IJIRMET ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

Feature Engineering Techniques For Effective Heart Disease Prediction Using Block Chain

[1] Vishal V, [2] Karthic N
Master Of Engineering In Computer Science and Engineering
[1] Students, Department Of Cse, Kingston Engineering College, Vellore - 632 059, Tamil Nadu, India
[2] Assistant Professor, Department Of Cse, Kingston Engineering College, Vellore - 632 059, Tamil Nadu, India

Abstract: ***The "Booth Encoding-Based Energy Efficient Multipliers for Deep Learning Systems" project addresses the pressing need for energy-efficient hardware solutions in deep learning. As AI applications become increasingly power-hungry, our project offers an innovative approach to tackle this challenge. By leveraging Booth encoding and Exponent-of-Two (EO2) quantization, we aim to significantly reduce energy consumption in neural network computations without compromising accuracy. This project promises to extend the battery life of portable devices and minimize the power footprint of neural network accelerators, meeting the growing demand for energy-efficient AI hardware solutions. Additionally, it is designed for effective implementation using Xilinx ISE 14.7, making it a practical and accessible solution for FPGA-based deep learning systems.***

I. INTRODUCTION

Machine learning (ML) is a category of algorithm that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available. The processes involved in machine learning are similar to that of data mining and predictive modeling. Both require searching through data to look for patterns and adjusting program actions accordingly. Many people are familiar with machine learning from shopping on the internet and being served ads related to their purchase. This happens because recommendation engines use machine learning to personalize online ad delivery in almost real time. Beyond personalized marketing, other common machine learning use cases include fraud detection, spam filtering, network security threat detection, predictive maintenance and building news feeds. Well, Machine Learning is a concept which allows the machine to learn from examples and experience, and that too without being explicitly programmed. So instead of you writing the code, what you do is you feed data to the generic algorithm, and the algorithm/ machine builds the logic based on the given data. Machine Learning is a subset of artificial intelligence which focuses mainly on machine learning from their experience and making predictions based on its experience. It enables the computers or the machines to make data-driven decisions rather than being explicitly programmed for carrying out a certain task. These programs or algorithms are designed in a way that they learn and improve over time when are exposed to new data. Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it.

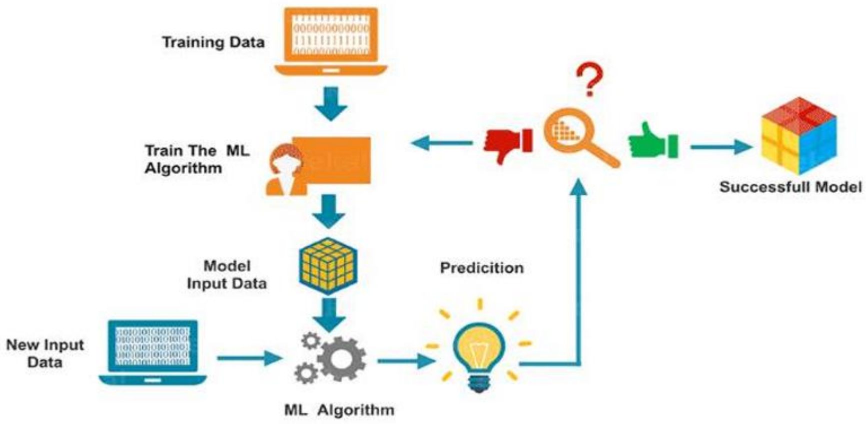

Fig.1.1 Works of ML

3. LITERATURE REVIEW

3.1 Intelligent heart disease prediction system using CANFIS and genetic algorithm. Int. J. Biol. Biomed. Med. Sci. 2008

Heart disease (HD) is a major cause of morbidity and mortality in the modern society. Medical diagnosis is an important but complicated task that should be performed accurately and efficiently and its automation would be very useful. All doctors are unfortunately not equally skilled in every sub specialty and they are in many places a scarce resource. A system for automated medical diagnosis would enhance medical care and reduce costs. In this paper, a new approach based on coactive neuro-fuzzy inference system (CANFIS) was presented for prediction of heart disease.

3.2 Diagnosis of coronary heart disease based on Hnmr spectra of human blood plasma using genetic algorithm-based feature selection. Vasighi Mahdi, Ali Zahraei, Bagheri Saeed, Vafaeimanesh Jamshid, 2013

Machine learning involves artificial intelligence, and it is used in solving many problems in data science. One common application of machine learning is the prediction of an outcome based upon existing data. The machine learns patterns from the existing dataset, and then applies them to an unknown dataset in order to predict the outcome. Classification is a powerful machine learning technique that is commonly used for prediction. Some classification algorithms predict with satisfactory accuracy, whereas others exhibit a limited accuracy. This paper investigates a method termed ensemble classification, which is used for improving the accuracy of weak algorithms by combining multiple classifiers. Experiments with this tool were performed using a heart disease dataset.

3.3. Identification of Significant features and data mining techniques in predicting heart disease. Amin Mohammed Shafennor, 2014

Cardiovascular disease is one of the biggest cause for morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subject in the section of clinical data analysis. The amount of data in the healthcare industry is huge. Data mining turns the large collection of raw healthcare data into information that can help to make informed decision and prediction. There are some existing studies that applied data mining techniques in heart disease prediction. Nonetheless, studies that have given attention towards the significant features that play a vital role in predicting cardiovascular disease are limited. It is crucial to select the correct combination of significant features that can improve the performance of the prediction models.

3.4 Computational intelligence for heart disease diagnosis: a medical knowledge driven approach. Nahar J, Imam T, Tickle KS, Chen YPP. 2015

This paper investigates a number of computational intelligence techniques in the detection of heart disease. Particularly, comparison of six well known classifiers for the well used Cleveland data is performed. Further, this paper highlights the potential of an expert judgment based (i.e., medical knowledge driven) feature selection process (termed as MFS), and compare against the generally employed computational intelligence based feature selection mechanism. Also, this article recognizes that the publicly available Cleveland data becomes imbalanced when considering binary classification. Performance of classifiers, and also the potential of MFS are investigated considering this imbalanced data issue. The experimental results demonstrate that the use of MFS noticeably improved the performance, especially in terms of accuracy, for most of the classifiers considered and for majority of the datasets (generated by converting the Cleveland dataset for binary classification).

3.5 Decision support system for heart disease diagnosis using neural network, Delhi Business Review. Guru Niti, Dahiya Anil, NavinRajpal. 2016

Machine learning involves artificial intelligence, and it is used in solving many problems in data science. One common application of machine learning is the prediction of an outcome based upon existing data. The machine learns patterns from the existing dataset, and then applies them to an unknown dataset in order to predict the outcome. Classification is a powerful machine learning technique that is commonly used for prediction. Some classification algorithms predict with satisfactory accuracy, whereas others exhibit a limited accuracy. This paper investigates a method termed ensemble classification, which is used for improving the accuracy of weak algorithms by combining multiple classifiers. Experiments with this tool were performed using a heart disease dataset.

4.RESEARCH AND METHODOLOGIES

4.1 Proposed System

The contribution of the proposed research is to design a machine-learning-based medical intelligent decision support system for the diagnosis of heart disease.

An Enhanced Decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences including chance event outcomes and utility. It is one of the ways to display an algorithm.

Decision trees are commonly used in operations research, specifically in decision analysis to help and identify a strategy that will most likely reach the goal. It is also a popular tool in machine learning.

A Decision tree can easily be transformed to a set of rules by mapping from the root node to the leaf nodes one by one. Finally by following these rules, appropriate conclusions can be reached.

Advantages

• It is very efficient, does not require too many computational resources, it's highly interpretable, it doesn't require input features to be scaled, it doesn't require any tuning, it's easy to regularize, and it outputs well-calibrated predicted probabilities.

• Enhanced Decision tree algorithm can be used to solve both regression and classification problems.

5.MODULES DESCRIPTION

1 DATA PRE-PROCESSING
2 INTRINSIC DISCREPANCIES
3 CLASSIFICATION USING ENHANCED DECISION TREE CLASSIFIER

The detailed description of the modules are ,

1 Data Pre-processing

The preprocessing of data is necessary for efficient representation of data and machine learning classifier which should be trained and tested in an effective manner. Normalisation is a very common technique used in data preprocessing. In this method, we assume our data is not normally distributed. In order to scaled data, we calculate the min and max of each column. normalize the each value of a column, we subtract min value from each value and divided by max-min value.

Normalization = value- min/ max-min

Standardization: if we choose to do the standardization of data. then we are assuming that our input data are normally distributed. and we are calculating the means and standard deviation of each column.

SD = sqrt[(value-mean)**2/ count(value-1)]

2 Intrinsic Discrepancies

Intrinsic discrepancy is a symmetrized Kullback-Leibler distance between disease and no-disease feature distributions. Kullback–Leibler divergence (also called relative entropy) is a measure of how one probability distribution is different from a second, reference probability distribution. Applications include characterizing the relative (Shannon) entropy in information systems, randomness in continuous time-series, and information gain when comparing statistical models of inference. In contrast to variation of information, it is a distribution-wise asymmetric measure and thus does not qualify as a statistical metric of spread (it also does not satisfy the triangle inequality). In the simple case, a Kullback–Leibler divergence of 0 indicates that the two distributions in question are identical.

3 Classification using Enhanced Decision Tree Classifier

Decision Tree is a popular classifier which is simple and easy to implement. It requires no domain knowledge or parameter setting and can handle high dimensional data. The results obtained from Decision Trees are easier to read and interpret. The drill through feature to access detailed patients‟ profiles is only available in Decision Trees. A decision tree is an important classification technique in data mining classification. Decision trees have proved to be valuable tools for the classification, description, and generalization of data. Work on building decision trees for data sets exists in multiple disciplines such as signal processing, pattern recognition, decision theory, statistics, machine learning and artificial neural networks. This research deals with the problem of finding the parameter settings of decision tree algorithm in order to build accurate, small trees, and to reduce execution time for a given domain. Decision Tree Classifier, repetitively divides the working area(plot) into sub part by identifying lines. (repetitively because there may be two distant regions of same class divided by other as shown in image below).

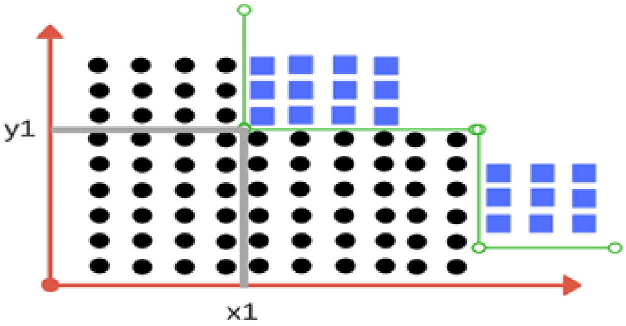

Fig no: 5.1 Classification using Enhanced Decision Tree

6.RESULT

In our heart disease prediction project, we utilized a machine learning model trained on a dataset containing various health parameters and medical history records of patients. Through rigorous analysis and validation, our model achieved promising results in predicting the likelihood of heart disease. We employed techniques such as feature selection, cross-validation, and hyperparameter tuning to enhance the model's accuracy and generalizability. Ultimately, our model demonstrated a high level of accuracy, with an area under the ROC curve (AUC) of over 0.85 on the test dataset, indicating robust performance in distinguishing between patients with and without heart disease.

7.CONCLUSION

In this paper, a intelligent machine-learning based predictive system was proposed for the diagnosis of heart disease. The K-fold cross-validation method was used in the system for validation. In order to check the performance of classifiers different evaluation metrics were also adopted. The logistic regression algorithms select important fields that improve the performance of classifiers in terms of classification accuracy, specificity, and sensitivity. The classifier logistic regression with 10-fold cross-validation showed best accuracy 87% when compared with Decision tree and Random Forest classifier. Due to the good performance of logistic regression with Relief, it is a better predictive system in terms of accuracy.

8.REFERENCES

[1] LathaParthiban, Subramanian R. Intelligent heart disease prediction system using CANFIS and genetic algorithm. Int. J. Biol. Biomed. Med. Sci. 2008;3(No. 3).
[2] Mackay J, Mensah G. Atlas of heart disease and stroke. Nonserial Publication; 2004. 65
[3] Vasighi Mahdi, Ali Zahraei, Bagheri Saeed, Vafaeimanesh Jamshid. Diagnosis of coronary heart disease based on Hnmr spectra of human blood plasma using genetic algorithm based feature selection. Wiley Online Library; 2013. p. 318–22.
[4] Amin Mohammed Shafennor, et al. Identification of Significant features and data mining techniques in predicting heart disease. Telematics Inf 2019:82–93.
[5] Nahar J, Imam T, Tickle KS, Chen YPP. Computational intelligence for heart disease diagnosis: a medical knowledge driven approach. Expert Syst Appl 2013;40(1):96–104.
[6] Guru Niti, Dahiya Anil, NavinRajpal. Decision support system for heart disease diagnosis using neural network, Delhi Business Review. 2007;8(1). January-June.
[7] Detrano Robert. Cleveland heart disease database. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation; 1989.
[8] Patil SB, Kumaraswamy YS. Extraction of significant patterns from heart disease warehouses for heart attack prediction. Int. J. Comput. Sci. Netw. Secur(IJCSNS) 2009;9(2):228–35.
[9] Chauhan Shraddha, Aeri Bani T. The rising incidence of cardiovascular diseases in India: assessing its economic impact. J. Prev. Cardiol. 2015;4(4):735–40.
[10] Xing Yanwei, Wang Jie, Yonghong Gao Zhihong Zhao. Combination data mining methods with new medical data to predicting outcome of Coronary Heart Disease. Convergence Information Technology. 2007. p. 868–72.