

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/372126412>

Machine Learning Algorithms for Disease Diagnosis using Medical Records: A Comparative Analysis

Conference Paper · May 2023

DOI: 10.1109/ICIEM59379.2023.10165850

CITATIONS

0

READS

90

6 authors, including:



Ashulekha Gupta

Graphic Era (Deemed to be University)

77 PUBLICATIONS 327 CITATIONS

SEE PROFILE



Chaithra .N

Jain University

10 PUBLICATIONS 11 CITATIONS

SEE PROFILE



Janhvi Jha

Jain University

8 PUBLICATIONS 7 CITATIONS

SEE PROFILE



Anu Sayal

Taylor's University

25 PUBLICATIONS 45 CITATIONS

SEE PROFILE

Machine Learning Algorithms for Disease Diagnosis using Medical Records: A Comparative Analysis

Ashulekha Gupta

Department of Management Studies
Graphic Era (deemed to be University)
Dehradun, India
ashulekha1806@gmail.com

Anu Sayal

School of Accounting and Finance
Taylor's Business School, Taylor's
University
Malaysia
anu.sayal.07@gmail.com

Chaithra N

Department of CSE (AI & ML)
JAIN (deemed to be University)
Bangalore, India
chaithra3nats@gmail.com

Veethika Gupta

Department of Economics
Doon University
Dehradun, India
veethikagupta28@gmail.com

Janhvi Jha

Department of CSE (AI & ML)
JAIN (deemed to be University)
Bangalore, India
jahnvi.jak@gmail.com

Minakshi Memoria

Department of CSE, UIT
Uttaranchal University
Dehradun, India
minakshimemoria@gmail.com

Abstract— The development of machine learning algorithms has revolutionized the medical data categorization industry through the introduction of artificial intelligence. The development of mathematical models employing statistical theory to draw conclusions from samples has proven to be a fruitful application of machine learning. Due to the enormous amounts of patient data, machine learning algorithms have been implemented into the medical industry to make crucial diagnostic decisions. Particularly for busy institutions, accurate diagnosis can be difficult and time-consuming. By analyzing medical imaging data and learning from labelled examples, machine learning algorithms have demonstrated tremendous potential for automatically classifying and identifying diseases. This study evaluates the performance of four prominent machine learning algorithms in various medical contexts, including cardiac care, trauma units, breast cancer diagnosis, etc. Using the Heart Disease UCI dataset as an example, this study evaluates the accuracy of various cardiovascular disease prediction algorithms using accuracy ratings and confusion matrices. It insinuates that machine learning algorithms, such as logistic regression, random forest, deep neural networks, and gradient boosting, have the potential to improve healthcare decision-making processes and demonstrates the importance of machine learning algorithms in this field.

Keywords— *Machine learning, Models, Neural Networks, Random forest*

I. INTRODUCTION

Machine learning techniques have been extensively utilized in intelligent healthcare systems, diagnosis, and prognosis over the past few decades. The conventional medical approach is based on the expertise of the physician, which is acquired through extensive examination of diverse patient symptoms and verified diagnoses over time. Nevertheless, even in such situations, ensuring accuracy remains uncertain. The progress of computer technology has enabled the accumulation and retention of extensive amounts of data, including electronic medical records. The analysis of large datasets in the field of healthcare requires the aid of computers, especially when performing intricate data queries, as it is beyond the capacity of health professionals to conduct such analyses manually. Hence, the field of intelligent

healthcare systems holds great significance and value in research. The implementation of an intelligent medical system has the potential to aid physicians in improving the accuracy of diagnoses and providing patients with more effective treatment options. Additionally, it may assist individuals in better understanding and preparing for their future healthcare needs. Machine learning (ML) techniques have the potential to automate complex manual tasks, such as text and speech analysis, traditionally performed by physicians. These techniques have been applied to identify and classify patients' emotions based on doctors' assessments [1].

In recent years, the integration of machine learning (ML) algorithms in health diagnostics has gained significant attention due to its potential to enhance clinical decision-making [2]. In recent years, the field of machine learning has gained significant popularity, with a multitude of algorithms being widely adopted. This research paper will discuss three commonly used machine learning algorithms: logistic regression, random forests, and neural networks. In recent years, algorithms have been utilized in various fields, including image recognition and natural language processing, with encouraging outcomes. In this paper, we will explore the strengths and weaknesses of these popular ML algorithms and their applications in various domains. Logistic regression has been widely utilized in the development of clinical prediction models for the detection of lung cancer and cardiac disease due to its efficacy in binary classification challenges. Random forests have gained popularity due to their capability of combining classifiers, making them a valuable tool for addressing multiclass classification issues such as the diagnosis of diabetes mellitus. Neural networks have demonstrated potential in patient classification and diagnosis due to their resemblance to the structure of the human nervous system. The implementation of these techniques has been observed in diverse settings, such as organizing patient records in cardiac care and trauma units, diagnosing breast cancer, and identifying neuromuscular disorders. Machine learning (ML) algorithms have been extensively studied in medical research to enhance patient outcomes and offer crucial insights that can assist in decision-making procedures. The purpose of this research paper is to examine four different approaches for assessing the Heart Disease UCI dataset to

determine the likelihood and frequency of individuals developing heart disease. The objective of this research paper is to investigate the effectiveness of various predictive models, including logistic regression, random forest classification, deep neural networks, and gradient boosting, in forecasting the likelihood of cardiovascular disease development in patients. In evaluating the effectiveness of models, various measures are utilized, including the accuracy score and confusion matrix [3,4].

I. ALGORITHMS

A. Logistic Regression

Logistic regression is a commonly used classification method when the response variable is binary. The logistic function was originally developed in the 19th century for the purpose of modeling population growth. When attempting to create a logistic regression model of high quality, the researcher encounters a significant challenge in the form of conflicting objectives. It has been observed that despite the commonality of data and objectives among modelers, there exists a significant variance in the solutions obtained. This can be attributed to the choices made by the modelers at different stages of the modeling process. The utilization of LR analysis has gained significant attention as a statistical approach in medical research within the last 20 years [5]. Prediction models have been shown to be useful in aiding physicians and patients in making informed decisions in various scenarios, including diagnostic testing, therapy initiation or discontinuation, and lifestyle modifications.

In the study conducted by the authors of [6], a prospective cohort was utilized to develop a clinical prediction model for lung cancer. The accuracy of the model was subsequently assessed using internal data. The logistic regression model underwent development, assessment, and validation procedures utilizing bootstrapping and optimism corrected methods. These statistical strategies are widely recognized as effective approaches for addressing missing covariate data. The TREAT model (Thoracic Research Evaluation And Treatment model) utilizes easily accessible data to assess the probability of malignancy in a lung nodule, as observed by examining surgeons. The TREAT model was utilized to identify the optimal candidates for low-dose CT scans and to evaluate the probability of malignancy in a recently detected lung nodule. This was achieved by employing a set of pre-established candidate factors that were sourced from verified models published in the literature. Restricted cubic splines with three and five knots were employed to evaluate the potential non-linear association between the continuous predictors and lung cancer outcome. A likelihood ratio test was conducted in order to evaluate the linearity of the relationship. The model underwent verification, and its ability to accurately distinguish between lung cancer and benign conditions exhibited minimal indications of overfitting.

Total bilirubin, albumin, alkaline phosphatase, and alanine aminotransferase were picked and categorized as biomarkers in [7] based on monitoring data from patients with hepatobiliary disease. Assessing the likelihood ratio and precision of a model of quaternary logistic regression, the researchers developed a ternary ungrouped logistic model, conducted numerous saliency and accuracy tests, and then modified the model based on the outcomes.

In their study, Miranda et al. utilized an intelligent computational model and machine learning technique to

develop a non-invasive prediction system for early cardiac disease detection. The employed methods included logistic regression and the stochastic gradient descent algorithm. The authors' findings are documented in reference [8]. A dataset consisting of 303 individuals was extracted from the UCI repository for clinical analysis. The results indicate that logistic regression exhibited superior performance compared to stochastic gradient descent with respect to accuracy, precision, F-measure, and recall.

B. Random Forest

The Random Forests Classifier (RFC) is a widely recognized and effective method in the field of machine learning. It is particularly useful for addressing high-dimensional classification and skewed issues [9].

The authors of [10] propose utilizing a Random Forest classifier (RF) as a means of predicting the behavioral response of individuals with diabetes, particularly older individuals, to changes in their physical and mental surroundings. The present investigation employs the Pima Indian diabetes database, which is located at the UCI Machine Learning Lab, to execute various test parameters for the random forest classifier. According to recent research, RF has the potential to effectively diagnose diabetes mellitus, provided that the patient provides the requisite characteristic value. The article [11] discusses the use of different machine learning methods, including Naïve Bayes, Multi-Layer Perceptron, Random Forest, Random Tree, and Modified J48, for the early detection of diabetes for the same dataset as used by [10]. The analysis indicates that the Modified J48 Classifier has the highest accuracy when compared to other techniques. In order to categorize the multi-class issues, [12] suggests a unique hybrid classification method based on C4.5 decision tree classifier and one-against-all technique. Datasets in the fields of dermatology (UCI), image segmentation (UCI), and lymphography (UCI) are all part of the UCI machine learning database. Classification accuracies of 84.48%, 88.79%, and 80.11% were obtained for the dermatological dataset, the image segmentation dataset, and the lymphography dataset, respectively, while utilizing 10-fold cross validation in this study.

The random forest-based approach is effective because it utilizes an aggregate method of classifiers, which can improve performance by taking advantage of the unique strengths of each individual classifier [13].

C. Neural Networks

The use of augmented reality (AR) in eCommerce enables b Computers are utilized in clinical applications for diagnosis, monitoring, and treatment purposes. Artificial neural networks (ANNs) have been widely used in medical research to assist in the analysis, modelling, and interpretation of intricate clinical data. In most clinical applications of artificial neural networks (ANNs), patients are categorized into distinct groups based on their measured characteristics. The artificial neural network (ANN) is a computational model designed to replicate the parallel processing of the human brain. Artificial neural networks (ANNs) are composed of interconnected processing elements, or neurons, that operate in parallel. The components were modelled after the nervous systems of humans and animals for this research paper. The functionality of a network is primarily influenced by the interconnections among its nodes, like the natural world. Each layer within a neural network is responsible for processing a

unique subset of elements. The placement of layers in a neural network typically involves positioning the input layer before the output layer. In some neural networks, there may be one or multiple layers of hidden units situated between the input and output layers. Neural networks can be trained to perform specific tasks by adjusting the weights (connections) between their nodes [14].

Artificial neural networks (ANNs) have shown potential in classifying patients and making diagnoses, as diagnosing a medical condition can be considered a pattern classification query. Artificial neural networks (ANNs) can integrate knowledge-based methods with data-driven analytical techniques, such as decision and classification theory, to offer valuable insights that can assist in decision-making processes. The effectiveness of artificial neural networks (ANN) in addressing medical concerns has been well-documented through various applications [15]. The study utilized real medical records of patients in the coronary care unit (CCU) and trauma unit to train an artificial neural network (ANN) model for the purpose of categorizing the medical records into specific groups, as documented in [16]. In reference [17], the diagnostic performance of various ANN models in detecting breast cancer was assessed. The present study employs self-organizing maps (SOMs) as a model selection technique to detect breast cancer, as demonstrated in reference [18]. In [19], the utilization of two artificial neural network (ANN) models, namely supervised and unsupervised, for the examination of electromyography (EMG) characteristics in the identification of neuromuscular disorders was explicated. The paper presented by [20] outlines a diagnostic approach for pulmonary illnesses. The study utilized clinical data to evaluate the efficacy of the system in treating diverse diseased organs. The use of an artificial neural network system to detect COVID-19 is proposed in [21]. The proposed method employs chest X-ray images and is based on Convolutional CapsNet. [22].

Cervical cancer is the second most prevalent malignancy in women and the leading cause of death. To detect cervical cancer in its earliest stages, a PAP smear that searches for malignant uterine cells is required. Historically, the examination of cells was performed using human eyes and a microscope, which was susceptible to error. As a simplified alternative, PAPnet, an application based on neural networks, has been developed. It is acknowledged that this network provides more precise stain analysis than conventional methodologies. Additionally, attempts have been made to identify ovarian cancer using artificial neural networks. Compared to other types of cancer, lung cancer is the leading cause of mortality around the globe. CT scans have been the standard method for detecting lung cancer, but a more advanced method is urgently needed. Chest radiographs are analyzed by a computer to facilitate the challenging task of classifying test results using only human judgment. Recently, CT scan images have been analyzed and classified using Neural Networks. Utilizing a virtual neural network approach, also known as Neural Embedded based Detection, permits the early detection of lung cancer via the examination of specimen images. It has been demonstrated that ANN is more precise than traditional PSA values for the early detection of prostate cancer [23].

A complex neural network with increased connectivity between its neurons is commonly known as a deep neural network (DNN), multilayer perceptron (MLP), or fully

connected deep network (FCDN). Deep neural networks are currently being studied and utilized in various aspects of healthcare imaging infrastructure and facilities. These include disease prediction, diagnosis, monitoring of disease recovery, drug development, health management, and health imaging. It is common to encounter resource inequality, access to care issues, doctor-patient relationship challenges, tension, and fatigue in the healthcare imaging setting. In traditional medicine, doctors' experience and patients' condition are given more importance, which can sometimes lead to a lower level of precision. A medical assessment model that utilizes integrated deep neural networks and medical big data, along with input from authoritative experts and medical literature, can address the limitations of human doctors' experience and machinery. This can lead to the creation of an interactive and intelligent medical platform that is based on information. This approach can be achieved through artificial intelligence. The speed and precision of diagnostic procedures are greatly enhanced by this [24].

D. Gradient boosting

Tabular datasets are commonly analyzed using gradient boosting, which is a highly favored machine learning algorithm. The tool is capable of detecting nonlinear relationships between the target and features of your model.

In their study, the authors employed XGBoost algorithm to effectively classify tumors into early and advanced stages. This approach was implemented to its fullest extent in order to achieve optimal results. The study in question can be found in reference [25]. The research findings indicate that the XGBoost model proposed exhibited comparable or superior accuracy in predicting cancer stage compared to existing models. Additionally, the model demonstrated the ability to extract significant features from multi-omics data on a large scale.

II. A COMPARATIVE STUDY OF MACHINE LEARNING MODELS FOR PREDICTING HEART DISEASE

The aim of this study is to utilize the Heart Disease UCI dataset to forecast the likelihood of an individual developing heart disease. The dataset was preprocessed using exploratory data analysis techniques before developing the model [26]. This involved generating a correlation matrix and plotting the age distribution of the data as in Fig 1 and Fig 2. The study revealed noteworthy associations between heart disease and various characteristics including Chest Pain Type, Maximum Heart Rate, Slope, Exercise-Induced Angina, Oldpeak, and Vessel Color. However, the correlation between heart disease and Fasting Blood Sugar was found to be the least significant. To mitigate the issue of dummy variable trap, categorical features were incorporated into the dataset and transformed into new columns through one-hot encoding during the training process. In order to simplify data analysis, the target column was substituted and the encoded prior category columns were eliminated. The data was scaled to a range of 0 to 1 following the production of NumPy values and the segregation of data into training and testing sets, as part of the research methodology [27].

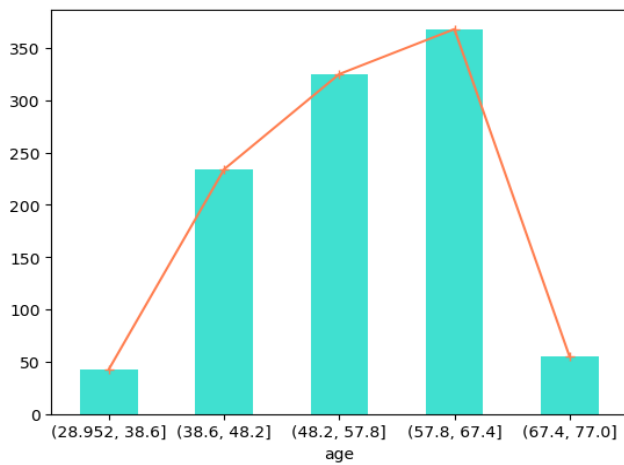


Fig. 1. Age Distribution of Heart Disease UCI Dataset

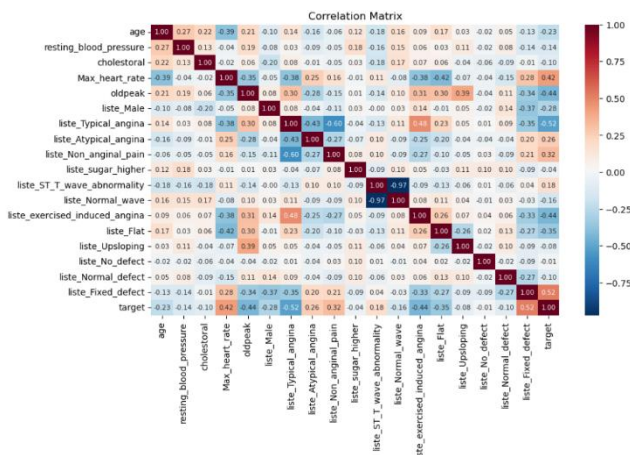


Fig. 2. Correlation Matrix

The present study examines the application of the logistic regression model to data related to cardiovascular disease [28]. The initial step involves importing essential libraries for the research, including pandas for data processing, `train_test_split`, `LogisticRegression` for implementing the regression, and `accuracy_score` and `confusion_matrix` for assessing the model's performance. The dataset was divided into a training set and a testing set using the `train_test_split` function in the code. In the context of machine learning, it is common practice to allocate 80% of the available data for training purposes, with the remaining 20% reserved for assessment. The exclusion of the target column from the training set is a common practice as it serves as the dependent variable in the research study. The input data was calibrated to the logistic regression model using the `fit()` function in the code. In the research paper, predictions for the test data were generated using the `predict()` function of the model. The process of generating a confusion matrix and accuracy score involves comparing the predicted values of a model with the actual values of a test set. This is a crucial step in evaluating the performance of a model. The results comprise of the accuracy metric and the confusion matrix, which are essential evaluation measures in research studies. The objective of this research paper is to utilize logistic regression to predict the occurrence of cardiac disease in patients, taking into account multiple variables including age, gender, cholesterol levels, and other health indicators. The logistic regression model is employed to determine the probability threshold based on patient characteristics, which is subsequently utilized to

classify patients as either having or not having heart disease. The assessment of the model's classification performance can be accomplished by examining the accuracy score and the confusion matrix.

Subsequently, the Heart Disease UCI dataset is subjected to the Random Forest classification technique. The study commences by importing the essential components of scikit-learn. For the purpose of ensuring replicability, the dataset has been partitioned into training and testing sets in an 80:20 ratio, utilizing the `train_test_split()` method with a random state of 42. The `RandomForestClassifier()` function was utilized to create a Random Forest classifier consisting of 100 trees. The `fit()` method was then employed to adjust the model to the training data. Subsequently, the classifier utilizes the `predict()` function to anticipate the target variable of the testing set, and records the outcome as `rf_y_predict`. The confusion matrix and accuracy of the model on the testing set were calculated using scikit-learn's `confusion_matrix()` and `accuracy_score()` methods. The accuracy score and confusion matrix are two commonly used metrics in evaluating predictive models. The study employs the Random Forest algorithm to classify data from the Heart Disease dataset. The performance of the model is assessed through the utilization of a confusion matrix and an accuracy score.

The next model is a deep neural network model which contains multiple layers of fully connected nodes, each with its own activation function. This comprises the neural network architecture constructed using the Python Keras module for data transformation. Backpropagation and stochastic gradient descent were employed during model training to minimize the loss function and optimize the model parameters. To avoid falling into the dummy variable pitfall, categorical variables were re-encoded as discrete columns using one-hot encoding during training. The prior category columns were expunged because they were encoded, and the target column was reinstated to facilitate data partitioning. We scaled the data to a range of 0 to 1 after producing the numpy values and segregating the data into training and testing sets. The two-layer deep neural network was trained with 100 backpropagation iterations and 32-element batches. The confusion matrix and precision score indicated that the model accurately predicted all instances in the test set. Based on a binary categorization of the input characteristics, the model's output indicates whether or not a given patient has cardiac disease. The model assigns each class a probability score between 0 and 1, and employs a threshold value to make its ultimate prediction. This output can be used by medical professionals to better assess patients and decide on appropriate treatments [29].

The fourth model for the analysis of the UCI Heart Disease dataset will involve the application of the Gradient Boosting technique [30, 31, 33]. The dataset was partitioned into a training set and a testing set utilizing the `train_test_split` function from the `sklearn.model_selection` library. The initialization of the Gradient Boosting model is initiated by using `GradientBoostingClassifier()` in this research paper. The `fit` function is utilized to train the model with the provided training set in the research paper. The testing set is predicted using the `predict` technique, and the model's performance is evaluated through the `confusion_matrix` and `accuracy_score` functions, as part of the research methodology [32, 34, 35]. The research paper discusses the utilization of "Gradient Boosting," a machine learning technique that combines

several weak prediction models, including decision trees, to create a powerful model. The model is constructed through a phased algorithmic approach, where each subsequent iteration is designed to rectify the errors of the previous model. The ultimate forecast outcome is determined by calculating the weighted average of the predictions made by each model. Modifying the hyperparameters of the Gradient Boosting method can improve the effectiveness of the model. The hyperparameters under consideration are the number of estimators and the learning rate.

The objective of this study is to evaluate the precision of various models employed for the classification of cardiac disease in the UCI dataset. The accuracy ratings of Logistic Regression, Random Forest, Deep Neural Networks, and Gradient Boosting are visually depicted as a bar chart in the research paper. The chart titled "Accuracy Scores for Heart Disease Classification Models" displays the accuracy ratings of various heart disease classification models. The x-axis indicates the designated accuracy rating of each model, while the y-axis represents the actual accuracy rating. The y-axis of the graph is limited to a range of 0.7 to 1.0. The generated code produces a Receiver Operating Characteristic (ROC) curve for each model, facilitating a comparative analysis of their respective efficacy. The FPR and TPR can be determined by utilizing the `roc_curve` function from the `sklearn.metrics` library, which is commonly used in research papers. The AUC is computed using the `auc` function to determine the area under the ROC curve. Distinct graphs displaying the ROC curve and AUC for each model are presented, with corresponding labels superimposed. This information is relevant to the research paper. The graphical representation being referred to is commonly recognized as a "Receiver Operating Characteristic (ROC) Curve" and is frequently used in research studies.

The algorithm generates visual depictions of the accuracy scores and ROC curves for each model (refer Fig 1 and Fig 2), facilitating a straightforward comparison of their respective strengths. The evaluation of the Heart Disease UCI dataset models' effectiveness on the test set is based on their respective scores, which indicate the percentage of correct predictions made by each model. The findings suggest that the Random Forest algorithm achieves the highest precision level (0.99), trailed by Gradient Boosting (0.97), a Deep Neural Network (0.94), and Logistic Regression (0.81).

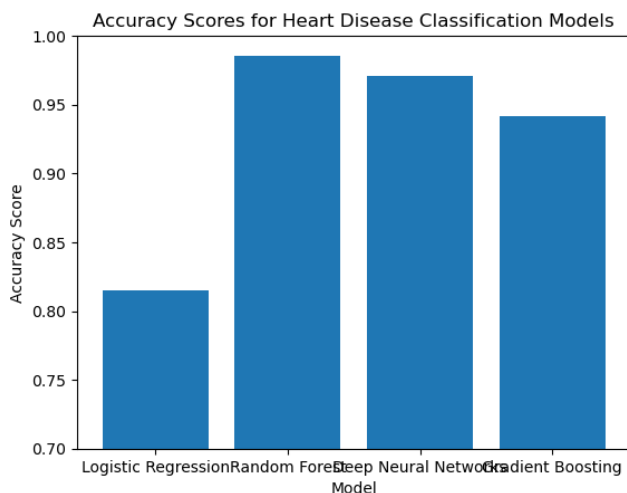


Fig. 3. Accuracy Scores for Heart Disease Classification Models

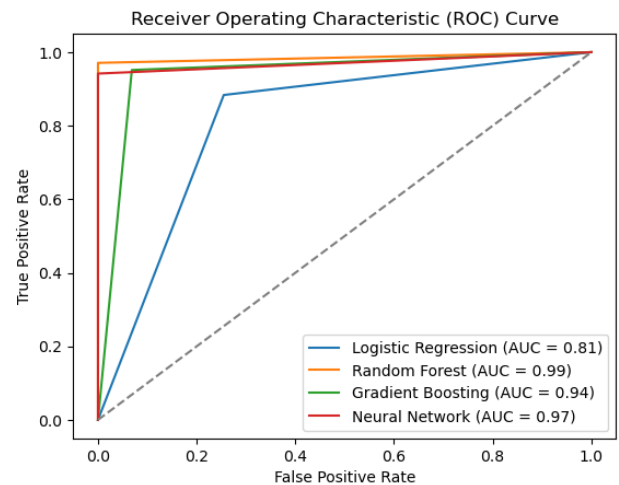


Fig. 4. ROC Curve for Classification Model

III. CONCLUSION

The application of several machine learning models in the diagnosis of various illnesses has been thoroughly surveyed in this work. It was found that researchers have published a large number of models over the years using algorithms like logistic regression, random forest, etc. to build predictive models that could analyze the onset of different cancers, heart diseases, pulmonary infections, diabetes, and more with a moderate level of accuracy. This has shown to be very useful in the early detection of the condition described, allowing for more effective therapy. Additionally, we came to the conclusion that random forest offered the highest accuracy via our analysis of a UIC dataset.

REFERENCES

- [1] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, "Machine learning with applications in breast cancer diagnosis and prognosis," *Designs*, vol. 2, no. 2, p. 13, 2018.
- [2] P. Oza, P. Sharma, and S. Patel, "Machine learning applications for computer-aided medical diagnostics," *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*, pp. 377–392, 2021.
- [3] D. Pavithra, "A study on machine learning algorithm in medical diagnosis," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 4, pp. 42–46, 2018.
- [4] T. H. Rafi, R. M. Shubair, F. Farhan, M. Z. Hoque, and F. M. Quayyum, "Recent advances in computer-aided medical diagnosis using machine learning algorithms with optimization techniques," *IEEE Access*, vol. 9, pp. 137847–137868, 2021.
- [5] D. Bertsimas and A. King, "Logistic regression: From art to science," *Statistical Science*, vol. 32, no. 3, 2017.
- [6] S. A. Deppen, J. D. Blume, M. C. Aldrich, S. A. Fletcher, et al., "Predicting lung cancer prior to surgical resection in patients with lung nodules," *Journal of Thoracic Oncology*, vol. 9, no. 10, pp. 1477–1484, 2014.
- [7] X. Gai and Y. Zhang, "Diagnosis of hepatobiliary disease based on logistic regression model," *IOP Conference Series: Materials Science and Engineering*, vol. 490, p. 062084, 2019.
- [8] E. Miranda, F. M. Bhatti, M. Aryuni, and C. Bernando, "Intelligent computational model for early heart disease prediction using logistic regression and stochastic gradient descent (a preliminary study)," *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAD)*, 2021.
- [9] L. Breiman, *Random Forests*, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [10] M. Butwall and S. Kumar, "A data mining approach for the diagnosis of diabetes mellitus using random forest classifier," *International Journal of Computer Applications*, vol. 120, no. 8, pp. 36–39, 2015.

- [11] R. D. Devi and P. R. Vijayalakshmi, "Performance analysis of data mining classification algorithms for early prediction of diabetes mellitus 2," *International Journal of Biomedical Engineering and Technology*, vol. 36, no. 2, p. 148, 2021.
- [12] K. Polat and S. Güneş, "A novel hybrid intelligent method based on c4.5 decision tree classifier and one-against-all approach for multi-class classification problems," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1587–1592, 2009.
- [13] S. Benbelkacem and B. Atmani, "Random forests for diabetes diagnosis," 2019 International Conference on Computer and Information Sciences (ICCIS), 2019.
- [14] Q. A. Shaye, G. E. Refae, and S. Yaseen, "Artificial Neural Networks for medical diagnosis using biomedical dataset," *International Journal of Behavioural and Healthcare Research*, vol. 4, no. 1, p. 45, 2013.
- [15] A. Quteishat and C. P. Lim, "Application of the fuzzy min-max neural networks to medical diagnosis," *Lecture Notes in Computer Science*, pp. 548–555, 2008.
- [16] C. P. Lim, R. F. Harrison, and R. L. Kennedy, "Application of autonomous neural network systems to medical pattern classification tasks," *Artificial Intelligence in Medicine*, vol. 11, no. 3, pp. 215–239, 1997.
- [17] T. Kıyan and T. Yildirim, "Breast Cancer Diagnosis Using Statistical Neural Networks," *International XII. Turkish Symposium on Artificial Intelligence and Neural Networks – TAINN 2003*, vol. 4, no. 2, pp. 1149–1153, 2004.
- [18] D. West and V. West, "Model selection for a medical diagnostic decision support system: A breast cancer detection case," *Artificial Intelligence in Medicine*, vol. 20, no. 3, pp. 183–204, 2000.
- [19] C. S. Pattichis, C. N. Schizas, and L. T. Middleton, "Neural network models in EMG diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 5, pp. 486–496, 1995.
- [20] G.-P. K. Economou, C. Spiropoulos, N. M. Economopoulos, N. Charokopos, D. Lymberopoulos, M. Spiliopoulou, E. Haralambopulu, and C. E. Goutis, "Medical diagnosis and artificial neural networks: A medical expert system applied to pulmonary diseases," *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, 1994.
- [21] S. Toraman, T. B. Alakus, and I. Turkoglu, "Convolutional capsnet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks," *Chaos, Solitons & Fractals*, vol. 140, p. 110122, 2020.
- [22] G. Marques, D. Agarwal, and I. de la Torre Díez, "Automated medical diagnosis of COVID-19 through EfficientNet Convolutional Neural Network," *Applied Soft Computing*, vol. 96, p. 106691, 2020.
- [23] M. Desai and M. Shah, "An anatomization on breast cancer detection and diagnosis employing multi-layer Perceptron neural network (MLP) and Convolutional Neural Network (CNN)," *Clinical eHealth*, vol. 4, pp. 1–11, 2021.
- [24] L. Ma and T. Yang, "Construction and evaluation of Intelligent Medical Diagnosis Model based on Integrated Deep Neural Network," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–10, 2021.
- [25] B. Ma, F. Meng, G. Yan, H. Yan, B. Chai, and F. Song, "Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data," *Computers in Biology and Medicine*, vol. 121, p. 103761, 2020.
- [26] K. Gangal, "Heart disease dataset UCI," Kaggle, 03-Aug-2021. [Online]. Available: <https://www.kaggle.com/datasets/ketangangal/heart-disease-dataset-uci>. [Accessed: 23-Apr-2023].
- [27] N. Pathan, "UCI_HEART_DISEASE," Kaggle, 27-Dec-2022. [Online]. Available: <https://www.kaggle.com/code/nawazkhanpathan/uci-heart-disease>. [Accessed: 23-Apr-2023].
- [28] "Access anytime anywhere," Cleveland Clinic, 2023. [Online]. Available: <https://my.clevelandclinic.org/>. [Accessed: 23-Apr-2023].
- [29] b3d1rhan, "Heart_diseases_deep_neural_network_model," Kaggle, 05-Aug-2021. [Online]. Available: <https://www.kaggle.com/code/b3d1rhan/heart-diseases-deep-neural-network-model>. [Accessed: 23-Apr-2023].
- [30] S. S. Negi, A. Gupta, et al., "Machine Learning Based Hybrid Technique for Heart Disease Prediction," 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM), Dehradun, India, 2022, pp. 1–6, doi: 10.1109/ICACCM56405.2022.10009219.
- [31] R. Kumar, M. Memoria, A. Gupta and M. Awasthi, "Critical Analysis of Genetic Algorithm under Crossover and Mutation Rate," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2021, pp. 976–980, doi: 10.1109/ICAC3N53548.2021.9725640.
- [32] A. Saxena, D. MISRA, R. Ganesamoorthy, J. L. Arias Gonzales, H. A. Almashaqbeh and V. Tripathi, "Artificial Intelligence Wireless Network Data Security System For Medical Records Using Cryptography Management," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 2555–2559, doi: 10.1109/ICACITE53722.2022.9823615.
- [33] Sharma, V., & Manocha, T. (2023). Comparative Analysis of Online Fashion Retailers Using Customer Sentiment Analysis on Twitter. Available at SSRN 4361107
- [34] G. Sharma, V. Tripathi, M. Mahajan and A. Kumar Srivastava, "Comparative Analysis of Supervised Models for Diamond Price Prediction," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 1019–1022, doi: 10.1109/Confluence51648.2021.9377183.
- [35] A. Gupta, R. Parmar, P. Suri and R. Kumar, "Determining Accuracy Rate of Artificial Intelligence Models using Python and R-Studio," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 889–894, doi: 10.1109/ICAC3N53548.2021.9725687.