STM JOURNALS

Journal of
Electronic Design Technology

ISSN: 2229-6980 (Online)
ISSN: 2321-4228 (Print)
Volume 16, Issue 1, 2025
January–April
DOI (Journal): 10.37591/JoEDT

https://journals.stmjournals.com/joedt

Research JoEDT

Advancing Healthcare Systems: A Machine Learning Approach to Multi-Disease Prediction

Ankit Sharma[1], A.N. Kshirsagar[2,*], Anish Kannawar[1], Abhishek Mishra[1]

Abstract

The integration of machine learning algorithms in healthcare has revolutionized the way we approach disease prediction and diagnosis. An attempt to employ machine learning techniques to forecast numerous diseases is presented in this study. A diverse dataset containing patient records, medical history, and relevant features for various diseases was used to develop predictive models. Feature selection and normalization were among the preprocessing methods used to clean and prepare the data. To train and assess the prediction models, a variety of machine learning methods were used, including k-Nearest Neighbors (k-NN), Random Forest, Decision Trees, and Support Vector Machines (SVM). The primary objective was to enhance healthcare by providing timely and accurate predictions for conditions such as diabetes, heart disease, cancer, and respiratory illnesses. A user-friendly interface for medical practitioners incorporates the predictive models that are produced.

Keywords: Machine learning, disease prediction, predictive modeling, decision trees, random forest, support vector machines, k-nearest neighbors

INTRODUCTION

One of the industries with the highest data collecting and processing concerns is healthcare. A significant amount of multidimensional data about patients is produced because of the digital age and technical improvements, including clinical factors, hospital resources, disease diagnostic data, patient records, and medical equipment. To extract knowledge for efficient decision-making, the vast, dense, and complicated data must be processed and assessed. Finding hidden patterns in medical data sets is very much possible with the help of medical data mining. Utilizing a variety of data mining technologies and machine learning techniques has transformed healthcare companies by spotting important patterns and finding correlations and links among numerous variables in gigantic databases. Medical professionals and doctors need to be well-versed in all relevant diagnostic criteria, patient history, and a combination of drug therapy for effective treatment. However, since they make decisions instinctively, mistakes are possible. Their cognitive abilities are constrained due to several issues. Significant attempts are being made by data mining and machine learning techniques to cleverly convert accessible data into useful information to increase the effectiveness of the diagnostic procedure. A lot of research has been done to look at how machine learning may be used to create diagnostic tools. It was found that machine learning algorithms could detect with 91.1% accuracy, outperforming the highest-skilled doctor's diagnosis accuracy of 79.97% [1]. To extract characteristics for the best possible sickness diagnosis, prediction, prevention, and therapy, machine learning techniques are specifically applied to datasets of illnesses.

***Author for Correspondence**
A.N. Kshirsagar
E-mail: amit.kshirsagar_skncoe@sinhgad.edu

[1]Student, Department of Electronics and Telecommunication Engineering, Smt. Kashibai Navale College of Engineering (affiliated to Savitribai Phule Pune University), Vadgaon, Pune, Maharashtra, India
[2]Professor and Head, Department of Electronics and Telecommunication Engineering, Smt. Kashibai Navale College of Engineering (affiliated to Savitribai Phule Pune University), Vadgaon, Pune, Maharashtra, India

Received Date: December 23, 2024
Accepted Date: February 01, 2025
Published Date: February 08, 2025

Citation: Ankit Sharma, A.N. Kshirsagar, Anish Kannawar, Abhishek Mishra. Advancing Healthcare Systems: A Machine Learning Approach to Multi-Disease Prediction. Journal of Electronic Design Technology. 2025; 16(1): 1–6p.

LITERATURE SURVEY

A Bayesian classifier assumes that each attribute (such as traits or data points) is independent of the others and uses probability to create predictions. It first examines the likelihood of each category before displaying any data. It then uses these probabilities to determine which category the data most likely falls into when fresh data is received. It organizes the relationships between the features and aids in making this prediction using a model known as a Bayesian network. In the end, the sum of these probabilities determines the most likely outcome. Kidney disease was predicted using the SVM and Naive Bayes techniques [2].

Several authors tried to classify different kidney disease stages using the specified ANFIS algorithm. The goal of the project was to develop an efficient classification algorithm employing a variety of evaluation criteria, including execution time and accuracy. The Naive Bayes algorithm performed better since it gave results faster, despite the SVM algorithm having higher classification accuracy. The findings demonstrate that SVM performs better at predicting renal disease than the Naive Bayes Approach. A fuzzy KNN classifier was employed in one study to predict heart disease, increasing accuracy by lowering data uncertainty. The fuzzy approach outperformed the standard KNN in terms of producing precise predictions when the data was divided into two sections for testing and training. In a different study, diabetes and breast cancer were diagnosed using adaptive Support Vector Machines (SVM), a type of machine learning. This technique changed a few parameters to achieve 100% accuracy in both diagnoses. Finally, a different strategy coupled the C4.5 classification algorithm with K-means clustering, which groups comparable data, to predict type 2 diabetes. To ensure the accuracy of the predictions, a technique called k-fold cross-validation was used. With a classification accuracy of 88.38% utilizing the hybrid technique, the model produced positive results that could be very helpful for physicians in making wise therapeutic decisions regarding diabetes [3–5].

Grampurohit and Sagarnal reported an automated method for resolving challenging questions regarding the prognosis of cardiac illness [6]. This intelligent system was developed using the Naive Bayes methodology to deliver faster, better, and more accurate results. It might help medical professionals make clinical decisions concerning heart attacks. A pacemaker may be added to the order, SMS capabilities added, and mobile applications created for Android and iOS.

A novel method built on the ANN algorithm was developed for the prediction of heart illness. The 13 most important clinical criteria were taken into consideration by the researchers as they created a categorization-based interactive prediction method using an artificial neural network algorithm. The recommended strategy demonstrated 80% accuracy in predicting cardiac problems and can be highly helpful for medical professionals [7–10].

Xie *et al.* explain the key methods behind using deep learning for predicting diseases. They point out the limitations of current prediction approaches, like issues with accuracy or data quality, but also emphasize the advantages that deep learning offers in predicting various health conditions. Looking ahead, they believe deep learning could play an even bigger role in healthcare, assisting doctors in making more accurate predictions and improving treatments. Essentially, they aim to show that deep learning is already valuable in medicine and has the potential to become even more impactful in the future [11].

MOTIVATION OF SYSTEM

The motivation behind the *"Advancing Healthcare Systems: A Machine Learning Approach to Multi-Disease Prediction"* project is to revolutionize healthcare by leveraging advanced technology to predict the likelihood of various diseases in individuals. Through the utilization of machine learning algorithms, the system seeks to deliver precise and prompt forecasts, facilitating early intervention and customized treatment approaches. This program aims to save healthcare expenditure, promote preventive care, and eventually increase people's longevity and general well-being.

IMPLEMENTATION DETAILS OF MODULE

The first step involves gathering and organizing vital health information, such as a patient's background, medical history, and test results. To prepare the data for machine learning, preprocessing procedures are next applied to clean it up and organize it. To increase the accuracy and efficiency of the model, the most crucial elements that influence predictions are selected in the feature selection step. The model is then taught to produce predictions using machine learning algorithms including neural networks, logistic regression, and support vector machines. These algorithms learn from the patterns present in the training data and develop a predictive capability. The implementation of multiple disease prediction models using machine learning has the potential to revolutionize preventive healthcare. By providing accurate and timely predictions, these models can enable early intervention and personalized treatment strategies, leading to improved patient outcomes and resource allocation in the healthcare system. The flow chart of the proposed system is shown in Figure 1.

Support Vector Machine

Import the Dataset

- *Purpose:* To train the machine learning model, the dataset must first be loaded.
- *Explanation:* This entails utilizing a Python library such as Pandas to import the data. Additionally, datasets from databases, Excel, and other sources can be imported.

Examine the Information to Determine How they Appear

- *Purpose:* Understanding the dataset's structure and contents is crucial before processing.
- *Explanation:* This makes it easier to find any missing values, data distribution issues, or features that require transformation.

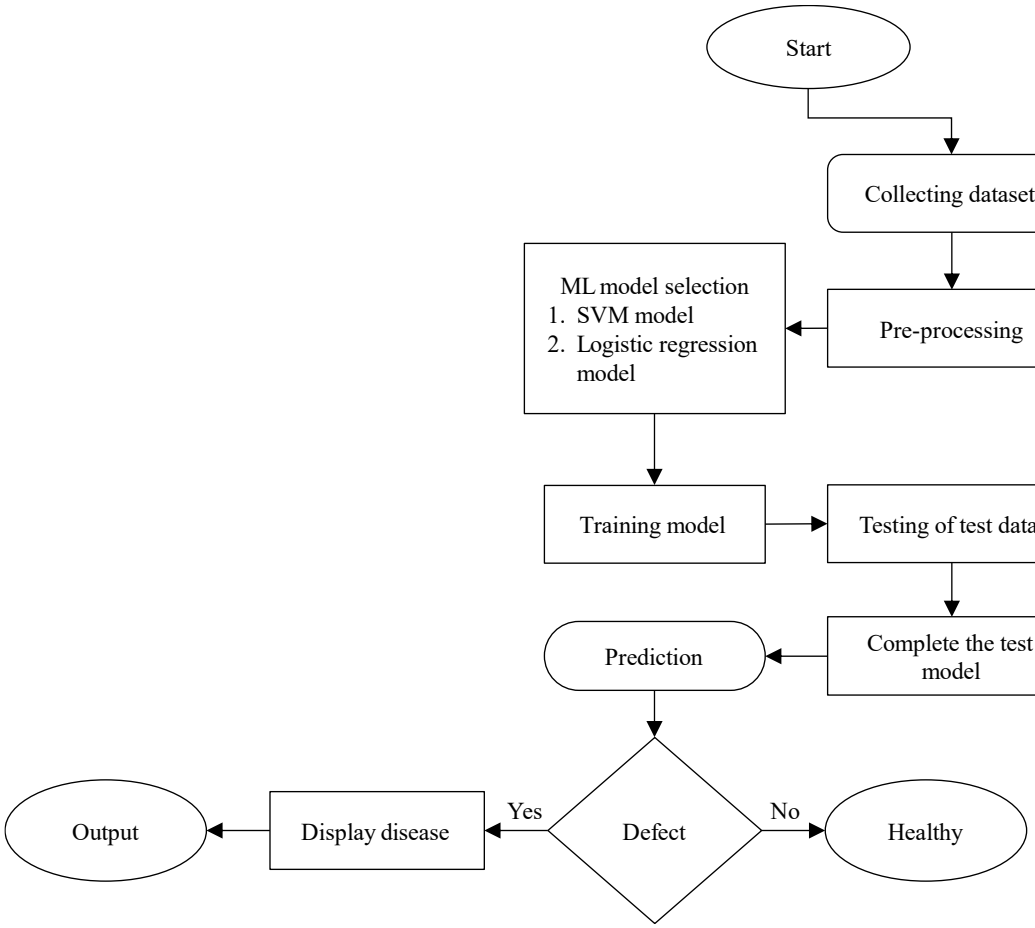

Figure 1. Flow chart of the proposed system.

Pre-process the Data

- *Purpose:* To make sure the model works correctly, data processing and cleansing are crucial.
- *Explanation:* The dataset is prepared for training by pre-processing, which guarantees that all the data is in a format suitable for the machine learning algorithm.

Split the Data into Attributes and Label

- *Purpose:* Since SVM is a supervised learning technique, the data must be divided into labels (y) and features (x).
- *Explanation:* In this case, the target variable (y), which is what we are attempting to predict, is separated from the input features (X).

Divide the Data into Training and Testing Sets

- *Purpose:* To assess the model's performance, the dataset is divided into two subsets: one for testing and one for training.
- *Explanation:* Training (80%) and testing (20%) portions of the dataset are separated by the train_test_split function. Training data is used to build the model, and testing data is used to assess the model's performance.

Train the SVM Algorithm

- *Purpose:* The model is developed and trained using the training data in this step.
- *Explanation:* After initializing the Support Vector Machine, the fit technique uses the training data to train the model. Although a linear kernel is utilized in this example, the dataset may call for the employment of other types, such as a radial basis function (RBF).

Make Some Predictions

- *Purpose:* The model can forecast fresh or untested data after it has been trained.
- *Explanation:* Using the test dataset (X_test), the model generates predictions. To evaluate the correctness of the model, these predictions are contrasted with the real labels (y_test).

Evaluate the Results of the Algorithm

- *Purpose:* Accuracy, precision, recall, and F1 score are among the metrics used to assess the model's performance.
- *Explanation:*
  - *Accuracy:* The proportion of cases with accurate classifications.
  - *Confusion matrix:* explains true negatives, false negatives, false positives, and true positives.
  - *Classification report:* summarizes the key performance indicators for classification, namely precision, recall, and F1-score.

Logistic Regression

1. *Step 1: Data collection: gather dataset.*
   Purpose: Getting the pertinent dataset with the labels (output or target data) and features (input data) is the initial stage. This dataset may originate from several sources, including databases, web scraping, CSV files, and APIs.
2. *Step 2: Preprocessing: handle missing values, scale features.*
   Purpose: Preprocessing is the procedure that cleans and transforms raw data into a format that may be used. For effective model training, the data must be prepared.
3. *Step 3: Split data: separate the test and train sets.*
   Purpose: To train the model and test it, you must separate your dataset into two sections. This guarantees that your model may be tested on previously unseen data.
4. *Step 4: Initialize parameters: set initial weights and bias.*
   Purpose: To reduce errors, machine learning models frequently have parameters (weights and biases) that must be changed during training. In the framework of logistic regression, these parameters stand for the bias and weights.

5. *Step 5: Define sigmoid function: maps features to probabilities.*
   Purpose: The sigmoid function in logistic regression maps every real-valued number into a range between 0 and 1. It determines the probability that a given input is a member of a certain class.
6. *Step 6: Compute cost function: measure model error.*
   Purpose: The model's performance is measured by the cost function, sometimes referred to as the loss function. In logistic regression, the logistic loss function (binary cross-entropy), which determines the discrepancy between the actual labels and the anticipated probabilities, is usually used as the cost function.
7. *Step 7: Gradient descent: update parameters to minimize error.*
   The goal of gradient descent is to reduce the cost function through optimization using weights and bias modifications. In the direction of the negative gradient, it adjusts the settings to reduce the mistake.
8. *Step 8: Iterate: repeat steps 5–7 until convergence.*
   Purpose: Iterative gradient descent is used to reach the ideal parameters. The procedure of calculating the cost and adjusting the parameters is repeated until the cost converges or ceases to vary substantially.
9. *Step 9: Assess performance on the test set to evaluate the model.*
   Purpose: Testing the model's performance on unknown data (the test set) is essential once it has been trained. For this, evaluation criteria such as F1-score, recall, accuracy, and precision are employed.
10. *Step 10: Prediction: use trained model for new data.*
    Purpose: Following training and assessment, the model is prepared to forecast new, untested data.

Output

This method was probably developed as an online tool to help anticipate medical problems using data and machine learning. Users' data is likely evaluated, and their risk of developing diabetes is calculated using an ML model (such as logistic regression, decision trees, or neural networks). The output of the proposed system is shown in Figure 2.

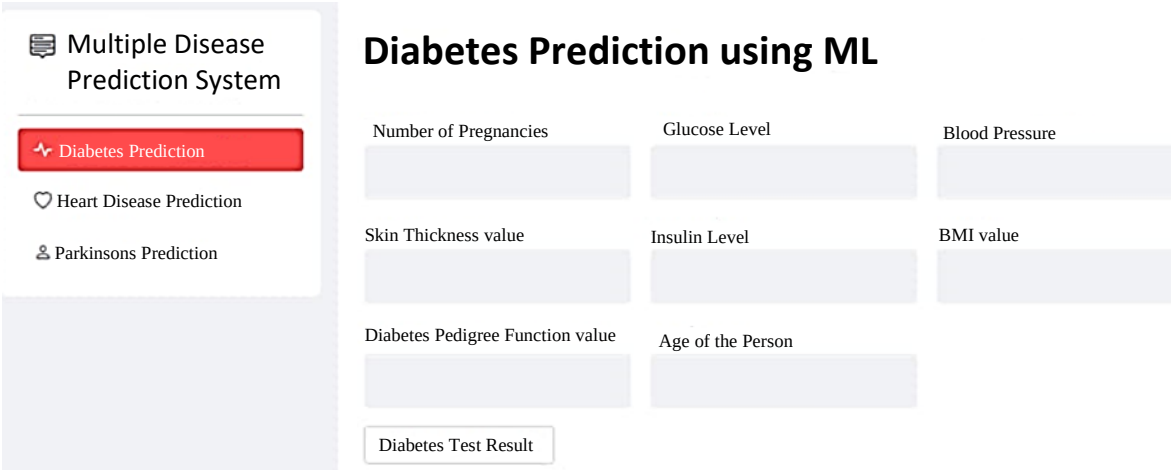

Figure 2. Output for Multiple disease prediction system.

CONCLUSION

The purpose of the disease prediction using machine learning algorithm project is to provide users with a smart and interactive system that can predict potential diseases based on their health information. Utilizing machine learning techniques, the initiative seeks to evaluate user input and offer insightful information about their health risks. The project can be beneficial in several ways. It provides users with a convenient tool to assess their health risks and obtain preliminary predictions for potential diseases. This can help users become more proactive in monitoring their health, seeking appropriate

medical guidance, and taking necessary precautions. It can also be a useful tool for people who do not have easy access to medical facilities or professional medical guidance. Overall, the project aims to contribute to promoting individual health awareness and preventive healthcare practices. Further, one can extend the scope of image-processing diseases like malaria and pneumonia. For the existing diseases, prediction accuracy can be improved by fine-tuning models by having a huge input dataset. The information that comes from studying huge data can be used by machine learning to produce insightful results. A chatbot feature in UI could be added as well for ease of communication between the user and system and to provide the user with a subtle UI UX experience.

REFERENCES

1. Kazemi Y, Mirroshandel SA. A novel method for predicting kidney stone type using ensemble learning. Artif Intell Med. 2018 Jan 1; 84: 117–26.
2. Barakat N, Bradley AP, Barakat MN. Intelligible support vector machines for diagnosis of diabetes mellitus. IEEE Trans Inf Technol Biomed. 2010 Jan 12; 14(4): 1114–20.
3. Patil TR, Sherekar SS. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int J Comput Sci Appl. 2013 Apr; 6(2): 256–61.
4. Princy RJ, Parthasarathy S, Jose PS, Lakshminarayanan AR, Jeganathan S. Prediction of cardiac disease using supervised machine learning algorithms. In 2020 IEEE 4th international conference on intelligent computing and control systems (ICICCS). 2020 May 13; 570–575.
5. Deepika P, Sasikala S. Enhanced model for prediction and classification of cardiovascular disease using decision tree with particle swarm optimization. In 2020 IEEE 4th international conference on electronics, communication and aerospace technology (ICECA). 2020 Nov 5; 1068–1072.
6. Grampurohit S, Sagarnal C. Disease prediction using machine learning algorithms. In 2020 IEEE international conference for emerging technology (INCET). 2020 Jun 5; 1–7.
7. Ratnakar S, Rajeswari K, Jacob R. Prediction of heart disease using genetic algorithm for selection of optimal reduced set of attributes. Int J Adv Comput Eng Netw. 2013 Apr; 1(2): 51–5.
8. Liang H, Tsui BY, Ni H, Valentim CC, Baxter SL, Liu G, Cai W, Kermany DS, Sun X, Chen J, He L. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019 Mar; 25(3): 433–8.
9. Ahamed BS, Arya MS, Nancy V AO. Prediction of type-2 diabetes mellitus disease using machine learning classifiers and techniques. Front Comput Sci. 2022 May 10; 4: 835242.
10. Arumugam K, Naved M, Shinde PP, Leiva-Chauca O, Huaman-Osorio A, Gonzales-Yanac T. Multiple disease prediction using Machine learning algorithms. Mater Today: Proc. 2023 Jan 1; 80: 3682–5.
11. Xie S, Yu Z, Lv Z. Multi-disease prediction based on deep learning: a survey. Comput Model Eng Sci. 2021 Aug 10; 128(2): 489–522.

