computing is the liberation of computing services such as servers, networking, analytics, storage, intelligence, database, software and more, over the other cloud. When an industry buys in third-party cloud

services as either a public or hybrid cloud offering, it is likely they will not be provided with a full services description, giving details how the platform works, and the security processes the seller operates. This lack of service precision makes it difficult for customers to wisely evaluate whether their data is being stored and processed securely at all times. Surveys have shown that around 85% of IT managers are slightly confident that company data is being stored securely by their cloud storage. Using public or hybrid cloud assistance/discounts can expose a business to security vulnerabilities caused by other users of the same cloud infrastructure.

The liability is upon the cloud vendor to see that this does not occur, yet no vendor is perfect. It is almost possible that a security issue caused by another client system in the same cloud will affect every other client. Cloud encryption is the development of encoding or converting data before it's transferred to cloud storage. Encryption uses mathematical algorithms to transform data (plaintext), file, may it be a text, code or image, to an unreadable form (ciphertext) that can mask it from unauthorized and malicious users. It is the easiest and most vital way to make sure that cloud data can't be stolen, breached, and read by someone with a bizarre motive. Cloud storage providers encrypt data and give encryption keys to the users. These keys are used to securely decipher data when required. Decryption modifies the concealed data back into readable data.

There are two methods used to code and decode data, and these methods are universally evolved as the field of information technology changes its ways of data protection and privacy security. These are also called encryption algorithms. These methods are as follows:

1. Symmetric algorithm

In this system, encryption and decryption keys are the same, which makes it best for closed systems and individual users. These keys are used for reliable communication. This is also known as the secret key algorithm

and is usually used for major data encryption. This is doubtlessly and rapidly performed by hardware and faster than the asymmetric method.

2. Asymmetric algorithm

In this method, two keys are used (private and public) and they are mathematically associated together. It is called asymmetric as they keys/keys are coupled with each other, but they are not the same. The private key must be kept concealed and secret, yet the public key can be shared with anyone.

In symmetric key algorithms, the same key/lead may be used to encipher /decipher a plain text. Subsequently, these algorithms are simple and computationally efficient. Examples of such algorithms: AES, SHA 128, SHA 256, MD5, Whirlpool, Blowfish. On the other hand, asymmetric encryption algorithms need two keys for encryption and decryption, thereby making the generation of these keys computationally high and counterproductive for encrypting large data. For instance, RSA is a n Asymmetric encryption algorithm. With the advancement of technology and computational power, hackers have developed different models to ambush these primary encryption algorithms. One of the outcomes to deal with this problem is to build a synthesis model that consists of multi-level encryption algorithms.

The ideology here is to combine the strength of multiple basic encryption algorithms together to build a complex, sophisticated encryption technique. Two challenges are facing these techniques: computation efficiency and security. The initial one is needed when sending massive data and the latter ensures that the combination between these steps will lead into a complete system that is secured. Many multi-level encryption approaches were put- forward in literature, but they were lagging from computational or security problems or being specified for a certain type of networks.

In this paper, we propose a multi-level encryption framework to encrypt the file based on its size that is secure and computationally efficient. The framework comprises using a symmetric encryption algorithm AES, SHA etc. for encrypting and decrypting data.

Most cloud storage suppliers grip the encryption key on behalf of the user, which requires him or her to impetuously trust the company and will not misuse the access to your files, release the key to hackers, or give it to prying government authorities. Furthermore, anyone with physical access to your phone or laptop could easily get to files on the cloud, because most of these facilities leave you registered in by default. A handful of companies with cloud backup, such as iDrive, authorize the user to create their own private key, encrypting data on their personal computer prior uploading it to the cloud. However, this is reserved for highly valuable business-tier cloud backup, but not the average individual user.

Cloud encryption is needed because its main aim is to secure and protect confidential information or file as it is transmitted through the Internet and other computer systems. Best way to analyze an organization's security and privacy status is through the CIA triad. This embraces for Confidentiality, Integrity, and Availability. By conventional method, the platform of information technology the availability of the data and its integrity is well-defined. IT does not provide enough impression on classified data. That is why, companies/institutions/organizations should prefer cloud encryption. Besides, encryption is not just about protection, securing data and its privacy: rather its core, digital data is significantly transferring, and encryption is needed to perform the transmission in a safe way. Users make sure that their particulars and data is secured when transmitted to another user and that the other user is who they intend to provide the data to and not any malicious attackers.

1. Accomplished statistics security of all times

when data or the information is being transferred or secured, that's when encryption works. This is a flawless solution no matter what is being done

with the data. Even if data is weak and liable to vulnerabilities while being transferred from elsewhere to somewhere else. Encryption verifies the security during this process.

2. Protection of privacy

Encryption defends liable data like personal information of every individual user. This enables privacy and inconspicuousness, alleviating chances of surveillance by government agencies, criminals, and cyber attackers.

3. Part of compliance

Encryption is one of the most reliable tools to share and save data as it follows with the restrictions proposed by an organization.

4. Collective devices protection

Several types of communication devices have become essential of our lives nowadays. Transfer of data from one device to another constitute high threat and vulnerability, hence encryption aid in protecting data across multiple devices.

5. Maintains integrity

Hackers benefit from modifying the information by swindling, just by stealing data. It is possible for these hackers to swap and amend ciphered data. However, the recipient of the information has the potential to identify if it is corrupted, allowing for a prompt response and solution to the attack.

Cloud Encryption Challenges Even though we agree that encryption is the best tool for data protection in the cloud, it's better to also acknowledge that there is no absolute approach when it comes to privacy and security, Besides any other tool to conflict vulnerabilities, threats in the information superhighway, there are challenges that an organization or user may face by using encryption. These are as follows:

1. Loss of data

The limitations about encryption: if a user loses track of his decryption keys and has no backup of the data, it's similar to closely losing data and wrecking it.

2. Encryption functions like a password

When operation on encryption to protect devices, files, and disks, the cue is a password that's chosen by the user. Human passwords are effortlessly easy to seize and hack unlike other solutions like AES-256 (Advance Encryption Standard) which involves long random keys.

3. Complexity of encryption

For a daily-basis user, some encryption programs are complex and they may end up using it inappropriately. This could end up as a failure in encrypting the data which they wanted to secure and encrypting data but that they did not want to encode. The synthesis of deciphering extracts the processor time in the computer. The more muddled up of the encoding, the longer it takes to process.

II TYPES OF ALGORITHMS:

1. AES:

The Advanced coding customary (AES) is AN agreement chosen by the U.S. government to fortify classified data. AES is run in package and hardware throughout the planet to cypher protected information. it's crucial for presidency network security, cyber security and electronic information protection. AES could be a kind of bilateral algorithmic program. It applies the same key for each coding and secret writing. AES work with block cipher. AES includes 3 block ciphers: AES-128, AES-192 and AES-256. AES-128 operates on a 128-bit key length to cipher and decode a group of messages, whereas AES-192 operates on AN 192-bit key length and AES-256 a 256-bit key length to cipher and decode messages. Each cipher codes and decodes data in blocks of 128 bits victimization cryptologic keys of 128, 192 and 256 bits, accordingly.

There are a unit ten rounds for 128-bit keys, twelve rounds for 192-bit keys and fourteen rounds for 256-bit keys. A spherical incorporates of many process steps that embrace substitution/replacement, transposition and mixture of the input plain text to rework it into the ultimate yield of cipher text.

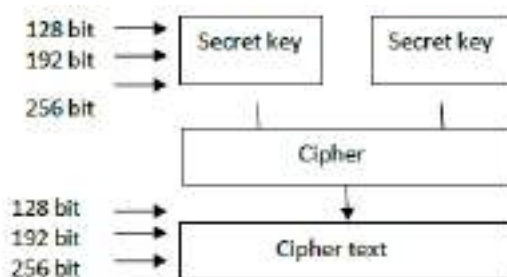


Fig 1 AES algorithm

Each spherical has for transformation.

- Sub bytes
- Shift rows
- Mix columns
- Add spherical keys

2. SHA

In cryptography, SHA-1 (Secure Hash algorithmic program 1) could be a cryptologic hash perform that takes the information And fabricates a 160-bit (20-byte) hash price referred to as a messagedigest – undoubtedly used as an hex variety, forty digits long. Secure Hash Algorithms, conjointly referred to as SHA, is a member of cryptologic functions designed to stay information safe. It works by transfiguring the information employing a hash function: AN algorithmic program includes of bitwise operations, standard additions, and compression functions. The hash perform creates a fixed-size string that appears obscurity about to the original. These algorithms area unit made public to be unidirectional functions, therefore reworkinginto their several hash values, that is much not possible to rework them back to the initialinformation. a couple of algorithms area unitSHA-1, SHA-2, and SHA-3, every of that were linearly designed with sturdy codingin response to hacker attacks. SHA-0, as an example, is currently noncurrent thanks to themanifested vulnerabilities.SHA-256 is one inall the most effective hash functions to SHA-1, and it's one in all the secured and it have well- defined hash functions on the market. SHA- 256 isn't a lot of complicated to code than SHA 128. The 256-bit key makes it a best partner-function for AES.

3. MD5:

MD5 is far faster than different class of message digest that extracts the plain text of 512 bit blocks that is moreover classified into sixteen blocks, every of thirty two bit and produces the 128 bit message digest that could be a set of 4 blocks, every of thirty two bits. MD5 produces the message digest through 5 steps i.e. padding, append length, divide input into 512 bit blocks, initialize chaining variables a method blocks and four rounds, uses contrastive constant it in every iteration.

Step1: Append Artifact Bits

Padding suggests that computing additional bits to the initial message. In MD5 original message is cushiony in order that its length in bits is congruent to 448 modulo 512. Artifact is completed specified the full bits area unit sixty-four less being a multiple of 512 bits length.

Padding is completed though the length of the initial message is already congruent to 448 modulo 512. In Artifactbits, the sole initial bit is one and also the remainder of the bits area unit zero.

Step 2: Append Length

After Artifact, sixty-four bits area unit compact at the top that is employed to document the length of the particular input. Modulo 2^{64} , at this locus, the ultimate message contains a length multiple of 512 bits.

Step 3: Initialize MD buffer

A four-word buffer (A, B, C, D) is operated to enumerate the values for the message digest. Here A, B, C, D area unit 32-bit registers and area unit initialized.

Step 4: process message in 16-word block

MD5 uses the auxiliary functions that confiscates the input as 3 32-bit variety and produces a 32-bit output. These functions employs logical operators like OR, XOR, NOR.

4. Whirlpool

Whirlpool could be a cryptologic hash perform that was derived from sq. and Advanced coding customary. it's a block cipher hash perform and invented once block cipher. It considers but 2^{256} bits length input and converts it into 512 bit hash. {the initial the primary} model of whirlpool is termed Whirlpool-0 and switched to Whirlpool-T once it's first revision in

2001. Consistent with this style the S-box is modified and become easy to use in hardware.

Every block cipher in whirlpool could be a 8*8 matrix. The state of the perform changes in ever spherical by victimization four operations:

- 1) Mix Rows(MR)
- 2) Substitute Bytes(SB)
- 3) Add Spherical Key(AK)
- 4) Shift Columns(sc)

Hash price is calculated by victimization the formula:

State = MR*AK*SC*SB(State)

5. Blowfish

Blowfish is a coding technique designed by Bruce Schneier in 1993 as another to DES coding Technique. it's unco quicker than DES and provides an honest coding rate with no constructive scientific discipline technique found until date. it's one in all the primary, secure block ciphers not subject to any patents and thence honestly on the market for anyone to use.

1. blockSize: 64-bits
2. keySize: 32-bits to 448-bits variable size number of subkeys: eighteen [P-array]
3. number of rounds: sixteen
4. number of substitution boxes: four

Based on the file dimensions any three algorithmic program are determined and cipher it. Once coding, the encrypted file are hoarded within the cloud.

III PROPOSED IMPLEMENTATION

Step 1: Before upload the file, the user have to register for login by entering basic information about the user.



Fig 2 Registration page

Step 2: After register, the user can login and then upload their files which they want to encrypt and store in the Cloud.



Fig 3 Login page



Fig 4 User page

Step 3: After uploading the file, based on the file size randomly three algorithm will be selected by the system and encrypted it by three times by using different algorithms.



Fig 5 Encrypted file

Step 4: During the encryption process, the user will get some secret key values or private key values. If the user wants to decrypt the file they have to enter that secret key. If the user doesn't have the secret key they can't access that file. If the user enters that correct key values that file will be decrypted.



Fig 6 Decrypted file

In this project we developed an algorithm to secure the user data in more efficient way. We have proposed high secured techniques to protect the user data which they want to be upload in the cloud without any third party access. Encryption is regarded as one of the most effective approaches for data security. In this system, we are using more than 3 algorithms which are more efficient for data encryption.

IV PERFORMANCE

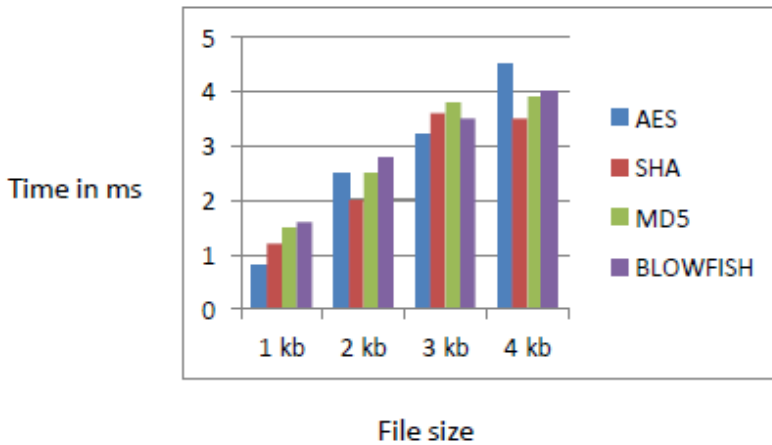


Fig 7 Comparison of Encryption Algorithm

Success of the execution test or analysis is required to evaluate the efficiency and security. Every encryption algorithm has some strength and weaknesses. In this performance analysis it describes the encryption time in milliseconds based on file size.

V CONCLUSION

Cloud computing is the arrangement of assets or administrations given through the web to the clients on their interest by cloud specialist co-ops. Since every single association is moving its information to the cloud, it implies it utilizes the capacity administration given by the cloud supplier. Consequently, it is compulsory to ensure that information is protected against unapproved access, alteration or disavowal of administrations and so forth. Distributed computing can turn out to be safer utilizing cryptographic calculations. Cryptography is the workmanship or study of keeping messages secure by changing over the information into non-intelligible structures. In any case, the current cryptographic calculations are single-level encryption calculations. Digital crooks can undoubtedly break single-level encryption. Consequently, we propose a framework which utilizes staggered encryption and decoding to give greater security to Cloud Storage. As our proposed calculation is a Multi-level cryptographic calculation. Regardless of whether some interloper (unapproved client)

gets the information unintentionally or purposefully, he/she should need to unscramble the information at each level which is a troublesome assignment without a secret key. It is normal that utilizing staggered encryption will give more security to Cloud Storage than utilizing single level encryption. Based on this process the message can send securely in cloud.

REFERENCES


- [1]S. Ahmad, K. M. R. Alam, H. Rahman, and S.Tamura, "A comparison between symmetric and asymmetric key encryption algorithm based decryption mixnets," in Proceedings of the IEEE International Conference on Networking Systems and Security, 2015.
- [2]S.A.M. Diao, M.A.K. Hatem, and M.H. Mohiy(2010). "Evaluating The Performance of Symmetric Encryption Algorithms" International Journal of Network Security,2010, 10(3), pp.213-219
- [3]Priti V. Bhagat, Kaustubh S. Satpute and Vikas R.Palekar "Reverse Encryption Algorithm: A Technique for Encryption &Decryption" International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 2 Issue 1 January 2013,pp 90-95.
- [4]Z. Hercigonja, D. Gimnazija, and C. Varazdin, "Comparative analysis of cryptographic algorithms and advanced cryptographic algorithms," International Journal of Digital Technology & Economy, vol. 1, no. 2, pp. 1 -8, 2016.
- [5]Gagandeepshahi, Charanjitsingh "Cryptography and its two Implementation Approaches" International Journal of Innovative Research in Computer and Communication Engineering ,Vol.1, Issue 3,May 2013,PP 668-672.
- [6]Kuldeep Singh, Rajesh Verma, RitikaChehal "Modified Prime Number Factorization Algorithm (MPFA) For RSA Public Key Encryption"International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231 -2307, Volume-2, Issue-4, September 2012, pp 204-206
- [7]Rostyslav Barabanov, Stewart Kowalski and Louise Yngström, "Information SecurityMetrics", DSV Report series No 11 -007, Mar 25, 2011

[8]Pallavi Vaidya and S. K. Shinde, "Application for Network Security Situation Awareness", in International Conference in Recent Trends in Information Technology and Computer Science (ICRTITCS - 2012), IJCA, ISSN: 0975 – 8887, 2012.

[9]SunJun Liu, Le Yu and Jin Yang, "Research on Network Security Situation Awareness Technology based on AIS", inInternational Journal of Knowledge and Language Processing, ISSN: 2191 -2734, Volume 2, Number 2, April 2011.

[10] Disina, A. H., Pindar, Z. A., & Jamel, S., "Enhanced caeser cipher to exclude repetition and withstand frequency cryptanalysis," Journal of Network and Information Security, 2015.

AUTHORS PROFILE

	<p>G.L.V.Prasad Associate Professor, Department of Information Technology, QIS College of Engineering and Technology Email: glv.prasad19@gmail.com</p>
	<p>K.Susmitha Student Department of Information Technology QIS College of Engineering and Technology Email: susmithasusmi1245@gmail.com</p>



CH.V.Sowmya
Student
Department of Information Technology
QIS College of Engineering
and Technology
Email: ch.guitar@gmail.com



T.Manasa
Student
Department of Information Technology
QIS College of Engineering
and Technology
Email: thotamanasa14@gmail.com



N. Divya
Student
Department of Information Technology
QIS College of Engineering
and Technology
Email: divyanalamothu77@gmail.com

ONLINE VOTING PROCESS FOR GOVERNMENT WITH SECURITY

K. Srinath¹, CH. Akhil², N. Satish³, G. Kalyani⁴, S. Bhanu Prasanthi⁵

¹Associate Professor, Department of Information Technology , QIS college of Engineering & Technology, Ongole

^{2,3,4,5} IV BTech Students, Department of Information Technology , QIS college of Engineering & Technology, Ongole

ABSTRACT

Government activities like Elections are used for the people to choose their candidate or their preferences in a representative democracy or other form of Government. For that purpose, we are developing this application, which can be used to vote from their living location based on the Voter Id using Mobile Services and stores the data in our own database. By using the user id and password citizens can log into the system and collect the necessary information from the home page. For this every user need to provide his Voter Id. By clicking on the submit button they will get their own details.

1. INTRODUCTION

★ Government Elections are used for the people to choose their candidate or their preferences in a representative democracy or other form of Government.

★ Every citizen who are 18 years old and above has right to vote.

★ Thus, we need elections and people choose their candidate by voting. For this we are using Ballot Boxes and EVM machine system.

★ For this we developed a Mobile based Application to vote easily from home.

EXISTING SYSTEM

- Now a days Election Commission are using Ballot boxes and EVM machines for voting purpose.

- People are facing many problems like; they have to go to their native places and have to wait at the polling booth's.

- Not only that, some people are unable to vote because of their busy works and some people are staying far from their native places

Online voting system is an online voting technique. In this system people who are authorized by the admin can cast his/her vote online without going to any physical polling station. There are many voting procedures which are being used for Voting purpose, such as ballot paper, EVM machine but all these procedures require more time and more man power so to eliminate all these drawbacks we provide an online voting system which provides features such as accuracy, convenience, flexibility, privacy and verifiability. Our online voting system provides user a platform where he can register himself to which can occur in tradition voting schemes.

Scope of Study As we all know that there are many organizations that conduct elections for the positions like "Group leader, Project leader, Employee of the month, and for some minor changes in working environment etc. In that case, online voting can very helpful to conduct vote. People can cast their vote from anywhere. As colleges conduct elections for positions like president, vice president etc. for many college societies like CSI, Trinity etc, and other management posts for students and online voting system can be used on any cases like these efficiently it can be customized according to client need on any type of elections.

This is a system that can be used by user to cast vote in an election. All the voters have to login and click on cast vote to his/her chosen candidates to submit his/her vote. The research development and testing are done on LAN. On other hand online voting software is been in research for many years, researched cases of wrong implementations reported in recent years. These factors are need to be resolved so public can cast their vote in a secured and fitting environment. online voting is a voting software in which any user can use his/her voting rights from anywhere. Online voting application contains.

As we are looking at the existing system, they are just providing online voting. As we knew that Government of India contain multiple elections. So we are implementing our system such that voter can select election and submit their vote region/ward wise. After studying existing system we

observed that they are not providing state wise, region wise voting facility. So its difficult to vote because there is no restriction, so voter can also cast his/her vote to those candidate who is not belonging from his/her area. In proposed system we are implementing that voter can cast his/her vote only those candidate who's belonging from his/her region/ward. We will display only those candidate who are belonging from that particular voters ward. So it will also help to conduct small election such as Gram Panchayat Election or Nagar Sevak Elaction. We are making our voting system user friendly.

SCOPE

- i. Increasing number of voters as individuals will find it easier and more convenient to vote.
 - ii. Less effort and less labor intensive, as the primary cost and focus primary on creating, managing, and running a secure web voting portal.
 - iii. The system can be used anytime and from anywhere by the Voters.
 - iv. No one can cast votes on behalf of others and multiple times.
 - v. Saves time and reduces human intervention.
 - vi. The system is flexible and secured to be used.
 - vii. Unique Identification of voter through Aadhar number.
 - viii. Improves voting with friendly Interface.
 - ix. No fraud vote can be submitted
-
- a) users details
 - b) users Names with ID and password.
 - c) users vote in a database.
 - d) sum of total number of votes.
 - e) result panel
 - f) chatbot to help users
 - g) a unique user id given my administration Various operational works proposed in the system information of the user in database.

PROBLEM BACKGROUND

In the recent times there are many literatures on online voting has been developed. While online voting has been an area of research in the recent years, there are efforts made to make online voting system more secure. The use of insecure Internet, and the resulting security Breaches have been reported recently. So, the main issue now to resolve these security breaches such as denial of service attack.

PROBLEM STATEMENT

Our online voting system will make all voting process easy because in this system we will provide chatbot, which will help every user during the voting process. If any user has any kind of issue during the process the chatbot will provide efficient solution for that issue. Our voting system will make the whole voting process cost efficient. Our voting system will give instant and unbiased poll result. Our voting system will help us to keep track of voter. And our system is time efficient.

RESEARCH OBJECTIVE

The main objective of this study is to make a step forward in the direction of online voting platform by providing all the essential security levels. The objective of this study is to make the voting process easy, less time taking, and secure. Online voting system eliminate the bogus voting

LITERATURE SURVEY:

To make the voting process very easy and efficient wireless and web technologies are used. The online- voting system has the possibility of secure, easy and safe way to capture and count the votes in the election. The author in online voting system based on adhaar id" uses adhaar id as key of authentication, system is efficient in terms of time and provides security the system is great improvement over traditional system but the main problem resides in this system is that of authentication, the authentication technique used is not that efficient as biometric is not used. The paper "Secure Authentication for Online Voting System" presents non traceability and integrity of the votes, smart card has been used to avoid

multiple votes casted by users, biometric is being used for authenticating voters. The author has introduced smart card for biometric identification and voter id card to be used at the time of casting vote. They are using smart card and voter id card at the time of election which is not feasible as anything can happen to those cards thus relying completely upon cards is not a good idea. And the use of various cards makes the system costly now each and every voter need to have these additional cards. Also, it may take reasonable amount of time to generate so many cards. All voting system generated priority though have met various features, which a voting system may consist but the main problem one could find in this system is that little "online" word, despite all techniques they have used to make system robust there is always a chance of malpractice when your system is online. In online voting system powered by biometric security "the author has used personal identification number, thumb impression and secret key altogether for authentication of the voter. Techniques such as cover image creation, secret key expansion have been used for securely sending data to server and then further authenticating voters. This system is quite robust; it takes care of authentication as well as security of voter's data stored in server. The main problem with such systems is that despite using various security techniques they won't be able to manage such huge amount of data by user. deletion of wrong information. Each information is submitted to administration.

PRODUCT PERSPECTIVE

The product is an election conducting tool with a simple GUI and a Chatbot in it. This system is developed using php. Though product is stand-alone. It requires a XAMPP server. Our is a voting system by which any Voter can use his/her voting rights from anywhere. And for the smooth processing of voting system has an integrated chatbot. It can guide users at any stage of process to make accessibility easy

PROPOSED SYSTEM

The online voting system will manage the voter information by which voter can login and use his voting rights. There is a data base which

is maintained by the ELECTION COMMISSION OF INDIA complete data of voter with complete information is stored at the time of registration voter will be asked.

Full name, age, Aadhar card no, mobile no. email id, finger prints and verified the details by administrator. At the time of requesting vote, voter will be asked to enter his Aadhar id. Then voter will be authenticated, and he can give vote from one of the candidates from the list. If voter already has AADHAR Id then he/she don't need to register, else before voting he/she need register himself/herself in AADHAR database.

SYSTEM ARCHITECTURE



ELECTRONIC VOTING SYSTEMS

The Council of Europe recommendations defined electronic voting (e-Voting) as “the use of electronic means in at least the casting of the vote” (Krimmer, et al., 2007). Electronic voting is a term encompassing several different types of voting, embracing both electronic means of casting a vote and electronic means of counting votes. Electronic voting systems are complex distributed systems, whose components range from general-purpose PCs to optical scanners and touch-screen devices, each running some combination of commercial off-the-shelf components, proprietary firmware, or full-fledged operating systems.¹² It is a fundamental demand of countries to enhance their election system . Now due to rapid emergence of technologies in computer and

telecommunication world e-Voting based systems are to be introduced that lessens all the traditional manual election systems' problems. With the introduction of e-Voting systems our elections processes and social lives are going to be easy, efficient and low-cost. Now in this system voters can cast their votes from anywhere in world. E- voting system must meet security requirements such as confidentiality, integrity, fairness, forgery attack, verifiability and so on.

The paper "Secure Authentication for Online Voting System" presents non traceability and integrity of the votes, smart card has been used to avoid multiple votes casted by users, biometric is being used for authenticating voters. The author has introduced smart card for biometric identification and voter id card to be used at the time of casting vote. They are using smart card and voter id card at the time of election which is not feasible as anything can happen to those cards thus relying completely upon cards is not a good idea. And the use of various cards makes the system costly now each and every voter need to have these additional cards. Also it may take reasonable amount of time to generate so many cards. All voting system generated priority though have met various features, which a voting system may consist but the main problem one could find in these system is that little "online" word, despite all techniques they have used to make system robust there is always a chance of malpractice when your system is online.

In "online voting system powered by biometric security" the author has used personal identification number, thumb impression and secret key altogether for authentication of the voter. Techniques such as cover image creation, secret key expansion have been used for securely sending data to server and then further authenticating voters. This system is quite robust; it takes care of authentication as well as security of voter's data stored in server. The main problem with such systems is that despite using various security techniques they won't be able to manage such a huge amount of data that they may encounter during election periods their system is online and they may face congestion during casting votes

The author in online voting system based on adhaar id "uses adhaar id as key of authentication, system is efficient in terms of time and provides

security the system is great improvement over traditional system but the main problem resides in this system is that of authentication, the authentication technique used is not that efficient as biometric is not used..

INTERNET VOTING SYSTEM

Internet voting is defined as an election system that utilizes the internet to ensure access to a website or domain which makes use of electronic ballots. In this regard, the electronic ballots allow voters to transmit their voted ballot to election officials over the Internet. It is a secure system that allows for eligible voters to cast their votes from any (remote)location. Features include:

- Highly secure platform
- Supports multiple languages
- Simple and user-friendly voting interface.
- Support multiple voting channels: web browsers; mobile phones and PDA

Merits of online voting system

Electronic Ballots:

Electronic voting systems may use electronic ballots to store votes casted in computer memory. Systems which use them exclusively are called DRE Voting Systems. When electronic ballots are used,exhaustion of supplied ballots never arise. Additionally, the need to print paper ballots are eliminated which is often a major chunk of the costs incurred. When administering elections in which ballots are required by law to be offered in multiple languages, electronic ballots can be programmed to provide the ballots in multiple languages for a single

Lower Costs:

A major edge over the paper – based voting systems, electronic voting systems save money by reducing personnel expense, management overheads and the overall costs incurred. It should be noted though that the initial cost of building the electronic voting system may be high but it is on the long term cheaper as it is only built once in a star. Contrast to the

paper – based voting system where ballots would have to be printed for every election even if the elections were cancelled as seen in the recent cancellation of the Nasarawa State Governorship Bye – Election as scheduled based on a court judgment, with less than 2 weeks to the election. Thus, it has rendered all the paper ballots printed totally useless and the costs incurred on logistics all wasted.

Convenience and Accessibility:

With a well – designed and user – centered Electronic Voting System, the voters can be guaranteed to cast their votes with ease and minimal technical skills needed. It is also possible to design a variety of ballot question formats to suit the different categories of voters that might use the system. It can also be fitted with adaptive technologies in order to provide accessibility for the voters with a form of disability. And voter participation would be increased as a direct result of enhancing the ease of use and accessibility of the voting system.

Efficient Results Collation:

With an electronic voting system, results of an election can be counted and displayed immediately. And it is particularly useful in the event that there is a very large amount of votes to be counted, which can be very tedious, error-prone and time –consuming if the ballots are counted by hand. The time used to collate and declare results is reduced significantly using electronic means.


Elimination of Duplicate or Multiple Voter Registrations:





The deployment of biometric integration, specifically fingerprint recognition, which leverages on the fact that no two persons possess the same features for a given biometrics characteristics, can be used to eliminate multiple voter registration and laid the foundation for the enshrinement of the One.

References:

- [1] Android Developer Guide:
<http://developer.android.com/guide/index.html>
- [2] Android API: <http://developer.android.com/reference/packages.html>
- [3] Android Fundamentals:
<http://developer.android.com/guide/topics/fundamentals.html>.
- [4] The Java Tutorials:
<http://downloadlnw.oracle.com/javase/tutorial/index.html>.
- [5] Android Native Development Kit:
<http://developer.android.com/sdk/ndk/index.html>
- [6] Android User Interfaces:
<http://developer.android.com/guide/topics/ui/index.html>.
- [7] Common Tasks: <http://developer.android.com/guide/appendix/faq/com montasks.html>.
- [8] Sample Source Code:
<http://developer.android.com/resources/samples/get.html>.

Authors profile:

	<p>K. Srinath Associate Professor Department of Information Technology QIS College of Engineering and Technology Email: srinathits@gmail.com</p>
---	--

	<p>N. Satish Student Department of Information Technology QIS College of Engineering and Technology Email: nethisatish99@gmail.com</p>
	<p>CH. Akhil Student Department of Information Technology QIS College of Engineering and Technology Email: akck1114@gmail.com</p>
	<p>S. Bhanu Prasanthi Student Department of Information Technology QIS College of Engineering and Technology Email: prasanthisatyavolu1999@gmail.com</p>
	<p>G. Kalyani Student Department of Information Technology QIS College of Engineering and Technology Email: gkalyanichowdary210@gmail.com</p>

INTEGRATING SECURE NETWORK CODING TECHNIQUES TO STORE THE DATA SECURELY IN CLOUD WITH DATA DYNAMICS

Dr.R.SuneethaRani¹, D.Goutham², D.Gopibabu³, K.Anilkumar⁴,
R.BV.Saikumar⁵, V.Aravindh⁶

¹Associate Professor & HOD, Department of Information Technology,
QIS college of Engineering & Technology, Ongole.

^{2,3,4,5,6}Students, Department of Information Technology, QIS college of
Engineering & Technology, Ongole.

Abstract:

Cloud service providers offer stockpiling re-appropriating office to their customers. In a secure cloud stockpiling (SCS) convention, the respectability of the customer's information is kept up. In this work, we develop an openly evident secure cloud stockpiling convention in view of a secure organization coding (SNC) convention where the customer can refresh the rethought information on a case by case basis. To the best of our insight, our plan is the primary SNC- based SCS convention for dynamic information that is secure in the norm model and gives protection saving reviews in a freely undeniable setting. Besides, we examine, in subtleties, about the (im)possibility of giving an overall development of an productive SCS convention for dynamic information (DSCS convention) from a subjective SNC convention. Also, we change an existing DSCS conspire (DPDP I) to help privacy preserving reviews. We additionally contrast our DSCS convention and different SCS plans (counting the adjusted DPDP I conspire). At last, we sort out certain impediments of a SCS conspire built utilizing a SNC convention

1 Introduction

CLOUD stockpiling is as a rule broadly embraced because of the fame of cloud registering. Notwithstanding, ongoing reports [1], [2] demonstrate that information misfortune can happen in cloud stockpiling providers (CSPs). Consequently, the issue of checking the honesty of the information in cloud stockpiling, which we alluded to as secure cloud stockpiling (SCS), has pulled in a great deal of consideration. On the other hand, organizing coding, which was proposed to improve the organization

limit, likewise deals with the issue of trustworthiness checking. A moderate switch may deliberately dirty codewords, which brings about deciphering disappointments at the endpoints. Checking the trustworthiness of codewords is alluded to as the secure organization coding issue. Various scientists have examined secure cloud stockpiling and secure organization coding autonomously. Answers for the previous issue, e.g., [3], [4], [5], were proposed as of late. Conversely, the last territory has been analyzed for more than ten years, e.g., [6], [7]. Secure cloud stockpiling. This issue was first proposed by Juels and Kaliski [3] and Ateniese et al. [4]. Two principle elements are engaged with these conventions: a client and a cloud stockpiling supplier. A client re-appropriates the information to the cloud who vows to store the information. The client at that point affirms the information respectability by associating with the cloud utilizing a secure cloud capacity convention. The inspiration of information trustworthiness checking lies in a few elements. To start with, because of the helpless administration of the cloud, the client's information could be lost because of framework disappointments (equipment or programming). To cover the mishap, the cloud may decide to mislead the client. Second, the cloud has a gigantic monetary motivating force to dispose of the information which is once in a while gotten to by the client. Disregarding some piece of the information makes a difference the cloud to diminish its expense. Third, a cloud could likewise be hacked and the information could be altered. Fourth, a cloud may carry on malevolently on account of different conceivable government pressures. Without a secure cloud stockpiling convention, the event of these episodes might be covered up by the cloud and gone undetected. The primary element of a secure cloud stockpiling convention is that the client can check the information uprightness without having the real information. Conventional procedures dependent on hash, message verification codes (MACs), and advanced marks anyway require the client to store the information locally. A few conventions (e.g., [5], [8], [9], [10]) are freely unquestionable, i.e., anybody other than the client can check the information uprightness; other conventions are secretly unquestionable since just the client with the mysterious key can check the information

uprightness. A more definite overview is conceded to Section 2. Secure organization coding. This issue was first proposed by Cai and Yeung [6] and Gkantsidis and Rodriguez [7]. Organization coding is a steering worldview where a switch in the network conveys encoded information parcels, which are an element of got information bundles, rather than the customary store-and-forward approach. Encoding can build the organization limit with respect to multicast assignments. Direct coding, in which a switch conveys a straight mix of got information bundles, is end up being adequate to accomplish the expanded limit [11], [12]. This is particularly helpful in agreeable networks. Notwithstanding, this worldview likewise causes extreme security concerns. On the off chance that an encoded bundle is altered unlawfully, this adjustment can immediately spread to the entire organization on the grounds that a switch encodes every got bundle, including the contaminated ones. This assault is otherwise called the contamination assault. The contamination of the codewords could bring about information misfortune when information collectors endeavor to unravel the information. Accordingly, at the point when security is a basic worry in an organization coding empowered organization, the information beneficiaries and switches need to check whether an information parcel is contaminated. Basically, the secure organization coding issue is likewise a kind of information uprightness checking issue. Numerous arrangements are embraced from customary information respectability procedures, for example, cryptographic hash capacities [13], MACs [14] or advanced marks [15], [16]. The fundamental thought is that each codeword in the organization should be validated by checking if the codeword has been adjusted illicitly. The test is that the parcels in the organization are straightly joined by switches furthermore, the new bundles likewise should be confirmed. All current answers for secure organization coding depend on a few homomorphic property of the fundamental cryptographic methods. Organization coding procedure [2, 21] fills in as an option in contrast to the traditional store-and-forward directing strategy utilized in a correspondence organization. In an organization coding (NC) convention, each middle hub (all hubs aside from the source and target hubs) in the organization

joins the approaching bundles to yield another parcel. The organization coding conventions appreciate significantly better throughput, productivity and adaptability contrasted with essentially transferring an approaching bundle for what it's worth. Nonetheless, these conventions are defenseless to contamination assaults brought about by a pernicious moderate hub that infuses invalid parcels in the organization. These invalid bundles produce all the more such parcels downstream. In the most noticeably awful case, the objective hub can't unravel the first document sent to it by means of the organization. Secure organization coding (SNC) conventions give countermeasures to determine this issue utilizing some cryptographic natives. In a SNC convention, the source hub verifies every one of the bundles to be sent through the organization. For the verification of the parcels, a little tag is connected to each packet. In a new work, Chen et al. [14] investigate the connection between a secure cloud stockpiling (SSCS) convention for static information and a secure organization coding (SNC) convention. They show that, given a SNC convention, one can build an SSCS convention utilizing the SNC convention. Be that as it may, for static information, the customer (information proprietor) can't play out any update (inclusion, cancellation or alteration) effectively on her information after she transfers them to the cloud server. This limitation makes a SSCS convention lacking in many cloud applications where a customer needs to refresh her information habitually. Clearly, a guileless method to refresh information in this situation is to download the entire information record, play out the necessary updates what's more, transfer the record to the server once more; however this strategy is profoundly wasteful as it requires gigantic measure of transfer speed for each update. In this way, further examinations are required towards an effective (and more nonexclusive) development of a secure cloud stockpiling (DSCS) convention for dynamic information utilizing a SNC convention.

2. Related Work

Verification of retrievability was proposed Juels and Kaliski [3] to empower a customer to confirm if the information moved to the cloud is

whole. The essential thought is that the client implants a few uncommon confirmation data (i.e., "sentinels" [3]) in the information at sporadic positions. Albeit the confirmation data isn't identified with the information and is produced haphazardly, the cloud doesn't have the foggiest idea about the particular places of them. The client can complete the evaluating by asking the cloud to send back the information at some arbitrary positions, which either contain the uncommon validation information or the client's typical information. Be that as it may, one disadvantage of this approach is that the absolute number of the "sentinels" is limited; accordingly, reviewing must be done a limited number of times. Ateniese et al. [4] proposed the possibility of provable information ownership which utilizes homomorphic verification information. Generally, calculation should be possible on a gathering of information blocks, with the end goal that another authenticator can be registered from similar calculation on their verifications. The client can at that point reviews the cloud stockpiling by requesting that the cloud send back some calculation of the haphazardly picked information blocks also, a verification of the registered outcome. In the event that the confirmation is right, the cloud stores the client's information unblemished. To make the inquiry and reaction smaller, two conventions in view of pseudorandom capacities and a blending based mark individually were proposed with formal security verifications [8]. One of them is reached out to help privacy preserving outsider evaluating [5]. A comparable convention which likewise uses blending is subsequently proposed [10]. In light of a unique responsibility convention, a convention which plans to diminish the correspondence cost is proposed [9]. There is additionally some fascinating work dependent on number-hypothetical hash capacities [17]. All the above conventions are planned in an advertisement hoc way. Interestingly, we give a precise and nonexclusive configuration approach.

3. Secure Cloud Storage

System Model, Threat Model, and Design Goals We model a secure cloud storage system as demonstrated in Fig. 1. There are two elements: client and cloud. In practice, a client could be an individual, a company, or

an organization utilizing a PC or a cell phone, and so forth; a cloud could be any CSP, e.g., Amazon S3, Dropbox, Google Drive, and so on. The client initially

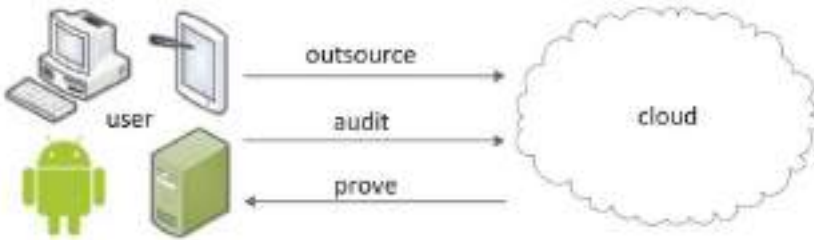


Fig. 1. A secure cloud storage system.

re-appropriates its data to the cloud. Later, the client periodically plays out an audit on the respectability of re-appropriated data. The client can then check whether the verification got back from the cloud is valid or not, meaning that the data remains intact, or obtaining a proof that the data has been tampered which will conceivably cause some further action (which is out of our scope), like legal action or data recuperation. Similar to past work [3], [4] and as motivated earlier in this paper, we model the cloud as potentially malicious. We assume the communication between the client and the cloud is authenticated, which can be finished by standard strategies. Subsequently, we can concentrate on the client and the cloud however not communication. A secure cloud storage system that enables a client to check the respectability of the rethought data is required to be:

- i. Correct. On the off chance that the cloud undoubtedly stores the entire re-appropriated data, the cloud can always demonstrate to the client that the data remains intact.
- ii. Secure. On the off chance that the client's data is damaged, the client can identify with high probability in the audit inquiry, even on the off chance that the cloud attempts to cover the occasion.
- iii. Efficient. The computation, storage, and communication cost of both the client and the cloud ought to be as small as conceivable.

High-Level Protocol Specification Presently we abstract a framework for the secure cloud storage issue. Clearly, to confirm whether the cloud lies to an audit inquiry, the client needs to have some restricted data on its side which is processed according to a certain security level.

Signify the restricted intel by K , the security level by a whole number λ , and the client's data by F . The client utilizes K to deal with F . The prepared data, indicated by F_0 , contains authentication information and is at that point moved to the cloud. On getting an audit question q from the client, the cloud utilizes the put away data F_0 to generate a verification G showing that the data is intact. The client at that point checks regardless of whether G is valid. Indicate the client's verification result by d . All the more specifically, a secure cloud storage protocol contains five productive algorithms KeyGen; Outsource; Audit; Demonstrate; Verify The prepared data contains some authentication information of the data F and is then shipped off the cloud.

Understanding Security

This subsection presents a reasonable security meaning of the secure cloud storage protocol by abstracting its realworld usage bit by bit. The way to understand the security is to characterize the meaning of security exactly. To start with, we need to understand the capability of a malicious cloud. In practice, the cloud has the prepared data F_0 Other than that, the cloud can see a great deal of audit inquiries and its confirmation reactions. It is also reasonable that the cloud can know if the client accepts a proof reaction. This is because if the client dismisses the evidence, the client may sue the cloud or follow some other cure actions; if the client accepts the verification, there are no such actions. Another important issue is the number of audit questions and verification results the cloud can get. Our definition allows the malicious cloud to see polynomially many (in security parameter) such questions and verification results; which can cover the client's periodical audit in practice

4. DPDP I: A Dynamic Provable Data

Possession Scheme Erway et al. [18, 17] propose two productive and completely dynamic provable data possession schemes: DPDP I (based on rank-based authenticated skip records) and DPDP II (based on rank-based RSA trees). We consider just the DPDP I scheme here.

Blockless Verification in DPDP I

May there be a key generation algorithm KeyGen that delivers a public key $pk = (N, g)$, where $N = pq$ is an item of two large primes and g is a component of Z_N with large request. Assume the initial data record comprises of \tilde{m} blocks $b_1, b_2, \dots, b_{\tilde{m}}$. For each square b , the customer processes a tag $T(b) = g^b \pmod N$. Presently, the customer fabricates a rank-based authenticated skip list M on the tags of the squares and uploads the data, tags and the skip rundown to the cloud server. The inclusion, erasure and modification operations are performed in a similar fashion as examined. There is no secret key associated with the DPDP I scheme. Although Erway et al. try not to claim unequivocally the public verifiability of the DPDP I scheme, we see that the scheme can be made openly verifiable by just making the metadata dM of the forward-thinking skip list and the value \tilde{m} public. During an audit, the verifier chooses I , a random l -component subset of $\{1, 2, \dots, \tilde{m}\}$, and generates a challenge set $Q = \{(i, v_i)\}_{i \in I}$, where each v_i is a random value. The verifier sends the challenge set Q to the server.

Modified DPDP I to Make Audits Privacy Preserving

The secure cloud storage scheme for dynamic data examined offers privacy- safeguarding audits where an outsider auditor (TPA) cannot learn about the actual data while auditing. Allow us to investigate whether the scheme DPDP I gives this facility. As in the original scheme [18] the server sends the aggregated square $B = \prod_{i \in I} v_i b_i$ to the verifier (or TPA) where $|I| = l$. Presently, a TPA can obtain the b_i values by settling a system of linear equations. Hence, the audits in the original scheme are not privacy-protecting. Notwithstanding, it isn't hard to make these audits privacy-protecting. We adjust the systems associated with an audit as follows. As previously, the verifier sends the challenge set $Q = \{(i, v_i)\}_{i \in I}$ to the server. The server processes an aggregated square $B = \prod_{i \in I} v_i b_i$. Presently, the server picks a random value r , and it processes $B_0 = B + r$ and $R = g^r \pmod N$. The server sends $\{T(b_i)\}_{i \in I}, B_0, R$ and verifications $\{\Pi(i)\}_{i \in I}$ to the verifier.

5. Performance Analysis

In this segment, we examine about the proficiency of our DSCS protocol and compare this scheme with other existing SCS protocols achieving provable data possession guarantees. We also recognize a few limitations of a SNC-based SCS scheme (for static or dynamic data) compared to the DPDP I scheme.

Efficiency

The computational expense of the algorithms in our DSCS protocol is dominated by the expense of exponentiations (modulo N) required. To generate the value x in an authentication tag for each vector (in the algorithm *Outsource*), the customer has to play out a multi-exponentiation and calculate the e -th base of the outcome. The server requires two multi-exponentiations to calculate the value of x . To check a proof utilizing the algorithm *Verify*, the verifier has to play out a multi-exponentiation and a solitary exponentiation. As referenced, because of the properties of a skip list, the size of each verification Π (related to the rankbased authenticated skip list), the time needed to generate Π and the time needed to confirm Π are $O(\log m)$ with high probability.

Comparison among PDP Schemes

As our DSCS protocol gives provable data possession (PDP) guarantees, we compare our scheme with some other PDP schemes found in the literature. Presently, we examine about a couple of limitations of our DSCS protocol compared to DPDP I (specifically), since the two of them are secure in the standard model, handle dynamic data and offer public verifiability. In the DSCS protocol, the audits are privacy-safeguarding, that is, an outsider auditor (TPA) cannot gain information on the data actually put away in the cloudserver.

6. Conclusion:

In this work, we have proposed a DSCS protocol based on an SNC protocol. To the best of our knowledge, this is the first SNC-based DSCS protocol that is secure in the standard model, enjoys public verifiability and offers privacy-preserving audits. We have also discussed about some

properties an SNC protocol must have such that an efficient DSCS protocol can be constructed using this SNC protocol. We have modified an existing DSCS scheme (DPDP I [18]) to make its audits privacy-preserving. We have analyzed the efficiency of our DSCS construction and compare it with other existing secure cloud storage protocols achieving the guarantees of provable data possession. Finally, we have identified some limitations of an SNC-based secure cloud storage protocol. However, some of these limitations follow from the underlying SNC protocols used. A more efficient SNC protocol can give us a DSCS protocol with a better efficiency.

7. References

- [1] S. Agrawal and D. Boneh. Homomorphic MACs: MAC-based integrity for network coding. In *Applied Cryptography and Network Security - ACNS 2009*, pages 292–305, 2009.
- [2] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.
- [3] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, L. Kissner, Z. N. J. Peterson, and D. X. Song. Provable data possession at untrusted stores. In *ACM Conference on Computer and Communications Security, CCS 2007*, pages 598–609, 2007.
- [4] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik. Scalable and efficient provable data possession. In *International Conference on Security and Privacy in Communication Networks, SECURECOMM 2008*, pages 9:1–9:10, 2008.
- [5] N. Attrapadung and B. Libert. Homomorphic network coding signatures in the standard model. In *Public Key Cryptography - PKC 2011*, pages 17–34, 2011.
- [6] M. Bellare and P. Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *ACM Conference on*

Computer and Communications Security, CCS 1993, pages 62–73, 1993.

[7] D. Boneh, D. M. Freeman, J. Katz, and B. Waters. Signing a linear subspace: Signature schemes for network coding. In *Public Key Cryptography - PKC 2009*, pages 68–87, 2009.

[8] K. D. Bowers, A. Juels, and A. Oprea. HAIL: A high-availability and integrity layer for cloud storage. In *ACM Conference on Computer and Communications Security, CCS 2009*, pages 187–198, 2009.

[9] K. D. Bowers, A. Juels, and A. Oprea. Proofs of retrievability: Theory and implementation. In *ACM Cloud Computing Security Workshop, CCSW 2009*, pages 43–54, 2009.

[10] D. Cash, A. Kulkarni, and D. Wichs. Dynamic proofs of retrievability via oblivious RAM. In *Advances in Cryptology - EUROCRYPT 2013*, pages 279–295, 2013.

[11] D. Catalano, D. Fiore, and B. Warinschi. Efficient network coding signatures in the standard model. In *Public Key Cryptography - PKC 2012*, pages 680–696, 2012.

[12] N. Chandran, B. Kanukurthi, and R. Ostrovsky. Locally updatable and locally decodable codes. In *Theory of Cryptography Conference, TCC 2014*, pages 489–514, 2014.

[13] D. X. Charles, K. Jain, and K. E. Lauter. Signatures for network coding. *International Journal of Information and Coding Theory*, 1(1):3–14, 2009.

[14] F. Chen, T. Xiang, Y. Yang, and S. S. M. Chow. Secure cloud storage meets with secure network coding. In *IEEE Conference on Computer Communications, INFOCOM 2014*, pages 673–681, 2014.

[15] R. Curtmola, O. Khan, R. C. Burns, and G. Ateniese. MR-PDP:

multiple-replica provable data possession. In IEEE International Conference on Distributed Computing Systems - ICDCS 2008, pages 411–420, 2008.

[16] Y. Dodis, S. P. Vadhan, and D. Wichs. Proofs of retrievability via hardness amplification. In Theory of Cryptography Conference, TCC 2009, pages 109–127, 2009.

[17] C. C. Erway, A. Kupcuk, C. Papamanthou, and R. Tamassia. Dynamic provable data possession. In ACM Conference on Computer and Communications Security, CCS 2009, pages 213–222, 2009.

[18] C. C. Erway, A. Kupcuk, C. Papamanthou, and R. Tamassia. Dynamic provable data possession. ACM Transactions on Information and System Security, 17(4):15, 2015.






[19] R. Gennaro, J. Katz, H. Krawczyk, and T. Rabin. Secure network coding over the integers. In Public Key Cryptography - PKC 2010, pages 142–160, 2010.

[20] M. T. Goodrich, R. Tamassia, and A. Schwerin. Implementation of an authenticated dictionary with skip lists and commutative hashing. In DARPA Information Survivability Conference and Exposition (DISCEX) II, pages 68–82, 2001.

[21] T. Ho, R. Koetter, M. Médard, D. R. Karger, and M. Effros. The benefits of coding over routing in a randomized setting. In IEEE International Symposium on Information Theory - ISIT 2003, page 442, 2003.

[22] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. IEEE Transactions on Information Theory, 52(10):4413–4430, 2006.

AUTHORS PROFILE

	<p>Dr.R.Sunitha Rani Associate Professor and HOD, Department of Information Technology QIS College of Engineering and Technology, Ongole</p>
	<p>V.Aravind Student Department of Information Technology QIS College of Engineering and Technology, Ongole</p>
	<p>K. AnilKumar Student Department of Information Technology QIS College of Engineering and Technology, Ongole</p>
	<p>R. B. V. SaiKumar Student Department of Information Technology QIS College of Engineering and Technology, Ongole</p>
	<p>GopiBabu Student Department of Information Technology QIS College of Engineering and Technology, Ongole</p>



D. Gowtham
Student
Department of Information Technology QIS
College of Engineering and Technology,
Ongole

**CDA GENERATION AND INTEGRATION FOR HEALTH
INFORMATION EXCHANGE BASED ON CLOUD COMPUTING
SYSTEM- UNDER CYBER SECURITY.**

**Dr. G.Lakshmi Vara Prasad¹, K.Ganesh², V.Amarnath³, N.Manideep⁴,
K.Sandeep⁵**

¹Associate Professor, Department of Information Technology ,QIS College
of Engineering & Technology, Ongole.

^{2,3,4,5} IV B.Tech Students, Department of Information Technology ,QIS
College of Engineering & Technology, Ongole.

ABSTRACT

Security is biggest problem in cloud computing. Successful deployment of Electronic Health Record helps improve patient safety and quality of care, but it has the prerequisite of interoperability between Health Information Exchange at different hospitals. The Clinical Document Architecture (CDA) developed by HL7 is a core document standard to ensure such interoperability, and propagation of this document format is critical for interoperability. Unfortunately, hospitals are reluctant to adopt interoperable HIS due to its deployment cost except for in a handful countries. A problem arises even when more hospitals start using the CDA document format because the data scattered in different documents are hard to manage. In this paper, we describe our CDA document generation and integration service based on cloud computing, through which hospitals are enabled to conveniently generate CDA documents without having to purchase proprietary software. Our CDA document integration system integrates multiple CDA documents per patient into a single CDA document and physicians and patients can browse the clinical data in chronological order. In order to enhance the security of network communication, the WAN adopts the 128-bit. Advanced Encryption Standard (AES-128) and utilizes two session keys: network session key and application session key, for encrypting/decrypting data between end devices and network/application servers. The proposed method provides the features of mutual authentication, confidentiality and message integrity. We preventing session hijacking, sql injection and bruteforce attack.

1. INTRODUCTION

Cloud security is the protection of data stored online via cloud computing platforms from theft, leakage, and deletion. Methods of providing cloud security include firewalls, penetration testing, obfuscation, tokenization, virtual private networks (VPN), and avoiding public internet connections. Electronic Health Record (EHR) is longitudinal assortment of electronic health data for and concerning persons, wherever health data is outlined as data touching on the health of a person or health care provided to a person and it will support of economical processes for health healthcare delivery. CDA (Clinical Document Architecture) by Health Level Seven could be a major normal for clinical documents. CDA could be a document markup normal that specifies the structure and linguistics of 'clinical documents' for the aim of exchange. The primary version of CDA was developed in two001 and unleash 2 came enter 2005. several CDA-based comes are with success completed in several countries . Active works square measure being done on up linguistics ability supported open HER. to confirm ability of pelt along, the amount of HIS that supports CDA must be sufficiently giant. However, the structure of CDA is extremely complicated and therefore the production of correct CDA document is difficult to realize while not deep understanding of the CDA normal and enough expertise with it. additionally, the HIS development platforms for hospitals vary thus greatly that generation of CDA documents in every hospital invariably needs a separate CDA generation system. Also, there's a resistance towards new systems unless it's completely necessary for provision of care. As a result, the adoption rate of EHR is extremely low aside from a couple of handful countries like New Seeland or Australia. The US Government runs the significant Use Program to enhance potency in attention and patient safety. This program was launched as a section of incentives to boost the EHR adoption rate for HER adopting hospitals. Even inside a town, a patient might head to totally different hospitals, and the personal medical records are distributed across every hospital. Medical errors may be avoided if the connected medical data of the patient may be retrieved properly and with efficiency. Thus the effective communication and

availability of medical data among hospitals is associate degree important issue. In recent years, there are plenty of researches on totally different architectures for sharing medical records supported web technologies. so as to reuse the sharable data, several medical standards are proposed for the sleek exchange of electronic medical records following the particular standards. Health Level Seven (HL7) organization has worked hard to produce a comprehensive framework for the exchange and sharing of clinical data (such as discharge summaries and progress notes). Clinical Document Architecture (CDA), Release, became associate degree yankee National Standards Institute (ANSI)— approved HL7 standard in two000 and CDA unharness 2 in 2005. The HL7 standard clearly defines the design of the changed data. According to CDA R2, CDA documents use the Extensible nomenclature (XML) format for a good variety of applications. XML could be a versatile text format that's easy to use over the web and thus several applications are antecedently printed with it. Distributed computing alludes to the applications which are conveyed as an administration over the web, equipment and programming frameworks to the server farms. The distributed computing has three noteworthy administrations as said underneath.

2. Proposed System

In proposed framework we presented a CDA documents generation framework and integration framework that produces CDA records. Distinctive engineer stages through server will be actualized by CDA document generates on various frameworks. For information we utilize patient subtle elements and by that points of interest we can get to the information for CDA document generation. At the point when the customer taps the catch 'create CDA' the initially transmitted to CDA generation in the cloud server by means of CDA generation interface and CDA record is produced. It has a specialist to guarantee CDA record to approve the CDA report by Utilizing the service provider at our cloud server. Distributed computing based CDA generation and combination framework has a couple articulated points of interest over existing activities. To start with healing facilities don't need to buy legitimacy

programming to produce and coordinate CDA documents and bear the cost as some time recently. The important data transmitted from one server to another may be attacked, falsified, or stolen easily. In this proposed system, we are providing security for information in cloud by using session hijacking and brute force. In this system we build the proficient method for producing the CDA design for the created and coordinated CDA Documents for the utilization of patients. And also we are providing security to those data. In this system we are preventing session hijacking, bruteforce attack, sql injection.

Advantages of CDA

- It helps when the patient is in an emergency and the medical history needs.
- CDA document integration system integrates multiple CDA documents per patient into a single CDA document.
- Interoperability between hospitals help improve patient safety and quality of care.
- Provide security
- Preventing session hijacking, brute force attack, sql injection.

3. System Architecture

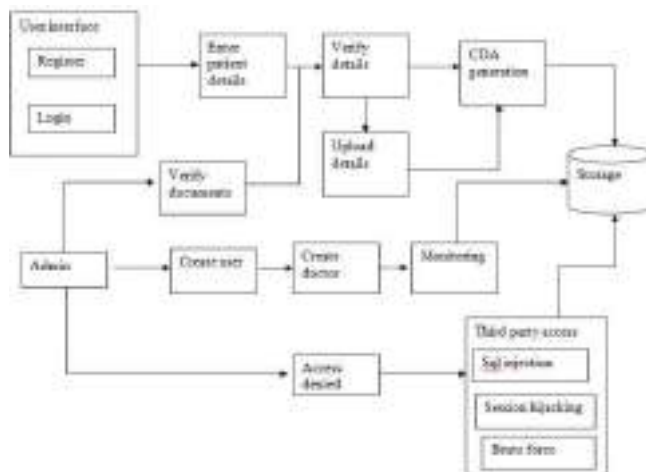


Fig 1 System architecture

Cloud computing is defined as using a network of remote servers,

hosted in the Internet that helps to store, manage, and process data, rather than a local server or a personal computer. It refers to the applications delivered as services over the Internet and software in the data centers that provide those services. The user pays fee depending on the amount of resources allocated, such as network, server, storage, applications and services. Currently, three major types of cloud computing services exist: Paas, Saas, Iaas. In Our project we using cloud storage to store the CDA generated files and patient's details. The generated CDA document is inspected by the CDA Validator to check the CDA standards. After inspection, the CDA document is returned to the recipient hospital. The hospital, the CDA documents to be integrated are processed through our CDA Integration Interface. The CDA Integration API in cloud relays each CDA document sent to the cloud to the CDA Parser, that converts each input CDA document to an XML object and analyzes the CDA header and groups them by each patient ID. The CDA Document Integrator integrates the given multiple CDA documents into a single CDA document. In this System we prevent System hijacking and brute force attack. If any third party try to access the data from storage the admin will deny that access. So no one can access the file without admin give permission.

LOF algorithm

Local Outlier Factor (LOF) is an algorithm proposed by Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng and Jörg Sander in 2000 for finding anomalous data points by measuring the local deviation of a given data point with respect to its neighbours.

LOF shares some concepts with DBSCAN and OPTICS such as the concepts of "core distance" and "reachability distance", which are used for local density estimation. The local outlier factor is based on a concept of a local density, where locality is given by k nearest neighbors, whose distance is used to estimate the density. By comparing the local density of an object to the local densities of its neighbors, one can identify regions of similar density, and points that have a substantially lower density than their neighbors. These are considered to be outliers.

KNN algorithm

KNN calculates the distance between the test data and the input and gives the prediction according. The k-nearest neighbors (KNN) algorithm is a simple algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slower as the size of that data in use grows. K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If $K = 1$, then the case is simply assigned to the class of its nearest neighbor.

Distance functions

Euclidean	$\sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Manhattan	$\sum_{i=1}^n x_i - y_i $
Minkowski	$\left(\sum_{i=1}^n x_i - y_i ^p \right)^{1/p}$

It should also be noted that all three distance measures are only valid for continuous variables. In the instance of categorical variables the Hamming distance must be used. It also brings up the issue of standardization of the numerical variables between 0 and 1 when there is a mixture of numerical and categorical variables in the dataset.

4. IMPLEMENTATION • CDAGeneration

In this module patient health informations are send to the cloud server. Now the cloud server will generate unique id for every users based on patient name, father name, date of birth and etc. If already id exist then the patient details will be appended with patients clinical history else new

CDA document will be generated.

The integrated CDA document is checked for error in Validator, and the result is returned as string to the hospital that requested CDA document integration. This is because the CDA Integration System and the CDA Generation System are separate different entities, and a new CDA document is made after document integration, hence it is mandatory to determine whether the new document complies with the CDA document integration, mainly whether there is any missing element, or the format is wrong. Error messages are returned when the missing element is found. And then the received string is converted to a CDA document file and saved. The validation process happening at the CDA Validator is based on the CDA schema. An error is returned when a required field has been left blank or the wrong data type has been used.

- **Monitoring patient**

In this module the new patient enter into hospital no need to give details about the disease and symptoms. The patient history already maintained in cloud server so we can get the patient histories by using key it is retrieve from patient personal details. The patient histories maintained in document which is contains patient clinical histories (hospital name, disease, prescription). Doctor view the patient information such as disease, symptoms etc. If it necessary patient is advised to take lab test. Lab Technician provides test result to patient. Based on test result, Doctor suggests prescription to the patient, and also patient health history should be maintained in appropriate hospital database. Doctor can view patient health history before he suggests prescription to the patient. These action will be monitor by admin. Admin will monitor patients details and doctor details.

5. Conclusion

We build up a proficient method for producing the CDA design for the created and coordinated CDA Documents for the utilization of patients. We are going to implement the security in cloud using session hijacking and brute force. Our cloud computing based CDA generation and integration

system has a few pronounced advantages over other existing projects. CDA documents increases, interoperability is achieved, but it also brings a problem where managing various CDA documents per patient becomes inconvenient as the clinical information for each patient is scattered in different documents. The CDA document integration service from our cloud server adequately addresses this issue by integrating multiple CDA documents that have been generated for individual patients. Thus, using our system the time is saved for the doctors in taking medical decisions at emergency times and deliver the correct health care as the medical records are in chronological order.

6. REFERENCE

- [1] M. Eichelberg, T. Aden, J. Riesmeier, A. Dogac, and Laleci, "A survey and analysis of electronic healthcare record standards," *ACM Comput. Surv.*, vol. 37, no. 4, pp. 277–315, 2005.
- T. Benson, *Principles of Health Interoperability HL7 and SNOMED*. In New York, NY, USA: Springer, 2009.
- [2] J. Lahteenmaki, J. Leppanen, and H. Kaijanranta, "Interoperability of personal health records," in *Proc. IEEE 31st Annu. Inter. Conf. Eng. Med. Biol. Soc.*, pp. 1726–1729, 2009.
- [3] R. H. Dolin, L. Alschuler, C. Beebe, P. V. Biron, S. L. Boyer, D. Essin, E. Kimber, T. Lincoln, and J. E. Mattison, "The HL7 Clinical Document Architecture," *J. Am. Med. Inform. Assoc.*, vol. 8, pp. 552–569, 2001.
- [4] R. H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F. M. Behlen, P. V. Biron, and A. Shabo, "The HL7 Clinical Document Architecture," *J. Am. Med. Inform. Assoc.*, vol. 13, no. 1, pp. 30–39, 2006.
- [5] K. Huang, S. Hsieh, Y. Chang, F. Lai, S. Hsieh, and H. Lee, "Application of portable cda for secure clinical-document exchange," *J. Med. Syst.*, vol. 34, no. 4, pp. 531–539, 2010.
- [6] C. Martinez-Costa, M. Menarguez Tortosa, and J. Tom, "An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes," *J. Biomed. Inform.*, vol. 43, no. 5, pp. 736–746, Oct. 2010.
- [7] MR. Santos, MP. Bax, and D. Kalra, "Building a logical HER architecture based on ISO 13606 standard and semantic web technologies,"

Studies Health Technol. Informat., vol. 160, pp.161– 165, 2010.

[8] K. Ashish, D. Doolan, T. Scott, and D. W. Bates, “The use of health information technology in seven nations,” *Int. J. Med. Informat.*, vol. 77, no. 12, pp. 848–854, 2008.


[9] K. Ashish, “Meaningful use of electronic health records the road ahead,” *JAMA*, vol. 304, no. 10, pp. 1709– 1710, 2010.

[11] S. M. Huff, R. A. Rocha, T. Fiers, W. D. Bidgood, A. W. Forrey, W. G. Francis, W. R. Tracy, D. Leavelle, F. Stalling, B. Griffin, P. Maloney, D. Leland, L. Charles and K. Hutchins, “Development of the logical observation identifier names and codes (loinc) vocabulary,” *J. Am. Med. Inform. Assoc.*, vol. 5, pp. 276– 292, 1998.

[12] J. D. D’Amore, D. F. Sittig, A. Wright, M. S. Iyengar, and R. B. Ness, “The promise of the CCD: Challenges and opportunity for quality improvement and population health,” in *Proc. AMIA Annu. Symp. Proc.*, pp. 285– 294, 2011.

[13] KS X 7504 Korean Standard for CDA Referral Letters (Preliminary Version)

Author Profile:

	<p>Dr. G. Lakshmi Vara Prasad Associate Professor Department of Information Technology QIS College of Engineering and Technology, Ongole</p> <p>Email: glv.prasad19@gmail.com</p>
---	--

	<p>V. Amarnath Student Department of Information Technology QIS College of Engineering and Technology, Ongole</p> <p>Email: amarnadhchowdaryvemula@gmail.com</p>
	<p>K. Ganesh Student Department of Information Technology QIS College of Engineering and Technology, Ongole</p> <p>Email: ganeshkolli6782@gmail.com</p>
	<p>N. Manideep Student Department of Information Technology QIS College of Engineering and Technology, Ongole</p> <p>Email: nethimanideep2727@gmail.com</p>
	<p>K. Sandeep Student Department of Information Technology QIS College of Engineering and Technology, Ongole</p> <p>Email: sandeepvardhankorrapati143@gmail.com</p>

UTILIZING HIDDEN SEARCH PATTERNS AND ACCESS PATTERNS IN SEARCHABLE ENCRYPTION SCHEME

R. Suneetha Rani¹, P. Haritha², J. Mounika³, M. Haritha⁴, M. JayaSankar⁵,
K. Adithya⁶

¹Associate Professor & HOD, Department of Information Technology, QIS
College of Engineering & Technology, Ongole.

^{2,3,4,5,6} Students, Department of Information Technology, QIS College of
Engineering & Technology, Ongole.

Abstract:

Dynamic searchable encryption techniques permit a customer to perform searches and updates over encoded information put away in the cloud. Notwithstanding, existing investigations show that the overall dynamic searchable symmetric encryption (DSSE) conspire is powerless against factual assaults because of the spillage of search designs what's more, access designs, which is unfavorable to ensuring the clients' protection. Albeit the conventional Oblivious Random Access Machine (ORAM) can shroud the access design, it additionally causes huge correspondence overhead and can't shroud the hunt design. These constraints make it hard to send the ORAM technique in genuine cloud conditions. To overcome this limit, a DSSE conspire called obliviously rearranged frequency network DSSE (OSM-DSSE) is proposed in this paper to access the scrambled information obliviously. The OSM-DSSE plot acknowledges productive hunt and update tasks in light of a rate lattice. Specifically, a rearranging calculation utilizing Paillier encryption is joined with 1-out-of-n obliviously move (OT) convention and nearby differential protection to muddle the hunt targets. Plus, a formalized security examination and execution examination on the proposed plot is given, which shows that the OSM-DSSE conspire accomplishes high security, effective pursuits, and low stockpiling overhead. Additionally, this plan not just totally shrouds the pursuit furthermore, access designs yet additionally gives versatile security against vindictive assaults by enemies. Besides, test results show that the OSM-DSSE conspire acquires 3-4x preferable execution proficiency over the state-of-workmanship arrangements.

1. Introduction

The ascent of cloud administration gives tremendous advantages to society and the IT business. Capacity as-a-Service is one of the most well-known cloud administrations accessible, which permits the customer to store information online distantly and access information all over the place, diminishing the expense of information the board and upkeep. Regardless of the benefits, Storage- as-a-Service additionally brings huge security and protection issues. Once information is re-appropriated, a customer loses the capacity to control the information. Likewise, touchy data might be altered or then again took by a malignant client. Albeit the customer can encode information with standard encryption plans (e.g., AES) to guarantee classification, fundamental activities (e.g., search/update) on the encoded information couldn't be performed. Furthermore, significant computational overhead is brought about, which enormously decreases the advantages of the cloud administration. To tackle the above issues, in 2000, Song et al. [1] first proposed the idea of searchable symmetric encryption (SSE). As another encryption crude, ocean rchable encryption empowers the client to look for a catchphrase over the ciphertext. Notwithstanding, the application was restricted to look on static scrambled information and was incapable to oppose the straightforward enemy assault. In 2003, Goh et al. [2] officially characterized the safe list and fostered a security model called the "semantic security" for versatile particular catchphrase assaults. Nonetheless, the precision of question result was restricted because of the utilization of the Bloom channel. In 2006, Curtmola et al. [3] proposed two new security models called "versatile security" and "non-versatile security", presenting a singlekeyword-search SSE with a conventional security definition. Because of the limits of the SSE proposed before and the quandary between guaranteeing client protection and productive information use on the cloud, Kamara et al. [4] presented the dynamic searchable symmetric encryption (DSSE) strategy, which empowered the client to perform search and update procedure on encoded information. The overall searchable encryption calculation improves the inquiry productivity at the expense of releasing some data about documents or inquiries to the server, for

example, the inquiry design and the access design. It is by and large recognized that the searchable encryption conspire is secure except if it doesn't uncover client information and inquiry data other than the data unveiled by the spillage profile. Nonetheless, in reality, a foe can abuse these spillages to dispatch factual assaults to recuperate the client information and question data. For example, Islam et al. [5] and Cash et al. [6] first and foremost abused access design spillage and earlier information about the dataset to recuperate the client's question data. Liu et al. [7] abused the inquiry example to dispatch assaults and acquired clients' question data. Zhang et al. [8] totally uncovered the customer's question and recuperated client information and inquiry data through the document infusion assault. Simon et al. [9] utilized both access what's more, search design spillages to recuperate the watchwords of questions. Along these lines, a significant course for future research is to zero in on the concealment of data exposure, as opposed to setting it as default. For the spillages and assaults depicted above, albeit a few arrangements have been proposed, the vast majority of the examination centers around forward-secure and backwardsecure techniques [10–12]. The Oblivious Random Access Machine (ORAM) can address the issue of access design spillage [13–15], however it was unreasonable for wide-spread reception. Garg et al. [16] abused ORAM furthermore, jumbled RAM (Random Access Memory) to stow away the hunt design. Kamara et al. [17] proposed an overall plan for smothering pursuit design spillage. The joined organized encryption (SE) in light of ORAM made the plan more productive than ORAM, however the plan was static. Additionally, the as of late proposed writing presents differential protection components [18] and hashing methods [19] to jumble access designs. In standard, arrangements that don't release any data to the server can be based on incredible methods, for example, secure two-party figuring, full homomorphic encryption (FHE), and so on, however they are regularly illogical. To accomplish safer searchable encryption, both the hunt design and the access design should be covered up. The above-proposed arrangements tackle either the search design spillage or the access design spillage yet not both. Albeit the technique proposed by Hoang et al. [20, 21] abused appropriated

information constructions to stow away the hunt design or the access design, it is typically important to consider whether there is conspiracy between servers.

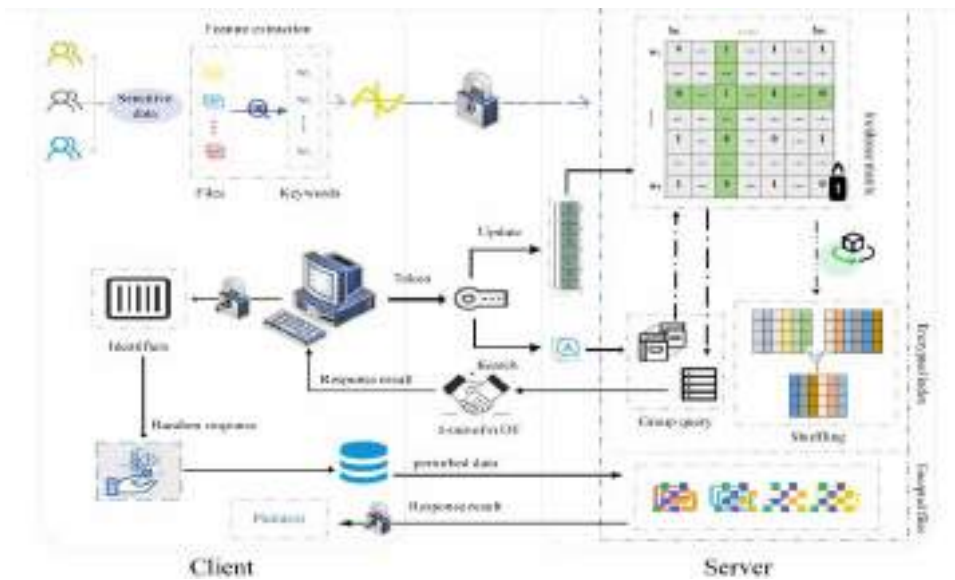
2. Related work

Melody et al. [1] first proposed the idea of static searchable encryption, making search scrambled information conceivable. Curtmola et al. [3] proposed the idea of versatile security and proposed the primary plan with ideal hunt intricacy $O(\#DB [w])$, where $\#DB [w]$ is the quantity of archives containing the watchword w . Numerous upgrades have been made in their ensuing work [25]. Pursue and Kamara proposed organized encryption to help inquiries on arbitrary data-structure. Kamara et al. [4] proposed the idea of dynamic searchable encryption, making searchable encryption not, at this point restricted to static tasks. Albeit ensuing examination endeavors zeroed in on viability [17], dynamics [10], restriction [12], security [14–19], and complex capacities [20–22], they actually experience the ill effects of releasing some significant data. Assailants can utilize these spillages to assault and recuperate information and cause more genuine data spillages. As of now, a few arrangements have been proposed to manage these spillages and assaults, yet these explores basically centered around forward-secure and backward-secure properties. Forward-secure alludes to the capacity to break the linkability of recently added information and inquiry catchphrases; in reverse-secure implies that the server is no longer ready to coordinate and recover the erased information.

Stefanov et al. [10] proposed the primary answer for help forward-secure property, however it displays a straight time intricacy for the hunt. Bost et al. [11] proposed a scheme depending on natives, for example, compelled pseudorandom capacities and puncturable encryption to accomplish fine control of the rival's force, forestalling the foe from assessing capacities on chosen inputs, or decoding explicit ciphertexts for forward and in reverse security. Sun et al. [12] proposed the first reasonable, non-intuitive in reverse-secure SSE scheme utilizing symmetric penetrated encryption. Be that as it may, the forward-secure and in reverse-secure

strategies principally focus on the data spillage in the update stage, without considering the data spillage of the access design and the hunt design. Subsequently, the issue of data spillage was not totally tackled. The oblivious random access machine (ORAM) can shroud the access design by befuddling each access interaction to make it indistinct from random access. The access design alludes to the arrangement of activities what's more, memory addresses.

The ORAM was first proposed by Goldreich et al. [13] to guarantee that any information block in memory didn't for all time live at a physical address and that two accesses are random. Goldreich et al. additionally



OSM-DSSE Model 1

proposed an ORAM model, giving a square root (Square-Root) and a layered arrangement. Zhang et al. [14] proposed a technique dependent on ORAM access design insurance in the distributed storage climate. Garg et al. [16] proposed a TWORAM plan that diminishes the customer stockpiling overhead while shrouds the record access design with ORAM. Nonetheless, the investigates have appeared that utilizing ORAM to dispense with data spills leads to high overhead and low execution proficiency.

3. Overview of OSM-DSSE Scheme

3.1. System model

Our system uses the client-server model. The client separates the catchphrases of the record and develops an occurrence network between the watchwords and the record, scrambles the occurrence lattice and the document, and sends them to the server. The client issues search and update solicitations to the server. The server stores the encoded frequency lattice and reactions to the client's search and update demands. Note that we consider a semi-genuine (legit yet inquisitive) server. During the entrance, despite the fact that information documents are encoded, the cloud server may attempt to infer other delicate data from clients' pursuit demands.

In this manner, albeit the server can loyally follow the convention, it can learn data.

3.2. System objective

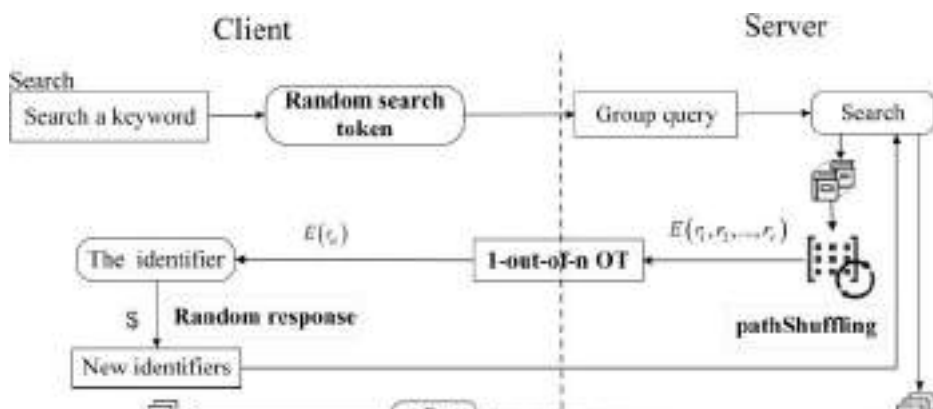
We will likely successfully play out a security ensured catchphrase search and record update on a scrambled cloud database. The principle goals of this system are as follows:

- Hide the entrance design: We use Paillier encryption to rearrange the rate framework. This calculation can randomize the position of watchwords in the frequency framework, befuddle access ways, and shroud access designs.
- Hide the pursuit design: Based on the gathering question, the server uses the two-level guide to acquire the objective information block containing different bits of information. What's more, it executes a proficient 1-out-of-n OT convention with the client to acquire the objective. In this cycle, the 1-out-of-n OT convention makes the server incapable to recognize which watchword the client is looking for. The client additionally doesn't have a clue about the server's different messages aside from the looked through watchword. This convention secures the protection of the client and server at the same time. Furthermore, the rearranging is performed after each search, which makes the line position of the catchphrase in the occurrence lattice change, and it likewise can

change over the deterministic token into an irregular token (express hunt design). The enemy can't dispatch an assault by investigating the hunt recurrence.

4. Construction of OSM-DSSE

Initially, the client creates public boundaries by DSSE. KeyGen. These boundaries incorporate the symmetric key KF to encode documents and the critical KI to scramble list. Besides, the client builds a safe arbitrary occurrence framework by DSSE. Build Index. the client produces an irregular key $k2$.



Hidden Search

The client extricates the catchphrase set $W = (w_1, \dots, w_m)$ from the document set $F = \{f_1, \dots, f_n\}$ (each record has a one of a kind identifier (id_1, \dots, id_n)). The places of every catchphrase a document in the frequency framework are controlled by the pseudorandom work G and the hash tables T_f and T_w . At that point, the client separates the catchphrase set W into information squares of size v . The last information block is cushioned up to v components if vital, and it is numbered as $(1, 2, \dots, dm/v)$. The client develops a two-level guide $\Omega (M_w, A)$ and encodes the key of M_w . All the while, the client builds the word reference D . Ultimately, the client encodes the record by ϵ .Enc, and sends the protected irregular occurrence grid I , encoded record C , two-level guide Ω , and address map table M_f to the server. Then, the client saves locally.

5. Performance analysis

5.1. Storage overhead

Client storage: The client keeps two hash tables and a word reference. Storage expenses of the two hash tables are corresponding to the quantity of watchwords and records, i.e., $O(m)$ and $O(n)$, where m , n addresses the quantity of watchwords and records, individually. The storage cost of the word reference is relative to the quantity of catchphrases, i.e., $O(m)$. The occurrence network is a $m \times n$ -dimensional lattice with a storage cost of $O(m \cdot n)$. The two-level guide Ω comprises of two sections: a location map table M_w and an cluster A . Storage of M_w is relative to the number of squares. Expecting that the information are isolated into t blocks, the storage cost of M_w is $O(t)$. The size of exhibit A_n is identified with the quantity of columns of the occurrence framework, and the storage cost is $O(m)$. The storage cost of M_f is identified with the quantity of documents, and the storage cost is $O(n)$. In this way, the absolute storage cost of the server is $O(m \cdot n + m + n + t)$.

5.2. Communication overhead

In the arrangement stage, the client sends the scrambled rate network and the encoded document to the server. The correspondence overhead is $O(m \cdot n + n c_i)$, where $m \cdot n$ is the size of scrambled rate network and c_i is the size of each scrambled record. In the inquiry stage, the client sends the $m \times m$ disarray network to the server, and the correspondence overhead is $O(m^2)$. The server returns a scrambled information square of size v with a correspondence overhead of $O(v)$. In the update stage, the client sends the $m \times 1$ section framework to the server, and the correspondence overhead is $O(m)$.

5.3. Computational overhead

The client principally produces a stage network, an inclining framework, and an encoded disarray network. Both the stage network and the inclining framework are $m \times m$ measurements, so the disarray network is of measurement $m \times m$. What's more to the stage grid and the askew lattice, there are m bits of information that are not 0. The leftover numbers

are each of the O , and the computational expense of creating O is insignificant. Along these lines, the computational expense of creating the stage framework and the inclining network is $O(m)$

6. Experimental outcomes

In the analysis, the performance of search and update activities of the proposed scheme is assessed and contrasted and existing schemes. The ideal opportunity for making a frequency framework of various sizes is assessed to represent how the size of the dataset impacts the construction season of rate network. As demonstrated in Fig. 5, the construction time is 10.114 s for a rate grid of 103×103 , When the size of the rate grid surpasses 103×103 the time to develop the scrambled frequency network increment quickly. For instance, it requires around 20 minutes to develop a 104×104 occurrence grid with 108 information. Since the occurrence grid is just developed once during the arrangement stage, the relationship among the hunt, date time, and the size of the occurrence grid is mostly examined.

7. Conclusion

This article proposes a searchable encryption scheme named OSM-DSSE to hide the search and access patterns. An effective shuffling algorithm based on Paillier is proposed to shuffle the incidence matrix, so that the position of the row in the incidence matrix is changed. This scheme combines the 1-out-of- n OT protocol and the differential privacy strategy based on random response to realize random data access. Besides, the security of the proposed scheme is formally analyzed, showing that the proposed scheme provides an adaptive semantic security that can against selective adversaries. Furthermore, the OSM-DSSE achieves approximately 3-4x execution speed than existing schemes. In the future, the optimal block size will be investigated and the scenarios with different security levels will be updated.

References

1. Song, D.X., Wagner, D., Perrig, A. Practical techniques for searches on encrypted data. In Proc. - IEEE Symp. Secur. Privacy. SP 2000. S&P 2000,

pages 44–55. IEEE, 2000.

2. Goh, E.J. et al. Secure indexes. *IACR Cryptol. ePrint Arch.*, 2003:216, 2003.
3. Curtmola, R., Garay, J., Kamara, S., Ostrovsky, R. Searchable symmetric encryption: Improved definitions and efficient constructions. 2006.
4. Kamara, S., Papamanthou, C., Roeder, T. Dynamic searchable symmetric encryption. In *In Proc ACM Conf Computer Commun Secur*, pages 965–976, 2012.
5. Islam, M.S., Kuzu, M., Kantarcioglu, M. Access pattern disclosure on searchable encryption: ramification, attack and mitigation. In *Ndss*, volume 20, page 12. Citeseer, 2012.
6. Cash, D., Grubbs, P., Perry, J., Ristenpart, T. Leakage abuse attacks against searchable encryption. In *In Proc ACM Conf Computer Commun Secur*, pages 668–679, 2015.
7. Liu, C., Zhu, L., Wang, M., Tan, Y.A. Search pattern leakage in searchable encryption: Attacks and new construction. *Inf Sci*, 265:176–188, 2014.
8. Zhang, Y., Katz, J., Papamanthou, C. All your queries are belong to us: The power of file- injection attacks on searchable encryption. In *In Proc. USENIX Secur. Symp. (USENIX Security 16)*, pages 707–720, 2016.
9. Oya, S., Kerschbaum, F. Hiding the access pattern is not enough: Exploiting search pattern leakage in searchable encryption. *arXiv preprint arXiv:2010.03465*, 2020.
10. Stefanov, E., Papamanthou, C., Shi, E. Practical dynamic searchable encryption with small leakage. In *NDSS*, volume 71, pages 72–75, 2014.
11. Bost, R., Minaud, B., Ohrimenko, O. Forward and backward private searchable encryption from constrained cryptographic primitives. In *In Proc ACM Conf Computer Commun Secur. ACM*, pages 1465–1482, 2017.
12. Sun, S.F., Yuan, X., Liu, J.K., Steinfeld, R., Sakzad, A., Vo, V., Nepal, S. Practical backward- secure searchable encryption from symmetric puncturable encryption. In *In Proc ACM Conf Computer Commun Secur*, pages 763–780, 2018.
13. Goldreich, O., Ostrovsky, R. Software protection and simulation on oblivious RAMs. *J. ACM*, 43(3):431–473, 1996.

14. Zhang, J., Ma, Q., Zhang, W., Qiao, D. Tskt- oram: A two-server k-ary tree oram for access pattern protection in cloud storage. In In: Proc IEEE Mil Commun Conf MILCOM. IEEE, pages527–532. IEEE, 2016.
15. Demertzis, I., Papadopoulos, D., Papamanthou, C., Shintre, S. {SEAL}: Attack mitigation for encrypted databases via adjustable leakage. In Proc. USENIX Secur. Symp., 2020.
16. Garg, S., Mohassel, P., Papamanthou, C. Tworam: Round-optimal oblivious ram with applications to searchable encryption. IACR Cryptol. ePrint Arch., 2015:1010, 2015.
17. Kamara, S., Moataz, T., Ohrimenko, O. Structured encryption and leakage suppression. In In Lect. Notes Comput. Sci, pages 339–370. Springer, 2018.
18. Chen, G., Lai, T.H., Reiter, M.K., Zhang, Y. Differentially private access patterns for searchable symmetric encryption. In IEEEProc IEEE INFOCOM , pages 810– 818. IEEE, 2018.
19. Patel, S., Persiano, G., Yeo, K., Yung, M. Mitigating leakage in secure cloud-hosted data structures: Volumehiding for multi-maps via hashing. In Proc ACM Conf Computer Commun Secur, pages 79–93, 2019.
20. Hoang, T., Yavuz, A.A., Guajardo, J. Practical and secure dynamic searchable encryption via oblivious access on distributed data structure. In ACM Int. Conf. Proc. Ser. ACM, pages 302–313, 2016.
21. Hoang, T., Yavuz, A.A., Durak, F.B., Guajardo, J. Oblivious dynamic searchable encryption on distributed cloud systems. In Lect. Notes Comput. Sci. Springer, pages 113–130. Springer, 2018.
22. Akavia, A., Gentry, C., Halevi, S., Leibovich, M. Setup free secure search on encrypted data
D. Faster and postprocessing free. PoPETs, 2019(3):87–107, 2019.
23. Paillier, P. Public-key cryptosystems based on composite degree residuosity classes. In Lect. Notes Comput. Sci., pages 223–238. Springer, 1999.
24. Zhang, Z., Wang, K., Lin, W., Fu, A.W.C., Wong, R.C.W. Repeatable oblivious shuffling of large outsourced data blocks. In SoCC - Proc. ACM Symp. Cloud Comput., pages 287–298, 2019.
25. Wang, C., Cao, N., Li, J., Ren, K., Lou, W. Secure ranked keyword

search over encrypted cloud data. In In Proc Int Conf Distrib Comput Syst, pages 253–262. IEEE, 2010.

26. Cao, N., Wang, C., Li, M., Ren, K., Lou, W. Privacy-preserving multi-keyword ranked search over encrypted cloud data. In IEEE Trans Parallel Distrib Syst, volume 25, pages 222–233. 2013.

27. Sun, W., Wang, B., Cao, N., Li, M., Lou, W., Hou, Y.T., Li, H. Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking. In ASIA CCS - Proc. ACM SIGSAC Symp. Inf., Comput. Commun. Secur., pages 71–82. 2013

Authors Profile

	<p>Dr. R.Suneetha Rani Associate Professor and HOD, Department of Information Technology, QIS College of Engineering and Technology, Ongole</p>
	<p>K.Adithya Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: kodeadithya37@gmail.com</p>

	<p>P.Haritha Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: pinnakaharitha8@gmail.com</p>
	<p>J.Mounika Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: mounichowdary57@gmail.com</p>
	<p>M.Jaya Sankar Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: jayasankar14660@gmail.com</p>
	<p>M. Haritha Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: harithamalempati9492@gmail.com</p>

IOT BASED DIGITAL FILTERING SYSTEM FOR COVID - 19

^[1]N.Suresh, ^[2]B.Avinash, ^[3]B.V.Ahalya, ^[4]M.Niharika, ^[5]M.Mounika
, ^[6]B.Sruthilaya

¹Associate Professor, Department of Information Technology, QIS College of Engineering & Technology, Ongole.

^{2,3,4,5,6}Students, Department of Information Technology, QIS College of Engineering & Technology, Ongole.

Abstract: The COVID-19 pandemic is quickly impacting the public, disrupting the global economy and movement. The most recent trend was facial protection with the mask. Many utility providers may wish that citizens to maintain masks properly in order to use their facilities. As a result, mask detection plays a crucial role in supporting world civilization. In addition to the mask, regular body temperature checks are required to ensure that the COVID-19 virus is present in the body. In that situations, we need adequate sanitation with that in mind. We will provide a simpler approach to achieve this objective by utilizing TensorFlow, Keras, OpenCV, and Scikit-Learn, and a couple of simple machine learning packages and we use the ESP01 (esp8266 family), MLX90614 (Temperature sensor) for sensing temperature and utilization the PIR sensor. The disinfectant can be sprayed it is possible to combine these three things using raspberry pi model b board and using certain wires and devices such as the camera and all sensors.

Key words: IOT, Face masks, deep learning, TensorFlow, Keras, OpenCV, and Scikit-Learn, MLX90614, ESP01, PIR, raspberry pi.

I. INTRODUCTION

Corona virus disease 2019 (COVID-19) has affected over 20 million people worldwide, resulting in over 0.7 million deaths, according to the World Health Organization's (WHO) official Situation Report – 205 [1]. Patients with COVID-19 have recorded a wide variety of symptoms, from minor signs to serious disease. Patients with COVID-19 have recorded a wide variety of symptoms, from minor signs to serious disease. All of these

being respiratory issues such as shortness of breath or trouble breathing. COVID-19 infection can cause severe complications in elderly people with lung disease, as they tend to be at a greater risk [2]. 229E, HKU1, OC43, and NL63 are some of the most common human coronaviruses that infect people all over the world. Viruses like 2019-nCoV, SARS-CoV, and MERS-CoV infect animals and transform into human coronaviruses before infecting humans [3]. Infectious beads can be spread from people with respiratory issues to everyone who comes into contact with them. The environment of a tainted person will affect touch transmission because virus-carrying droplets can land on his neighbouring surfaces [4]. Wearing a clinical mask is important for preventing such respiratory viral infections, such as COVID-19. The public should know whether to wear the mask for source protection or COVID-19 aversion. The use of masks has the potential to reduce susceptibility to danger from a noxious individual during the "pre-symptomatic" phase, as well as stigmatise individuals who wear masks to prevent the spread of virus. Health goggles and respirators for health care assistants are prioritised by the WHO [4]. Face mask recognition has since become a critical role in today's global society. Face mask identification involves identifying the orientation of a person's face and then assessing whether or not they are wearing a mask. The problem is related to general object recognition, which is used to identify different types of objects. Face identity is the process of categorising and identifying a particular category of entities, namely faces. It has a wide range of uses, including autonomous driving, schooling, and surveillance [5]. The simple Machine Learning (ML) packages such as TensorFlow, Keras, OpenCV, and Scikit-Learn are used in this paper to present a simpler approach to fulfil the above purpose. The internet of things (IoT) is used to link electronic devices to the internet. Vehicle and temperature measurement devices Computers, as well as other technological equipment such as controllers, software, and network networking facilities that allow for data processing and sharing, are examples of computers. In the twenty-first century, there has been a major paradigm change toward IoT as a burgeoning discipline with many possibilities and varied prospects for growth and development [11]. Having access to the internet encourages us

to operate the systems that have become integral part of daily lives and lives in an efficient way. The Internet allows you to link and network various types of equipment, such as sensors and fitness trackers. The internet encourages wholesome and ideal tracking schemes using closed circuit cameras in the modified scenario following the September 11, 2001 attack on the United States, where surveillance has assumed supreme significance in proposed model defence and survival [12]. Allows us to upload input and output to the Internet, which uses cloud furnishing. The knowledge gathered in this way is available through the internet for tracking and study from anywhere on the planet [13]. People are increasingly relying on embedded systems to manage and track the factors impacting the environment in order to reduce human effort and participation. The importance of temperature and humidity in studying and interpreting nature cannot be overstated. IoT enters the picture here by dramatically improving the mechanism's performance and systematically reducing human intervention and, as a result, total spending [14]. Practically any aspect of exercise includes temperature and humidity monitoring schedules. However, in any region, knowing the exact temperature value and its significance is critical for monitoring [15]. Temperature perception is used in various sectors, such as the pharmaceutical industry, as the driving force behind these control devices, which can use computerised and simple temperature sensors [16]. Temperature values are estimated using resistors, semiconductors, and thermistors. These components are present within the sensor to obtain the temperature based on the situation. Our system's primary objective is to track live temperature and humidity at a low cost [17]. Raspberry Pi is an observational device or controller that is used to save money in the cloud. Python is the programming language that Raspberry Pi makes use of. HTU 211D sensors are temperature sensors that are used in this experiment [18]. This provides an opportunity to calculate temperature and the optimal fundamental position through use of HTU 21 sensors, that are easier and quicker to use. Using connecting cables, the sensor is connected to the Raspberry Pi. The temperature sensor HTU 211D sensors were examined, saved, and shown in the Raspberry Pi unit [19]. IoT-based technologies are

used to track all of the electrical and electronic equipment that are found in both homes and enterprises. Furthermore, the IoT devices' saved data can be accessed from any location [20]. Wherever in the planet, the sensor analyses the graphical representation of the detected data in any user-defined format. An IoT-based Arduino with a Raspberry Pi microcontroller is used in this project. Monitoring humidity and temperature with Arduino is a fun and safe operation. Furthermore, by using data stored on the internet to calculate the actuator, this modular method obtains more values [21]. A USB line serial interface is needed to link the Arduino board to the Raspberry Pi for any application [22]. Wireless solutions for different health management applications have been facilitated by technological advancements in the area of disease prevention and patient health maintenance. The soundness of the human cardiovascular system is closely related to heart rate, which is a very important health parameter. It refers to a host of mental states, such as biological workload, burnout and task concentration, sleepiness, and the active state of the autonomic nervous system. This was normal growth with reduction in a vein as plasma was driven over heart's regular contractions. According to a recent poll, cardiovascular disease is one of the leading causes of death in a number of countries. In comparison, coronary disease disables many million individuals [23]. This is due to a delay in delivering treatment to those who are affected. So, if services are allocated to early diagnosis and prevention of heart disease, there is a better chance of a decrease in coronary disease-related deaths than increased post-hospitalization care. As a result, novel methods are needed to shorten the time between diagnosis and care. One approach is to keep track of the patients. In addition, the movement toward self-sufficiency has boosted demand for personalised non-hospital-based services. Heart disease affects the elderly the most of the time. They frequently live alone, and no one is able to keep an eye on them 24 hours a day [24]. A self-monitoring system, including a pulse frequency VDU, allows us to track the pulse frequency in actual period. The first models consisted of a chest-mounted control box with a series of electrode leads [25]. A microprocessor is used in newer versions of the heart rate sensor to constantly monitor the ECG and calculate the heart rate and other critical

parameters. The majority of previous work relies on the PPG (photoplethysmography) technique and employs a PIC controller. A new device has been proposed as a way of making surveillance networks more efficient and cost efficient. Sensors are used to monitor the pulse rate and temperature of patients in this proposed system. The RASPBERRY PI MODEL - B controller system is used to store data for transmission temporarily [26].

II. Related Work

A face is identified from an image that contains multiple attributes in the face detection process. Face identification analysis necessitates expression recognition, face recording, and pose prediction, according to [21]. The challenge is to recognise the face in a single picture. Face recognition is a challenging task due to the fact that faces vary in height, form, colour, and other characteristics and are not timeless. It becomes a difficult task when an invisible image is obstructed by something not facing the eye, and so on. According to the authors of [22], occlusive face recognition faces two main challenges: 1) the lack of large datasets for both masked and unmasked expressions, and 2) the absence of facial expression in the protected section. Some absent expressions can be regained and the ascendancy of facial signs can be mitigated to a large degree using the locally linear embedding (LLE) algorithm and dictionaries educated on an enormous pool of masked faces, synthesized mundane features. The size of the input image is a strict limit for convolutional neural networks (CNNs) in computer vision, according to research published in [11]. To overcome the inhibition, the common procedure is to reconfigure the images before fitting them into the network. The task's key challenge is to accurately detect the face from the picture and then determine whether or not it has a mask on it. The suggested method should be able to track a face and a mask in motion in order to execute surveillance activities.

PROPOSED SYSTEM:

Detecting the effect of a mask on the face using deep learning algorithms and libraries such as TENSORFLOW, KERAS, and OPENCV.

Use IoT devices with sensors such as the ESP01 (esp8266 family), MLX90614 to determine a person's temperature. SENSOR FOR MEASURING TEMPERATURE WITHOUT CONTACT (Temperature sensor). With the aid of a PIR sensor, we can sense a person's movement and spray sanitizer. Using IoT devices, we're attempting to combine all three elements into a single product.

A. TensorFlow

TensorFlow, Emotion analysis, linear model, visual intellect retrieval, computer vision, text summarization, database processing, computational drug development, and fault detection all use a programming interface for expressing machine learning algorithms. TensorFlow is used at the backend of the proposed model's Sequential CNN architecture (which consists of many layers). It's also used in data analysis to reshape the data (image).

B. Keras

Keras provides fundamental reflections and building units for the construction and transportation of machine learning arrangements at fast iteration rates. TensorFlow's scalability and cross-platform features are fully used. Layers and templates are Keras' primary data structures. Keras is used to incorporate all of the layers in the CNN model. It aids in the compilation of the overall model, as well as the translation of the class vector to the binary class matrix in data processing.

C. OpenCV

Open CV is an image processing and artificial intelligence development kit that can be used to recognise faces, recognise objects, group gestures in videos, trace progressive modules, follow eye gestures, record camera behaviour, remove red eyes from flash photos, perceive environment, and set ultra violet filters. In order to resize and colour convert data images, the proposed approach makes use of OpenCV's features.

III. Methodology

Suspects of standard operating procedures (SOPs) are increasingly increasing in developed countries where citizens do not narrowly obey the rules of the government. To stop the spread of COVID-19, we proposed IoT-based rapid screening models that can be publicly located to disinfect the virus and divide people's face masks into three groups, as previously stated. Furthermore, our proposed model will recognize and distinguish between various types of face masks, such as N-95 or surgical masks.

Walkthrough Gate including IoT-based Smart Screening and Sanitization

Walkthrough Gate including IoT-based Smart Screening and Sanitization is a system which consists of multiple modules. The device is a multi-sensor, multi-purpose smart device designed to minimize the risk of COVID-19 carriers spread. In countries, like Pakistan and other developing countries, where governments did not apply lockdown and economic activities are entirely not stopped, people are allowed to go to the market, and they can travel according to some standard operating procedures (SOP's). The virus could spread through their clothing, shoes, hands, or other surfaces in this case. Furthermore, people do not bother to obey the government's defined SOPs or to consider social distancing. The architecture of our proposed WGSSS is thoroughly clarified. When anyone steps through from WGSSS, the temperature testing module performs a contactless temperature check as the first move. As is suspected, if the temperature reaches 99 degrees Fahrenheit or higher, a snapshot of the man is taken and stored in a database, along with his temperature and health status. At the same time, our mask detection module will determine whether or not anyone is wearing a mask. Meanwhile, the person in charge of the entrance will divert the perpetrator to a COVID-19 testing facility. Our proposed IoT-based device can be located at the entrance towards any public location, hospital, or crowded area where COVID-19 can be located.

Body Temperature Detection

Our proposed IoT-based SSDWG is split into two modules and has

a number of capabilities. Since [6] temperature or fever are some of the most frequent symptoms of COVID- 19, the most popular component of SSDWG tracks the temperature of the human body in real-time in a contact-free way, and it stores the individual's body temperature, along with an image of a suspicious person in our system, we can monitor the most vulnerable individuals through documented temperature data, and we can track the most vulnerable persons through recorded temperature data.

In our proposed SSDWG [33], we use the MLX90614 temperature measuring sensor. It's a non-contact digital infrared thermometer for measuring human body temperature, with a microcontroller that can communicate wirelessly. Since it does not violate the social distance SOPs, we suggested contact-free temperature testing mechanisms. A suspected picture of a body temperature that is not normal is captured using an image processing module.

Our temperature sensor and image capture module were mounted on the front side of SSDWG in our proposed temperature sensor is a contactless sensor that detects the temperature of the human body. The temperature sensor sends a signal to the alarm whether the human body temperature is measured to be 100.4 Fahrenheit or 38.0 Celsius, or higher, and the body temperature and picture taken from CCTV are stored in the database [34]; the authors suggested a full device to auto-detect routine examinations on a tele ophthalmology network. To extract information from the image, they combined the exploration of pathology profiles with methods for detecting lesions.

Mask Detection Module

The second module of SSDWG detects people wearing masks when they walk through it. VGG16, MobileNetV2, Inception V3, ResNet-50, and Convolutional Neural Network (CNN) are five pre-trained deep learning models that we used to detect face mask wear in three classes: face with proper mask (FWPM), face with inappropriate mask (FWIPM), and face without a mask (FWOM). Furthermore, our suggested model divides and classifies face masks into two categories: N-95 and surgical masks. Transfer learning was used to fine-tune It's faster and cheaper to use these five pre-

trained models rather than starting from scratch with dynamically initialised weights. The authors [13] used a deep neural network to study vibration signals generated by walking people on the floor to detect human localization in big buildings in a similar way. To avoid the output layers of these models from being conditioned or updated after we fed our dataset, we made them non-trainable by freezing the weights and other trainable parameters in each layer. We have added an output layer to our dataset to practise on. In our current model, this output layer will be the only one that could be trained. We used the Adam optimizer with a learning rate of 0.0001, cross-entropy for our loss, and precision for our matrix. An input layer, an output layer, and several opaque layers make up CNN's architecture. Convolutional layers, pooling layers, completely linked layers, and normalisation layers are common hidden layers (ReLU). The network receives two types of images presence and not presence of a face mask (proper and improper) withan 224 by 224 pxls and three GBR channels. The first layer, as seen in Equation (1), is a convolutional layer with the Relu function, with a kernel size of 224 by 224 and 64 outputchannels.

$$\text{Relu}(x) = \max(0, x) \quad (1)$$

If the relu function receives some negative feedback, it returns 0; but, if the value of x is positive, it returns that value. The architecture's next layer is max pooling, which entails shrinking the picture stack. The window size for pooling an image is (112 112, 64 output channels). The window is then categorized depending on how well the strides appear, with the maximum priority for each window being registered. We have a normalisation layer with the max-pooling layer called the Rectified Linear Unit (Relu) operation, which is seen in Equation (1) and changes all negative values in the filtered image to 0. The Relu layer raises the model's non-linear properties, and this process is replicated on all of the filtered images. We had a convolutional layer on the third layer, similar to the first layer, that processed images with a kernel size of (56 56) and processed RGB to 256 output channels. This is a type of layer called as purling layer with a size of 77 of the kernel before the fully connected layer. With the softmax function seen in Equation, we have a flattened layer with a totally model. To exit from our device, A individual must stand in front of the temperature

sensor for three seconds; the connected layer (2).

$$P = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

where x_i denotes the cumulative input obtained by unit I and p_i denotes the image projection likelihood for class i . Any value in this layer receives a vote in order to predict the image type. Each intermediate layer votes on phantom secret categories, and the completely linked layers are often stacked together. In effect, each additional layer helps the network to learn ever more complex combinations of features in order to improve decision-making [35]. Finally, we have a performance in the form of a mark with the words "proper face mask," "improper face mask," and "no face mask."

Dataset

For mask identification and classification, there is no suitable dataset. The available databases are often noisy or artificially developed, making real-time mask detection systems impossible to create. The dataset of photographs of artificial face masks, which does not constitute real-world representation, was not used. As a result, we used a massive dataset for our proposed model, which totaled 149,806 photographs obtained from various sources. To doing the mask detection on the face and arrangement module, we used 3 datasets: GitHub MAFA ,Face-Net, as well as a 3rd dataset from Bing. The Github (MaskedFace-Net) dataset includes a total of 133,783 photographs, with 67,049 images belonging to individuals who properly wear a face mask and 66,734 images belonging to individuals who do not properly wear a face mask. From the (Face-Net) list, we had mainly used 34,456 face from front images for masked faces correctly and 33,106 for poorly frontal images face masked . We used the MAFA dataset [38], which includes 30,811 images with various orientations and occlusion degrees, as well as another massive dataset from Github (FMLD) [37]. We used the Haar cascade classifier [32] to remove features from the MAFA dataset by using the Haar Wavelet technique with a 24 24 window scale to pick just 12,000 frontal face images and break the dataset into 9600 images for training and 2400 images for research. The Bing dataset comprises 4039 images, which we divided into 3232 training images and 807 testing images.

The bing-images library [39], which can directly fetch images using URLs, was used to collect face masks images from Bing. The photographs of a face with a correct mask (FWPM), a face with an inappropriate mask (FWIPM), and a face without a mask were combined from the Github dataset [37] and the Bing dataset (FWOM).

Fine-Tuning with Transfer Learning

Due to its superior success over other algorithms, DNN (Deep Neural Network) has made considerable strides in image recognition. DNN development and preparation from the ground up is a lengthy and time-consuming process that necessitates a lot of computing capacity and energy. To reduce complexity, the neural network's learned parametric weights are transferred to the current model using a transfer learning technique. In our proposed system, we used five pre-trained models, with the Inception V3 and VGG-16 possessing higher precision than the others. The models which are capable of individual 1000 styles, but we only need three: improper face mask, face with out a mask, and face with proper mask. For doing that , we had created a fully connected layer with 3 output features, rather than the one thousand in the previous fully connected layer of the models. The model classifier' tool, a six-layer stack, was used to access the VGG-16 and Inception V3 classifiers. We also disabled convolutional layer testing because we only practised the completely linked layer. Our loss function (cross-entropy) and optimizer were also specified. The rate of learning a StepLR object which starts at 0.001 and comes down by a 0.1 factorial every seven epochs. By going through all of the batch training at each epoch, loss measured, and going to network weights adjustment accordingly. Methods backward() and optimizer.step() Subsequently, we evaluated the performance over the test dataset. At the end of each epoch, we showed the network progress (loss and accuracy). The precision indicated how many of our forecasts were accurate.

Convolutional Neural Network (CNN)

To stop model over-fitting in CNN, we used a similar data augmentation strategy. It portrays the con – neural network architecture, which having more than one and less than two convolutional layers with

kernel size of 200 was divided into 3X3 and with the kernel size of 100 divided into 3X3. A new layer was added called as flattening layer which transforms a 2-D matrix with features of a vector fed into connected in fully neural network classifier which links the fully and convolutional connected layers. A layer of dense with 50 neurons, single neurons, and one output with the next layer follow the top layer of neural networks. Finally, a new layer called dense layer with two neurons gives the classification production, for the people who are protecting their face with/without having any protection.

VI. CONCLUSION

At the end of this project, we briefly summarized the work's inspiration. The model's learning and success role was then shown. The approach has achieved a reasonable level of precision using simple machine learning methods and simpler techniques. It can be used for a variety of things. Given the Covid-19 crisis, wearing a mask could become mandatory in the near future. Many government agencies will need people to wear masks properly in order to use their facilities. The implemented paradigm would make a significant contribution to the universal health care system. Other sensors such as the ESP01 (esp8266 family), MLX90614 CONTACT-FREE TEMPERATURE MEASUREMENT SENSOR (Temperature sensor), and the PIR sensor can be used to sense the temperature and spray the sanitizer. It may be enhanced in the future to detect whether or not an individual is wearing the mask properly. The model should be developed further to detect whether the mask is virus-prone or not, i.e. whether it is surgical, N95, or not.

Limitations

For mask identification and classification, there is no suitable dataset. Since the datasets available are often noisy or artificially developed, they are unsuitable for designing a real-time mask detection method. We, on the other hand, spent a significant amount of time gathering and pre-processing the appropriate face mask files. Additionally, Only frontal face images were selected from the noisy images using the Haar Wavelet technique.

Future enhancement

In upcoming days, by making a new network modeled framework which uses previously trained models using the neural network methods to know the physical contact (maintaining the social distance) among one another as a precautionary measure against the transmission of the virus in real-time COVID-19. A plain translucent face mask can also be used for those with a hearing disability, according to the CDC (Centers for Disease Control and Prevention). As a result, we'll concentrate on detecting and classifying clear face masks. According to World health organization suggestion, sternutation and cough was the major signs of COVID-19; but, in the upcoming, we'll use neural network models to find the people who are having cough and sternutation, which will lead in the detection of COVID-19 spread.

VI. REFERENCES

- [1] W.H.O., "Coronavirus disease 2019 (covid-19): situation report, 205". 2020
- [2] "Coronavirus Disease 2019 (COVID-19) – Symptoms", Centers for Disease Control and Prevention, 2020. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/symptomstesting/symptoms.html>. 2020.
- [3] "Coronavirus – Human Coronavirus Types – CDC", Cdc.gov, 2020. [Online]. Available: <https://www.cdc.gov/coronavirus/types.html>. 2020.
- [4] W.H.O., "Advice on the use of masks in the context of COVID-19: interim guidance", 2020.
- [5] M. Jiang, X. Fan and H. Yan, "RetinaMask: A Face Mask detector", arXiv.org, 2020. [Online]. Available: <https://arxiv.org/abs/2005.03950>. 2020.
- [6] B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing", International Journal of Computer Sciences and Engineering, vol. 6, no. 10, pp. 712-714, 2018. Available: 10.26438/ijcse/v6i10.712714.
- [7] F. Hohman, M. Kahng, R. Pienta and D. H. Chau, "Visual Analytics

- in Deep Learning: An Interrogative Survey for the Next Frontiers,” in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 8, pp. 2674-2693, 1 Aug. 2019, doi:10.1109/TVCG.2018.2843369.
- [8] C. Kanan and G. Cottrell, “Color-to-Grayscale: Does the Method Matter in Image Recognition?”, PLoS ONE, vol. 7, no. 1, p. e29740, 2012. Available: 10.1371/journal.pone.0029740.
- [9] Opencv-python-tutroals.readthedocs.io. 2020. Changing Colorspaces — Opencv-Python Tutorials 1 Documentation. [online] Available at:https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_colorspaces/py_colorspaces.html. 2020.
- [10] M. Hashemi, “Enlarging smaller images before inputting into convolutional neural network: zero- padding vs. interpolation”, Journal of Big Data, vol. 6,no. 1, 2019. Available: 10.1186/s40537-019-0263-7 . 2020.
- [11] S. Ghosh, N. Das and M. Nasipuri, “Reshaping inputs for convolutional neural network: Some common and uncommon methods”, Pattern Recognition, vol. 93, pp. 79-94, 2019. Available: 10.1016/j.patcog.2019.04.009.
- [12] R. Yamashita, M. Nishio, R. Do and K. Togashi, “Convolutional neural networks: an overview and application in radiology”, Insights into Imaging, vol. 9, no. 4, pp. 611-629, 2018. Available: 10.1007/s13244- 018-0639-9.
- [13] “Guideto the Sequential model - KerasDocumentation”, Faroit.com, 2020. [Online]. Available: <https://faroit.com/keras-docs/1.0.1/gettingstarted/sequential-model-guide/>. 2020.
- [14] Nwankpa, C., Ijomah, W., Gachagan, A. and Marshall, S., 2020. Activation Functions: Comparison Of Trends In Practice And Research For Deep Learning. [online]arXiv.org. Available at: <https://arxiv.org/abs/1811.03378>. 2020.
- [15] K. Team, “Keras documentation: MaxPooling2D layer”, Keras.io, 2020. [Online]. Available: https://keras.io/api/layers/pooling_layers/max_pooling2d/. 2020.
- [16] “prajnasb/observations”, GitHub, 2020. [Online]. Available:

<https://github.com/prajnasb/observations/tree/master/experiments/data>. 2020.

[17] "Face Mask Detection", Kaggle.com, 2020. [Online]. Available: <https://www.kaggle.com/andrewmvd/face-mask-detection>. 2020.

[18] "TensorFlow White Papers", TensorFlow, 2020.[Online]. Available: <https://www.tensorflow.org/about/bib>. 2020.

[19] K. Team, "Keras documentation: About Keras", Keras.io, 2020. [Online]. Available:<https://keras.io/about>. 2020.

[20] "OpenCV", Opencv.org, 2020. [Online]. Available: <https://opencv.org/>. 2020.

Author Profile:

	<p>Nannuri Suresh Associate Professor, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: nannurijs@gmail.com</p>
	<p>Bolineni Avinash, Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: bavinashongole@gmail.com</p>
	<p>Boina Venkata Ahalya Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: venkataahalya203@gmail.com</p>

	<p>Maddireddy Niharika Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: niharika.maddireddy19@gmail.com</p>
	<p>Muthineni Nounika Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: mouni14306@gmail.com</p>
	<p>Bodapati Sruthilaya Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: bsruthilaya02022000@gmail.com</p>

AUTOMATIC E-TIMETABLE GENERATING SYSTEM.

D.Ashok¹ , CH.Sowmya² ,K.Susmitha³ ,N.Divya⁴ ,T.Manasa⁵.

¹Associate Professor, Department of Information Technology, QIS College of Engineering & Technology, Ongole.

^{2,3,4,5} Students, Department of Information Technology, QIS College of Engineering & Technology, Ongole.

ABSTRACT

The manual system of preparing timetable in colleges is very time consuming and tedious task which usually ends up with various classes clashing either at identical room or with same teachers having more than one class at a time. Due to manual approach, proper use of resources is neither effective nor efficient. To overcome all these problems, we propose to make an automated system with computer assisted timetable generator.

INTRODUCTION

A great deal of time is devoted by the teaching personnel in generating and managing timetables. This project aims at the development of a tool which will allow institutes generate timetables for schools and colleges without any hindrance, directly from raw schedule. While generating a timetable, the availability of teachers and other resources is considered by this timetable generation software. Furthermore, timetables can be changed according to our necessity, depending on the availability of students, technicians, teachers, substitutes, classrooms and lessons.

The difficulty faced during timetabling can be represented as a constraint satisfaction problem with loose parameters and many constraints. These constraints can be replicated in a format which can be managed by the scheduling algorithm in an organized manner. The scheduling involves allowing for a many a pair wise constraints using which tasks can be accomplished simultaneously. For example, while scheduling classes in an organization, the same faculty member teaching two courses cannot be assigned the same time slot. On the other hand, two different courses to be attended by the same group of students also should not clash.

In order to deal with the timetabling issue, we are putting forward a system which would mechanically generate timetable for the different courses of the institute. Courses and lectures will be scheduled in accordance with all the possible constraints and the given inputs and thus, a timetable will be generated.

EXISTING SYSTEM

Normally timetable generation done manually. As we know all organizations have its own timetable, managing and maintaining these will not be difficult considering workload with this scheduling will make it more complex. Creating the timetable based on the organization requirements manually is very difficult and is very time consuming.

METHODOLOGY OF PROPOSED SYSTEM

Automatic timetable generator is a java-based software used to generate timetable automatically. It will help you to manage all the periods automatically and also will be helpful for faculty, admin, class students. In order to deal with the timetabling issue, we are putting forward a system which would mechanically generate timetable for the different courses of the institute. Courses and lectures will be scheduled in accordance with all the possible constraints and the given inputs and thus, a timetable will be generated. The system will allow interaction between the staff and students.

IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus, it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

MODULES

1. Faculty
2. Student
3. Admin

Module Description:

Faculty:

In this at first faculty will register and login with valid credentials. Then they can select the subjects and view preferred subjects and then view Assigned subjects and view the timetable and finally logout.

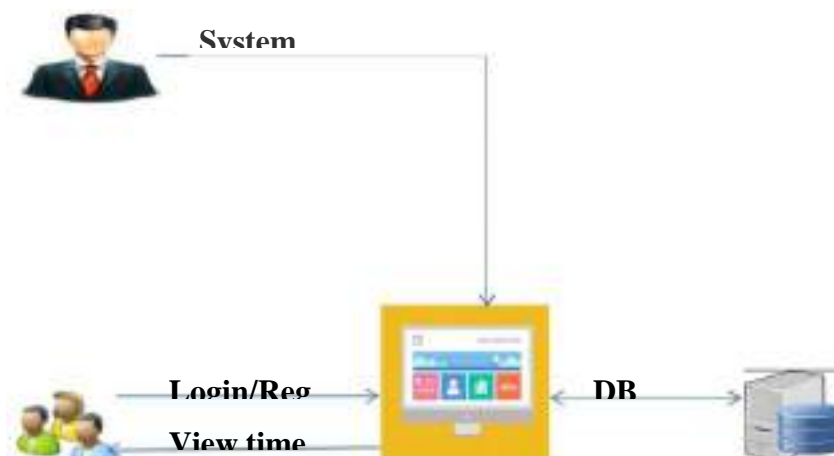
Student:

In this at first student will register and login with valid credentials. Then view the timetable along with faculty subjects and then logout.

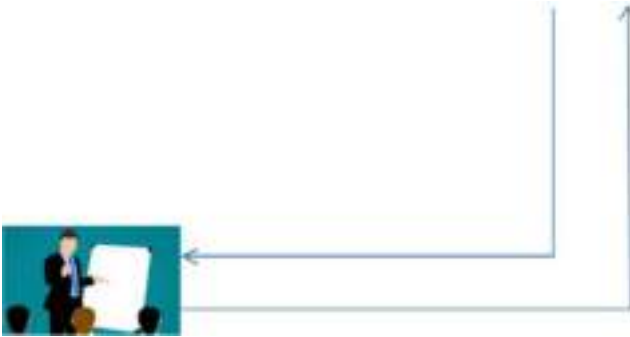
Administrator:

In this at first admin will login with valid credentials. Then admin can view faculty and add subjects. Admin can view the subjects and then generate timetable and assign faculty and finally logout.

SYSTEM ARCHITECTURE



STUDENTS AUTOMATIC E-TIME TABLE GENERATING SYTEM DATABASE



CONCLUSION

The application will make the procedure of timetable generation easier consistently which may otherwise need to be done using spread sheet manually which might lead to constraints problem that are strenuous to establish when timetable is generated physically. The purpose of the algorithm is to generate a timetable schedule mechanically. The algorithm includes many techniques, aimed at improving the efficiency of the search operation. It also addresses the chief hard constraints. Most of the non-rigid soft constraints are also productively handled. Keeping in mind the generality of the algorithm operation, it can further be modified to more particular scenarios, e.g. University, examination scheduling, etc. A number of hours which are spent on creating a fruitful timetable can be reduced ultimately through the mechanization of the timetable issue. The most fascinating future direction in the evolution of the algorithm lies in its addendum to constraint propagation.

FUTURE SCOPE

This project implements complex algorithms to generate automated time table for institutions ,colleges, Universities. The future work can be to create SMS alert for all Students and Faculty regarding holiday notification and exam notification and we can also create facility to faculty to apply for leaves and to upload E-notes and etc.,

We can able to show attendance percentage of students in their respective accounts in web sever and also availability of resources of

institutions like labs, class rooms, etc.

REFERENCES:

- [1] Anirudha Nanda, Manisha P. Pai, and Abhijeet Gole (August 2012), An Algorithm to Automatically Generate Schedule for School Lectures Using a Heuristic Approach, International Journal of Machine Learning and Computing, Vol. 2, No.2
- [2] E.K.Burke, J.P.Newall, Solving Examination Timetabling Problems through Adaptation of Heuristic Orderings
- [3] Masri Ayob, Salwani Abdullah and Ariff Md Ab Malik (September 2007), A Practical Examination Timetabling Problem at the University Kebangsaan Malaysia, IJCSNS International
- [4] D. Datta, Kalyanmoy Deb, Carlos M. Fonseca, —Solving Class Timetabling Problem of IIT Kanpur using Multi- Objective Evolutionary Algorithm. || KanGAL 2005.
- [5] Edmund K Burke, Barry McCollum, Amnon Meisels, Sanja Petrovic, Rong Qu, —A Graph-Based Hyper-Heuristic for Educational Timetabling Problems. || European Journal Operational Research, 176: 177-192, 2007.
- [6] Awad and Chinneck, —Proctor Assignment || at Carleton University (1998).
- [7] Cheang B., Li H., Lim A. and Rodrigues B. 2003, —Nurse Rostering Problems: A Bibliographic Survey. || European Journal of Operational Research, 151(3) 447-460.
- [8] Easton K., Nemhauser G. and Trick M. 2004, —Sports Scheduling. || In: Leung J.(ed.) in Handbook of Scheduling: Algorithms, Models, and Performance Analysis. Chapter 52, CRC Press.
- [9] S. and Burke E.K. 2004. University Timetabling In: Leung J. (ed.) —Handbook of Scheduling: Algorithms ||, —Models, and Performance Analysis. || Chapter 45. CRC Press.
- [10] W. Legierski, —Constraint-based Techniques for the University Course Timetabling Problem ||, CPDC, (2005), pp.59-63.
- [11] S. Abdullah, E. K. Burke and B. McCollum, —A Hybrid Evolutionary Approach to the University Course Timetabling Problem ||, Proceedings of the IEEE Congress Evolutionary Computation, Singapore, (2007).

ANDROID MALWARE DETECTION USING MACHINE LEARNING

K.Sreenath¹ D.Akhila² G.Jahnavi³ N.Preethi⁴ T.UdayaBhanu⁵,
V.Srikanthreddy⁶

¹Associate Professor, Department of Information Technology, QIS College of Engineering & Technology, Ongole.

^{2,3,4,5,6}Students, Department of Information Technology, QIS College of Engineering & Technology, Ongole.

Abstract: Malware is one of the major issues regarding the operating framework or in the software world. The android framework is also going through the same issues. We have seen other Signature based malware location strategies were utilized to recognize malware. Yet, the strategies couldn't recognize obscure malware. In spite of various discovery and analysis procedures are there, the discovery accuracy of new malware is as yet a crucial issue. In this paper, we study and feature the current identification and analysis techniques utilized for the android malicious code. Along with contemplating, we propose Machine learning algorithms that will be utilized to analyze suchmalware and also we will do semantic analysis. We will be having a data set of authorizations for malicious applications. Which will be compared with the consents extracted from the application which we want to analyze. Eventually, the client will actually want to perceive how much malicious authorization is there in the application and also we analyze the application through remarks.

1. Introduction

Malware is only the short name for malicious software, in general, alluded to many types of threatening or interruption creating software, spyware, Trojan horses, backdoor, and rootkits. The main aim of malware is to damage, steal, upset or then again do some bad actions. Malware is sufficiently incredible to contaminate any sort of figuring machine running application, and thecounteraction of malware is in effect all around read for personal PCs (PC). A Smartphone gadget the location procedures utilized is lagging far behind as compared tothe fast development of the versatile

population is being Some new study has shown that there are about 2.1 million android applications are there in the market. Because of increase in usage of the android framework has prompted more rollout of android malware. This malware is spreading in the market by the outsiders creating applications. The Google android market also doesn't vow to guarantee that all the applications recorded are without threat. There are also such reports about Trojans applications that whenever downloaded, their malicious code is also installed and cannot be easily identified by Google's advancements during publication in the Google android market. The android threats incorporate banking Trojans, spyware, bots, root abuses, SMS fraud, phishing and fake installer. Android Apps are unreservedly available on Google Playstore, the official Android app store as well as outsider app stores for clients to download. Because of its open source nature and popularity, malware authors are increasingly zeroing in on creating malicious applications for Android operating framework. Despite various attempts by Google Playstore to ensure against malicious apps, they actually discover their way to mass market and cause harm to clients by abusing personal information related to their telephone directory, mail accounts, GPS location information and others for abuse by outsiders or else take control of the telephones distantly. Consequently, there is need to perform malware analysis or figuring out of such malicious applications which present genuine threat to Android platforms. Broadly speaking, Android Malware analysis is of two types: Static Analysis and Dynamic Analysis. Static analysis basically includes analyzing the code structure without executing it while dynamic analysis is examination of the runtime behavior of Android Apps in constrained climate. Offered in to the always increasing variants of Android Malware presenting zero-day threats, a productive mechanism for discovery of Android malwares is required. In contrast to signature-based approach which requires regular update of signature database, machine-learning based approach in combination with static and dynamic analysis can be utilized to identify new variants of Android Malware presenting zero-day threats. In [5], broad yet lightweight static analysis has been performed achieving a good discovery accuracy of 94% utilizing Support Vector Machine algorithm. Nikola Milosevic et al. [6]

introduced static analysis based classification through two strategies: one was consent based while the other included representation of the source code as a bag of words. Another approach based on recognizing most significant consents and applying machine learning on it for evaluation has been proposed in [7-11]. The main commitment of the work is decrease of feature measurement to not exactly half of original feature-set utilizing Genetic Algorithm with the end goal that it tends to be taken care of as contribution to machine learning classifiers for training with diminished intricacy while maintaining their accuracy in malware classification. In contrast to exhaustive technique for feature choice which requires testing for 2^N various combinations, where N is the number of features, Genetic Algorithm, a heuristic searching approach based on wellness work has been utilized for feature choice. The upgraded feature set obtained utilizing Genetic algorithm is utilized to train two machine learning algorithms: Backing Vector Machine and Neural Network. It is noticed that a respectable classification accuracy of over 94% is maintained while working on a much lower feature measurement, along these lines, lessening the training time intricacy of classifiers.

2. Background Work

There are various ways and techniques through malware or any malicious record can enter your framework or application. A portion of the basic strategies of malware getting interfered into the framework are as per the following: -

- Penetration: Penetration procedures generally utilized for malware applications for installation activation and running on the android framework are repackaging, updating and downloading.
- Repackaging: It is among the regular procedures for malware engineers to install malicious applications on an android platform. Repackaging approach for popular applications and abuse them as malware. The engineer downloads such sorts of application and recodes them and adds their own malicious code and uploads that application to the official Android app store or on the various markets.

- **Updating:** This procedure is considerably more hard for distinguishing malware. The malware engineer may in any case utilize repackaging however instead of encoding deliver code to the application the designer may incorporate an update part that will able to download malicious code at the run time.
- **Downloading:** This is the most traditional attacking strategy. The malware designer needs to attract the client to download intriguing and attractive applications

3. Literature Review

In “Android Malware Detection Using Machine Learning on Image Patterns “[1], the paper was distributed in the year 2018. They have played out the malware discovery with the assistance of 300 malware records and 300 kindhearted apk documents, also they managed to generate just 183 malware and 300 favorable gray-scale images. The other 117 malware samples were unable to generate into images because the apk documents were tainted or either that records didn't contain classes.dex document. Also, the accuracy was a lot less in all the algorithms they utilized. They have recognized with the help of three diverse classifier strategies namely the k nearest neighbor (KNN), Random Forest (RF), and Decision Tree(DT).

In “Android mobile security by detecting and classification of malware based onpermissions using machine learning algorithms “[2], the paper was distributed in the year 2017. They had utilized diverse machine learning algorithms like Naive Bayes,j48, random woodland, Multiclass classifier and multilayer perceptron to recognize android malware and evaluate the performance of each algorithm. Here they executed a framework for classifying android applications with the assistance of the machine learning strategies to check whether it is a malware or normal application. For validating their framework they have gathered 3258 samples of android apps and those have to be extracted for each application, extract their features and have totrain the models going to be evaluated with the assistance of classification accuracy and time taken for the model.

In “An Android Behaviour-Based Malware Detection Method using Machine Learning”[3], the paper was distributed in the year 2016. They have proposed a Robotium program in an Android sandbox that can trigger any android application automatically and screen its behavior. The program has a UIIdentification automatic trigger program that can tap the portable applications in a meaningful request. The program was able to perform largerscale tests. They also attempted to construct a choice model utilizing behavior that has gathered with the assistance of the random woodland algorithm. It has had the option to decide if the obscure application is malware and also shows its certainty value. They could store the outcome and also the certainty value of the obscure apk document in their database.

In “RanDroid: Android malware detection using random machine learning classifiers”[4], the paper was distributed in the year 2018. They have proposed the android malware identification framework with the help of consents, APIs, and also with the presence of distinctive key apps information, for example, the dynamic code, Reaction code, native code, cryptographiccode, database, and so on as the feature to train and fabricate classification model just by utilizing various machine learning methods which can automatically recognize malicious Android apps(Malware) from the legitimate ones.

4. System Analysis

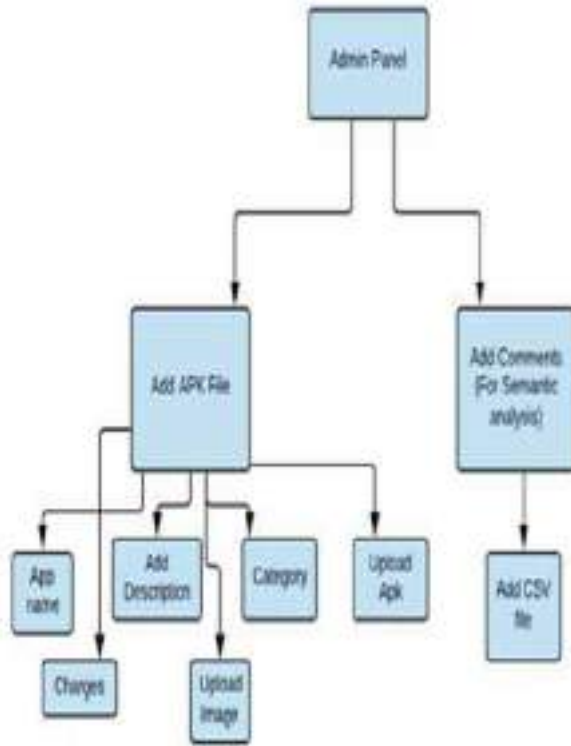
Existing System

The main contribution of the work is reduction of feature dimension to less than half of original feature-set using Genetic Algorithm such that it can be fed as input to machine learning classifiers for training with reduced complexity while maintaining their accuracy in malware classification. In contrast to exhaustive method of feature selection which requires testing for 2^N different combinations, where N is the number of features, Genetic Algorithm, a heuristic searching approach based on fitnessfunction has been used for feature selection. The optimized feature set obtained using Genetic algorithm is used to train two machine learning algorithms: Support Vector Machine and Neural Network. It is

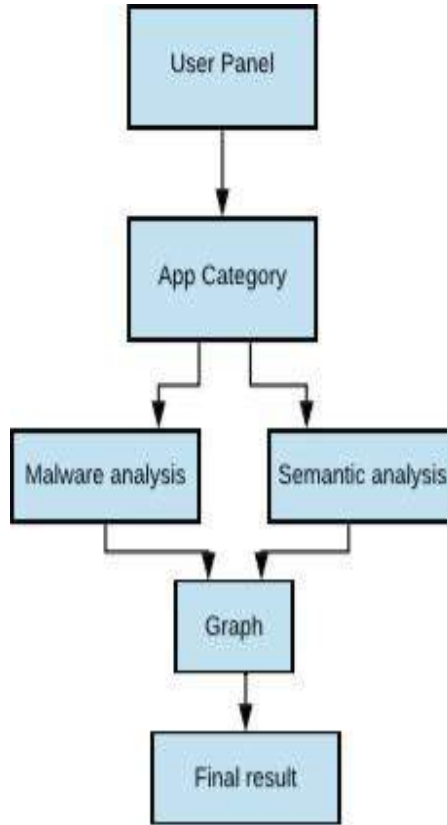
observed that a decent classification accuracy of more than 94% is maintained while working on a much lower feature dimension, thereby, reducing the training time complexity of classifiers.

Proposed system :-

Two set of Android Apps or APKs: Malware/Goodware are reverse engineered to extract features such as permissions and count of App Components such as Activity, Services, Content Providers, etc. These features are used as feature vector with class labels as Malware and Goodware represented by 0 and 1 respectively in CSV format. To reduce dimensionality of feature-set, the CSV is fed to Genetic Algorithm to select the most optimized set of features. The optimized set of features obtained is used for training two machine learning classifiers: Support Vector Machine and Neural Network. In the proposed methodology, static features are obtained from AndroidManifest.xml which contains all the important information needed by any Android platform about the Apps. Androguard tool has been used for disassembling of the APKs and getting the static features. In our system, we have implemented an admin panel as well as a user panel. In the admin panel admin have the access to upload the apk files and its details along with its categorization and also the admin can upload the comment that can be used for semantic analysis. In the user-panel the user can see the select the category of the application and can see its details like pricing description name. User can see the malicious percentage of the application. And the processed output of the semantic analysis will be displayed to the user in the form of graph and the user will get a proper review of the application.



Admin Panel 1



User Panel 1

5. Results

The Malware Detection can recognize a widerange of consents in view of the which it hasbeen asked and furthermore which of the consents which it has been taken of course. Likewise, the semantic investigation is been utilized to get the appropriate remarks resultant it in to get if the application is beenappropriate or not. The consent based examination and furthermore the semantic investigation gives the appropriate yield withthe goal that the client can utilize those specific applications or on the other hand not.

6. Conclusion

In our work, we propose a framework for authorization

examination and semantic investigation. Our framework is additionally used to identify malware authorizations dependent on an application by contrasting it and a dataset. This proposed framework can be applied in the fields of the security framework and furthermore for the n clients like a malware discovery programming. Nonetheless, there are limits in our framework. The authorizations which we are characterizing are according to our however it can vary from clients to clients. The consents which the client likes that it's anything but a malware-based can be malware for some other client. Future work will contain the improvement of that.

References

- [1] Darus, Fauzi Mohd, Salleh Noor Azurati Ahmad, and Aswami Fadillah Mohd Ariffin. "Android Malware Detection Using Machine Learning on Image Patterns" 2018 Cyber Resilience Conference (CRC). IEEE, 2018.
- [2] Vrama, P. Ravi Kiran, Kotari Prudvi raj, and KV Subba Raju. "Android mobile security by detecting and classification of malware based on permissions using machine learning algorithms." 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE, 2017.
- [3] Chang, Wei-Ling, Hung-Min Sun, and Wei Wu. "An Android Behaviour-Based Malware Detection Method using Machine Learning." 2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). IEEE, 2016.
- [4] Koli, J. D. "RanDroid: Android malware detection using random machine learning classifiers." 2018 Technologies for Smart-City Energy Security and Power (ICSESP). IEEE, 2018.
- [5] S. Arshad, M. A. Shah, A. Wahid, A. Mehmood, H. Song, and H. Yu, "SAMADroid: A Novel 3-Level Hybrid Malware Detection Model for Android Operating System," *IEEE Access*, vol. 6, pp.4321–4339, 2018.
- [6] T. Kim, B. Kang, M. Rho, S. Sezer, and E. G. Im, "A Multimodal Deep Learning Method for Android Malware Detection using Various Features," vol. 6013, no. c, 2018
- [7] A. Martin, F. Fuentes-Hurtado, V. Naranjo, and D. Camacho, "Evolving Deep Neural Networks architectures for Android malware

classification," *2017 IEEE Congr. Evol. Comput. CEC 2017 - Proc.*, pp. 1659–1666, 2017.

[8] X. Su, D. Zhang, W. Li, and K. Zhao, "A Deep Learning Approach to Android Malware Feature Learning and Detection," *2016 IEEE Trust.*, pp. 244–251, 2016.

[9] K. Zhao, D. Zhang, X. Su, and W. Li, "Fest : A Feature Extraction and Selection Tool for Android Malware Detection," *2015 IEEE Symp. Comput. Commun.*, pp. 714–720, 4893.





[10] A. Feizollah, N. B. Anuar, R. Salleh, and A. W. A. Wahab, "A review on feature selection in mobile malware detection," *Digit. Investig.*, vol. 13, pp. 22–37, 2015.

[11] A. Firdaus, N. B. Anuar, A. Karim, M. Faizal, and A. Razak, "Discovering optimal features using static analysis and a genetic search based method for Android malware detection *," vol. 19, no. 6, pp. 712–736, 2018.

[12] A. V. Phan, M. Le Nguyen, and L. T. Bui, "Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems," *Appl. Intell.*, vol. 46, no. 2, pp. 455–469, 2017.

Authors Profile

	<p>K. Sreenath Associate Professor, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: srinathits@gmail.com</p>
	<p>D. Akhila Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: akhilaakhila74936@gmail.com</p>

	<p>G. Venkata Sai Jahnvi Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: gontlajahnvi@gmail.com</p>
	<p>N. Preethi Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: preethinalluri123@gmail.com</p>
	<p>T. Naga Udaya Bhanu Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: bhanu.thottempudi@gmail.com</p>
	<p>V. Srikanth Reddy Student, Department of Information Technology, QIS College of Engineering and Technology, Ongole Email: vallabasrikanthreddy@gmail.com</p>

A COMPARISON APPROACH IN IMAGE OPTIMIZATION TECHNIQUES USING MNIST HANDWRITTEN DIGIT DATASET

N.SaiKiran¹, N.NagaMounika², N.V.Sahithi³, G.Vineela⁴,

T.Navya Haritha⁵, V.Rakesh⁶

¹AssociateProfessor, Department of Information Technology, QISCET,
Ongole

^{2,3,4,5,6} IV B.Tech Students, Department of Information Technology
,QISCET, Ongole

ABSTRACT: Convolutional neural networks have achieved progressive performance in a variety of computer vision problems such as image classification. The task of hand written digit recognition using a classifier and it has great importance and use such as online handwriting recognition on computer tablets. The performance of deep neural network strained for high-level computer vision tasks such as classification degrades under noise, blur and other imperfections present in raw image data. Cleaning up raw data using convention allow-level image processing does not essentially improve performance. Hence the proposed work is to identify the best optimization techniques for finding the optimal solution in image classification. This project evaluates various optimization techniques for fine-grained image classification. The main objective of this project is to implement various optimization techniques for better analysis of pattern in images, where the neural networks are pre-trained with dataset to boost the performance with higher accuracy. The results will demonstrate the effectiveness of optimization technique over MNIST dataset. The experimental findings will helps to obtain rained models with improved accuracy.

INTRODUCTION

Image Processing:

Image process is usually a stimulating field because it offers improved pictorial info for human interpretation and process of image knowledge for storage, transmission, and illustration for machine perception. Image process may be a technique to boost raw pictures received from

cameras/sensors placed on satellites, house probes and aircrafts or footage taken in traditional day- after-day life for varied applications. This field of image process considerably improved in recent times and extended to numerous fields of science and technology. In image process the main deals with image acquisition, Image improvement, image segmentation, feature extraction, image classification etc.

Image Classification:

1.4 Convolutional Neural Network:

Individual neurons in the section of brain respond to stimulate only in a restricted region of the visual field known as the receptive field. The different neurons overlap, together they make the entire visual field. This effectively means that the certain neurons were activated and it is a certain attribute in the visual field, for example, horizontal edge. Convolutional Neural Network (CNN) is a class of deep, feed- forward traditional artificial neural networks. The machine learning models that extend the traditional neural network by adding increased depth and additional constraints to the early layers.

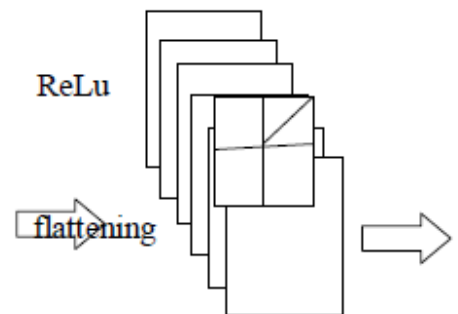
The work has focused on tuning their architecture to achieve maximum performance on benchmarks such as the Image Net Large Scale Visual Recognition Challenge (ILSVRC). They can trace their origins and that is back to the early 1980s with Fukushima's Neocognitron. More directly, they were shown to be highly effective in the 1990s. When they used for handwritten digit recognition and eventually in industry for automated check readers. The technique of Google researchers used is called Convolutional Neural Networks (CNN) and it is also one of the type of advanced artificial neural network. It differs from the regular neural networks in terms of the flow of signals between neurons. The neural networks passes signals were the input-output channel in a single direction, without allowing signals to loop back into the network. This is called a feed-forward.

Convolutional Neural Networks (**ConvNets** or **CNNs**) are the category of Neural Networks. The neural networks have to proven very effective in areas such as image recognition and classification. ConvNets have

been successfully identify the faces, objects and traffic signs apart from powering vision in robots and self driving cars. The ConvNets are an important tool for most machine learning practitioners. The understanding of ConvNets and learning to use them for the first time can sometimes be an intimidating experience. The primary purpose of these blog post is to develop an understanding of how the Convolutional Neural Networks work on images.

A **Convolutional Neural Network(ConvNet/CNN)** is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0



Convolutional Neural Network processes image through following layers

- Convolutional Layer – Used to detect features
- Non-Linearity Layer –Introducing non-linearity to the system
- Pooling (Downsampling) Layer –Reduces the amount of weights and controls overfitting
- Flattening Layer – Prepares information for Classical Neural Network
- Fully-Connected Layer – normal Neural Network used for classification.

Input image
Pooling layer
layer

Convolutional layer
Input to final

Fig: Convolutional neural network

PROBLEM IDENTIFICATION

- There are different challenges faced while attempting to solve this problem. The handwritten digits are not always of the same size, thickness, or orientation and position relative to the margins.
- Our goal was to implement a pattern classification method to recognize the handwritten digits provided in the MNIST data set of images of hand written digits (0-9).
- The general problem to predict would face in this digit classification problem was the similarity between the digits like 1 and 7, 5 and 6, 3 and 8, 9 and 8 etc.
- The main objective of this project, that varying learning rate known as parameter fine tuning can improve result, where in a networks are pre-trained on a dataset can boost the performance of these networks.

SYSTEM ARCHITECTURE

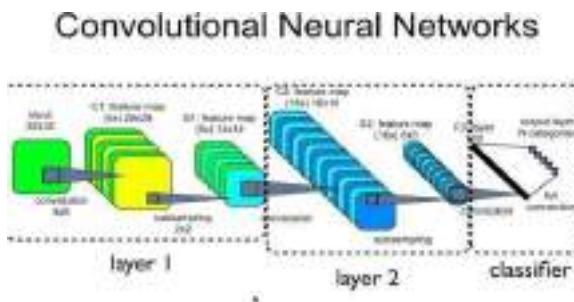


Fig: Overall Architecture

METHOD DESCRIPTION

MNIST dataset:

Hand writing recognition of characters has been around since the 1980s. The task of handwritten digit recognition, using a classifier, has great

importance and use such as – online handwriting recognition on computer tablets, recognize zip codes on mail for postal mail sorting, processing bank check amounts, numeric entries in forms filled up by hand (for example - tax forms) and so on.

The goal is to implement a pattern classification method to recognize the handwritten digits provided in the MNIST data set of images of hand written digits (0- 9). The data set used for our application is composed of 1500 training images and 500 testing images, and is a subset of the MNIST data set [1] (originally composed of 60,000 training images and 10,000 testing images). Each image is a 32x32 grayscale (0-255) labeled representation of an individual digit.

The general drawback we have a tendency to foresee we might face during this digit classification drawback was the similarity between the digits like one and seven, 5 and 6, 3 and 8, nine and eight etc. conjointly folks write an equivalent digit in many alternative ways in which - the digit '1' is written as '1', '1', '1' or '1'. equally seven is also written as seven, 7, or 7. Finally the individuality and selection within the handwriting of various people conjointly influences the formation and look of the digits.

Example MNIST dataset is,



Fig: 8.1.1 Example MNIST

Dataset - Load Data

The MNIST data-set is about 12 MB and will be downloaded automatically if it is not located in the given path. The MNIST data-set has now been loaded and consists of images and class-numbers for the images. The data-set is split into 3 mutually exclusive sub-sets.

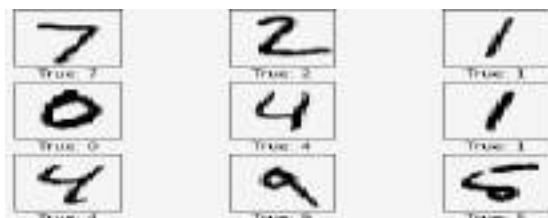


Fig: Example Dataset

Convolutional Layers:

This is the basic idea of image processing in the first convolutional layer for handwritten images. The input image depicts the number 7 and four copies of the image are shown here, so we can see more clearly how the filter is being moved to different positions of the image. For each position of the filter, the dot-product is being calculated between the filter and the image pixels under the filter, which results in a single pixel in the output image. So moving the filter across the whole input image leads to a brand new image being generated.

The red filter-weights mean that the filter has a positive reaction to black pixels in the input image, while blue pixels means the filter has a negative reaction to black pixels.

In this case it appears that the filter recognizes the horizontal line of the 7-digit, as can be seen from its stronger reaction to that line in the output image.

Convolutional Layer 1:

Create the first convolutional layer. It takes `x_image` as input and creates `num_filters1` different filters, each having width and height equal to `filter_size1`. Finally we wish to down-sample the image so it is half the size by using `2x2` max-pooling.

Convolutional Layer 2:

Create the second convolutional layer, which takes as input the output from the first convolutional layer. The number of input channels corresponds to the number of filters in the first convolutional layer.

Flatten Layer:

The convolutional layers output 4-dim tensors. We now wish to use these as input in a fully-connected network, which requires for the tensors to be reshaped or flattened to 2-dim tensors.

Fully-Connected Layer 1:

Add a fully-connected layer to the network. The input is the flattened layer from the previous convolution. The number of neurons or nodes in the fully-connected layer is `fc_size`. ReLU is used so we can learn non-linear relations.

Fully-Connected Layer 2:

Add another fully-connected layer that outputs vectors of length 10 for determining which of the 10 classes the input image belongs to. Note that ReLU is not used in this layer.

Predicted Class:

The second fully-connected layer estimates how likely it is that the input image belongs to each of the 10 classes. However, these estimates are a bit rough and difficult to interpret because the numbers may be very small or large, so we want to normalize them so that each element is limited between zero and one and the 10 elements sum to one. This is calculated using the so-called softmax function and the result is stored in `y_pred`.

Cost-function to be optimized:

To make the model better at classifying the input images, we must somehow change the variables for all the network layers. To do this we first need to know how well the model currently performs by comparing the predicted output of the model `y_pred` to the desired output `y_true`.

The cross-entropy is a performance measure used in classification for image optimization. The cross-entropy may be a continuous operate that's perpetually positive and if the anticipated output of the model specifically matches the required output then the cross-entropy equals zero. The goal of optimization is therefore to minimize the cross-entropy so it gets as close to zero as possible by changing the variables of the network layers.

TensorFlow has a built-in function for calculating the cross-entropy in image classification. Note that the `tf.nn.softmax` calculates the `tf.nn.softmax` internally so we tend to should use the output of `layer_fc2` directly instead of `y_pred` that has already had the softmax applied.

Create the first convolutional layer. It takes `x_image` as input and creates `num_filters1` different filters, each having width and height equal to `filter_size1`. Finally we wish to down-sample the image so it is half the size by using `2x2` max-pooling.

Convolutional Layer 2:

Create the second convolutional layer, which takes as input the output from the first convolutional layer. The number of input channels corresponds to the number of filters in the first convolutional layer.

Flatten Layer:

The convolutional layers output 4-dim tensors. We now wish to use these as input in a fully-connected network, which requires for the tensors to be reshaped or flattened to 2-dim tensors.

ALGORITHM DESCRIPTION

Step 1: The first step is to import the classes and functions needed.

Step 2: To initialize the random number generator to a constant to ensure that the results of your script are reproducible.

Step 3: To normalize the pixel values to the range 0 and 1 and one hot encode the output variables.

Step 4: The first hidden layer is a convolutional layer called a Convolution2D. The layer has 32 feature maps, which have the size of 5×5 and a rectifier activation function. This is the input layer, expecting images with the structured outline above [pixels][width][height].

Step 5: Next we define a pooling layer that takes the max called MaxPooling2D. It is configured with a pool size of 2×2 with optimized methods.

Step 6: The next layer is a regularization layer using dropout called Dropout. It is configured to randomly exclude 20% of neurons in the layer and it is order to reduce overfitting.

Step 7: Next is a layer that converts the 2D matrix data to a vector called Flatten. It allows the output to be processed by standard fully connected layers in convolutional layer.

Step 8: Next a fully connected layer with 128 neurons and rectifier activation function.

Step 9: Finally, the output layer has 10 neurons for the 10 classes and a softmax activation function to output probability-like predictions for each class.

OPTIMIZATION ALGORITHMS:

The improvement techniques will effectively cut back the coding time whereas holding the standard of the retrieved. Numerous

improvement techniques area unit explained below:

Adagrad optimizer:

It merely permits the training Rate $-\eta$ to adapt supported the parameters. thus it makes massive updates for sporadic parameters and little updates for frequent parameters. For this reason, it's well-suited for managing thin knowledge.

It uses a unique learning Rate for each parameter θ at a time step supported the pastgradients that were computed for that parameter.

Previously, we tend to performed associate update for all parameters θ promptly as each parameter $\theta(i)$ used constant learning rate η . As Adagrad uses a unique learning rate for each parameter $\theta(i)$ at each time step t , we tend to initial show Adagrad's per-parameter update, that

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} \cdot g_{t,i},$$

we tend to then vectorize. For brevity, we set $g(t,i)$ to be the gradient of the loss perform w.r.t. to the parameter $\theta(i)$ at time step t .

Momentum optimizer:

The high variance oscillations in SGD makes it hard to reach convergence, so a technique called *Momentum* was invented which *accelerates SGD* by navigating along the relevant direction and softens the oscillations in irrelevant directions. In other words all it does is add a fraction ' γ ' of the update vectorsof the past step to the current update vector.

$$\mathbf{V}(t) = \gamma \mathbf{V}(t-1) + \eta \nabla J(\theta).$$

and finally we update parameters by

$$\theta = \theta - \mathbf{V}(t).$$

The same thing happens with our parameterupdates —

1. It leads to faster and stableconvergence.
2. Reduced Oscillations

The momentum term γ will increase for dimensions whose gradients purpose within the same directions and reduces updates for dimensions whose gradients amendment directions. this implies it will parameter updates just for relevant examples. This reduces the unnecessary parameter

updates which leads to faster and stable convergence and reduced oscillations.

RMSProp optimizer:

RMSprop, or Root Mean sq. Propagation has a remarkable history. it absolutely was devised by the legendary Geoffrey Hinton, whereas suggesting a random plan throughout a Coursera category.

RMSProp additionally tries to dampen the oscillations, however in an exceedingly completely different manner than momentum. RMS prop additionally takes away the requirement to regulate learning rate, and will it mechanically. More so, RMSProp choses a special learning rate for every parameter. In RMS prop, every update is completed in keeping with the equations represented below. This update is completed individually for every parameter.

Adam optimizer:

Adam stands for adjustive Moment Estimation. adjustive Moment Estimation (Adam) is another methodology that computes adjustive learning rates for every parameter. additionally to storing AN exponentially decaying average of past square gradients like AdaDelta , Adam additionally keeps AN exponentially decaying average of past gradients $M(t)$, kindof like momentum: $M(t)$ and $V(t)$ are unit values of the primary moment that is that the

Mean and also the moment that is that the uncentered variance of the gradients severally. Adam works well in observe and compares favourably to different adjustive learning- method algorithms because it converges in no time {and the|and therefore the|and additionally the} learning speed of the Model is quiet quick and economical and also it rectifies each drawback that's two-faced in different optimisation techniques like vanishing Learning rate, slow convergence or High variance within the parameter updates that results in unsteady Loss operator.

Adadelta optimizer: It is an extension of AdaGrad that tends to get rid of the decaying learning Rate downside of it. Rather than accumulating all previous square gradients, Adadelta limits the window of accumulated

past gradients to some mounted size w . Instead of inefficiently storing w previous square gradients, the ad of gradients is recursively outlined as a decaying mean of allpast square gradients. The running average $E[g^2](t)$ at time step t then depends (as a fraction γ equally to the Momentum term) solely on the previous average and therefore the current gradient.

$$E[g^2](t) = \gamma \cdot E[g^2](t-1) + (1-\gamma) \cdot g^2(t)$$

We set γ to an identical price because the momentum term, around 0.9. Enhancements we've got done thus far —

1. We tend to are conniving completely different learning Rates for every parameter.
2. We tend to also are conniving momentum.
3. Preventing Vanishing (decaying) learning Rates.

RESULT AND DISCUSSION:

Num_iterations	Optimization techniques	Accuracy
75	RMSProp	98.45
85	Adam	98.96
65	Adagrad	88.33
74	Momentum	97.74
53	Adadelta	61.3

Table:Accuracy

From the above table Adam optimizer have the highest accuracy and Adadelta have the very lowest accuracy.

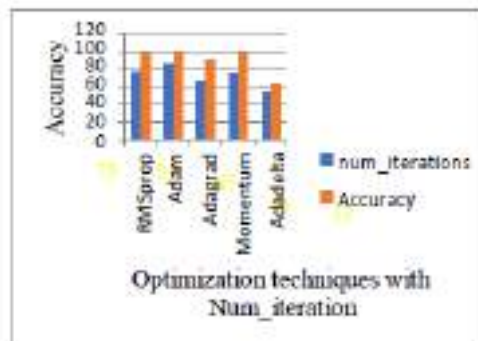


Fig: Accuracy

CONCLUSION & FUTURE ENHANCEMENT

Written digits (0-9). The main objective of this project, that varying learning rate known as parameter fine tuning can improve result, where in a networks are pre-trained on a dataset can boost the performance of these networks. It achieves much greater registration accuracy. The results will demonstrate the effectiveness of various optimization techniques over MNIST dataset. Finally, the uniqueness and variety in the handwriting digits of the different individuals are also influences the formation and appearance of the digits. The experimental findings will helps to train models with improved accuracy. In future, in our proposed system we compare only five Optimization Techniques. Further, we can use more Algorithms for finding the best optimization with highest accuracy.

REFERENCES

- [1] S.S Panda, etal[2012], " A Comparison Approach in Image Optimization Techniques", international journal of research in computer and communication technology, vol1, issue 1, pp [1-4]
- [2] Nicholas Becherer, John Pecarina, Scott Nykl, Kenneth Hopkinson[2017], " Improving Optimization of Convolutional Neural Networks through parameter fine- tuning", Neural Computing and Applications, pp 1-11
- [3] Ujjwalkam[2016], " An Intutive Explanation of Convolutional Neural Networks", on the data science blog
- [4] Nikola M. Zivkovic[2018], " Introduction In this project, the proposed system to identify the best optimization techniques for de-noising, de-blurring and it makes classification robust to realistic noise and blur. This project evaluates various optimization techniques for fine-grained image classification. The goal was to be implementing a pattern classification method to recognize the handwritten digits provided in the MINIST data set of images of hand to Convolutional Neural Networks", in artificial intelligence and machine learning technology
- [5] Xianli Zou, etal[2018], " Fast convergent capsule network with applications in MNIST", international publishing AG, pp [3-10]
- [6] Manju Devi, Uma Mehta[2016], " A Review on Various Techniques of Image Compression", International Journal Of Engineering And

- [7] Szegedy C, Liu W, Jia Y, Sermanet P[2014], "Going deeper with convolutions", arXiv Prepr.arXiv 1409.4842
- [8] Tianmei Guo, Jiwen Dong,etal[2017], " Simple convolutional neural network on image classification", IEEE 2nd international conference on Bigdata analysis (ICBDA)
- [9] I.Sato, H. Nishimura, K. Yokoi[2015], "APAC: Augmented PAttern Classification with Neural Networks", arXiv:1505.03229v1
- [10] ShwetaV.Jain,etal[2013], "ImageOptimization and Prediction", in cornell university library arxiv.org/ftp/arxiv/1305.2828
- [11] Qianru Zhang, Meng Zhang, etal[2018], " Recent Advances in Convolutional Neural Network Acceleration", in neurocomputing the Research Grants Council of Hong Kong
- [12] K. He, X. Zhang, S. Ren, J. Sun[2016], "Deep residual learning for image recognition", in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp [770–778]
- [13] Jia Qu , Nobuyuki Hiruta, etal [2018] ," Gastric Pathology Image Classification Using Stepwise Fine-Tuning for Deep Neural Networks", Journal of Healthcare Engineering Volume 2018, <https://doi.org/10.1155/2018/8961781>
- [14] Bharath P.T, etal [2011], " A Review on Optimization of image processing techniquesusing Neural Networks", in information Science and Application
- [15] Krizhevsky A, Sulskever I, Hinton GE[2012], " ImageNet classification with deep convolutional neural networks", in Advances in neural information processing system on pp 1-9
- [16] B. Chitradevi, P.Srimathi [2014], " An overview on image processing techniques", International Journal of innovative Research in Computer and CommunicationEngineering vol-2 issue 11
- [17] Olga Russakovsky, Jia Deng [2015], " ImageNet Large Scale Visual RecognitionChallenge" arXiv:1409.0575v3 [cs.CV] on30 Jan 2015
- [18] Diederik P. Kingma, etal [2017], "Adam: A Method For Stochastic Optimization" arXiv: 1412.6980v9 [cs.LG]
- [19] Max Jaderberg, etal [2014], "Speeding up Convolutional Neural

Authors Profile

	<p>N. Sai Kiran, Associate Professor Department of Information Technology QIS College of Engineering & Technology, Ongole Gmail: neelam.saikiran534@gmail.com</p>
	<p>N. Naga Mounika Student Department of Information Technology QIS College of Engineering & Technology, Ongole Gmail: meghana.narahari11@gmail.com</p>
	<p>T. Navya Haritha Student Department of Information Technology QIS College of Engineering & Technology, Ongole Gmail: tellakulanavyaharitha@gmail.com</p>

	<p>N.V. Sahithi Student Department of Information Technology QIS College of Engineering & Technology, Ongole Gmail: sahithinuthanki27@gmail.com</p>
	<p>G. Vineela Student Department of Information Technology QIS College of Engineering & Technology, Ongole Gmail: Vineelagaddam12@gmail.com</p>
	<p>V. Rakesh Student Department of Information Technology QIS College of Engineering & Technology, Ongole Gmail: rakeshraki9989@gmail.com</p>