iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.45625

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

Clinical Decision Making using Machine Learning

Danushri N C[1], Boomika H P[2], L Lavanya[3], Vidhya T.D[4], *Mr .Hemanth kumar K*[5]
[1, 2, 3, 4] *East West Institute of Technology, Bangalore, Karnataka.*
[5]*Corresponding Author*

***Abstract:** Deaths due to cardiovascular diseases are increasing at an alarming rate. This led to nearly 2.1 million deaths in India in 2015. Heart disease is one of the deadliest causes of death worldwide and has a major impact on the lives of rural people. According to a recent study, cardiovascular disease mortality among rural Indians has surpassed urban Indians. Such numbers are alarming, especially when 68% of India's population lives in rural areas that have poor access to quality healthcare. This paper aims to provide a solution to this problem by introducing a new model*clinical*decision*support system*, abbreviated as CDSS, which*includes machine learning algorithms for the diagnosis of cardiovascular diseases. CDSS is intelligent enough to diagnose a patient's*disease* and help the doctor prescribe the*correct medication, reducing the cost and effort required to prescribe unnecessary treatment. *In this work, we applied correlation-based feature selection (CFS) and a multilayer perceptron classifier on a large heart disease dataset. The dataset used in this study is the "Cleveland Clinic Foundation Heart Disease Dataset" available at the UCI Machine Learning Repository. Our proposed model produced greater accuracy compared to other existing models used in this study. This system can be integrated into a public health care setting to help rural people get a correct, timely and cost-effective diagnosis.*

***Keywords:** Clinical decision support system, Cardiovascular disease, Feature selection, Cleveland dataset, Multilayer perceptron, neural network, Machine learning, rural healthcare*

I. INTRODUCTION

Cardiovascular disease or CVD is the leading cause of mortality in India. An average age of people with heart ailments is dropping drastically. Now-a-days, there are so many young Indians suffering from a heart attack. Factors such as stress,*hypertension, smoking, diabetes, obesity and physical*inactivity amplifies the risk of occurrence of cardiovascular diseases [1]. A World Health Organization (WHO) report says - “Cardiovascular diseases would be the largest cause of death and disability in India by 2020” [2]. If this trend continues, then approximately 23.6 million people will die from heart disease till 2030. (Fig2) Thus it is critical to minimize this risk and look for ways to identify areas*for improvement. Well, prediction of this condition in an early stage along with a suitable treatment*can*save a lot of lives. Unfortunately, due to complex interdependence agnosis becomes a challenging task. The condition of healthcare in the rural sector is even worse as compared to the facilities provided in the urban areas.(Fig.1) Shortage of skilled doctors, quality hospitals & lack of healthcare infrastructure makes it difficult for rural people to have proper access to quality primary healthcare. Data from the National Rural*Health Mission shows that nearly*8% of primary*health centers in rural India were functioning*without a doctor, while 61% of them had just one doctor *as of March 2017[3]. Therefore, there’s great need for development of medical diagnostics systems thus helping the physicians and other healthcare professionals in the diagnostic*process. By using such knowledge-driven healthcare systems, the medical practitioners are able to make informed decisions and plan effectively for the future; thereby reducing time spent on treating each patient and also saves cost by eliminating the cost of unnecessary treatments. In this paper, we have proposed a model of a clinical*decision support system for a public healthcare system to aid the physicians and other medical professionals in making better decisions, choosing appropriate treatments for patients and also help them in an early diagnosis of the disease. Our objective is to design, develop and additionally evaluating the CDSS that can be easily integrated into the work-flow of healthcare providers. Objective of our research is to predict the presence of heart disease more accurately with less number of attributes

We propose an improved model for neural network approach in medical diagnosis by using Correlation-based feature selection technique which saves lot of computational cost and time complexity, thus improving the overall*performance of the classifying algorithms. Organization of our paper is as follows. The section II specifies role of Clinical*Decision*Support*Systems and their categories. Section III contains the description of the dataset and the software tool considered for our work along with the related work in this field of study. Section IV describes the methodology adopted for the course of our research including pre-processing, feature selection and the MLP*classifier. Section V illustrates our proposed model and its workflow.

©IJRASET: All Rights are Reserved

Section VI includes the experimental*results and the comparative analysis between the proposed*work and other models for validation and finally Section VII concludes the paper with high accuracy rates of our proposed model along with directions for future work.

II. LITERATURE SURVEY

Over the years, the research has shown good prediction performances in cardiology. There have been various areas of medical research that have been pondered upon by the researchers widely like- ECG, myocardial infarction, and heart failure.

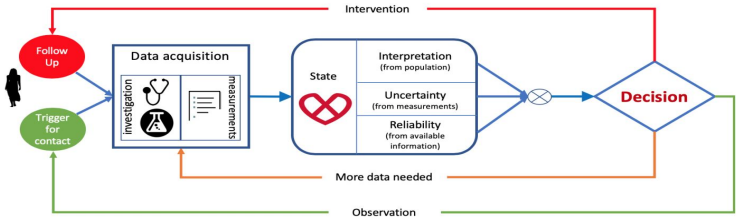
[5] Machine*learning algorithms have been used extensively by the researchers in the past to help healthcare professionals deliver good quality care and treatments. There have been various studies that are performed on diagnosing*heart-related diseases and improving their accuracy. Some of the works includeIn

[6], the authors used SVM,*artificial neural networks and the decision tree to predict survival of Coronary heart*disease using 502 cases. In [7] the results showed that Naïve*Bayes surpassed the other used algorithms whereas in [8], Neural Networks was determined as the best prediction algorithm compared to Decision Trees and Naïve Bayes. Genetic*Algorithm has also been used to obtain an optimal subset amongst patient attribute-values by optimizing the datasize that is sufficient for heart disease prediction.

[10] In [9]Tu*et al.*carried out a comparative*research comparing the efficiency of J4.8*Decision*Tree and*bagging algorithm to diagnose heart diseases. A hybrid*model was also suggested by some researchers by combining it with neural networks in order to achieve improved accuracy for prediction.

[11] Also, medical communities encourage the use of CDSS. In spite of various ongoing challenges in developing such systems and deploying them at rural stations, these systems have actually proven their reliability and accuracy repeatedly. [12] The motivation behind the study was the need to develop a computerized system to search the most-suitable machine learning technique for predicting the survival-rate of heart patients. These studies guided our way to design this framework. .

III. METHODOLOGY



A. *CNN Consists Of Four Components Namely*

1) *Convolution Layer:* has several filters that perform the convolution operation. Each image is treated as a matrix of pixel values. Consider a 5x5 image whose pixel values are either 0 or 1. There is also a 3x3 filter matrix. Move the filter matrix over the image and calculate the dot product to get the convolved element matrix.
2) *Pooling layer:* dimensionality of feature map. The corrected feature map is now passed through the pooling layer to create a pooled feature map. The pooling layer uses different filters to identify different parts of the image.
3) *Fully connected layers:* One node is connected to all the nodes in the previous layer in a fully connected layer. Additionally, more than one hidden layer may be used, and the output layer may use a different classification operator

a) A useful tool in predicting the probability of a binary outcome is the ROC curve.
b) It is a plot of false positive rate (x-axis) versus true positive (y-axis) for a number of different candidate thresholds between 0.0 and 1.0. In other words, it plots the false alarm rate versus the hit rate.
c) The true positivity rate is calculated as the number of true positives divided by the sum of the number of true positives and the number of false negatives. It describes how good the model is at predicting the positive class when the actual result is positive.
d) True positive ratio = true positive / (true positive + false negative)
e) True positive rate is also called sensitivity.
f) Sensitivity = True Positive / (True Positive + False Negative)
g) The false positive rate is calculated as the number of false positives divided by the sum of the number of false positives and the number of true negatives.

Our proposed method is experimented on a benchmark dataset available in the UCI Machine Learning Repository The main components of the proposed system include; input data collection, pre-processing, clustering, fuzzy modeling and diagnostics output. In addition, Table I shows the relationship between possible symptoms and physiological parameters.

B. Preprocessing

To remove noise and artifacts we have implemented; low pass filtering, removal of missing values (zero or negative), data sampling, inspection and removal of outliers from the data set and calculation of statistical/descriptive values such as; maximum, minimum, mean, median, mode, standard deviation and range to have a normalized data set during diagnosis

C. Training Process

The proposed CDSS adopted the two most common clustering mechanisms to classify data with high accuracy and reliability. Due to the nature of clinical data, which is complex and often incomplete, the proposed system used clinical data patterns to build a decision support model using a training and learning dataset in an unsupervised approach [15]. Two fuzzy clustering techniques including fuzzy c-means clustering (FCM) and fuzzy k-means clustering (FKM) were adopted for syste evaluation and validation. Data clustering/classification was performed after applying the most common pre-processing methods used in big data applications and healthcare system analysis.

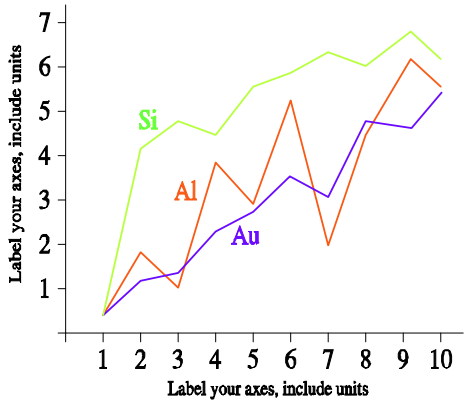


Fig. 1 A sample line graph using colors which contrast well both on screen and on a black-and-white hardcopy

Fig. 2 shows an example of a low-resolution image which would not be acceptable, whereas Fig. 3 shows an example of an image with adequate resolution. Check that the resolution is adequate to reveal the important detail in the figure.
Please check all figures in your paper both on screen and on a black-and-white hardcopy. When you check your paper on a black-and-white hardcopy, please ensure that:the colors used in each figure contrast well,the image used in each figure is clear,all text labels in each figure are legible.

D. Diagnosis Output

The proposed diagnostic system, based on self-organizing fuzzy logic modeling, has been proposed using vital signals. The system detects abnormal signs which are directly related to five key symptoms; Bradycardia, Tachycardia, Hypotension, Hypertension and Hypovolaemia.
Although more rules could be added but we limited them to 10 rules to reduce false alarms rate (false positive and false negative). In the last case (rule 10), the diagnosis result will show “High-vital-signs”. This is because when two or more out of the three vital signs go higher than the normal range then system check against the expert given rules, if no exact rule is found then the system will warn as ‘High vital signs’. After setting the output, we feed related data to the same symptom (Hypotension) as the input. Then the system can be trained with multiple data sets of each symptom. The initial training data set for the self-organizing fuzzy model was selected from the authentic diagnosis database called MIMIC II waveform [16]. We selected approximately 446 records from the MIMIC II database [16] for classifying symptoms. The system was trained in such a way that whenever the new patient’s data enters with similar symptom it generates an output based on the trained input datasets with early alerts/warnings. Table II shows comparison of the proposed CDSS with different clustering methods. It is also apparent that, the system can achieve better accuracy with the trained 446 records from MIMIC II database. The system was then tested with 30 hospitalized patient datasets.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

IV. SYSTEM RESULTS

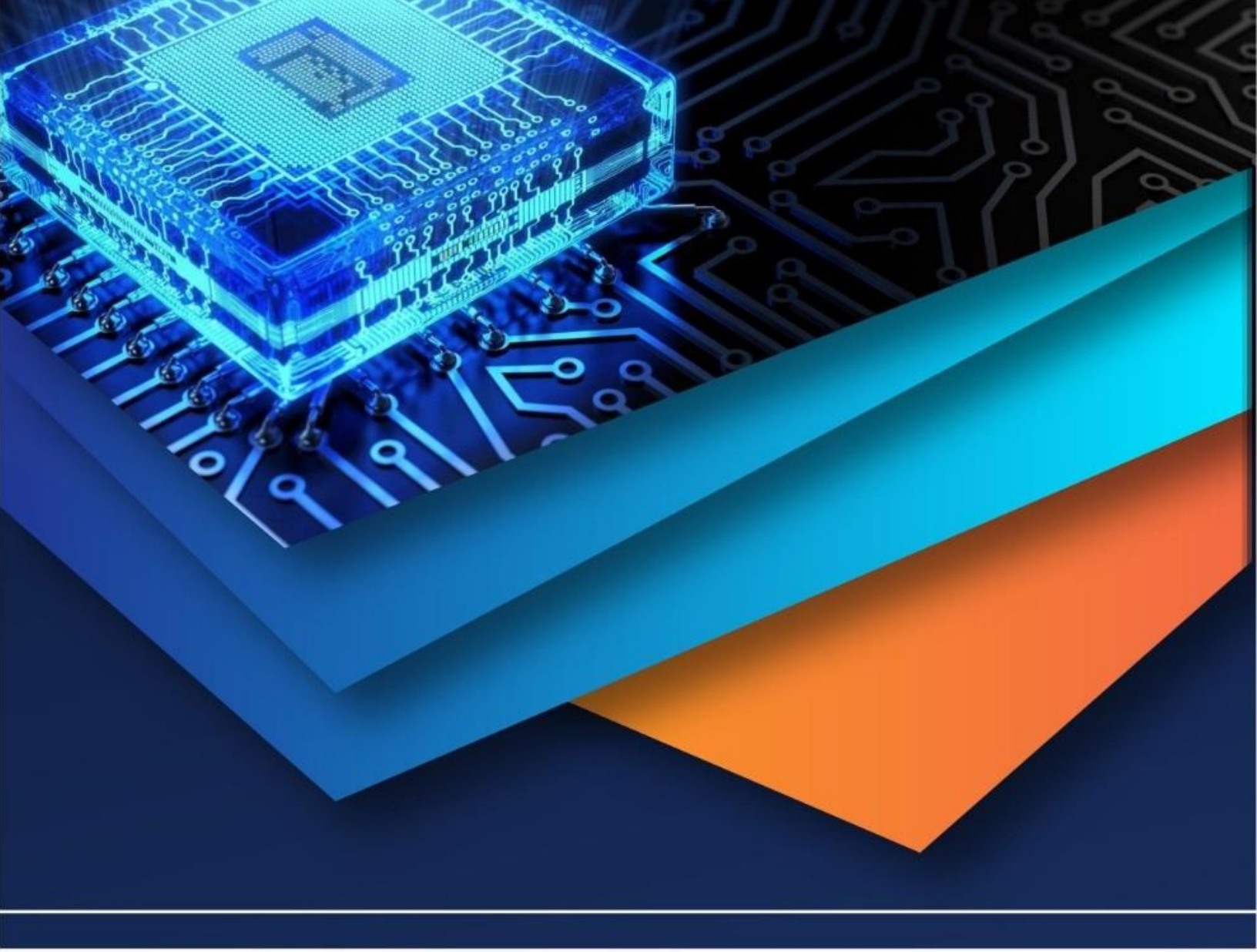
To measure the level of acceptance between the system generated outcome and the human expert's diagnosis, we used Kappa analysi (i.e. as the measure of how accurately the system can mimic human performance). The proposed system raised a total of 52 alarms and out of these, 47 alarms matched with the expert's diagnosis. The proposed system achieved an overall positive agreement (Po) with the accuracy of 95% and Kappa value of 91%. As a result, there were a total of five false positives generated, three of them were related to 'possible hypothermia'. This diagnostic was due to recording the ear temperature of the patient at below threshold value. While for this case,, the expert considered that the ear temperature value as boarder line of the threshold and would have delayed to see some more readings before considering this as an alert to a 'possible hypothermia'. We adjusted the proposed system to the assessment method undertook by the expert by taking the average of at least three values before generating an alert.

V. CONCLUSIONS

The clinical decision support system was trained offline for performance evaluation purpose using clinical normal threshold-based values and approximately abnormal valuedatasets. Two methods have been implemented and tested against MIMIC II dataset for medical diagnosis using HR, BP and PV. It is also proven that FCM can be used in this type of medical data where following the relationship between data and a particular physiological event is essential. The FCM algorithm will be modified using new fuzzy rules and membership functions to cluster more classes. The clustered data will be fed to the fuzzy neural module for self-organizing the limits, rules and memberships to detect several events.

REFERENCES

[1] N. Skyttberg, J. Vicente, R. Chen, H. Blomqvist, and S. Koch, "How to improve vital sign data quality for use in clinical decision support systems? A qualitative study in nine Swedish emergency departments," BMC medical informatics and decision making, vol. 16, p. 1, 2016.
[2] A. Wright, T.-T. T. Hickman, D. McEvoy, S. Aaron, A. Ai, J. M. Andersen, et al., "Analysis of clinical decision support system malfunctions: a case series and survey," Journal of the American Medical Informatics Association, p. ocw005, 2016.
[3] C. Y. Tsai, S.H. Wang, M.H. Hsu, and Y.C. J. Li, "Do false positive alerts in naïve clinical decision support system lead to false adoption by physicians? A randomized controlled trial," Computer Methods and Programs in Biomedicine, vol. 132, pp. 83-91, 2016.
[4] M. El-Bardini and A. M. El-Nagar, "Direct adaptive interval type-2 fuzzy logic controller for the multivariable anaesthesia system," Ain Shams Engineering Journal, 2011.
[5] L. Qiao and G. D. Clifford, "Suppress False Arrhythmia Alarms of ICU Monitors Using Heart Rate Estimation Based on Combined Arterial Blood Pressure and Ecg Analysis," in Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on, 2008, pp. 2185-2187.
[6] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," Fuzzy Sets and Systems vol. 1, pp. 3-28, 1978.
[7] A. Belard, T. Buchman, J. Forsberg, B. K. Potter, C. J. Dente, A. Kirk, et al., "Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care," Journal of clinical monitoring and computing, pp. 1-11, 2016.
[8] C. E. Butler, S. Noel, S. P. Hibbs, D. Miles, J. Staves, P. Mohaghegh, et al., "Implementation of a clinical decision support system improves compliance with restrictive transfusion policies in hematology patients," Transfusion, vol. 55, pp. 1964-1971, 2015.
[9] D. Blum, S. X. Raj, R. Oberholzer, I. I. Riphagen, F. Strasser, S. Kaasa, et al., "Computer-based clinical decision support systems and patient-reported outcomes: a systematic review," The PatientPatient-Centered Outcomes Research, vol. 8, pp. 397-409, 2015.
[10] C. Castaneda, K. Nalley, C. Mannion, P. Bhattacharyya, P. Blake, A. Pecora, et al., "Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine," Journal of clinical bioinformatics, vol. 5, p. 1, 2015.
[11] M. A. Rose, L. A. Hanna, S. A. Nur, and C. M. Johnson, "Utilization of Electronic Modified Early Warning Score to Engage Rapid Response Team Early in Clinical Deterioration," Journal for nurses in professional development, vol. 31, pp. E1-E7, 2015.
[12] M. M. Baig, H. Gholamhosseini, M. J. Connolly, and M. Lindén, "Advanced Decision Support System for Older Adults," in PHealth 2015: Proceedings of the 12th International Conference on Wearable Micro and Nano Technologies for Personalized Health 2–4 June 2015 Västerås, Sweden, 2015, p. 235.
[13] M. Mirza, H. GholamHosseini, S. W. Lee, and M. J. Harrison, "Detection and Classification of Hypovolaemia during Anaesthesia," in 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2011, Boston, MA, USA, 2011, pp. 357-360.
[14] M. Mirza, H. Gholam Hosseini, and M. Harrison, "Fuzzy Logicbased System for Anaesthesia Monitoring," presented at the 32nd Annual International Conference of the IEEE EMBC, Buenos Aires, Argentina, 2010.
[15] M. Sarkar and T.Y. Leong, "Fuzzy K-means Clustering with Missing Values," in American Medical Informatics Association (AMIA) Symposium, 2001, pp. 588-592.
[16] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13). PMID: 10851218; doi: 10.1161/01.CIR.101.23.e215



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)