IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁙ Impact Factor 8.102 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.12563

Driver Driving Performance Analysis And Risk Detection Using Deep Learning

Mrs.G.Geetha[1], J.Navin[2], P.Sanjeevi[3], M.Surya Sivaraj[4]

[1]Assistant Professor, Department of Computer Science and Engineering, DMI College of Engineering, Chennai, India

[2]B.E, Department of Computer Science and Engineering, DMI College of Engineering, Chennai, India

[3]B.E, Department of Computer Science and Engineering, DMI College of Engineering, Chennai, India

[4]B.E, Department of Computer Science and Engineering, DMI College of Engineering, Chennai, India

Abstract: Distracted driving is any activity that deviates an individual's attention from driving. Driver inattention and distraction are the main causes of road accidents, many of which result in fatalities. Driver distraction is a major cause of road accidents. Distracting activities while driving include text messaging and talking on the phone. Currently, distraction detection systems for road vehicles are not yet widely available or are limited to specific causes of driver inattention such as driver fatigue. Research efforts have been made to monitor drivers' attention states and provide support to drivers. Both invasive and non-invasive methods have been applied to track driver's attention states, but most of these methods either use exclusive equipment which are costly or use sensors that cause discomfort. The existing work of distracted driver detection is concerned with a limited set of distractions(Mainly cell phone usage).In this paper, a robust driver distraction detection system that extracts the driver's state from the recordings of an onboard camera using Deep Learning based Faster Region Convolution Neural Network (FRCNN).This project uses the state farm distracted driver detection, which contains four classes: calling, texting, looking behind, and normal driving The main feature of the proposed solution is the extraction of the driver's body parts, using deep learning-based segmentation, before performing the distraction detection and classification task. Experimental results show that the segmentation module significantly improves the classification performance. The average accuracy of the proposed solution exceeds 96% on our data set. The class activation map (CAM) of our proposed method is subjectively more reasonable, which would enhance the reliability and explain ability of the model.

Keywords: Alert Message, DD, FRCNN, Face Detection,

I. INTRODUCTION

A. Overview

Distracted driving is any activity that diverts attention from driving, including talking or texting on your phone, eating and drinking, talking to people in your vehicle, fiddling with the stereo, entertainment or navigation system — anything that takes your attention away from the task of safe driving. Distracted driving occurs when a motorist is not giving their complete attention to the road, other vehicles, and signs along the roadway. According to the National Highway Traffic Safety Administration, three types of distracted driving exist: Manual distraction: when a driver adjusts the radio, reaches for an item, pets their dog. etc. by removing their hands from the wheel. Visual distraction: when a driver rubber necks by an accident, checks text message(s), looks at their kids in the back seat, etc. by taking their eyes off the road. Cognitive distraction: when a driver daydreams, thinks about personal issues, considers their grocery list. etc. by taking their mind off of driving.

B. Distracted Driver Behaviors

Distracted driving comes in a variety of forms. These behaviors include instances when a driver is looking down or away from the road ahead for a period of time long enough to lose situational awareness of the forward driving scene like Day dreaming, using a cell phone, looking at something outside the vehicle, Activities of passengers, reaching for something on the dashboard, seat, or floor, Eating, drinking, or smoking, Changing the radio, climate control, or using a device in the car, Pets, insects, and objects moving inside the vehicle,Drowsy driving, talking on a cell phone, Texting, Smoking, using a tablet, reading paperwork, Programming an in-vehicle infotainment system.

C. Common Causes of Distraction

Mobile phones: A substantial body of research shows that using a hand-held or hands-free mobile phone while driving is a significant distraction, and substantially increases the risk of the driver crashing. Mobile Phones and Driving

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.12563

Factsheet outlines the law relating to the use of mobile phones in a vehicle, concerns around the use of both hands-held and hands-free mobile phones while driving and issues around mobile phones that employers are advised to consider.

Headphones: Although it has been noted that headphones can cause hearing damage, much less attention has been paid to the effects of headphones on the quality of driving. It is expected that drivers who are wearing headphones would need to shift their attention from what they hear in their headphones to external sound sources in certain situations, which could delay the speed of their response to external events. This is thought to be dangerous enough to form a risk in emergency situations. Headphones as a Driving Distraction Factsheet provides an overview of the evidence relating to headphones as a driving distraction.

Satellite navigation (sat nav) devices: Several different types of Sat Navs are available to drivers, many of which are built into the vehicle itself. Used well, a Sat Nav can help drivers plan routes and prevent them from making last minute lane changes or hesitating because they are not sure of the directions. However, a badly used Sat Nav can distract the driver and increase the risk of an accident. It is important that drivers understand how best to use them sat nav and learn not to use it when it may be dangerous to do so. Satellite Navigation Factsheet provides advice on making the best use of a sat nav and choosing the best device.

Infotainment Systems: Over the last ten years, there has been a huge increase in the digital technology available to motorists, allowing them to perform tasks that are unrelated to driving while they are behind the wheel. One of the biggest developments in this period has been the rise of infotainment systems. This refers to vehicle systems that combine entertainment and information delivery for drivers and passengers, often with the use of audio and touchscreens. Whilst infotainment systems can be handy for the driver, allowing them to carry out many tasks, there is also the potential that they can be a distraction – more so than traditional radio systems.

D. Problems Identified

A driver may be engaging in one or more distracted driving behaviours at any given time. The result is the same: heightened risk of a collision, injury, or even fatality. There’s little argument that distracted driving is one of the fastest growing threats to safe driving today and cause of collisions in the India Nearly 95% of serious traffic collisions are due to human error, with over 70% of commercial fleet collisions involving distracted drivers. The National Safety Council reports on a typical day, more than 700 people are injured in distracted driving crashes.

These statistics are true indicators of how safely drivers are using the roadways and why keeping commercial fleet drivers and bystanders safe on the road is becoming a bigger and costlier challenge for many fleet managers and safety leaders. Distracted driving detection systems can be used to prompt early warnings to alarm drivers of hazardous driving conduct, including using a cell phone to call or text, using navigation applications, or selecting radio frequencies or music. Distracted driving identification techniques are predominantly found in the driver’s facial expression, head activity, line of sight, or body activity.

The driver’s driving conduct and physiological state can be recognized through the visual following, target identification, movement acknowledgment, and different advancements. Detecting distracted driving has gained much attention from the research community, government agencies, and industry. The driver’s activity that diverts his attention during driving is a distraction like interaction with other passengers, making phone calls, and adjusting the multi-media and navigation tools. Most computer vision algorithms detect the driver’s behaviour using traditional computer vision and machine learning algorithms. Deep learning techniques have significantly improved the accuracy of vision-related tasks. Recent technological progress makes it possible for real-time algorithms to detect the distraction activity and assist and alert the distracted driver.

E. Deep Learning

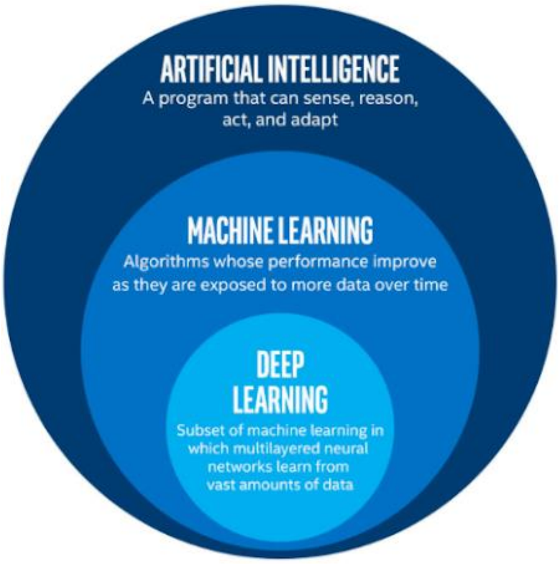
Artificial intelligence is a set of algorithms and intelligence to try to mimic human intelligence. Machine learning is one of them, and deep learning is one of those machine learning techniques. Deep Learning is a subset of Machine Learning that uses mathematical functions to map the input to the output. These functions can extract non-redundant information or patterns from the data, which enables them to form a relationship between the input and the output. This is known as learning, and the process of learning is called training.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.12563



Deep learning can be used on a variety of input data types including audio, video, text, images, radio waves and machine signals to create applications such as natural language processing, audio recognition, computer vision and target recognition. At scale, these applications can comb through massive amounts of data that would be impossible for a team of humans to process.

II. LITERATURE SURVEY

Suresh Kumar S, Pavithra S, Jemima Preethi M., et al. (2020), Cardiovascular disease is a leading cause of death worldwide, and early prediction and prevention of the disease can reduce mortality rates. Machine learning algorithms can be used for prediction, but the accuracy of different algorithms needs to be compared to identify the best-performing algorithm. The study used a dataset of 303 patients, with 14 features including age, sex, blood pressure, and cholesterol levels. Four machine learning algorithms - K-Nearest Neighbors, Decision Tree, Random Forest, and Support Vector Machine - were trained and tested using 10-fold cross-validation. The performance was evaluated using accuracy, precision, recall, and F1-score metrics. The Random Forest algorithm outperformed the other three algorithms, achieving an accuracy of 90.38%, precision of 91.26%, recall of 90.55%, and F1-score of 90.9%. This was been developed by Suresh Kumar in 2020.

Jyoti Yadav, Neha Bhargava, Gaurav Kumar, Alok Kumar Singh. et al. (2021), Cardiovascular disease is the leading cause of death globally, and early prediction can help in early intervention and better management. The authors conducted a systematic literature review of articles published between 2017 and 2020, using the PubMed, ScienceDirect, and IEEE Xplore databases. The selected articles were analyzed based on the machine learning algorithms used, dataset characteristics, performance metrics, and limitations The review found that machine learning techniques, including decision trees, random forests, support vector machines, and artificial neural networks, have been widely used for cardiovascular disease prediction. The study also identified several challenges, including data imbalance, overfitting, and lack of interpretability of models.

Durga Prasad Sharma, Dhirendra Pratap Singh, Alok Kumar Yadav, and Vaibhav Pandey. et al. (2020), Cardiovascular disease (CVD) is the leading cause of death globally, and its prediction is challenging due to the involvement of multiple risk factors. The study used a dataset of 12,201 patients from the National Health and Nutrition Examination Survey (NHANES) 2011-2012. The dataset was preprocessed and feature-selected, and then five machine learning algorithms, including logistic regression, decision tree, random forest, support vector machine, and K-nearest neighbor were trained and tested for predicting CVD risk. The study used only one dataset, which may limit the generalizability of the results. The study did not include some potential risk factors for CVD, such as physical activity and family history of CVD.

Ahmed Al-Mallah, Mouaz Al-Mallah, Yasar Albakri, and Fatima AI-Anazi. et al. (2019), Cardiovascular disease (CVD) is the leading cause of death worldwide. Early detection and prediction of CVD can improve patient outcomes and reduce healthcare costs. To develop a machine learning model for predicting CVD using clinical data. The authors

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁙ Impact Factor 8.102 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.12563

used a retrospective cohort study design and developed a machine learning model using three different algorithms: logistic regression, decision tree, and neural network. The authors found that the neural network algorithm outperformed the logistic regression and decision tree algorithms, with an accuracy of 87.3%. The most important features for predicting CVD were age, systolic blood pressure, and body mass index.

Muhammad Attique Khan, Muhammad Usman Ghani Khan, Naveed Iqbal Qureshi. et al. (2021), CVD is a major public health concern, and early identification and management of risk factors can improve patient outcomes. Traditional risk prediction models have limitations in accuracy and applicability to different populations. To develop a machine learning model for predicting CVD risk using big data analytics. The authors used a retrospective cohort study design and developed a machine-learning model using the Random Forest algorithm. They trained and tested the model using a dataset of 11,043 patients with 45 clinical features and demographic data. They also used feature selection and data pre-processing techniques to improve the model's performance. The authors found that the Random Forest model had an accuracy of 83.56% in predicting CVD risk. The most important features for predicting CVD risk were age, blood pressure, and cholesterol levels. The model was also able to identify subgroups of patients with a higher risk of CVD.

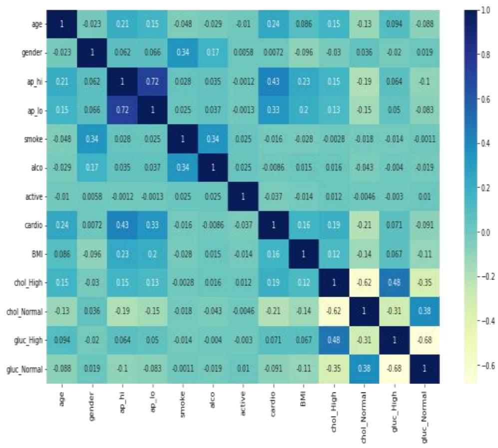
III. PROPOSED SYSTEM

The proposed methodology for web-based cardiovascular disease (CVD) prediction and its causes using Random Forest Algorithm. It involves collecting and pre-processing health and sensor data, training a Random Forest model for CVD prediction, developing an alert system, generating personalized recommendations, evaluating the system's performance, and deploying it on a web-based platform. Careful attention to data quality, model validation, and privacy and security concerns is essential for the success of the proposed system.

A.Dataset Introduction
In this project, we collected a heart disease dataset known as the Cleveland heart disease database from an online machine learning and data mining repository of the University of California, Irvine (UCI). it covers the role of various subsystems/modules/classes along with implementation details listing the code for the major functionalities.

B. Data set Preprocessing
Every dataset consists of various types of anomalies such as missing values, redundancy, or any other problem for removing this problem there is a need for a certain step called processing data. The pre-processing step is needed to overcome such a problem. There are three pre-processing steps:

Formatting**:** The data set used for implementation is taken from the UCI repository, it may contain certain attributes whose names are not clear in the (dataset name) and also contain certain unrelated attribute which is not useful for the greater performance of the proposed work. An attribute name "Thal" has been removed from the dataset by using the following command in R, Dataset that Null



Cleaning: This part of pre-processing belongs to removing or fixing missing out the entry in the data frame. A row containing these incomplete columned to be removed also for removing certain redundant entries in the data frame this step is recommended

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.12563

Sampling: Sampling is also done on the dataset to enhance the performance of the algorithm on the sample data set may lead the algorithm to take longer time.

C. Feature Selection

In feature selection, irrelevant features are eliminated and the most important or relevant features are applied to the network. Thus, if we supply all features to Random Forest, some features may be noisy and if they are learned in the training process, they may degrade the generalization of the network although the network will show good performance on the training data. That is why a large number of features are also considered one of the main causes of overfitting. Thus, searching out an optimal subset of features by eliminating noisy features can help Random Forest to show good performance on both training and testing data.In this module, we use the X^2 statistical model to eliminate irrelevant features. In the feature's elimination process, we compute X^2 statistics between each non-negative feature F_i and class i.e., y. The X^2 model performs an X^2 test that measures dependence between the features and class. Hence, the model is capable of eliminating those features which are more likely to be independent of class. Because these features can be regarded as irrelevant for classification.

D. Random Forest CVD Classification

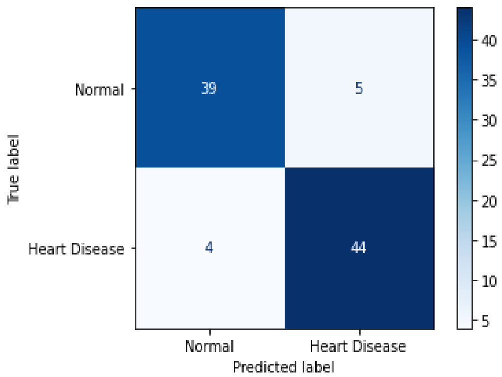
Random Forest is an ensemble learning algorithm that is widely used for classification and regression tasks. It is a combination of multiple decision trees, where each tree is built using a subset of the available data and a random selection of features. In the case of Cardiovascular Disease (CVD) prediction, the Random Forest algorithm can be used to build a classification model that predicts the likelihood of a person having CVD based on their age, gender, blood pressure, cholesterol levels, smoking habits, diabetes status, and family history of CVDs.

The Random Forest algorithm works as follows:

Random Sampling: A random sample of the available data is selected to build each decision tree in the forest. This ensures that each tree is built on a different subset of data, reducing over fitting and increasing the diversity of the model.
Random Feature Selection: At each node of the decision tree, a random subset of features is considered to split the data. This also helps to reduce overfitting and increase the diversity of the model.
Voting: Once all the decision trees are built, the prediction for a new data point is made by taking the majority vote of all the decision trees. This voting mechanism ensures that the model is robust to noise and outliers in the data.
The Random Forest algorithm has several advantages over other classification algorithms. It can handle large datasets with high-dimensional features and is less prone to overfitting. It is also easy to interpret the results and identify the most important features that contribute to the prediction.

E. Performance Analysis

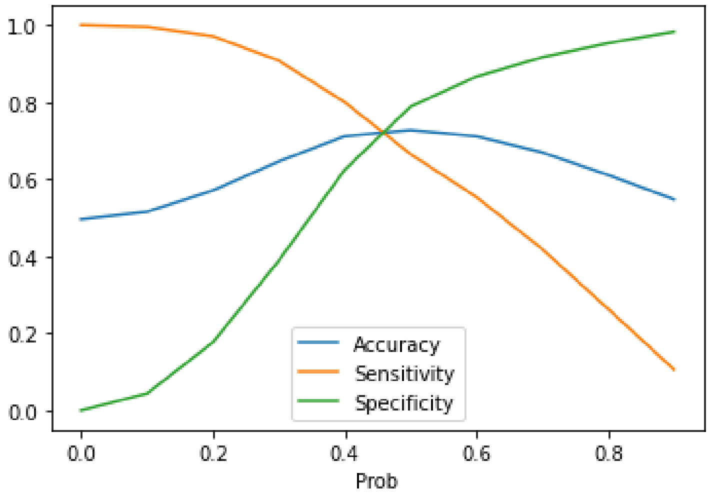
Evaluation Metrics To comprehensively evaluate the classification performance and effectiveness of our proposed method, we applied accuracy, recall, F1 score, precision, specificity, ROC, and AUC evaluation metrics. For the sake of expression of the significance and calculation formula of these evaluation metrics, we introduced the confusion matrix (See Table 5) first. The confusion matrix is a specific matrix used to visually present the performance of the algorithm. The confusion matrix of binary classification consists of two rows and two columns. Rows represent the true labels of the two classes in the dataset (denoted true). Columns represent the predicted label of the two classes acquired by the model (denoted as type). As shown in Table 5, the confusion matrix of binary classification includes four indicators: TN, FN, FP, and TP. The four indicators are defined as follows. We specified that the label of the positive class is 1 and the label of the negative class is 0.



International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.12563

F. Accuracy
Accuracy refers to the proportion of samples that can be correctly predicted by the model in all samples. The calculation equation of accuracy is as follows. TN, TP, FN, and FP refer to true negative, true positive, false negative, and false positive, respectively.

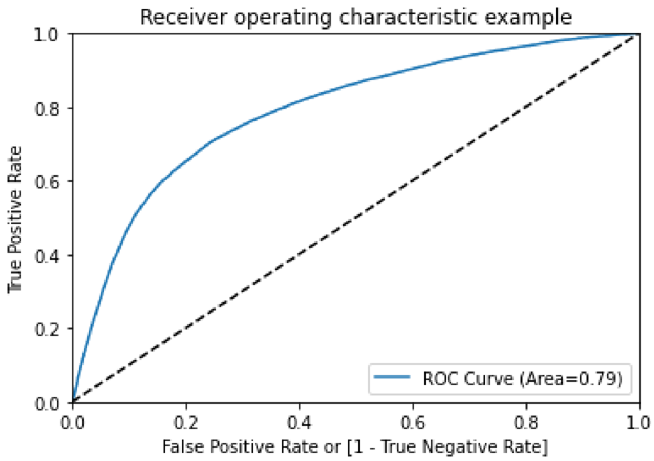
Accuracy = TN + TP TN + TP + FN + FP (1)



Accuracy is one of the most frequently used and most important model performance evaluation metrics. However, in the dataset with a class imbalance problem, due to the influence of majority class samples, the accuracy is often difficult to accurately measure the classification ability of the model. Therefore, in the dataset with a class imbalance problem, in addition to accuracy, more evaluation indicators need to be applied.

G. ROC and AUC
The area under the curve (AUC) is the area under the receiver operating characteristic (ROC) curve. The ROC curve is drawn with the false positive rate (FPR) as the x-axis and the true positive rate (TPR) as the y-axis. The ROC curve intuitively reflects the relationship between specificity and recall. The value of AUC is between 0 and 1, when the value of the x-axis (i.e., the false positive rate (FPR) of the model) is closer to 0, and the value of the y-axis (i.e., the true positive rate (TPR) of the model) is closer to 1, the value of AUC is closer to 1. The closer the AUC value is to 1, the higher the prediction performance of the classifier.

Classifiers	**Accuracy**	**Recall**	**F1**	**Precision**	**Specificity**	**AUC**
Random Forest	**0.97 ± 0.060**	**0.922 ± 0.068**	**0.980 ± 0.038**	**0.963 ± 0.041**	**0.881 ± 0.095**	**0.93 ± 0.05**



International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁘ Impact Factor 8.102 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.12563

IV. RESULT &DISCUSSION

The importance of the cardiovascular disease (CVD) prediction and alert system with sensor data using Random Forest lies in its potential to improve the early detection and management of CVDs, which are a leading cause of morbidity and mortality worldwide. CVDs are a major health concern globally, and their prevalence is expected to increase in the coming years due to factors such as aging populations and changes in lifestyle and dietary habits. Early detection and management of CVDs can significantly improve patient outcomes and reduce healthcare costs, but current diagnosis and management processes are often subjective and time-consuming. The proposed system addresses these challenges by leveraging sensor data and machine learning algorithms to predict the likelihood of CVDs and provide personalized recommendations and alerts to patients, guardians, doctors, and ambulance services in case of any emergency. By improving the accuracy and timeliness of CVD diagnosis and management, the system can improve patient outcomes and reduce healthcare costs. The best model is the random forest tree model. The accuracy is 96.7%, with an f1-Score, recall and precision of 97.5%

V. CONCLUSION

In this project, the machine learning-based support vector machine classification and prediction models were developed and evaluated based on the diagnostic performance of coronary heart disease in patients using sensitivity, specificity, precision, FScore, AUC, DOR, 95% confidence interval for DOR, and K-S test. The developed machine learning classification and prediction models were built with a multilayer perceptron equipped with linear and non-linear transfer functions, regularization and dropout, and a binary sigmoid classification using machine learning technologies to create a strong and enhanced classification and prediction model. The developed Random Forest-based classification and prediction models were trained and tested using the holdout method and 28 input attributes based on the clinical dataset from patients at the Cleveland Clinic. Based on the testing results, the developed machine learning models achieved diagnostic accuracy for heart disease of 83.67%, a probability of misclassification error of 16.33%, a sensitivity of 93.51%, a specificity of 72.86%, a precision of 79.12%, an F-score of 0.8571, AUC of 0.8922, the K-S test of 66.62%, DOR of 38.65, and 95% confidence interval for the DOR of this test of [38.65, 110.28]. These results exceed those of currently published research. Therefore, the developed machine learning classification and prediction models can provide highly reliable and accurate diagnoses for coronary heart disease and reduce the number of erroneous diagnoses that potentially harm patients. Thus, the models can be used to aid healthcare professionals and patients throughout the world to advance both public health and global health, especially in developing countries and resource-limited areas where there are fewer cardiac specialists available.

REFERENCES

[1] K. Polaraju, D. Durga Prasad, “Prediction of Heart Disease using Multiple Linear Regression Model”, International Journal of Engineering Development and Research Development, ISSN:2321-9939, 2017. [2] Marjia Sultana, Afrin Haider, “Heart Disease Prediction using WEKA tool and 10-Fold cross-validation”, The Institute of Electrical and Electronics Engineers, March 2017.
[3] Dr.S.Seema Shedole, Kumari Deepika, “Predictive analytics to prevent and control chronic disease”, January 2016.
[4] Ashok Kumar Dwivedi, “Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation”, Springer, 17 September 2016.
[5] Megha Shahi, R. Kaur Gurm, “Heart Disease Prediction System using Data Mining Techniques”, Orient J. Computer Science Technology, vol.6 2017, pp.457-466.
[6] Mr. Chala Beyene, Prof. Pooja Kamat, “Survey on Prediction and Analysis of the Occurrence of Heart Disease Using Data Mining Techniques”, International Journal of Pure and Applied Mathematics, 2018. [7] R. Sharmila, S. Chellammal, “A conceptual method to enhance the prediction of heart diseases using the data techniques”, International Journal of Computer Science and Engineering, May 2018.
[8] Jayami Patel, Prof. Tejal Upadhay, Dr. Samir Patel, “Heart disease Prediction using Machine Learning and Data mining Technique”, March 2017.
[9] Purushottam, Prof. (Dr.) Kanak Saxena, Richa Sharma, “Efficient Heart Disease Prediction System”, 2016, pp.962-969.
[10] K.Gomathi, Dr. D.Shanmuga Priyaa, “Multi Disease Prediction using Data Mining Techniques”, International Journal of System and Software Engineering, December 2016, pp.12-14.

