MACHINE LEARNING IMPLEMENTATION FOR SMART HEALTH RECORDS: A DIGITAL CARRY CARD

ABHISHEK KUMAR[*], TVM SAIRAM[**], VISHAL DUTT[***]

ABSTRACT

Machine learning is a high-level approach for any kinds of health care implementation in this real-world scenario. We need to implement machine learning methodologies to identify the best-predicted values related to the patients in their respected health condition and also need to analyze the previous health records. For that, we need to maintain a repository or the warehouse where we need to maintain digital data related to the patients and their treatment. For that in this article, we are proposing an application using which we can maintain the health records of the patients using a digital card which can be used only by the doctor, receptionist and the respected other departments in the hospital. We implemented this prediction methodology using the responsive web application and the remaining things like predictions and the statistical analysis with the advanced methodologies like machine learning and python. In this article, we impose some machine learning methodologies on our data and will find out the optimal solution for the process of treatment and the good maintenance of the digital records of the patients. The records we maintain will be in sequential order and there is a chance of modification of the data in further times on admin privileges. All these kind of information was stored in the form of application and we explained the procedure of machine learning usage in this application and the prediction models design and implementation.

KEYWORDS: Machine Learning, Healthcare, Statistical Analysis, Prediction, Methodology.

INTRODUCTION

Machine learning is widely used technology in the field of science and technology and the rapid growth of the requirements of the people in their daily life improves the utilization of advanced technologies and they used to rule our daily life. Perhaps, consider medical industry. We need to identify the patient health information and in this kind of situation, we need an application which can monitor the patient's health records from the client side and to the doctor's side. We need a proper channel to design application and have to maintain a model which can identify the predictions and insights of the data.

[*]Assistant Professor, CSE Department ACERC, Ajmer.

[**]Assistant Professor, Department of IT, ALIET, Vijayawada-520010.

[***]Lecturer, Computer Science Department Dec, Ajmer.

Correspondence E-mail Id: editor@eurekajournals.com

As a data scientist people look more into the insights of the data and the forms of data which we are gathering and the importance of maintaining every feature in our models. This paper will discuss the process of gathering information and also we need to identify the process of establishing a counter for data to get predicted. In this article, we are dealing with machine learning models like decision trees and random forest algorithms which are more likely to be used for different prediction models design and implementations. First of all, we need to identify them using some of the machine learning algorithms and maintain the accuracy rate using confusion matrix[1-5].

In this article the following things to be noted down for better research of this concept further. They are as follows:

DATA COLLECTION

We need to collect the data from some sources in our environment and in our circumstances. We need to create a repository which will be useful for maintaining the stability in gathering the information from other sources. In collection of information we may get noisy data and we need to remove them using pre-processing methodology. This process will maintain the stability and ambiguous information in the repository[6-9]. But the main question here is how to get the information from other sources and what are the sources. For this question, the answer will be in the next sub-section[10-12].

MEDICAL APPLICATION

We created an application using full stack development for digital health records maintenance. For this we need to perform a survey on other things which are there related to the information gathering of the health records. In the literature review section we will explain the complete survey information we gathered and based on the base paper we considered we are designing a best way to maintain the health records in the digital format. For that purpose we created an application with which we need to maintain the stability of data gathering.

BACKWARD ELIMINATION

It is a nested operator with which we need to carry out the process of eliminating the unwanted featured in the dataset and the information related to P and SL values will be generated based on the features we consider. Based on that information, if the value of the P is greater than the SL value, then the information can be eliminated from the repository. Because they can give the lowest accuracy of the model which we are using

MACHINE LEARNING

Machine learning is the main background of this prediction process and the data we acquired from the medical application. This application details can be informed in the later sections and the information we gathered can be used for the machine learning models for better prediction of what is going to happen for the patient in future and what are the main constrains the patients have to follow if there are any problems with their health condition[13-15].

FULL STACK DEVELOPMENT

A full stack medical application was designed and all the operations based on that applications are stored in the database. The application consists of the QR Code scanner which can be only accessed by the doctor and the receptionist in the hospital. The QR Code looks like as follows[16-17].

This QR Code contains the following information related to the patient and their treatment information:

a. Name

b. Gender

c.   Age
d.   Weight
e.   Height
f.   Contact info
g.   Habits
h.   Food Type
i.   Previous Disease information
j.   Duration of the suffering
k.   Current symptoms

l.   Lab reports
m.   Scanning Reports
n.   Medication History
o.   Current Medication
p.   Current Review updates
q.   Current Scan reports (If Any)
r.   Current Lab Reports(If Any)
s.   Next Review Updates

Figure 1.Medical card in digital format

Figure 1 explains the structure of the medical card and the information related to the patients can be stored in the QR Code which we are generating on the card.

LITERATURE REVIEW

Literature review consists of the information related to different research identifications in the field of medical. NCBI is the main repository for the medical information identification. All the medical records and the information related to different researches in the medical domain can be identified in NCBI repository. Here we gathered information related to some of the research works over the globe. They are as follows:

Jougen [1] and his team worked on the principle of identifying the comparison between different paper based and electronic health records of the patients. EHR also referred as Electronic Health Records is the base function of many other things like digital maintenance of the reports, files, care sheets etc. All the information which are on the paper can be converted to the digital records using EHR. EHR is a very powerful thought of mankind to maintain health records into digital format. Figure 2 represents the sample look of EHR applications[18-19].

Figure 2.EHR application scenario

Healththit [2] referring some of the interesting this on EHR models. Those are as follows:

An electronic health record (EHR) is a digital version of a patient's paper chart. EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users. While an EHR does contain the medical and treatment histories of patients, an EHR system is built to go beyond standard clinical data collected in a provider's[19-21] office and can be inclusive of a broader view of a patient's care. EHRs can:

- Contain a patient's medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results
- Allow access to evidence-based tools that providers can use to make decisions about a patient's care
- Automate and streamline provider workflow

Kathrin M. Cresswell [3] Deals with the format of EHR maintenance in the health records maintenance in this research work mentioned in the science direct. He was dealing with the inpatient clinical records maintenance of the different hospitals

These are the three things we gathered from different sources related to the information of the patient stored in different formats of the applications.

EXISTING SYSTEM

There are many different systems which are designed on the health care records. But the most effective thing in the medical domain is that there will be a large amount of data to be manipulated and the performance of the machine learning models will be high with the best features we include in our work. The existing systems of this kind of work doesn't consists of th eQR Code and data security implementations. The QR Code will help the doctor to identify the best way to treat the patient based on the past information. The current system of EHR are mainly focusing on how the data can be stored in the repository and not even considering the process of maintaining those data for high performance of predictions.

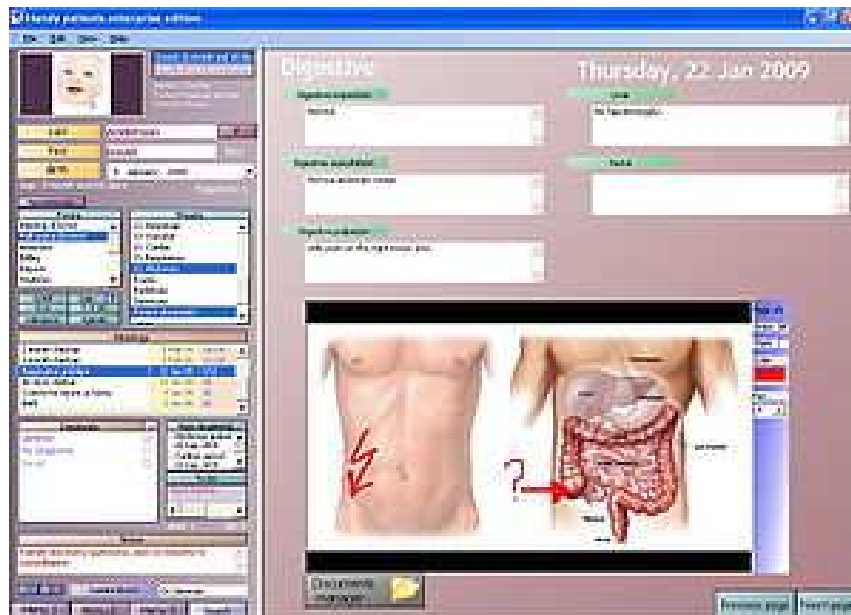Figure 3 and figure 4 explains the different forms of EHR applications we have currently[22].

Figure 3.Sample EHR application with data maintanance

Figure 4.Sample existing EMR application system

The problems with the existing system is we doesnot have any prediction model to implement and maintain the acquired data in an organizaed manner. In some cases the EHR are failed because of not having proper development of design and also because of not having a prediction architecture.[23,24,25]

We overcome the procedure with the two basic machine learning methodologies in the proposed system. Those are Decision Trees and Random Forest. RF is the advanced version of DT. Thet are well explained in the proposed architecture. [26,27]

PROPOSED SYSTEM

In the proposed system we are maintaining a model with decision trees and the random forest methodologies. Using which we need to consider the which methodology will give the highest accuracy for the prediction model. Basically we need to understand the basics of the machine learning models used in this proposed architecture.[28,29]

DECISION TREES

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

Figure 5 will explain the basic structure of decision trees in medical domain. This will give a basic idea only [30,31].
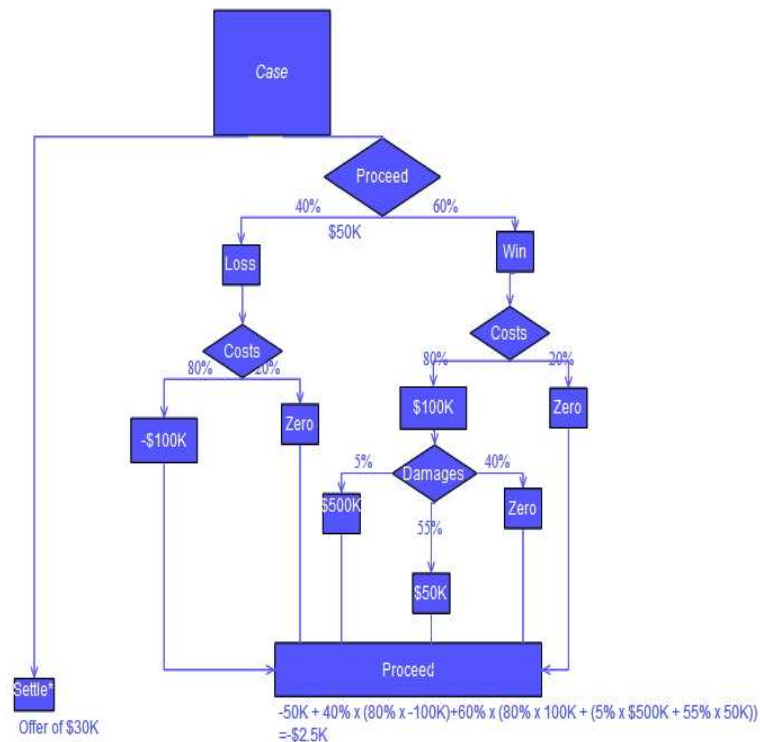

Figure 5.Decision Trees methodology

In this scenario we will identify the all the possibilities for identifying the basic model for the methods. The model consists of the best features based on the priorities and the priorities will be identified by the backward elimination process.[32,33]

RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Random decision forests correct for decision trees' habit of over fitting to their training set. Figure 6 will explain the structure of the random forest we are using[23].

Figure 7 will also state the importance of random forest algorithm as it is a combination of multiple decision trees.
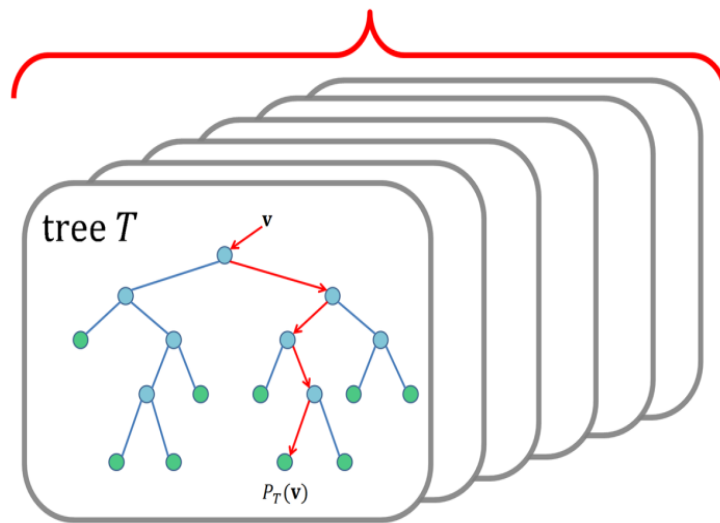
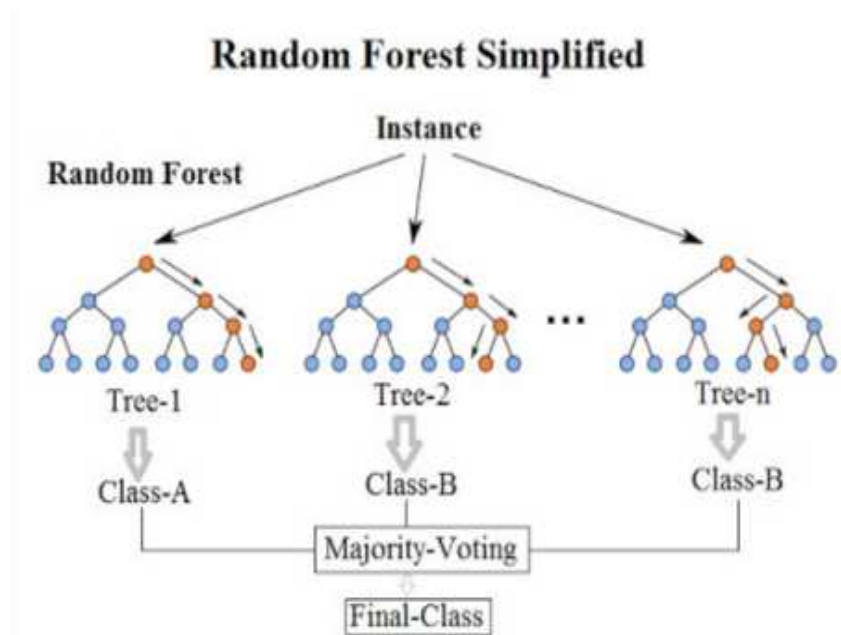**Figure 6.Random Forest implementation**

Figure 7.Group of trees will form Random Forest

PROPOSED ARCHITECTURE

The proposed architecture consists of the things which are the combination of the decision trees and random forest. Because in one part of the classification we are using DT and after the first classification we are implementing RF for the remaining things in the architecture.

Figure 8 will explain the architecture of the proposed architecture.
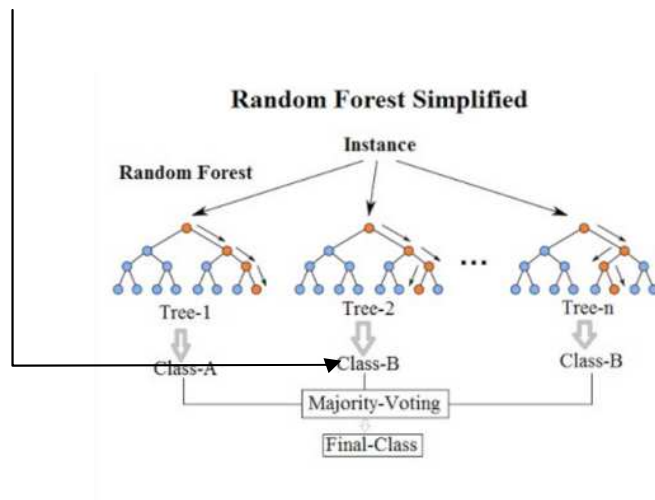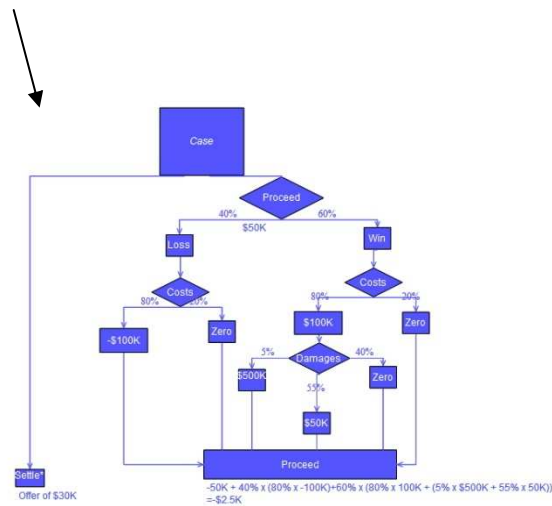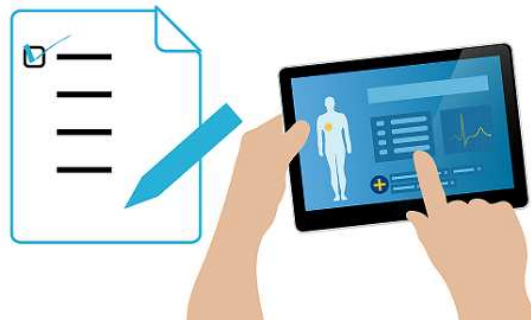
Figure 8.Proposed Architecture

Initially data will be generated with the EHR application and in the first phase they are sent to decision tree model and after acquiring the results we transfer the remaining data to the random forest for the better prediction result. Here we get the highest accuracy for the random forest than DT.

The results of the prediction model are as follows:

Figure 9 will represent the result between random forest and decision trees in identifying the result of the disease with the features gathered.
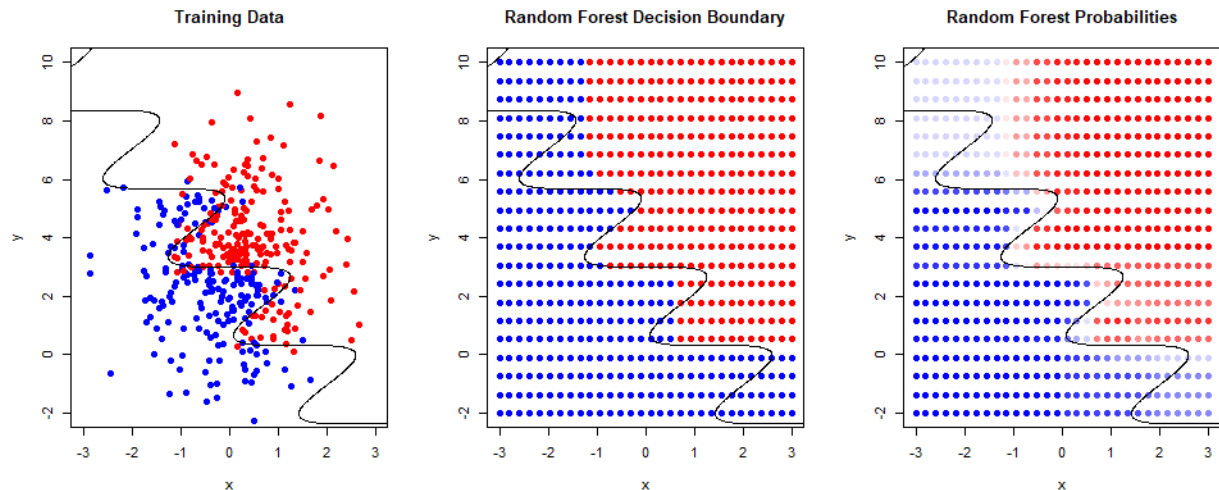
Figure 9.Results of Tranining data with Decision trees and the random forest decision boundary implementation

CONCLUSION

The random forest methodology will give the highest accuracy with the decision boundary design and implementation using the multiple decision trees. The decision tree is used to identify the features and predict a specific value and the random forest is the

 combination of more decision trees and we may get variable decision boundaries in the random forest. So based on our requirement we identified that pure random forest can get the accurate result of identifying the patient health information and decision trees will stand next to RF in the accuracy.

REFERENCES

[1]. R. T. Mylavarapu and B. K. Mylavarapu, "Huge information extraction techniques of Data Security," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 179-183, doi: 10.1109/ICICCT.2018.8473017.

[2]. B. K. Mylavarapu, "Implementing Machine Learning in Intensive Care Units: For Avoiding Over Medication," (2018) International Journal of Pure and Applied Mathematics, Volume 118 No. 20 2018, 4799-4811, URL: https://acadpubl.eu/hub/2018-118-21/articles/21f/33.pdf.

[3]. Kumar. Attangudi P. Perichappan, S. Sasubilli and A. Z. Khurshudyan, "Approximate analytical solution to non-linear Young-Laplace equation with an infinite boundary condition," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp.1-5. doi: 10.1109/ICOMET.2018.8346349.

[4]. Sreenivas Sasubilli,Kumar Attangudi Perichiappan Perichappan, P. Srinivas Kumar,Abhishek Kumar,"An Approach towards economical hierarchic Search over Encrypted Cloud", 2018 Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 125-129, DOI: 10.15439/2018KM38.

[5]. Sriramakrishnan Chandrasekaran; "A Machine Learning Implementation of Predicting the Real Time Scenarios in a better way" - International Journal of Pure and Applied Mathematics Volume 119 2018 Page 1301-1311; https://acadpubl.eu/hub/2018-119-15/4/737.pdf.

[6]. Gopinadh Sasubilli,Uday Shankar Sekhar, Ms.Surbhi Sharma, Ms.Swati Sharma, "A Contemplating approach for Hive and Map reduce for efficient Big Data Implementation" 2018 Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 131–135 DOI: 10.15439/2018KM20.

[7]. Comparing Paper-based with Electronic Patient Records: Lessons Learned during a Study on Diagnosis and Procedure Codes, Jurgrn 2003, Sept, NCBI https://www. healthit.gov/faq/what-electronic-health-record-ehr.

[8]. "Inpatient clinical information system", Kathrin M. Cresswell, 2017, Science Direct.

[9]. Knerr, S., L. Personnaz, and G. Dreyfus. "Single-layer Learning Revisited: A Stepwise Procedure for Building and Training a Neural Network."Springer. 68 (1990): 41-50. Web.

[10]. Liu, Cheng-Lin, Kazuki Nakashima, Hiroshi Sako, and Hiromichi Fujisawa. "Handwritten Digit Recognition: Investigation of Normalization and Feature Extraction Techniques". Pattern Recognition. 37.2 (2004):265-79. Web.

[11]. M. Chen, Z. Xu, K. Weinberger, and F. Sha. Marginalized denoising autoencoders for domain adaptation. In Proceedings of the 29th International Conference on Machine Learning, pages 767{774. ACM, 2012.

[12]. G. E. Dahl, M. Ranzato, A. Mohamed, and G. E. Hinton. Phone recognition with the mean-covariance restricted Boltzmann machine. In Advances in Neural Information Processing Systems 23, pages 469{477, 2010.

[13]. O. Dekel, O. Shamir, and L. Xiao. Learning to classify with missing and corrupted features. Machine Learning, 81(2):149{178, 2010.

[14]. A. Globerson and S. Roweis. Nightmare at test time: robust learning by feature deletion. In Proceedings of the 23rd International Conference on Machine Learning, pages 353{360.ACM, 2006.

[15]. I. J. Goodfellow, D.Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks. In Proceedings of the 30th International Conference on Machine Learning, pages 1319-1327. ACM, 2013.

[16]. G. Hinton and R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 313(5786):504-507, 2006.

[17]. G. E. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. Neural Computation, 18:1527-1554, 2006.

[18]. K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition In Proceedings of the International Conference on Computer Vision (ICCV'09). IEEE, 2009.

[19]. A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[20]. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classi_cation with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25, pages, 1106{1114, 2012.

[21]. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D.Jackel. Back propagation applied to handwritten zip code recognition. Neural Computation, 1(4):541{551, 1989.

[22]. Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, Z. Li, M.-H. Tsai, X. Zhou, T. Huang, and T. Zhang. Imagenet classi_cation: fast descriptor coding and large-scale svm training. Large scale visual recognition challenge, 2010.

[23]. A. Livnat, C. Papadimitriou, N. Pippenger, and M. W. Feldman. Sex, mixability, and modularity. Proceedings of the National Academy of Sciences, 107(4): 1452{1457, 2010.

[24]. B. K. Mylavarapu and R. T. Mylavarapu, "A Framework for Hierarchical Big Image Data," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 169-173.doi: 10.1109/ICICCT.2018.8473339.

[25]. R. T. Mylavarapu, "A Method for Approximated Deep Learning Towards Dynamic Sharing from Big-Data Analysis," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018, pp. 1-6.doi: 10.1109/RICE.2018.8509060.

[26]. Sreenivas Sasubilli, Kumar Attangudi Perichiappan Perichappan, P. Srinivas Kumar, Abhishek Kumar, An Approach towards economical hierarchic Search over Encrypted Cloud, pages 125-129; Annals of Computer Science and Information Systems, Volume 14. ISSN 2300-5963.

[27]. Sriramakrishnan Chandrasekaran; "A Machine Learning Implementation of Predicting the Real Time Scenarios in a better way" - International Journal of Pure and Applied Mathematics Volume 119 2018 Page1301-1311; https://acadpubl.eu/hub/2018-119-15/4/737.pdf.

[28]. Gopinadh Sasubilli,Uday Shankar Sekhar, Ms. Surbhi Sharma, Ms. Swati Sharma, "A Contemplating approach for Hive and Map reduce for efficient Big Data Implementation" 2018 Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 131–135 DOI: 10.15439/2018KM20.

[29]. Abhishek Kumar, K. Rawat, and D. Gupta, "An advance approach of pca for gender recognition," in Information Communication and Embedded Systems (ICICES), 2013 International Conference on. IEEE, 2013, pp. 59-63.

[30]. D Kumar, R Singh, A Kumar, N Sharma An adaptive method of PCA for minimization of classification error using Naïve Bayes classifier Procedia Computer Science, 2015. Elsevier, pp.9-15.

[31]. Kumar, A., & SAIRAM, T. (2018). Machine Learning Approach for User Accounts Identification with Unwanted Information and data. International Journal of Machine Learning and Networked Collaborative Engineering, 2(03), 119-127.

[32]. Rawat K., Kumar A., Gautam A.K. (2014) Lower Bound on Naïve Bayes Classifier Accuracy in Case of Noisy Data. In: Babu B. et al. (eds) Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012. Advances in Intelligent Systems and Computing, vol 236. Springer, New Delhi DOI: https://10.1007/978-81-322-1602-5_68.