Development & Deployment of web application for Prediction of Cardiovascular Disease (CVD) by using Machine Learning Algorithms

1. Sharanabasappa R Patil, 2. Dr. S A Madival

M.Tech. Scholar, Dept. of AI & DS, Faculty of Engineering & Technology (Co-Edu.), Sharnbasva University, Kalburagi, Karntaka-India, sharanhvac@gmail.com

Chairperson & Professor, Department of Computer Science & Engineering, Faculty of Engineering & Technology (Co-Edu.), Sharnbasva University, Kalburagi, Karntaka-India.

ABSTRACT: Heart disease is a danger to people's health because of its prevalence and high mortality risk. Predicting cardiac disease early using a few simple physical indications collected from a routine physical examination has become difficult. Clinically, it is critical and sensitive for the signs of heart disease for accurate forecasts and concrete steps for future diagnosis. The manual analysis and prediction of a massive volume of data are challenging and time-consuming. In this paper, a unique heart disease prediction model is proposed to predict heart disease correctly and rapidly using a variety of bodily signs. Predicting heart disease is regarded as one of the most difficult challenges in the health-care profession. To predict cardiac disease, researchers employed a variety of algorithms including LDA, RF, GBC, DT, SVM, and KNN, as well as the feature selection algorithm sequential feature selection. For verification, the system employs the K-fold cross-validation approach. These ten strategies were used to conduct the comparative study. The Dataset for Cleveland, Hungray, Switzerland, and Long Beach V, as well as the Dataset Heart Statlog Cleveland Hungary, were used to assess the models performance.

Keywords: *Cardiovascular disease, heart disease data set, mixed machine learning techniques; numerical features; categorical features; SVM, DT, RF, MLP, SMO, GBC, LDA, KNN, XGB, VC .*

DOI: 10.48047/ecb/2023.12.8.734

INTRODUCTION

The study of disease diagnosis is crucial in the realm of healthcare [1]. A disease is defined as any cause or set of conditions that lead to suffering, sickness, malfunction, or finally death of a human person. Individuals have the fundamental right to good health, according to WHO principles [2]. It is thought that proper health care services should be offered for frequent health checks. Heart disease is the leading cause of death in the world, accounting for nearly 31% of all deaths. Early detection and treatment of many cardiac disorders are highly challenging, especially in poor countries, due to a lack of diagnostic centers, skilled doctors, and other resources that affect the proper prognosis of heart disease [3]. Some prevalent risk factors, such as diabetes, high blood pressure, and excessive cholesterol, make it difficult to detect heart disease. Underlying disorders induce irregular cardiac rhythms and breathing difficulties, such as pulmonary cracks, improved jugular vein weight, and borderline edema [4]. Because the symptoms of cardiac disease are so varied, they must be treated with extreme caution. Failure to do so may have a negative impact on the heart [5]. According to the American College of Cardiology, there are 26 million individuals globally who have heart disease, and 3.6 million people are tested each year. Within a year 15–35 % of individuals with heart disease will die, and the rest will die in 4–5 years. Diseases can affect a person physically and mentally, and they can have a significant impact on how they live. The pathological process is defined as the study of the causes of disease.

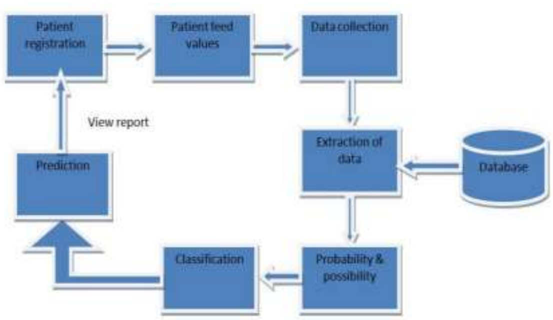

Fig.1: Example figure

Disease diagnosis is the most difficult procedure and, at the same time, a crucial phenomenon for a medical care expert to understand before reaching a judgment. The diagnostic procedure might be lengthy and difficult. Care specialists collect empirical facts to determine a patient’s disease to reduce the uncertainty in medical diagnosis health. Due to a mistake in the diagnostic process, the patient’s necessary therapy may be postponed or ignored, resulting in major health complications. Unfortunately, not every doctor is a specialist in every field of medicine. As a result, an independent verdict system was needed that combined human understanding with Machine Learning (ML) precision [7]. To get accurate outcomes from the diagnosis procedure at a lower cost, we need a good decision support system. For human specialists, classifying disease based on multiple factors is a difficult task, but machine learning techniques might assist to detect and treat such situations. Various machine learning approaches are currently being applied in medicine to accurately diagnose cardiovascular disease. ML is a component of computer science that enables computers to become more intelligent. Learning is a must for every intelligent system. Learning-based strategies in machine learning include sequential feature technique. One of the most significant is Artificial Intelligence (AI) technologies in the medical field is a rule-based intelligent system, which provides a collection of if-then rules in medical healthcare and works as a decision support system. AI-based autonomous techniques need very little human interaction are progressively adding intelligent systems in the medical business [8]. In recent years, medical aid software has been developed using computer technology and machine learning techniques as a support system for the early identification of cardiovascular disease [9]. Early detection of any heart-related disease can lower the chance of mortality [10].

2. LITERATURE REVIEW

Towards achievement of universal health care in Indiaby 2020: a call to action:

To sustain the positive economic trajectory that India has had during the past decade, and to honour thefundamental right of all citizens to adequate health care, the health of all Indian people has to be given the highestpriority in public policy. We propose the creation of the Integrated National Health System in India throughprovision of universal health insurance, establishment of autonomous organisations to enable accountable andevidence-based good-quality health-care practices and development of appropriately trained human resources, therestructuring of health governance to make it coordinated and decentralised, and legislation of health entitlementfor all Indian people. The key characteristics of our proposal are to strengthen the public health system as theprimary provider of promotive, preventive, and curative health services in India, to improve quality and reduce theout-of-pocket expenditure on health care through a well regulated integration of the private sector within thenational health-care system. Dialogue and consensus building among the stakeholders in the government, civilsociety, and private sector are the next steps to formalise the actions needed and to monitor their achievement. Inour call to action, we propose that India must achieve health care for all by 2020.

Changes in causes of death in systemic sclerosis, 1972–2002:

Background: Survival of scleroderma has changed since the renal crisis treatment has become possible. **Aims:** To document the changes in survival and organ system causes of mortality in systemic sclerosis (SSc) over the past 25 years in patients from a single medical centre.**Methods:** Consecutive patients evaluated at the University of Pittsburgh, Pittsburgh, Pennsylvania, USA between 1 January 1972 and 31 December 1996 were studied. Survival was determined in five 5-year time periods between 1972 and 1997. Causes of death included scleroderma-related (scleroderma renal crisis, pulmonary arterial hypertension, pulmonary fibrosis (PF), gastrointestinal (GI), heart and multiorgan failure) and non-scleroderma-related (cancer, atherosclerotic cardiovascular or cerebrovascular disease, infection, sudden death, other and unknown) causes. **Results:** The 10-year survival improved steadily from 54% to 66% during each of the time intervals. There was a significant improvement in survival for patients during 1982–91 compared with those during 1972–81 ($p<0.001$), even when patients with renal crisis were excluded ($p<0.005$). The frequency of deaths due to renal crisis significantly decreased over the 30-year time period, from 42% to 6% of scleroderma-related deaths ($p<0.001$), whereas the proportion of patients with scleroderma who died of PF increased from 6% to 33% ($p<0.001$). The frequency of pulmonary hypertension, independent of PF, also significantly increased during this time period ($p<0.05$). There were no changes in scleroderma GI- and heart-related deaths, nor in any of the non-scleroderma-related causes, although patients with scleroderma were less likely to die from scleroderma-related problems in the past 15 years. **Conclusion:** The change in the pattern of scleroderma-related deaths over the past 30 years demonstrates that the lung (both pulmonary hypertension and PF) is the primary cause of scleroderma-related deaths today. It is important that aggressive searches continue to develop better therapies for these severe pulmonary complications of SSc.

Predicting the likelihood of heart failure with a multi level risk assessment using decision tree:

Heart failure comes in the top causes of death worldwide. The number of deaths from heart failure exceeds the number of deaths resulting from any other causes. Recent studies have focused on the use of machine learning techniques to develop predictive models that are able to predict the incidence of heart failure. The majority of these studies have used a binary output class, in which the prediction would be either the presence or absence of heart failure. In this study, a multi-level risk assessment of developing heart failure has been proposed, in which a five risk levels of heart failure can be predicted using C4.5 decision tree classifier. On the other hand, we are boosting the early prediction of heart failure through involving three main risk factors with the heart failure data set. Our predictive model shows an improvement on existing studies with 86.5% sensitivity, 95.5% specificity, and 86.53% accuracy.

The science of clinical practice: Disease diagnosis or patient prognosis? Evidence about "what is likely to happen" should shape clinical practice:

Diagnosis is the traditional basis for decision-making in clinical practice. Evidence is often lacking about future benefits and harms of these decisions for patients diagnosed with and without disease. We propose that a model of clinical practice focused on patient prognosis and predicting the likelihood of future outcomes may be more useful. Disease diagnosis can provide crucial information for clinical decisions that influence outcome in serious acute illness. However, the central role of diagnosis in clinical practice is challenged by evidence that it does not always benefit patients and that factors other than disease are important in determining patient outcome. The concept of disease as a dichotomous 'yes' or 'no' is challenged by the frequent use of diagnostic indicators with continuous distributions, such as blood sugar, which are better understood as contributing information about the probability of a patient's future outcome. Moreover, many illnesses, such as chronic fatigue, cannot usefully be labelled from a disease-diagnosis perspective. In such cases, a prognostic model provides an alternative framework for clinical practice that extends beyond disease and diagnosis and incorporates a wide range of information to predict future patient outcomes and to guide decisions to improve them. Such information embraces non-disease factors and genetic and other biomarkers which influence outcome. Patient prognosis can provide the framework for modern clinical practice to integrate information from the expanding biological, social, and clinical database for more effective and efficient care.

Doctor AI: Predicting Clinical Events via Recurrent Neural Networks:

Leveraging large historical data in electronic health record (EHR), we developed Doctor AI, a generic predictive model that covers observed medical conditions and medication uses. Doctor AI is a temporal model using recurrent neural networks (RNN) and was developed and applied to longitudinal time stamped EHR data from 260K patients and 2,128 physicians over 8 years. Encounter records (e.g. diagnosis codes, medication codes or procedure codes) were input to RNN to predict (all) the diagnosis and medication categories for a subsequent visit. Doctor AI assesses the history of patients to make multilabel predictions (one label for each diagnosis or medication category). Based on separate blind test set evaluation, Doctor AI can perform differential diagnosis with up to 79% recall@30, significantly higher than several baselines. Moreover, we demonstrate great generalizability of Doctor AI by adapting the resulting models from one institution to another without losing substantial accuracy.

3. METHODOLOGY

The most challenging step in diagnosing a disease is also one that a medical professional must fully comprehend before making a decision. The diagnostic process may take a while and be challenging. Care professionals use empirical data to identify a patient's illness in order to lower the level of uncertainty in medical diagnosis. The patient's required therapy may be put off or disregarded as a result of a diagnostic error, leading to serious health issues. Sadly, not every physician has advanced training in every area of medicine. A method of impartial judgement that combines human comprehension with Machine Learning (ML) accuracy was therefore required. We require a good decision support system to obtain correct results from the diagnosis procedure at a cheaper cost.

Disadvantages:

1. The diagnostic process could be time-consuming and challenging.
2. The patient's required therapy may be put off or disregarded as a result of a diagnostic error, leading to serious health issues.

Researchers used a number of algorithms, including LDA, RF, GBC, DT, SVM, and KNN, as well as the feature selection technique sequential feature selection, to predict cardiac illness. The system uses the K-fold cross-validation technique for verification. The comparison research was conducted using these six methodologies.

Advantages:

1. Decision Tree Classifier sfs and Random Forest Classifier sfs generated the greatest and very similar accuracy values.
2. The SFS approach can reduce calculation time while enhancing the classification accuracy of the classifier.

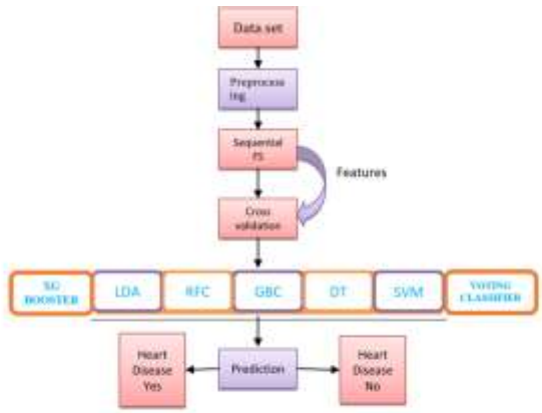

Fig.2: System architecture

MODULES:

In this project we have designed following modules

- Data exploration: using this module we will load data into system
- Processing: Using the module we will read data for processing
- Splitting data into train & test: using this module data will be divided into train & test
- Model generation: Build models: SVM, RF, DT J48, MLP, SMO-LinearSVC, Gradient boosting, LDA, KNN, XGBoost and Voting classifier. Algorithms accuracy calculated
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

4. IMPLEMENTATION

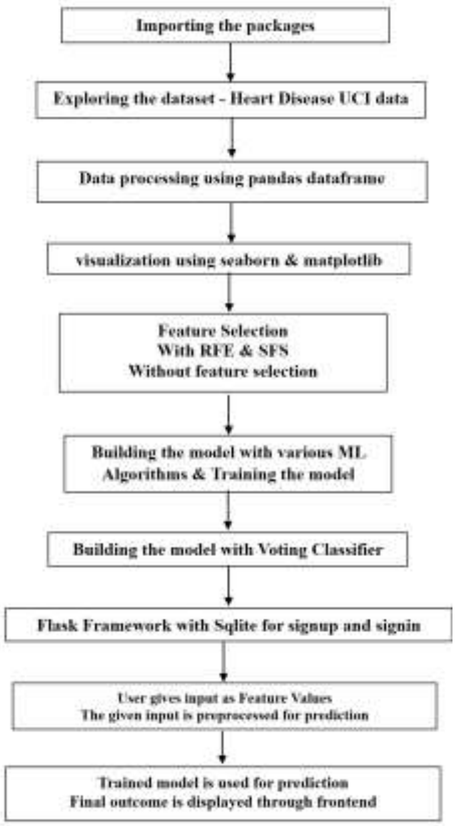

Fig:3 Work flow of the proposed system

SVM:

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

RF:

Random Forest (RF) algorithm is one of the best algorithms for classification. RF is able for classifying large data with accuracy. It is a learning method in which number of decision trees are constructed at the time of training and outputs of the modal predicted by the individual trees.

DT J48:

J48 is based on a top-down strategy, a recursive divide and conquer strategy. You select which attribute to split on at the root node, and then you create a branch for each possible attribute value, and that splits the instances into subsets, one for each branch that extends from the root node.

MLP:

MLPClassifier stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network. Unlike other classification algorithms such as Support Vectors or Naive Bayes Classifier, MLPClassifier relies on an underlying Neural Network to perform the task of classification.

SMO-LinearSVC:

Linear Support Vector Machine (Linear SVC) is an algorithm that attempts to find a hyperplane to maximize the distance between classified samples.

Gradient boosting:

Gradient boosting is a type of machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error.

LDA:

The Amazon SageMaker Latent Dirichlet Allocation (LDA) algorithm is an unsupervised learning algorithm that attempts to describe a set of observations as a mixture of distinct categories. LDA is most commonly used to discover a user-specified number of topics shared by documents within a text corpus.

KNN:

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point.

XGBoost:

XGBoost is a popular and efficient open-source implementation of the gradient boosted trees algorithm. Gradient boosting is a supervised learning algorithm, which attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models.

Voting classifier:

A voting classifier is a machine learning estimator that trains various base models or estimators and predicts on the basis of aggregating the findings of each base estimator. The aggregating criteria can be combined decision of voting for each estimator output.

5. EXPERIMENTAL RESULTS

Table:1 Precision, Recall, F1 Score & accuracy comparison of various ML models

Sl. No	Machine Learning Model (Algorithm)	Precision		Recall		F1-Score		Accuracy
		0	1	0	1	0	1	
1.	Support Vector Machine (SVM)	0.71	0.70	0.67	0.75	0.69	0.72	70.82%
2.	Decision Tree (DT)	0.84	0.86	0.85	0.84	0.84	0.85	84.68%
3.	Random Forest (RF)	1.00	1.00	1.00	1.00	1.00	1.00	100%
4.	Multilayer Perceptron (MLP)	0.84	0.86	0.85	0.85	0.85	0.85	85%
5.	Sequential minimal optimization (SMO)	0.90	0.81	0.78	0.92	0.83	0.86	84.70%
6.	Grading Booster Classifier (GBC)	0.91	0.89	0.88	0.92	0.90	0.90	89.95%
7.	Linear Discriminant Analysis (LDA)	0.90	0.80	0.71	0.92	0.82	0.86	84%
8.	K Nearest Neighbors (KNN)	1.00	1.00	1.00	1.00	1.00	1.00	100%
9.	XG Boost (XGB)	1.00	1.00	1.00	1.00	1.00	1.00	100%
10.	Voting Classifier (VC)	1.00	1.00	1.00	1.00	1.00	1.00	100%

Table:6.2 Precision comparison of various ML models

Sl No	Machine Learning Model (Algorithm)	Precision	
		0	1
1.	Support Vector Machine (SVM)	0.71	0.70
2.	Decision Tree (DT)	0.84	0.86
3.	Random Forest (RF)	1.00	1.00
4.	Multilayer Perceptron (MLP)	0.84	0.86
5.	Sequential minimal optimization (SMO)	0.90	0.81
6.	Grading Booster Classifier (GBC)	0.91	0.89
7.	Linear Discriminant Analysis (LDA)	0.90	0.80
8.	K Nearest Neighbors (KNN)	1.00	1.00
9.	XG Boost (XGB)	1.00	1.00
10.	Voting Classifier (VC)	1.00	1.00

Fig:4 Precision of ML Algorithms for Target

Table:6.3 Recall comparison of various ML models

Sl No	Machine Learning Model (Algorithm)	Recall	
		0	1
1.	Support Vector Machine (SVM)	0.67	0.75
2.	Decision Tree (DT)	0.85	0.84
3.	Random Forest (RF)	1.00	1.00
4.	Multilayer Perceptron (MLP)	0.85	0.85
5.	Sequential minimal optimization (SMO)	0.78	0.92
6.	Grading Booster Classifier (GBC)	0.88	0.92
7.	Linear Discriminant Analysis (LDA)	0.75	0.92
8.	K Nearest Neighbors (KNN)	1.00	1.00
9.	XG Boost (XGB)	1.00	1.00
10.	Voting Classifier (VC)	1.00	1.00

Fig:4 Recall of ML Algorithms for Target

Table:6.4 F-1 Score comparison of various ML models

Sl No	Machine Learning Model (Algorithm)	F-1 Score	
		0	1
1.	Support Vector Machine (SVM)	0.69	0.72
2.	Decision Tree (DT)	0.84	0.85
3.	Random Forest (RF)	1.00	1.00
4.	Multilayer Perceptron (MLP)	0.85	0.85
5.	Sequential minimal optimization (SMO)	0.83	0.86
6.	Grading Booster Classifier (GBC)	0.90	0.90
7.	Linear Discriminant Analysis (LDA)	0.82	0.86
8.	K Nearest Neighbors (KNN)	1.00	1.00
9.	XG Boost (XGB)	1.00	1.00
10.	Voting Classifier (VC)	1.00	1.00

Fig:5 F1 Score of ML Algorithms for Target

Table:6.5 Precision, Recall, F1 Score comparison of various ML models for Target (0)

Sl No	Machine Learning Model (Algorithm)	Target (0)		
		Precision	Recall	F1 Score
1.	Support Vector Machine (SVM)	0.71	0.67	0.69
2.	Decision Tree (DT)	0.84	0.85	0.84
3.	Random Forest (RF)	1.00	1.00	1.00
4.	Multilayer Perceptron (MLP)	0.84	0.85	0.85
5.	Sequential minimal optimization (SMO)	0.90	0.78	0.83
6.	Grading Booster Classifier (GBC)	0.91	0.88	0.90
7.	Linear Discriminant Analysis (LDA)	0.90	0.75	0.82
8.	K Nearest Neighbors (KNN)	1.00	1.00	1.00
9.	XG Boost (XGB)	1.00	1.00	1.00
10.	Voting Classifier (VC)	1.00	1.00	1.00

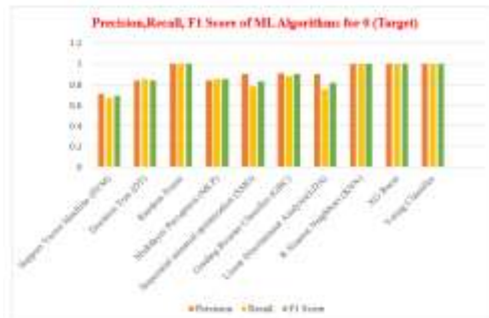

Fig:6 Precision, Recall, F1 Score of ML Algorithms for Target (0)

Table:6.6 Precision, Recall, F1 Score comparison of various ML models for Target (1)

Sl No	Machine Learning Model (Algorithm)	Target (0)		
		Precision	Recall	F1 Score
1.	Support Vector Machine (SVM)	0.70	0.75	0.72
2.	Decision Tree (DT)	0.86	0.84	0.85
3.	Random Forest (RF)	1.00	1.00	1.00
4.	Multilayer Perceptron (MLP)	0.86	0.85	0.85
5.	Sequential minimal optimization (SMO)	0.81	0.92	0.86
6.	Grading Booster Classifier (GBC)	0.89	0.92	0.90
7.	Linear Discriminant Analysis (LDA)	0.80	0.92	0.86
8.	K Nearest Neighbors (KNN)	1.00	1.00	1.00
9.	XG Boost (XGB)	1.00	1.00	1.00
10.	Voting Classifier (VC)	1.00	1.00	1.00

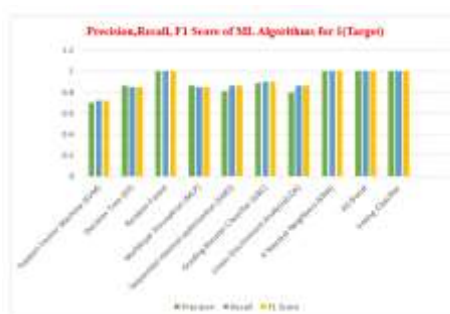

Fig:7 Precision, Recall, F1 Score of ML Algorithms for 1 (Target)

Table:6.7 Accuracy comparison of various ML models

Sl No	Machine Learning Model (Algorithm)	Accuracy
1.	Support Vector Machine (SVM)	70.82%
2.	Decision Tree (DT)	84.68%
3.	Random Forest (RF)	100%
4.	Multilayer Perceptron (MLP)	85%
5.	Sequential minimal optimization (SMO)	84.70%
6.	Grading Booster Classifier (GBC)	89.95%
7.	Linear Discriminant Analysis (LDA)	84%
8.	K Nearest Neighbors (KNN)	100%
9.	XG Boost (XGB)	100%
10.	Voting Classifier (VC)	100%

Fig:8 Accuracy of various ML Models

6. Web application (front End) Screenshots

Fig:9 Home page for heart disease prediction

Fig:10 Login page for heart disease prediction

Fig:11 Registration form for heart disease prediction

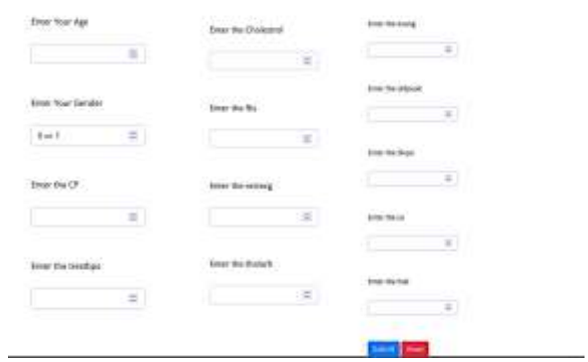

Fig:12 Registration form for heart disease prediction

Fig:13 Sample of output page

7. CONCLUSION

In this paper a prediction strategy based on hybrid intelligent machine learning was developed to diagnose mortality during follow-up. Data from a database of heart disease clinical records were used to evaluate the approach. To choose significant characteristics, one of the most challenging difficulties in medicine is predicting disease sickness. Researchers used a range of algorithms; including SVM, DT, RF, MLP, SMO, GBC, LDA, KNN, XGB, VC, as well as the feature selection approach SFS, to predict cardiac illness. The system uses a K-fold cross-validation technique for verification. These ten approaches were used in the comparison study. The KNN, DT, VC & XGB models got 100% accuracy. The web application was built by using Voting classifier algorithm to predict the cardiovascular disease.

8. FUTURE SCOPE

In this project a prediction strategy based on hybrid intelligent machine learning was developed to diagnose mortality during follow-up. Data from a database of heart disease clinical records were used to evaluate the approach. To choose significant characteristics, one of the most challenging difficulties in medicine is predicting disease sickness. Researchers used a range of algorithms; including SVM, DT, RF, MLP, SMO, GBC, LDA, KNN, XGB, VC, as well as the feature selection approach SFS, to predict cardiac illness. The system uses a K-fold cross-validation technique for verification. These ten approaches were used in the comparison study. The KNN, DT, VC & XGB models got 100% accuracy. The web application was built by using Voting classifier algorithm to predict the cardiovascular disease.

REFERENCES

[1] J. L. Scully, ‘‘What is a disease,’’ EMBO Rep., vol. 5, no. 7, pp. 650–653, 2004.

[2] K. S. Reddy, V. Patel, P. Jha, V. K. Paul, A. S. Kumar, and L. Dandona, ‘‘Towards achievement of universal health care in India by 2020: A call to action,’’ Lancet, vol. 377, no. 9767, pp. 760–768, Feb. 2011.

[3] V. D. Steen and T. A. Medsger, ‘‘Changes in causes of death in systemic sclerosis,’’ Ann. Rheumatic Diseases, vol. 66, no. 7, pp. 940–944, Jul. 2007.

[4] P. Ponikowski, A. A. Voors, S. D. Anker, H. Bueno, J. G. Cleland, and A. J. Coats, ‘‘ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC,’’ Eur Heart J., vol. 37, no. 27, pp. 200–2129, 2016.

[5] A. J. Aljaaf, D. Al-Jumeily, A. J. Hussain, T. Dawson, P. Fergus, and M. Al-Jumaily, ‘‘Predicting the likelihood of heart failure with a multi-level risk assessment using decision tree,’’ in Proc. 3rd Int. Conf. Technol. Adv. Electr., Electron. Comput. Eng. (TAEECE), Apr. 2015, pp. 101–106.

[6] P. Croft, D. G. Altman, J. J. Deeks, K. M. Dunn, A. D. Hay, H. Hemingway, L. LeResche, G. Peat, P. Perel, S. E. Petersen, R. D. Riley, I. Roberts, M. Sharpe, R. J. Stevens, D. A. Van Der Windt, M. Von Korff, and A. Timmis, ‘‘The science of clinical practice: Disease diagnosis or patient prognosis Evidence about ‘what is likely to happen’ should shape clinical practice,’’ BMC Med., vol. 13, no. 1, p. 20, Dec. 2015.

[7] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, ‘‘Doctor AI: Predicting clinical events via recurrent neural networks,’’ in Proc. Mach. Learn. Healthcare Conf., Los Angeles, CA, USA, Aug. 2016, pp. 301–318.

[8] K. Srinivas, B. Rani, and D. Govardhan, ‘‘Applications of data mining techniques in healthcare and prediction of heart attacks,’’ Int. J. Comput. Sci. Eng., vol. 2, pp. 250–255, Jan. 2010.

[9] S. Kaur, J. Singla, L. Nkenyereye, S. Jha, D. Prashar, G. P. Joshi, S. El-Sappagh, M. S. Islam, and S. M. R. Islam, ‘‘Medical diagnostic systems using artificial intelligence (AI) algorithms: Principles and perspectives,’’ IEEE Access, vol. 8, pp. 228049–228069, 2020.

[10] D. Ntiloudi, G. Giannakoulas, D. Parcharidou, T. Panagiotidis, M. A. Gatzoulis, and H. Karvounis, ‘‘Adult congenital heart disease: A paradigm of epidemiological change,’’ Int. J. Cardiol., vol. 218, pp. 269–274, Sep. 2016.

[11] L. Yahaya, N. D. Oye, and E. J. Garba, ‘‘A comprehensive review on heart disease prediction using data mining and machine learning techniques,’’ Amer. J. Artif. Intell., vol. 4, no. 1, pp. 20–29, Apr. 2020.

[12] H. Eyre, R. Kahn, and R. M. Robertson, N. G. Clark, C. Doyle, Y. Hong, T. Gansler, T. Glynn, R. A. Smith, K. Taubert, and M. J. Thun, ‘‘Preventing cancer, cardiovascular disease, and diabetes: A common agendafor the American Cancer Society, the American Diabetes Association, and the American Heart Association,’’ Circulation, vol. 109, no. 25, pp. 3244–3255, Jun. 2004.

[13] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, ‘‘Comparing different supervised machine learning algorithms for disease prediction,’’ BMC Med. Informat. Decis. Making, vol. 19, no. 1, pp. 1–16, Dec. 2019.

[14] I. Olaronke and O. Oluwaseun, ‘‘Big data in healthcare: Prospects, challenges and resolutions,’’ in Proc. Future Technol. Conf. (FTC), Dec. 2016, pp. 1152–1157.

[15] A. L’Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, ‘‘Machine learning with big data: Challenges and approaches,’’ IEEE Access, vol. 5, pp. 7776–7797, 2017.

[16] V. Sharma, S. Yadav, and M. Gupta, ‘‘Heart disease prediction using machine learning techniques,’’ in Proc. 2nd Int. Conf. Adv. Comput., Commun. Control Netw. (ICACCCN), Dec. 2020, pp. 177–181.

[17] A. M. Usman, U. K. Yusof, and S. Naim, ‘‘Cuckoo inspired algorithms for feature selection in heart disease prediction,’’ Int. J. Adv. Intell. Inform., vol. 4, no. 2, pp. 95–106, Jul. 2018.

[18] A. U. Haq, J. Li, M. H. Memon, M. Hunain Memon, J. Khan, and S. M. Marium, ‘‘Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection,’’ in Proc. IEEE 5th Int. Conf. Converg. Technol. (I2CT), Mar. 2019, pp. 1–4.

[19] A. Javeed, S. S. Rizvi, S. Zhou, R. Riaz, S. U. Khan, and S. J. Kwon, ‘‘Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification,’’ Mobile Inf. Syst., vol. 2020, pp. 1–11, Aug. 2020.

[20] D. C. Yadav and S. Pal, ‘‘Prediction of heart disease using feature selection and random forest ensemble method,’’ Int. J. Pharmaceutical Res., vol. 12, no. 4, pp. 56–66, Oct. 2020.