Decision Support Tool for Medical Prediction System Using Multiplexed Machine Learning Techniques

Selvakumar Subramanian[a]*, Mahesh Akuthota[b], Syed Thisin[c], Vinuthna Amireddy[d]

[d]*Visvesvaraya College of Engineering and Technology, Hyderabad, India, sselvakumar@yahoo.com*
[b]*Visvesvaraya College of Engineering and Technology, Hyderabad, India, maheshvcetcse@gmail.com*
[c]*Visvesvaraya College of Engineering and Technology, Hyderabad, India, syedthaisin@gmail.com*
[d]*Visvesvaraya College of Engineering and Technology, Hyderabad, India, vinuthnaamireddy@gmail.com*

Abstract: Medical information systems are used by large number of hospitals to control the patient data and clinical information. Typically, these systems produce enormous volumes of data in statistics, text, graphs, and images. Regrettably, these statistics are rarely utilized to assist clinicians in making clinical decisions. This problem can be solved using advanced techniques of machine learning. So, the main goal of the proposed study is to create a prototype of a Decision Support in Medical Prediction Systems using prominent and efficient machine learning techniques like Naive Bayes, ID3, and Compact Weighted Associative Classification. It can predict a patient's chance of developing cancer, HIV, diabetes, and heart attacks. The system's performance and accuracy are investigated and examined using various test plan scenarios. A group of related inquiries must be handled to achieve the appropriate decision-making solution from the revealing prediction system. One of the system's benefits is that it may be used as a module in a hospital management system. The inputs can be automatically given from the patient details, and they produce the exact results. Everyone can easily access it, even in their homes, when the system is spread online. The proposed approach was tested on the UCI machine learning dataset with the simulation environment. The objective of the research work is to create a web-based questionnaire application for Decision Support in Medical Prediction systems. It can retrieve hidden knowledge (patterns and correlations) linked with users' answers to the particular disease from a historical diseases database. This proposed system helps predict the patient's condition and achieve better results by comparing the performance of these algorithms and calculating their accuracy.

Keywords—Decision Support, Medical Prediction System, Machine Learning Techniques

1. Introduction

Decision Support in Medical Prediction System is designed and implemented to support clinicians' diagnoses. It is visible that by improving these systems, the medical diagnostic decisions quality can be increased. Machine learning yield a better way to retrieve the information buried in the data. In large and complex collections of data ML algorithms can find patterns hidden, which evade conventional statistical methods to analysis. With the increase of deadly diseases that threaten human health and life, Health Decision Support Systems carry on with their effectiveness in supporting healthcare professionals and physicians with paves way in clinical decision-making.

1.1. Medical prediction system

Many healthcare organizations (hospitals, medical centers) provide high-quality care at low prices. They can accomplish these effective services by making use of relevant system based health information and decision-making methods. In recent days most hospitals use hospital information systems to process their hospital patient health data. Typically, these systems generate massive volume of raw or noisy data in statistics, text, tables, images and charts. It's regret that, these noisy or unstructured data are seldom used to assist physicians in making critical clinical decisions. These data provide a plethora of secret knowledge that is largely untapped. The challenge of turning data with noise into useful knowledge or data that allows medical professionals to decide informed clinical decisions is always unresolved, which is the primary motivation for this study.

1.2. Classification and prediction, decision tree, weighted association rule mining

Classification is a definitive machine learning method based on categorizing a given set of data. Classification is helpful in classifying each item in a data into one of a predefined set of classes or category. Classification is used to forecast missing or numerical data values that are not available, than class labels. In spite of the fact that the concept of prediction can refer to predictions of both class label and numeric, it is the discovery of relationships between independent factors and dependent and independent variables that is the goal of prediction. The decision tree structured as tree structure that resembles structure of flowchart. Each branching node correlates to a test on an value of attributes, the test's result represented by the branch, and leaves of the tree represent class distributions or classes. Decision trees can conveniently be transformed to classification rules. Three types of nodes make up a decision tree, Decision nodes usually represented by squares, Circle denotes Chance nodes, triangles represented as End nodes. Weighted ARM operates the weighted support and also the Confidence Framework to derive Association rule from the repository of data. The classical ARM framework assumes that all items are of equal importance or significance within a transaction or record; their weight is similar, which is not always the case. A lot of existing algorithms are used to classify the datasets. This paper presents a new efficient Medical Prediction System using three types of Machine learning Techniques, namely Naïve Bayes Classification, ID3 Decision Tree Algorithm, and Compact Weighted Associative Classification. They predict the possibility of patients who may be affected by disease. To acquire better results, the performances of these three algorithms are compared, and their accuracy is calculated.

The remaining part of the paper is exhibited as follows. Section 2 explores the literature survey and machine learning algorithms. Section 3 discusses the experimental requirements for the proposed research work. Section 4 confers on the problem description. Design and methodology are discussed in section 5. The implementation work and the results of the experimentation are presented in section 6. The Conclusion and the proposed research's future work in section 8.

2. Literature Review

A comprehensive survey was made to acquire antecedent knowledge on decision support systems based on the domain of the application, design of the system, and system complexity; and further existing models are analyzed and discussed in this section. Many Classification algorithms are discussed in the literature, and some are better than others. Those different methods utilized for predicting the level of risk the diseases affected person were observed. Better results are achieved by comparing the performance of these algorithms and calculating their accuracies. Several prediction systems used to identify the risk level of the diseases were examined. Of these, some essential systems like Decision Support Heart Disease Prediction System were discussed, which uses various machine learning methods like ID3, Naive Bayes Classification, Neural Networks, Weighted Associative Classifiers and Feature Subset Selection with Genetic Algorithms. This system acts as the main milestone to think and implement the algorithm.

Francesco. et al. [2] experimented the problems in decision support systems groups with algorithm k-means. Together with a hierarchical process, the method achieves phenomenal decisions for the assigned tasks. Based on their importance alternatives are sorted and ordered to get the decision. Bezier fitting curve is used to retrieve the decision maker's preference of function. This experimentation work which uses knapsack model, selects alternatives based on the classes and constrains of budget, which gives remarkable performance. Likewise, Siddiqui. et. al reports a web-based group decision support system.

In the [5], machine learning algorithms applied for Medical decision support was developed for detecting seronegative in neonatal sepsis in a early stage; the purpose is non invasive predictive models for late-onset neonatal sepsis from electronic medical files and clinical data. Machine learning techniques can serve with the medical decision support system in both the diagnostic and clinical levels [8]. This author's method initially analyzes the milestones from previous records and particular applications of system based clinical decision support tools in human and veterinary medicine. The next step of the experimentation initiates a mechanical look at three archetypal learning algorithms like decision trees, neural network and Naive Bayes. These are usually used to power these clinical decision system support tools. And finally, the work focuses on the record sets that are experimented to train these machine algorithms and examines validation of data, representation of the data, transformation, and the efficient feature selection methods. In the examined work, the researcher can appreciate how the decision health support tools are employed in health care and gain deeper understanding into their innermost operating characteristics. The [9] research work suggests machine learning techniques on the basis of blood tests for the COVID-19 mortality risk prediction. Various Machine learning models such as neural networks, logistic regression method, random forests algorithm, XGBoost, decision trees algorithm, and SVM were trained and compared by detecting the model that achieves higher accuracy persistently over the duration that spans the disease. Feature neural network classification XGBoost predicts as early as 16 days with an accuracy of 90%, prior to the outcome has the best performing method. This experimentation provided solutions that will promote to initiate the healthcare decision making systems for specified clinical diagnostics and treatment accurately and reliably.

The experimentation [10] attempts to find to narrate the prevalence and the nature of medical expert involvement in developing, evaluating, and implementing decision support systems for health care. The work [10] use ML methods to analyze clinical data stored electronically, to assist doctors and clinical persons in prognostic and therapy predictive decision support system in the medical field. In the [10], the empirical studies of predictive decision support system utilizing medical data for health care professionals and physicians in the medical treatment environment in the data of five years. The studies that met eligibility criteria is summed upto be eighty. Medical expert engagement in the work seems most prevalent at the initial and later stages of design of the health decision system. Medical professional involvement is most pervasive when predictive decision support system specifications are developed or when implementations of the system are assessed. In spite of health care professional are less prevalent in developmental stages to evaluate medical data validation, model features selection, data pre-processing, or serve as a standard bench mark.

The [11] suggest a diabetes prediction decision support system based on ML algorithms. It evaluated conventional ML algorithms with algorithm of deep learning. The paper considered the most widely employed classifiers such as Random Forest and the Support Vector Machine algorithms. Furthermore, it employed a fully Convolutional Neural Network (CNN) for Deep Learning (DL) for diabetic prediction and detection of patients. The proposed method is experimented on the Pima Indians Diabetes database, consisting of 768 samples, which constitute eight features for each. 500 samples were marked as not a diabetic, whereas 268 were labeled as diabetic. On the whole, the accuracy obtained using DL, Support Vector Machine, and Random Forest methods shows 76.81%, 65.38%, and 83.67%. The research work outcome, insights, that the Random Forest was more efficient algorithm for prediction of diabetes than the SVM and deep learning methods.

In [12], the researcher examined a hybrid decision support system that can help in the early heart disease detection, based on the patient health record details. The research work utilized multivariate imputation by chained equations technique to handle the missing values. Ensembling the Genetic Algorithm and recursive feature elimination as a hybridized feature selection algorithm has been used to select appropriate features from the dataset. Additionally, Synthetic Minority Oversampling Technique abbreviated as SMOTE and standard scalar methods were utilized for pre-processing of data. In the final step of developing the hybrid system initiated, the authors used logistic regression, naive bayes, AdaBoost, random forest and support vector machine algorithms. It is visible that the system has given the most accurate results with a random forest classifier. It has achieved an 86.6% accuracy, which is superior to some of the existing heart disease prediction systems found in the literature. The [13] the authors suggested a decision support system based on data mining algorithm using a hybrid approach with a artificial neural network and decision tree to forecast the marketing strategies for an establishment.

This paper [22] discusses the comprehensive survey based on AI algorithms to predic various diseases such as Alzheimer, cancer, diabetes, chronic heart disease, tuberculosis, stroke and cerebrovascular, hypertension, skin, and liver disease. The work conducted an survey in the for medical imaging dataset for predictions and their feature extraction and classification process. The experimentation are analyzed using various quality parameters such as sensitivity, specificity, accuracy, recall, the area under curve precision, and prediction rate. This article [23] covers the present clinical position and issues related to tasks of interpretation and decision support. The author also analyze the challenges in the learning process, the system infrastructure, the auditability, the traceability, and the integration within medical diagnostic procedure in cardiovascular imaging.

3. Choice of algorithms and tools

After considering all these, it is decided that a single machine learning technique was insufficient to give good results. So, three methods like ID3 Decision Tree, Naive Bayes Classification, and Compact Weight Associative Classifiers are used to develop a novel algorithm. SQL Server is selected to store the transaction records.VB.NET Framework is used to implement the algorithm in Windows Platform. The Feature Enhancements have been developed with the VB.NET and SQL Server database.

3.1. Development of graphical user interface machine learning framework

Patient invoice, inventory, billing management, and the statistics generation are widely supported by large number of health care information systems. Health care decision support systems are employed in various hospitals; however, these are limited support systems. These support system can answer uncomplicated questions like "How many procedures resulted in hospital stays of more than ten days?" and "Identify the female cancer patients who are single and over 30 years old." Clinical decision assistance integrated with automated system based clinical records can lower clinical faults and patient safety improvement, reduces unnecessary variation in treatment, and patient outcomes betterment. This technique assures a promise since data modelling and analysis techniques, like the machine learning, can develop a deeper knowledge domain that the quality of healthcare discernment are enhanced dramatically.

4. Problem formation and data set

The proposed medical prediction system uses UCI Machine Learning Repository for the experimentation. The dataset used for the experimental setup utilizes 14 features. In these 14 features eight of the record are categorical, and six of the record are numeric.

Machine learning algorithms viz., Naive Bayes, Decision Trees, and Compact Weighted Associative Classifiers are utilized for the this prospective Decision Support methodology in Medical Prediction. The findings reveal that each strategy has a distinct advantage in achieving the objectives and established learning goals. Cancer, heart disease, Hepatitis, and diabetes are among the conditions for which data is collected. The Patients Age, Sex, Serum Cholesterol, Chest Pain Type, Resting BP, Results Resting ECG, Maximum Heart beat rate, Fasting Blood Sugar, Exercise-Induced Angina, which is a kind of chest pain due to lesser blood flow in to the heart, Old Peak reading, Slope of Peak Exercise, Thal, Number of Major Vessels are among the 13 variables collected in a Heart Dataset. A Cancer Dataset is collected with seven attributes Age, node-caps, breast-quad, Menopause, tumour-size, Inv-nodes, irradiate. A Diabetic dataset holds with seven attributes, like Body mass index, Age, plasma glucose concentration, serum insulin, diastolic blood pressure, triceps skin fold thickness, and the diabetes pedigree function.

This system is used to predict the patients' disease and achieves better results by comparing the performance of these algorithms and calculating their accuracies. The main focus of this proposed research work is to build a Decision Support in Clinical Prediction System as a web enabled questionnaire system. The end user response can uncover and extract hidden patterns and relationships, which are the knowledge associated with the specific disease from a large chronicled clinical disease database system.

5. Design and methodology of the decision support in medical prediction system

5.1. Overview

The approach includes two steps. Firstly, critical Clinical features are selected as the datasets. Secondly, an artificial Classification techniques algorithm is developed for classifying illness established on these clinical features. Proposed approach is efficient in predicting the patients risk level. The system utilized a novel approach that is used to classify diseases.

5.2. Basic Measures in Classification

In classification methods, there are two fundamental measures. Probability and the Maximum Posteriori Hypothesis are the two. A conditional probability, which is the probability of certain conclusion C, for a given some evidence E, and the C and E have a dependency connection.

$$\mathrm{P(C|E)} = \frac{\mathrm{P(E|C)P(C)}}{\mathrm{P(E)}}$$

There are m classes in the maximum posterior hypothesis, C1, C2,... Cm. X belongs to the class with the highest posterior probability, which the classifier will predict, conditioned on X, stated X's tuple. That is, given one jm, j I, P (Ci|X)>P (Cj|X) is only subjected that the naive Bayesian classifier predicts that the given tuple x belongs to the class Ci.

As a outcome, P (Ci|X) was maximized. The class Ci, the maximum posterior hypothesis is for P maximization of P (Ci|X). According to Bayes' theorem

$$\mathbf{P(C_i\,|X)} = \frac{\mathbf{P(X|Ci)P(Ci)}}{\mathbf{P(X)}}$$

5.3. Basic Measures in Decision Tree

There are two important measures in Classification methods. They are Entropy and Information Gain. To calculate the homogeneity of a sample an entropy formula is utilized. A fully homogeneous sample holds entropy of 0. Entropy of 1 is visible with an equally divided sample. Entropy(s) = + plog2 (p+) -plog2 (p-) for a negative and positive element samples. The generalized formula for the entropy is derived as:

$$\textbf{Entropy (S)} = \sum_{i=1}^{c} \mathbf{p_{i\,\log_2}\,p_i}$$

ID3 considers gain in information, as its attribute selection measure. This amount is measured by

$$\mathbf{Info_A(D)} = \sum_{\mathbf{j=1}}^{\mathbf{v}} \frac{\mathbf{|Dj|}}{\mathbf{|D|}} \times \mathbf{Info(Dj)}$$

The expression |Dj| / |D| are the weight of the partition j. The anticipated information needed to classify a tuple out-of D established on partitioning of A is InfoA(D). The variance in the initial information requirement, which is particularly based on the proportion of classes and the information gain, obtained by partitioning on A is the new need.

Gain (A) = Info (D)-InfoA (D)

5.4. Basic Measures In Association Rule Mining

The sensitivity, specificity, and accuracy measures can be calculated. These measures are defined as

$$\mathbf{Sensitivity} = \frac{\mathbf{t-pos}}{\mathbf{pos}}$$

The number of true positives, which are the correctly identified healthy samples is T-pos, while the number of healthy positive samples is pos.

$$\mathbf{Specificity} = \frac{\mathbf{t-neg}}{\mathbf{neg}}$$

T-neg is the number of true negatives which are the correctly classified sick samples a neg is the number of negative samples which are sick.

$$\mathbf{Accuracy} = \mathbf{Sensitivity}\frac{\mathbf{pos}}{(\mathbf{pos+neg})} + \mathbf{Specificity}\frac{\mathbf{neg}}{(\mathbf{pos+neg})}$$

6. Methodology of the proposed system

Two steps are followed in building Compact Weighted Associative classifier. In the first step, The complete weighted class association rules are created in the initial step, based on the weighted association rule method. The weight is considered an essential feature to find the weighted association rule. From this, the weights for each item can be derived. To evaluate the rule weighted confidence and weighted support measures are used here. The class association rule is most important if the weighted support is larger than the minimum. Further the rules that assure some threshold conditions are utilized to build the classifier after ranking the rules. As the rule ranking is executed, its followed by considering specifically the high-ranking rules are selected to generate a classifier, and the rest of the rules are pruned.

PSEUDOCODE:

a. Choose an information gain attribute from the given dataset.

b. Weights are assigned established on that chosen attribute.

c. Compute the weighted support.

d. Create a two-item set and so forth, if an items weighted support is more outstanding than minimum weighted support,

e. Calculate weighted confidence for all the item set after creating all the item sets.

f. Weighted Confidence and Weighted Support are ranked based on the rules.

g. Classify the dataset that are on test, using these rule set and acquire the classifier accuracy.

h.

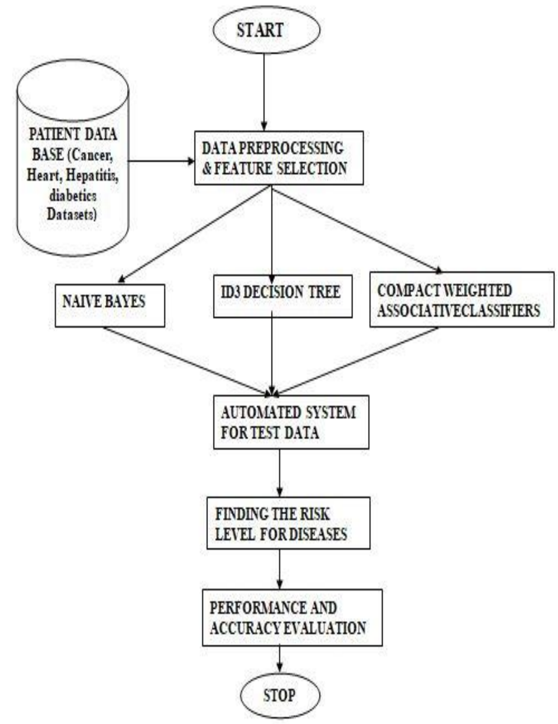


Fig. 1 - Flow Diagram of the Proposed Medical Prediction System

6.1. Modules Description

- Pre-processing of the Patients Dataset from Different Departments:

The UCI database provided a dataset containing medical attributes. Initially, the dataset includes 75 attributes. The pre-processing data facility enables the system to consolidate required information from different depots and remove the dirty information from redundant data. After the pre-processing method, the dataset is reduced to 13 attributes for the Heart dataset, seven attributes for cancer and Hepatitis, respectively.

- Implementation of the ID3Decision tree, Naive Bayes Classification, and CWAC:

This module implements the ID3Decision tree, Naive Bayes Classification, and CWAC algorithms using a trained dataset. The probabilities of all attributes are calculated using Naïve Bayes Algorithm, Information Gain for all attributes is calculated using ID3 Decision tree Algorithm, and the frequent itemsets using weighted support and confidence are calculated using CWAC.

- Implementation of the Medical Prediction System:

This module gets test data from the user and predicts the result about the risk level of the disease to the user. The result of prediction about the disease is given by the three algorithms, namely ID3Decision tree, CWAC, and Naive Bayes Classification.

- Performance Comparison of the Algorithms:

Finally, the performance analysis of the entire system is studied, and they are tested under various test plan cases. The accuracy of the algorithms is also calculated.

6.2. Database Description

Four tables are used to store the transaction records. Schema of the table is given below,

Table 1 - Heart dataset

Name	Type
Age	Numeric
Sex	Numeric
Chest Pain Type	Numeric
Serum Cholesterol	Numeric
Resting BP.	Numeric
Maximum Heart Rate	Numeric
Fasting Blood Sugar	Numeric
Angina caused by Exercise	Numeric
Previous peak	Numeric
Peak Exercise slope	Numeric
Number of Major Vessels	Numeric
Thal	Numeric

Table 2 - Cancer dataset

Name	Type
Age	Numeric
Size of the Tumour	Numeric
Menopause	Numeric
Breast-quad	Numeric
Presence of Node-caps	Numeric
Inv-nodes	Numeric
Irradiate	Numeric

Table 3 - Diabetics dataset

Name	Type
Age in years	numeric
Body mass index	numeric
Diastolic blood pressure	Numeric
Concentration of the plasma glucose	Numeric
Serum insulin	Numeric
Diabetes Pedigree Function	Numeric
Triceps skin fold thickness	Numeric

6.3. Data Sets and ARFF File Creation

In spite of the clinical decision support tool being deployed or its fundamental learning method, they all depend on input data to land at their classification predictions. Its visible that the quality of these input data holds a high influence on the performance and the process of the machine learning. Hence the data preparation and employing training many of time consume large sum of effort associated with clinical decision support tools. Further here the experimentation explores few primary topics related to these data optimizing and their use in classifier performance validation.

The Attribute Relation File Format, abbreviated as ARFF file is an ASCII text file that illustrate a list of instances sharing a set of attributes. Attribute Relation File Format files were developed by UCI Machine Learning Repository for the operation with the Weka machine learning software. Data of the patients records of heart, cancer, hepatitis, and diabetics dataset training sample is given in arff or CSV format as input in Weka 3.8.

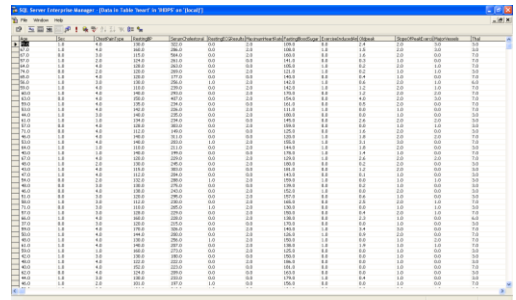


Fig. 2 - Datasets inserted in SQL Server.

7. Results analysis of the proposed system

7.1. Results Analysis of in the Weka Tool

The ARFF file is given as input in the Weka tool, and classification methods like Naive Bayes Classification ID3 Decision Tree are applied. The correctly classified, incorrectly classified instances and the confusion matrix are displayed.

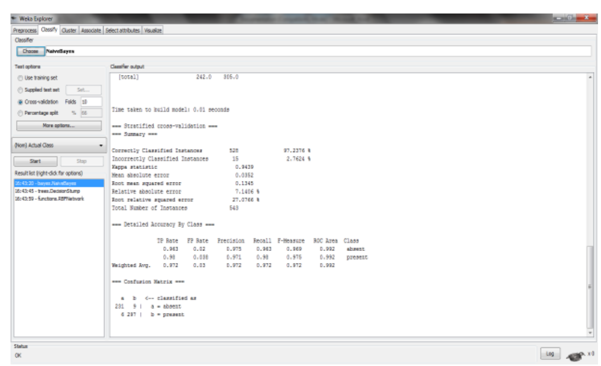


Fig. 3 - Result of Naïve Bayes in Weka

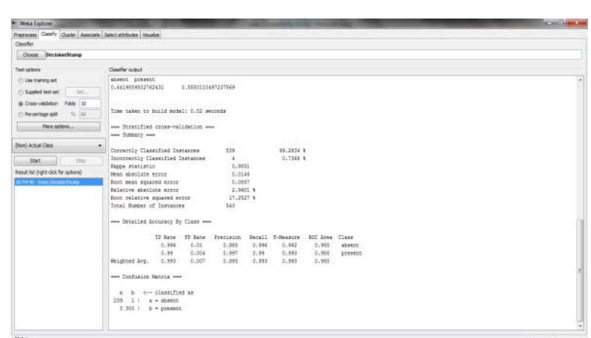


Fig. 4 - Result of ID3 decision tree in Weka

The Weka tool provides information about the correctly and incorrectly classified instances. Thus, the accuracy is calculated from it.

7.2.Results Of Existing And Proposed System

The proposed system uses Compact Weighted Associative Classifiers Algorithm with 13 attributes and 543 records. The performance analysis for the algorithms is done, and its accuracy is calculated and tabulated.

Table 4 - Accuracy Results of the proposed System

DATASETS	NAIVE BAYES	ID3 DECISION TREE	CWAC
Heart	92.84	92.67	98.54
Cancer	88.50	78.33	90.51
Hepatitis	69.27	80.36	86.14
Diabetics	80.10	83.94	91.44

The accuracy of the trained sample of the heart dataset is represented graphically.

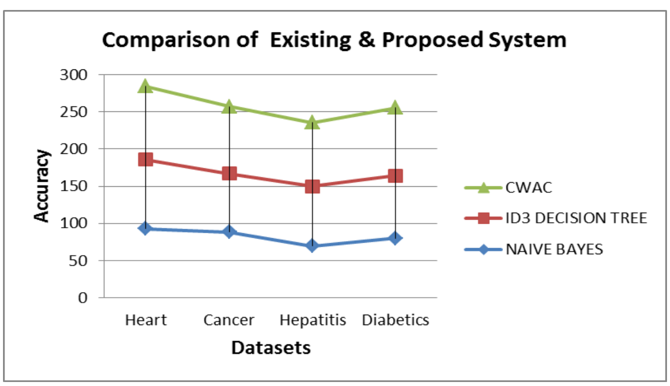


Fig. 5 - Comparison of Existing and Proposed System

7.3. Prediction Results by the Machine Learning framework

The framework for applying user test data is created using VB.NET, which can predict heart disease by the user. A few sample framework is depicted in the figures below.

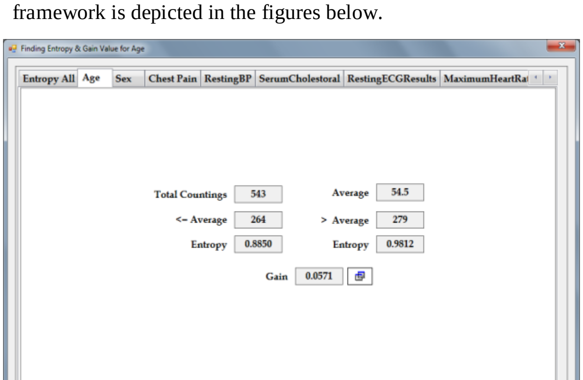


Fig. 6 - Finding Gain Value using ID3 Decision Tree

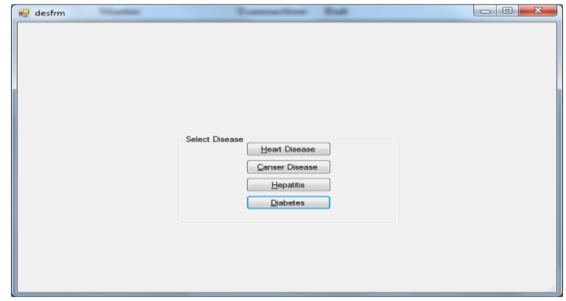


Fig. 7 - Selecting Disease type for Test Data

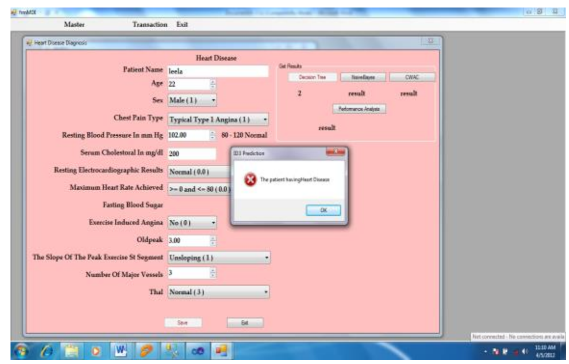


Fig. 8 - Prediction Results using ID3 Decision Tree with Heart test data

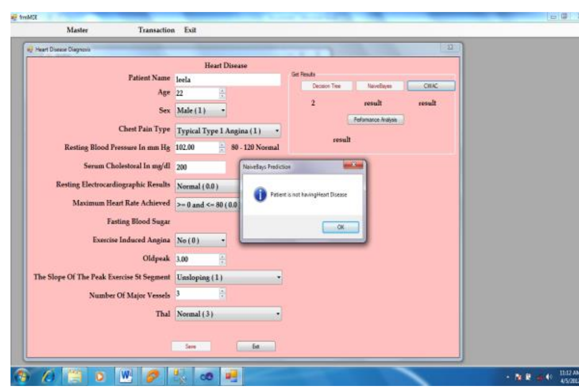


Fig. 9 - Prediction Results using Naive Bayes with Heart test data

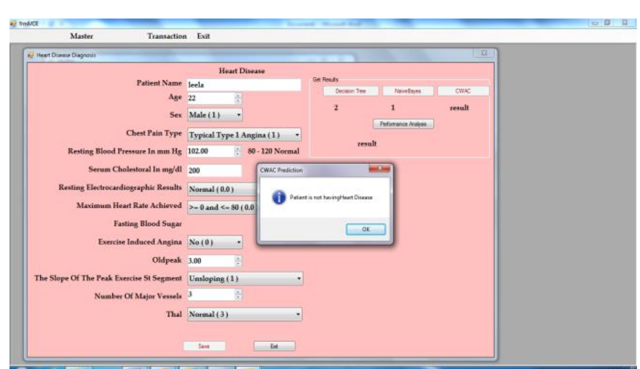


Fig. 10 - Prediction Results using CWAC with Heart test data

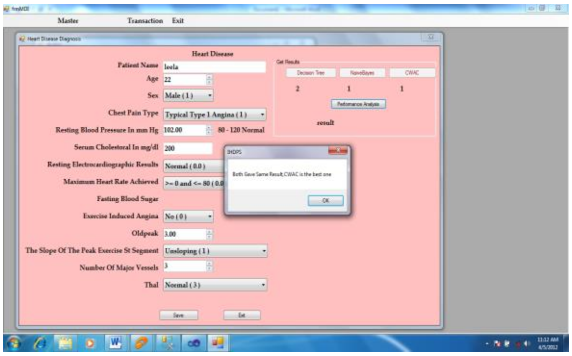


Fig. 11 - Performance Analysis of the System

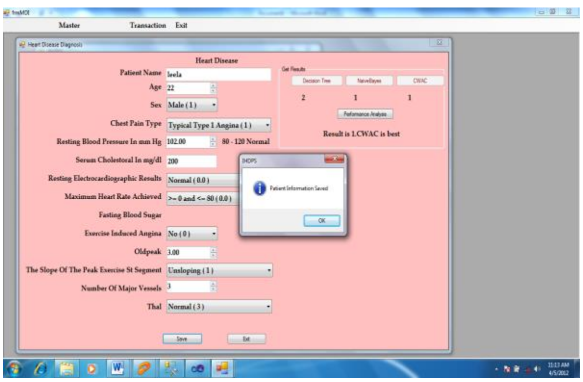


Fig. 12 - System of the Patient Information

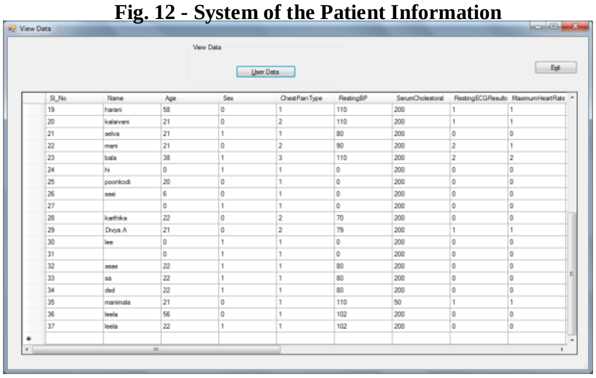
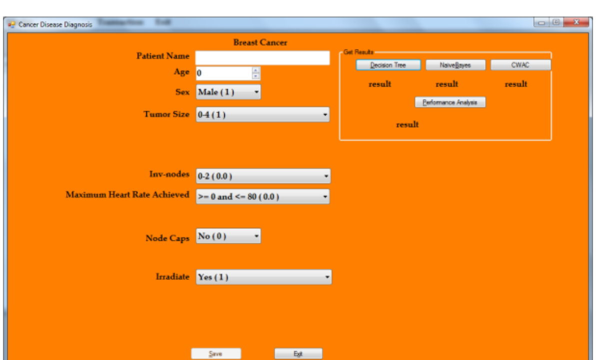


Fig. 13 - Test Data Loading in the System



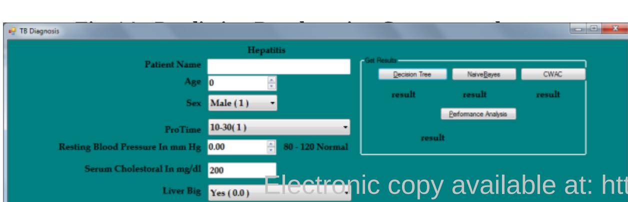


Fig. 15 - Prediction Results using Hepatitis test data

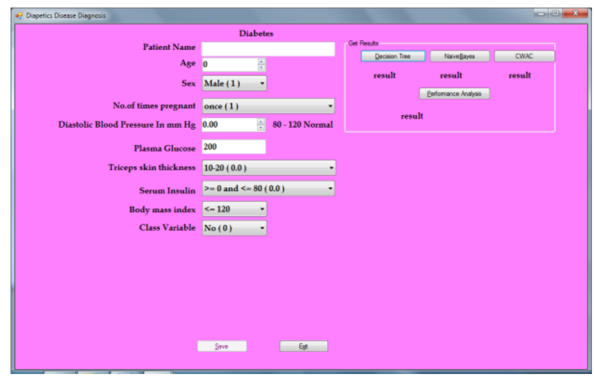


Fig. 16 - Prediction Results using Diabetics test data

8. Conclusion and future scope

Medical prediction system BY DSS is a complete automated research tool that behaves as a user-friendly tool. It exactly predicts the person will get heart disease or not. It is straightforward to update records and data with this software developed. All details which cause heart disease are considered in this paper.

The system can be improved and expanded in the future. It can, for example, include medical qualities other than the 13 listed. Different machine learning methods can be incorporated, such as Clustering, Time Series, and Association Rules. Continuous data can be employed rather than categorical data,. Another area to look at is using text analytics to harvest the massive amounts of unstructured data in healthcare systems. Integrating machine learning and text analytics would be another hurdle. The severity of the cardiac illness can be assessed using fuzzy learning algorithms. The Cleveland Heart Disease database provided a dataset containing medical attributes.

References

C.L.Blake, C.J.Mertz, “U.C.I. Machine Learning Repository," http://archive.ics.uci.edu/ml/datasets.php," 2021.

Francesco L. et al., "Requalifying public buildings and utilities using a group decision support system," Journal of Cleaner Production.164: 1081-1092, 2017.

Charly, "Data Mining for the Enterprise," IEEE International Conference on System Sciences, 7, 295-304, 2005.

Fayyad G., "Data Mining and Knowledge Discovery in Databases Implications from scientific databases," International Conference on Scientific and Statistical Database Management, USA, 2-11, 2003.

Mani S, Ozdas A, Aliferis C, Varol HA, Chen Q, Carnevale R, Chen Y, Romano-Keeler J, Nian H, Weitkamp JH. "Medical decision support using machine learning for early detection of late-onset neonatal sepsis," Journal of Am Med Inform Assoc. 326-36, DOI: 10.1136/amiajnl-2013-001854, 2014.

Han, J., Kamber, M, "Data Mining Concepts and Techniques," Morgan Kaufmann Publishers, 2006.

T. JH, "Data Mining and Data Warehousing," Prentice-Hall, 2007.

Away she A, et al., "Review of Medical Decision Support and Machine-Learning Methods," Vet Pathol, 56(4), 512-525, 2019, DOI: 10.1177/0300985819829524.

Akshaya Karthikeyan, Akshit Garg, Vinod PK, U Deva Priyakumar, "Clinician involvement in research on machine learning-based predictive clinical decision support for the hospital setting: A scoping review. J Am Med Inform Assoc. 2021 Mar 1;28(3):653-663. DOI: 10.1093/jamia/ocaa296.

Yahyaoui, et al., "A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques," International Informatics and Software Engineering Conference (UBMYK), 2019, pp. 1-4, DOI: 10.1109/UBMYK48245.2019.8965556.

Rani, P., Kumar, R., Ahmed, NMOS, et al. A decision support system for heart disease prediction based upon machine learning. J Reliable Intell Environ 7, 263–275, 2021. https://doi.org/10.1007/s40860-021-00133-6

Kumar, D. T. S., "Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm," Journal of Artificial Intelligence and Capsule Networks, issue 2(3), 185-193. doi:10.36548/jaicn, 2020.

Tang, MacLennan, "Data Mining with SQL Server," Indianapolis: Wiley, 2005.

Thuraisingham, B.: "A Primer for Understanding and Applying Data Mining," IT Professional, 28-31, 2002.

Yan L et al., "Consider the laboratory aspects in developing patient prediction models," Nature Machine Intelligence, 3-19, 2021. DOI: 10.1038/s42256-020-0220-3.

Safdar, S., Zafar, S., Zafar, N. et al. Machine learning-based decision support systems (DSS) for heart disease diagnosis: a review. Artif Intell Rev 50, 597–623 (2018). https://doi.org/10.1007/s10462-017-9552-8.

Jiawei Han: "Data Mining: Concepts and Techniques": Second Edition-Micheline Kamber, 2008.

Obenshain, "Application of Data Mining Techniques to Healthcare Data," Infection Control and Hospital Epidemiology, Vol. 02, No. 02, pp- 250-255, 2010.

Wilson PW et al., "Prediction of coronary heart disease using risk factor categories" Circulation, 2005.

G.Subbalakshmi, "Decision Support in Heart Disease Prediction System using Naive Bayes," Indian Journal of Computer Science and Engineering, Vol. 01, No. 03, pp- 260-270, 2010.

Kumar, Y., et al., "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," Journal of Ambient Intelligence and Humanized Computing, 2022, https://doi.org/10.1007/s12652-021-03612-z

Sanchez-Martinez S et. al., "Machine Learning for Clinical Decision Making: Challenges and Opportunities in Cardiovascular Imaging," Frontiers in CardiovascularMedicine,2022edicine 8:765693, 2022, doi: 10.3389/fcvm.2021.765693