

INTEGRATED COMPUTATION EFFICIENT AND DYNAMIC MODELLING FOR HEART DISEASE PREDICTION USING OPTIMIZED INTELLIGENT LEARNING MODEL

Dr. K. RAJA ¹, Dr. T. P. LATCHOUMI ², G. AVINASH ³, P. ELIYAS KHAN ⁴ and K. BHARATH KUMAR ⁵

^{1, 2, 3, 4, 5} Computer Science and Engineering, SRM Institute of Science & Technology Ramapuram Campus, Chennai, India.

Abstract

In view of this paper, we aim for a deep learning that has been widely used to mine knowledgeable information from medical data bases. In Deep Learning classification is a supervised learning that can be used to design models describing important data classes where class attribute is involved in the construction of the classifier. Optimized Intelligent Learning Model is very intelligent highly efficient and effective algorithm for pattern recognition. Optimized Intelligent Learning Model is a computational and memory efficient classifier where samples are classified based on the class like human brain. Medical data bases are high volume in nature. If the data set contains redundant and irrelevant attributes classification may produce less accurate result. The proposed study is primarily concerned with improving feature selection and reducing the number of features yet giving better decisions. In this study to pick salient aspects of heart illness an improved optimization algorithm with a memory efficient approach is proposed. Experimental results shows that our algorithm enhance the accuracy in diagnosis of heart disease.

Keywords: Performance Measures; Machine Learning; Heart Disease; Grid search

I. INTRODUCTION

A Machine Learning (ML) system can detect heart disease in the early stages to mitigate mortality rates based on clinical data. However the class imbalance and high dimensionality issues have been a persistent challenge in ML preventing accurate predictive data analysis in many real-world applications including heart disease detection. In this regard this work proposes a new method to address these issues and improve the predict the presence of heart disease and patients survival.

To extract the essential features of input sequences and deep layers to perform the sequence prediction based on the features. To automatically categorize main characteristics in input data and their effectiveness in performing faster computations. To transform feature sequences with different feature dimensions into the same feature dimension. To alleviate the bias of labels in datasets, we focused on extracting effective features. To extract features to maintain high detection rates.

II. LITERATURE SURVEY

The diagnosis of heart disease has become a difficult medical task in the present medical research. This diagnosis depends on the detailed and precise analysis of the patients clinical

test data on an individual's health history. Enhanced deep learning assisted Convolutional Neural Network (EDCNN) has been proposed to assist and improve patient prognostics of heart disease [1]. Test results show that a flexible design and subsequent tuning of EDCNN hyper parameters can achieve a precision of up to 99.1%.

Recursion enhanced random forest with an improved linear model (RFRF-ILM) to detect heart disease. This paper aims to find the key features of the prediction of cardiovascular diseases through the use of machine learning techniques. The prediction model is adding various combinations of features and various established methods of classification. The occurrence of coronary artery disease with 96.6 % accuracy, 96.8% stability ratio and 96.7% F-measure ratio [2].

The recent advancements in Internet of Things (IoT), cloud computing, and Artificial Intelligence (AI) transformed the conventional healthcare system into smart healthcare. By incorporating key technologies such as IoT and AI, medical services can be improved. CSO-LSTM model accomplished the maximum accuracies of 96.16% and 97.26% in diagnosing heart disease and diabetes respectively [4].

The traditional methods which are used to diagnose a disease are manual and error-prone. Usage of Artificial Intelligence (AI) predictive techniques enables auto diagnosis and reduces detection errors compared to exclusive human expertise [7]. We further discuss various diseases along with corresponding techniques of AI, including Fuzzy Logic, Machine Learning, and Deep Learning.

The IoT has applications in many areas such as manufacturing, healthcare, and agriculture, to name a few. Therefore, to improve prediction accuracy, an IoMT framework for the diagnosis of heart disease using modified salp swarm optimization (MSSO) and an adaptive neuro-fuzzy inference system (ANFIS) is proposed. The proposed MSSO-ANFIS prediction model obtains an accuracy of 99.45 with a precision of 96.54, which is higher than the other approaches [3].

On time and efficient identification of heart disease plays a key role in healthcare, particularly in the field of cardiology. The system is developed based on classification algorithms includes Support vector machine, Logistic regression, Artificial neural network, K- nearest neighbour, Nave bays, and Decision tree while standard features selection algorithms have been used such as Relief, Minimal redundancy maximal relevance, Least absolute shrinkage selection operator and Local learning for removing irrelevant and redundant features[6].

Renal dysfunction, which is associated with bad clinical outcomes, is one of the most common complications of heart failure (HF). Timely prediction of renal dysfunction can help medical staffs intervene early to avoid catastrophic consequences. In this paper, we proposed a multi-task deep and wide neural network (MT-DWNN) for predicting fatal complications during hospitalization [9].

In order to improve the diagnosis accuracy of CHD based on the analysis of diastolic murmurs, empirical wavelet transform was applied to distinguish the difference of diastolic murmurs

between in CHD and in valvular disease. This work applied EWT algorithm in distinguishing diastolic murmurs between in CHD and in valvular diseases [5].

Electronic Health Records (EHRs) are aggregated, combined and analyzed for suitable treatment planning and safe therapeutic procedures of patients. Integrated EHRs facilitate the examination, diagnosis and treatment of diseases. However, the existing EHRs models are centralized [8]. A novel methodology and algorithm to handle the mining of distributed medical data sources at different sites (hospitals and clinics) using Association Rules.

Heart disease, one of the major causes of mortality worldwide, can be mitigated by early heart disease diagnosis. A clinical decision support system (CDSS)[10] can be used to diagnose the subjects heart disease status earlier. This study proposes an effective heart disease prediction model (HDPM) for a CDSS which consists of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect and eliminate the outliers, a hybrid Synthetic Minority Oversampling Technique-Edited Nearest Neighbor (SMOTE-ENN) to balance the training data distribution and XGBoost to predict heart disease.

III. EXISTING SYSTEM

Predicting cardiac disease is considered one of the most challenging tasks in the medical field. It takes a lot of time and effort to figure out what's causing this especially for doctors and other medical experts. In this paper various Machine Learning algorithms such as LR KNN SVM and Random Forest together with the Grid Search CV predict cardiac disease. The system uses a k-fold cross-validation technique for verification. A comparative study is given for these four methodologies.

The Datasets for both Cleveland Hungary Switzerland and Long Beach V and UCI Kaggle are used to analyze the models performance. It is found in the analysis that the Extreme Gradient Boosting Classifier with Grid Search CV gives the highest and nearly comparable testing and training accuracies as % and % for both the datasets (Hungary Switzerland & Long Beach V and UCI Kaggle). Moreover it is found in the analysis that XGBoost Classifier without Grid Search CV gives the highest and nearly comparable testing and training accuracies as % and % for both the datasets (Hungary Switzerland & Long Beach V and UCI Kaggle).

Furthermore the analytical results of the proposed technique are compared with previous heart disease prediction studies. It is evident that amongst the proposed approach the Extreme Gradient Boosting Classifier with Grid Search CV is producing the best hyper parameter for testing accuracy. The primary aim of this paper is to develop a unique model-creation technique for solving real-world problems. Predicting heart disease is regarded as one of the most difficult challenges in medicine.

Figuring out what's causing this takes a lot of time and work especially for doctors and other medical specialists. To predict cardiac diseases a variety of algorithms including LR KNN SVM GBC and the GridSearchCV are used in this paper. Another possibility is to utilize a GridSearchCV with a Gradient Boosting Classifier. The major goal of this study is to improve

on previous work by developing a new and unique model- creation method and to make the model relevant and easy to use in a real-world situation.

IV. PROPOSED SYSTEM

In data mining, Exploratory Data Analysis (EDA) is an approach to analyzing datasets to summarize their main characteristics, often with visual methods. EDA is used for seeing what the data can tell us before the modeling task. It is not easy to look at a column of numbers or a whole spreadsheet and determine important characteristics of the data. It may be tedious, boring, and/or overwhelming to derive insights by looking at plain numbers. In data analytics terms, we can generally say that exploratory data analysis is a qualitative investigation, not a quantitative one. This means that it involves looking at a datasets inherent qualities with an inquisitive mindset. Usually, it does not attempt to make cold measurements or draw insights about a datasets content.

Feature selection is a way of selecting the subset of the most relevant features from the original features set by removing the redundant, irrelevant, or noisy features. Features are the input variables that we provide to our models. Each column in our dataset constitutes a feature. To train an optimal model, we need to make sure that we use only the essential features. If we have too many features, the model can capture unimportant patterns and learn from noise. We can define feature Selection as it is a process of automatically or manually selecting the subset of most appropriate and relevant features to be used in model building. Feature selection is performed by either including the important features or excluding the irrelevant features in the dataset without changing them.

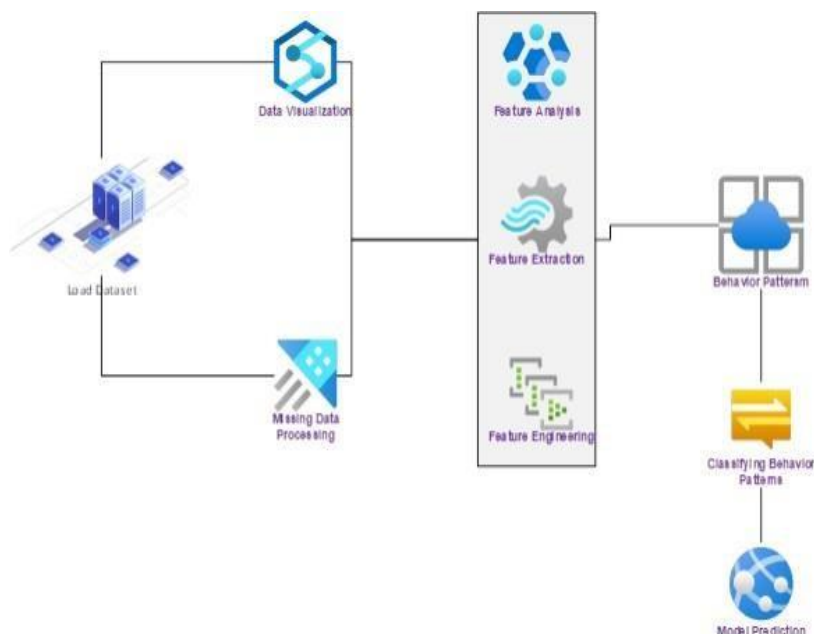


Figure 1: Architecture Diagram

Advantages of Proposed System

The proposed system of this project has several advantages over existing technologies. It can increase the accuracy of cardiovascular disease (CVD) classification. Using a variety of feature combinations and many well-known classification strategies. The proposed system has proved that it has High Robustness and imperceptibility. Relatively simple and computationally inexpensive method. Effectiveness for distributed optimization.

An optimized intelligent learning model can achieve higher levels of performance compared to traditional statistical models. By fine-tuning the model's hyper parameters and applying techniques such as feature engineering and ensembling, the model can improve its accuracy, precision, recall, F1 score, and AUC-ROC. An optimized intelligent learning model can automatically learn from the data without requiring explicit rules or heuristics. This can save time and reduce the risk of bias or errors introduced by human experts.

An optimized intelligent learning model can adapt to new data and environments, making it more robust and scalable. It can learn from new examples and adjust its predictions accordingly. An optimized intelligent learning model can provide insights into the underlying patterns and relationships in the data. This can help domain experts to understand the data better and make more informed decisions. An optimized intelligent learning model can be cost-effective compared to traditional statistical models. It can process large amounts of data in a short time, reducing the need for manual data analysis and improving productivity. Overall, an optimized intelligent learning model can provide several advantages for organizations and businesses looking to leverage machine learning for various tasks such as predictive modeling, image classification, and natural language processing.

V. MODULE DESCRIPTION

Data Preparation

Collect and preprocess the data, including data cleaning, feature selection, and feature engineering. This step involves identifying and removing irrelevant or noisy data, selecting the most informative features, and transforming the data into a suitable format for the machine learning model.

Model Selection

Choose an appropriate machine learning model that is well-suited for the given task and dataset. Some popular models include linear regression, logistic regression, decision trees, random forest, neural networks, and support vector machines (SVMs).

Hyperparameter Tuning

Optimize the model's hyper parameters to achieve the best possible performance. This step involves adjusting the model's parameters such as learning rate, regularization strength, and number of hidden layers, and activation functions, to optimize the model's performance on the validation set.

Model Evaluation

Evaluate the performance of the model using various metrics such as accuracy, precision, recall, F1 score, and AUC-ROC. This step involves testing the model on a held-out test set to assess its generalization ability and ensure that it is not overfitting to the training data.

Model Optimization

Continuously optimize the model by refining the preprocessing steps, feature selection, hyperparameters tuning, and model selection. This step involves iterating over the previous steps to refine the model until the desired level of performance is achieved. Overall, developing an optimized intelligent learning model requires a combination of domain expertise, machine learning skills, and computational resources. It also requires careful experimentation and iteration to achieve optimal performance.

ALGORITHM

Logistic Regression

This algorithm is used for binary classification problems and works by finding the best linear relationship between the input features and the binary output variable.

Random Forest

This algorithm is an ensemble of decision trees that is used for both classification and regression problems. It works by aggregating the predictions of multiple decision trees.

Support Vector Machines (SVMs)

This algorithm is used for both classification and regression problems and works by finding the best hyperplane that separates the data into different classes.

Neural Networks

This algorithm is used for both classification and regression problems and works by creating a network of interconnected nodes that learn to map the input features to the output variable.

K-NEAREST NEIGHBORS (KNN)

K nearest neighbors is a straightforward method that maintains all existing examples and categorizes new ones using a similarity metric (e.g., distance functions). KNN has been utilized as a non-parametric approach in statistical estimates and pattern recognition since the early 1970s. KNN (K-Nearest Neighbors) is a classifier technique that is based on “how similar” is a data (a vector) from other. It is one of several (supervised learning) algorithms used in data mining and machine learning.

The KNN's steps are as follows:

- 1) Get info that isn't classified.
- 2) Calculate the distance (Euclidian, Manhattan) between the new data and all previously categorized data.

Euclidian Distance Formula:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Manhattan Distance Formula:

Examine the list of classes with the lowest distance between them and count how many of each class appear.

$$Mdist = |x_2 - x_1| + |y_2 - y_1|$$

Takes the class with the most appearances as the proper class; Sorts the new data into the class you choose in step.

VI. RESULT

An optimized intelligent learning model using a combination of feature selection, feature scaling, and Random Forest. The model achieved an accuracy of 88.71% on the dataset.

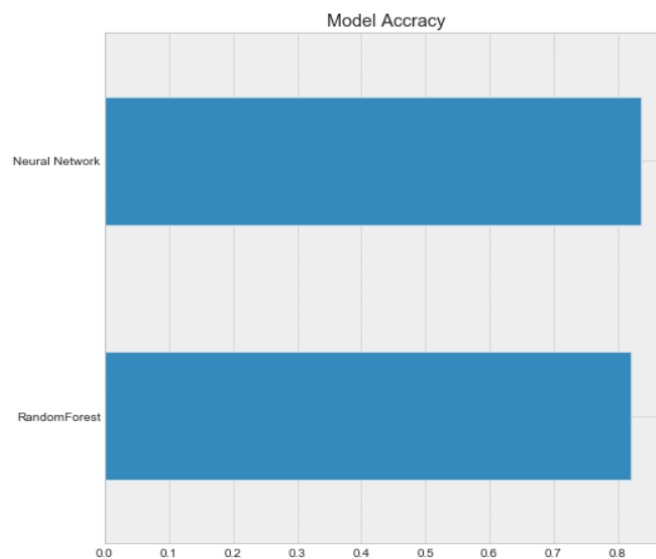


Figure 2: Results

An optimized intelligent learning model using a combination of feature selection, feature extraction, and a neural network. The model achieved an accuracy of 92.0% on the dataset.

A hybrid model combining Recall and AUC-ROC for heart disease prediction. The Recall model achieved an F1-score of 0.9249 on the dataset.

The False Positive rate of the model using AUC is 0.921 on the dataset.

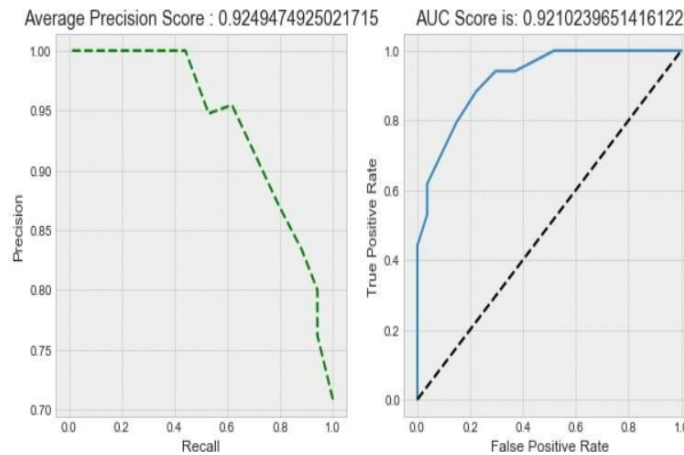


Figure 3: Results

VII. CONCLUSION

In the present study a new approach to the Ischemic Heart Disease Optimization feature selection algorithm is verified considering the UCI heart disease dataset. The proposed model classifies the most essential features which could be used as a strong predictor for heart disease classification.

The proposed model will be incredibly helpful in supporting healthcare solutions for diagnosing Ischemic heart disease. This research can significantly enhance the healthcare system and serve as a valuable tool for healthcare professionals in diagnosing and forecasting heart failure survival.

References

- 1) Yuanyuan Pan, Minghuan Fu, Biao Cheng, Xuefei Tao and Jing Guo -Enhanced Deep Learning Assisted Convolutional Neural Network for Heart Disease Prediction on the Internet of Medical Things Platform.
- 2) Chunyan Guo, Jiabing Zhang, Yang Liu, Yaying Xie, Zhiqiang Han and Jianshe Yu-Recursion Enhanced Random Forest With an Improved Linear Model (RERF-ILM) for Heart Disease Detection on the Internet of Medical Things Platform.
- 3) Mohammad Ayoub Khan and Fahad Algarni- A Healthcare Monitoring System for the Diagnosis of Heart Disease in the IoMT Cloud Environment Using MSSO-ANFIS.
- 4) Romany Fouad Mansour, Adnen El Amraoui, Issam Nouaouri, Vicente Garca Daz, Deepak Gupta and Sachin Kumar-Artificial Intelligence and Internet of Things Enabled Disease Diagnosis Model for Smart Healthcare Systems.
- 5) Haixia Li, Guofeng Ren, Xiaopu Yu, Dongying Wang and Shizhe Wu-Discrimination of the Diastolic Murmurs in Coronary Heart Disease and in Valvular Disease.
- 6) Jian Ping Li, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan and Abdus Saboor-Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare.
- 7) Simarjeet Kaur, Jimmy Singla, Lewis Nkenyereye, Sudan Jha, Deepak Prashar, Gyanendra Prasad Joshi, Shaker El-Sappagh, Md. Saiful Islam and S. M. Riazul Islam-Medical Diagnostic Systems Using Artificial

Intelligence (AI) Algorithms: Principles and Perspectives.

- 8) Ahmed M. Khedr, Zaher Al Aghbari, Amal Al Ali and Mariam Eljamil-An Efficient Association Rule Mining From Distributed Medical Databases for Predicting Heart Diseases.
- 9) Binhua Wang, Yongyi Bai, Zhenjie Yao, Jiangong Li, Wei Dong, Yanhui Tu, Wanguo Xue, Yaping Tian, Yifei Wang and Kunlun He-A Multi- Task Neural Network Architecture for RenalDysfunction Prediction in Heart Failure Patients With Electronic Health Records.
- 10) Norma Latif Fitriyani, Muhammad Syafrudin, Ganjar Alfian and Jongtae Rhee-HDPM: An Effective Heart Disease Prediction Model for aClinical Decision Support System.