Monitoring Health Records For Chronic Disease Prediction And Risk Stratication

Nehal Ghosalkar[1] and Kailas Devadkar[2]

Sardar Patel Institute of Technology, Mumbai, India

Abstract. The chronic diseases related to heart are the primary purpose behind countless deaths over the last couple of decades and has developed as the most dangerous illness, in India as well as in the entire world. In this way, there is a need of dependable, precise and reliable framework to analyze such ailments in time for appropriate treatment.The ongoing advances in innovation have encouraged the standard gathering and storage of medical information that can be utilized to help medicinal choices. Be that as it may, in many nations, there is a requirement to gather patients information in digitized structure. At that point, the gathered information are to be examined all together for a therapeutic choice to be made, regardless of whether it includes prediction of disease, its diagnosis or course of treatment. In this paper,dataset from UCI archive is utilized for Heart ailment analysis. The right nding execution of the programmed conclusion framework is assessed by utilizing characterization exactness, affectability and particularity examination. The investigation demonstrates that, the SVM with CNN-MDRP calculation have better decision for therapeutic infection analysis application.Early recognition of cardiovascular ailments and constant supervision of clinicians can lessen the death rate. A precision dimension of 96.77 percent exactness was found from the proposed framework.

Keywords: Data Mining, Machine Learning, Healthcare, CNN-MDRP, Chronic Heart Disease (CHD), Linear Support Vector Machine (LSVM)

1 Introduction

The use of machine learning techniques in medical eld is a subject of good investigation, that principally focuses on displaying some of the human activities or thinking forms and perceiving certain diseases from a scope of input sources. Diverse application zones are information revelation [10] and medicinal strength frameworks, that grasp genetic science and DNA analysis.

The Data Mining strategies can be used to bring down the mortality rate, to improve the accuracy in disease prediction and also primarily reducing its diagnosis time.[3] Human faces many issues associated with the chronic diseases the common reason behind its increase are improper living habits, decient exercise, unhealthy diet, and irregular sleeping [3]. Eightieth of individuals around the world, spend additional quantity on the diagnosis of chronic sickness [1].

Individuals offers additional aid for correct prediction of sickness [1]. In several regions, different diseases are caused due to the environmental factors and every individuals lifestyle [1].
Sometimes it may result in the incorrect decision concerning the disease prediction. Because of preliminary disease prediction, it will cut back the danger of sickness and patient gets diagnosed as early as possible. In a diagnosis drawback, whats required might be a lot of examples or characteristics that are illustrative of the considerable number of varieties of the diseases. The models must be picked precisely if the framework is to perform loyally and quickly. The very reality that theres no compelling reason to give a selected algorithm on an approach to recognize the ailment, shows a genuine preferred standpoint over the applying of Machine Learning methodologies to the present kind of issues. For instance, amid a chronic disease detection task, its vital for the prediction on healthy patients to be precised as high as possible, as a misclassication amid this class may prompt a healthy patient undergoing treatment for certain illness for reasons unknown. CHD have risen on the grounds that the perfect executioner in each urban and country territories in the vast majority of the nations. Its determined that, in a few cases because of wrong prediction it has resulted in patients health compromise.
In the vast majority of the developing nations specialists arent wide reachable for the correct diagnosis. Thus, such machine-driven framework will encourage to medicinal calling to help specialist for the right analysis well before.
The existing work has only considered about just structured information. For unstructured information, convolutional neural system (CNN) is utilized to remove content qualities naturally. Structured information is widely used for the Chronic disease prediction apart from unstructured data. However by the employment of a convolutional neural network, it becomes straightforward to accommodate unstructured knowledge additionally [1]. The convolutional neural network is deep learning algorithmic rule that extracts the options mechanically from the big dataset and gets the correct result [1].
Through the trial, we tend to make a determination that the execution of CNNMDRP is best than various existing ways. The rest ow of the paper is sorted out as follows: Section II quickly audits some previously proposed strategies in Heart disease diagnosis. Area III depicts the process for coronary illness nding; Section IV highlights the methodology utilized for CHD analysis. Also, the proposed CNN-MDRP calculation and the systems utilized are talked about nitty gritty in Section IV. The experimental results are given in Section V. Segment VI nally concludes the paper.

2 Literature Review

Theodora Brisimi, et.al (2018) presented in this paper [15], two new strategies: K - LRT, which is a probability based proportion technique , and a joint clustering and classication (JCC) strategy it recognizes shrouded quiet groups and adjusts classiers to each cluster. The prediction problem is formulated as binary

classication problem and think about an assortment of ML techniques, including support vector machines (SVMs), logistic regression, and decision tree.
Akhilesh Kumar Yadav, et.al (2013) presented in this paper [8], that dierent analytic tool has been used to extract information from huge datasets such as in medical eld where a huge data is available. The classication becomes inecient due to noise, high dimensional and missing values. Due to the dierent challenges have to face while performing data analytics clustering is used in replace of it. The foggy k-mean clustering based novel technique need to be developed is the main focus of authors. The proposed algorithm has been tested by performing dierent experiments on it that gives excellent result on real data sets. In real world problem enhanced results are achieved using proposed algorithm as compared to existing simple k-means clustering algorithm.
Min Chen, et.al (2017) proposed in this paper [7], a novel convolutional neural network based multimodal disease risk prediction (CNN-MDRP) algorithm. In order to make predictions related to the chronic disease that had been spread within several regions, various machine learning algorithms were streamlined here. A latent factor model was utilized to reconstruct the incomplete type of data present within the gathered data. A chronic disease of cerebral infarction was utilized in order to perform various experiments to predict the accuracy of proposed method. 94.8 percent of prediction accuracy was achieved here along with the higher convergence speed in comparison to other similar enhanced algorithms.
Sanjay Chakrabotry et.al, (2014) stated in this paper [9], that powerful tool clustering is used as dierent forecasting tools. The weather forecasting has been performed using proposed incremental K-mean clustering generic methodology. The weather category has been dened in dierent clusters and a new data is checked by incremental K means to group it into existing clusters. In the last the authors have performed dierent experiments to check the proposed approach correctness.
As different research are done by authors using various Machine Learning Algorithms, it is seen than predicting chronic diseases is a complex study.

3 METHODOLOGY AND DATA ANALYSIS

In this study, an effective machine learning algorithm was looked over some accessible algorithms in order to recognize the probability of having heart disease from a huge dataset. The well ordered structure methodologies of the proposed framework and the work process of the total framework have been referenced below The Design is isolated into three fundamental stages: Initial, Middle and Last stage. The Initial stage is identied with Data aggregation and Analysis. The next stage includes distinctive substages like FeatureSelection, Training the SVM Model and Measure the slipup estimations. Last stage incorporates the Visualization of the data. [3]

Data Imputation For patients examination information, therell be an oversized range of missing information because of human error. Along these lines, we will ll the organized data. Prior to data ascription, we tend to at rst decide inadequate medical data so as to alter or erase them to upgrade the data quality. At that point, we tend to utilize data incorporation for information pre-preparing. we will incorporate the healthcare data to guarantee data atomicity. For data attribution, we tend to utilize the latent factor model that is given to illustrate the recognizable factors as far as the inert factors.

Splitting the data into Training and Testing Set The inmate dataset information contains structured and unorganized data. The structured information includes the data from the laboratory and therefore the patients basic data like the patients age, their gender, previous medications and lifestyle etc. whereas the unorganized text information consists of the patients narration of his/her health problem, the doctors interrogation records and diagnosing etc.

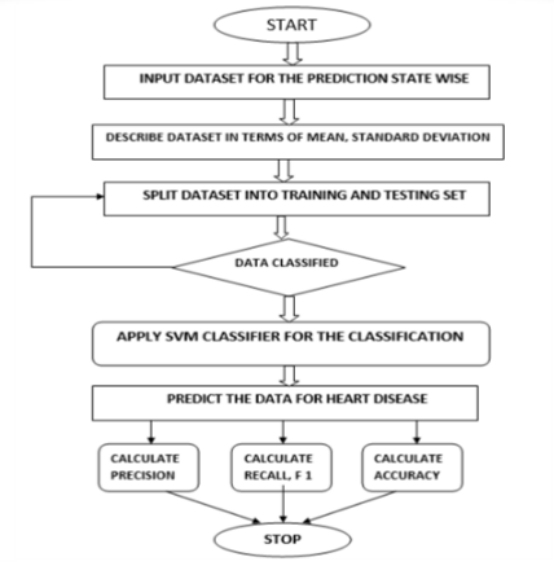

Fig. 1. System Architecture

4 Results

At first, sample size, network size, test measure, selection of model and feature extraction and classification are considered to be the key parameters for the in-

vestigation of a reasonable system configuration issues identified with learning and speculation. While experimenting it is watched that, right and complete information gathering strategy is the correct course for the choice of best classifier. For assessing speculation execution as for precision, affectability, and explicitness dataset is parceled into number of subsets (i.e training set and testing set) The experiment conducted on the training data consisting of medical symptoms of patients. The system uses a Linear SVM algorithm along with CNN-MDRP for prediction of disease based on their patient symptoms. We have achieved the performance and accuracy of Linear SVM algorithm as 96.77 percent which is accurate for prediction of symptoms from the diseases. The execution time is 0.128 seconds. As well as the performance of disease prediction with the help of unstructured data has increased compared from the previous work done.

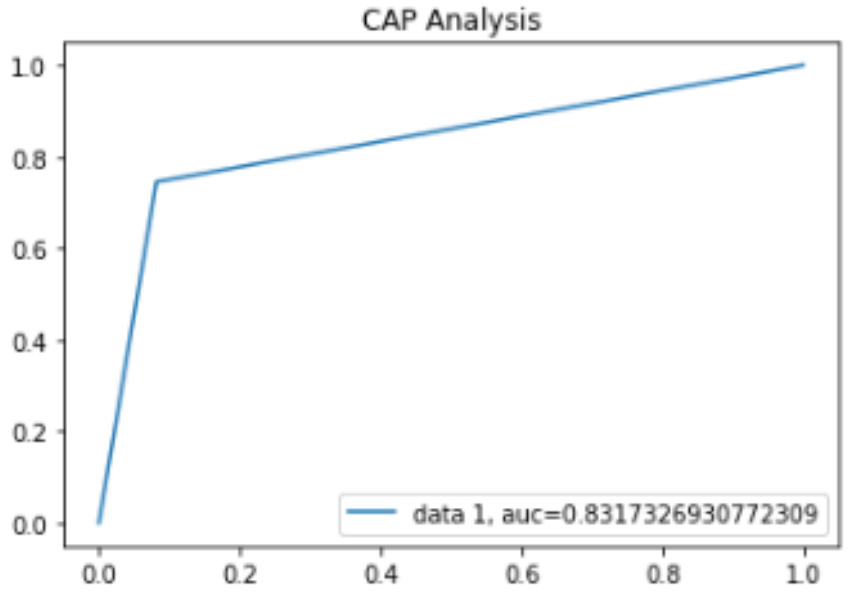

Fig. 2. ROC Curve

5 Conclusion

Chronic Disease diagnosis has turned out to be exceptionally credited with the advancement of innovation recently. Moreover the Data mining and specialized apparatuses have improved the medicinal practice execution to a more noteworthy degree. Here we have proposed a Convolutional Neural Network System

for the conclusion of CHD by method of Linear Support Vector Machine. Accordingly the conclusion of Heart ailment is done using distinctive information tests from assorted patients and the outcomes have meant that SVM with Linear kernel is great in the analysis of Heart related illness. The characterization accuracy, affectability, and particularity of the LSVM have been observed to be high in this manner making it a decent alternative for such types of finding.
Thus, in this paper, we leverage not only the structured data but also the unorganized text data of patients based on the proposed CNN-MDPR calculation. We find that by joining these two information, the exactness rate achieved is 96.77 percent, in order to better the prediction risk of CHD.This framework leads in low time utilization and negligible cost workable for illness forecast. In future work, we may include more ailment into it so the general public gets more advantages about this framework.

References

1. Abdelghani Bellaachia and Erhan Guven *Predicting Breast Cancer Survivability Using Data Mining Techniques.*Washington DC 20052, vol. 6,pp. 234-239, 2010.
2. Oyelade, O. J, Oladipupo, O. O and Obagbuwa *Application of k-Means Clustering algorithm for prediction of Students Academic Performance.* International Journal of Computer Science and Information Security, I.C, vol. 7, pp. 123-128,2010.
3. Azhar Rauf, Mahfooz, Shah Khusro and Huma Javed (2012) *Enhanced K-Mean Clustering Algorithm to Reduce Number of Iterations and Time Complexity* . Middle-East Journal of Scientic Research,, vol. 12,pp. 959-963 2012.
4. Osamor VC, Adebiyi EF, Oyelade JO and Doumbia S *Reducing the Time Requirement of K-Means Algorithm.* IEEE Transactions on Power Electronics, PLoS ONE, vol. 7, pp-56-62, 2012.
5. Pranjul Yadav, Michael Steinbach, Vipin Kumar and Gyorgy Simon *Mining Electronic Health Records (EHRs): A Survey.* ACM Computer Survey, Article 85, 2018.
6. Kajal C. Agrawal and Meghana Nagori *Clusters of Ayurvedic Medicines Using Improved K-means Algorithm.* International Conf. on Advances in Computer Science and Electronics Engineering, ,volume 23, 2013
7. Min Chen, Yixue Hao, Kai Hwang, Fellow, IEEE, Lu Wang, and Lin Wang (2017) *Disease Prediction by Machine Learning over Big Data from Healthcare Communities.* IEEE Transactions, vol. 15, pp- 215-227, 2017
8. Akhilesh Kumar Yadav, Divya Tomar and Sonali Agarwal, *Clustering of Lung Cancer Data Using Foggy K-Means.* International Conference on Recent Trends in Information Technology (ICRTIT), vol. 21, pp.121-126, 2013.
9. Sanjay Chakrabotry, Prof. N.K Nigwani and Lop Dey, *Weather Forecasting using Incremental K-means Clustering*, Proceedings of the IEEE Region 10 Conference, vol. 8,2014.
10. Chew Li Sa, Bt Abang Ibrahim, D.H., Dahliana Hossain, E. and bin Hossin, *Student performance analysis system (SPAS).* Information and Communication Technology for The Muslim World (ICT4M), 2014.