
MULTIPLE DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

J. Fasila Afreen*¹

*¹Student, Department Of CSE, Kings College Of Engineering, Pudukkottai,
Tamil Nadu, India.

DOI : <https://www.doi.org/10.56726/IRJMETS35252>

ABSTRACT

Artificial intelligence and machine learning play a significant part in today's society. We can find them everywhere, from autonomous vehicles to the medical industry. Patient data generated by the medical sector is produced in enormous quantities and is processed in numerous ways. Hence, using machine learning, we were able to develop a prediction system that can identify multiple diseases simultaneously. Several of the currently used methods are only able to predict one disease at a time, and that too with reduced accuracy. Poor accuracy can pose a major threat to a patient's health. Heart, liver, and diabetes are the three ailments we are now thinking about, but there may be many more added in the future. The system will output if the user has the disease or not when the user enters several disease-related parameters. Many people can benefit from this project because it allows for the monitoring of a person's health and the taking of the necessary actions to prolong life.

Keywords: Healthcare, Clinical Decision Support, Artificial Intelligence, Deep Learning, Machine Learning, Diseases Diagnosis, Medical Image Processing.

I. INTRODUCTION

Data is a valuable resource in this digital age, and vast amounts of data were produced across all industries. All patient-related data is included in the healthcare sector's data. Here, a general architecture for disease prediction in the healthcare sector has been put out. Several of the current models focus on just one disease for each analysis. One analysis, for example, might be performed for skin illnesses, cancer, and diabetes. There isn't a mechanism in place that can analyze multiple diseases at once. As a result, we are focusing on providing users with rapid and precise disease predictions based on the symptoms they enter. Hence, using Django, we propose a method that may be used to forecast various diseases. We will examine the diseases of malaria, diabetes, and the heart in this system. Later, many more illnesses could be added. We'll utilize machine learning techniques and Django to put numerous disease prediction systems into action. In order to preserve the model's behavior, Python pickling is employed. The significance of this system analysis is that it considers all the factors that contribute to the development of the diseases under study, making it possible to detect them more effectively and precisely. A python pickle file will be used to store the behavior of the final model.

II. METHODOLOGY

A multiple disease prediction system using machine learning methodology could be built using supervised learning algorithms such as Support Vector Machines (SVM), Decision Trees, and Random Forests. The system would be trained on a dataset containing medical records of patients with various diseases. Features such as patient age, gender, medical history, lifestyle, and environmental factors would be used as inputs to the model. After the model has been trained, it can be used to predict which disease a patient is likely to have based on the available data. The model can also be used to identify risk factors associated with a particular disease in order to help diagnose and treat it. The system could also utilize natural language processing to analyze text documents and extract relevant features. Finally, the system could use deep learning algorithms, such as convolutional neural networks, to identify complex patterns in large-scale medical data. Due to its correlation, we conducted experiments on three diseases: heart, diabetes, and liver. The first step is to import the UCI dataset, PIMA dataset, and Indian liver dataset, respectively, for the datasets for heart disease, diabetes disease, and liver disease. After the dataset has been loaded, each inputted piece of data is visualized. Following pre-processing the data for visualization, which involves looking for outliers, missing values, and scaling the dataset, the data is divided into training and testing. Subsequently, we used the knn, xgboost, and random forest algorithms on the training dataset before applying what we learned about the classification method to the

testing dataset. We will use our information to select the algorithm that will produce the highest accuracy for each of the disease.

Decision Tree

This particular supervised machine learning technique focuses primarily on categorization issues. Creating a training model that may be used to forecast the class or values of the desired value is the primary goal of employing decision trees. To do this, basic decision procedure inference is learned from previously collected data (training data). Using the decision tree approach, we begin by predicting the class at the tree's root. We combine the root characteristic's values with the trait of the data. We move to the next node with the branch that is parallel to that value on the basis of differentiation. The decision Tree divides the symptoms in this system according to their classification, which reduces the complexity of the dataset.

Random Forest

A random forest can be constructed by merging N decision trees, and then it can be used to make predictions for each tree that was produced in the first step. Supervised machine learning algorithms include Random Forest. Although it is utilized for both classification and regression, classification problems are its main focus. The use of Random Forest is fairly simple, as is the implementation. If we need to create a model quickly, Random Forest is the ideal replacement. A large number of decision trees are built during the training phase of Random Forest, an ensemble learning technique. By using voting, it determines which answer is best. Several decision trees make up Simple Random Forest. A forest of trees is produced. The accuracy rate is precisely related to the number of trees in the forest, which also solves the overfitting issue. Because Random Forest is not dependent on overfitting and is insensitive to dataset noise, it performs well when applied to real-world issues. It is very effective and offers superior performance to other tree-based algorithms. Bootstrap aggregation or bagging is commonly used for tree learning.

KNN

KNN is a supervised learning algorithm that classifies fresh data points based on metrics of similarity. It is a non-parametric method, which means that it makes no assumptions about the underlying data. KNN is a lazy learner approach as well because it doesn't do any training and just keeps the training dataset without trying to derive any discriminative functions from it. KNN is based on the concept of the feature similarity approach, which assumes that homogenous objects exist nearby. Also, it is an approach for instance-based learning in which the function is roughly local. This approach is utilized when there are non-linear decision-making divisions between classes and it can handle vast amounts of data.

XGBoost

XGBoost is an open-source machine-learning library for gradient boosting. It is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost is used for supervised learning tasks such as classification and regression. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solves many data science problems in a fast and accurate way. XGBoost is an implementation of gradient-boosted decision trees designed for speed and performance. It is a scalable, portable, and accurate library for large-scale machine learning. XGBoost uses a more regularized model formalization to control over-fitting, which gives it better performance. It also has tools for parallel processing, allowing it to exploit the power of multiple computers to perform computations in a fraction of the time. It has a wide range of hyperparameters that can be tuned to improve the model's performance.

MODELING AND ANALYSIS

The patient can foresee the disease using the system. The user adds the input for the specific disease, and the output is displayed based on the trained model of the user input.

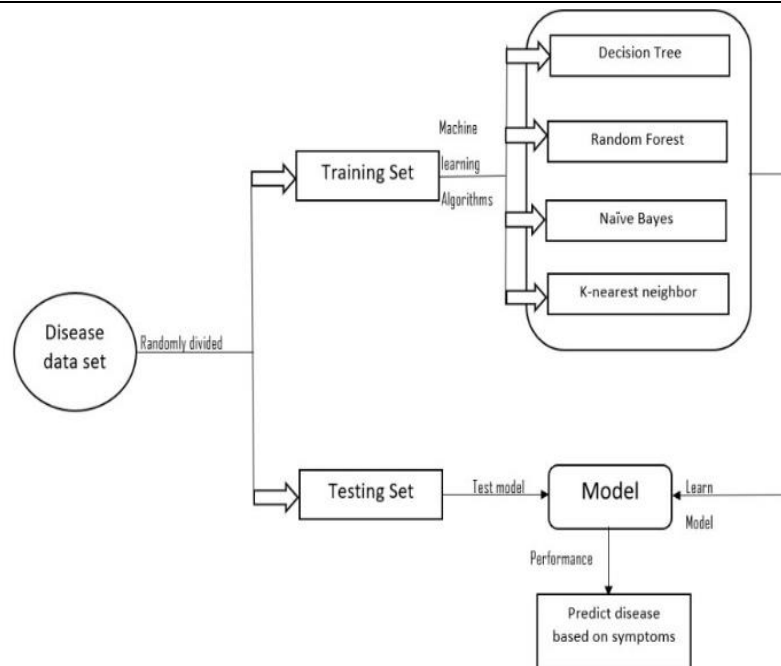


Figure 1: System Architecture.

III. RESULTS AND DISCUSSION

The fields of disease diagnosis and prediction have a great deal of promise to be revolutionized by DL and ML techniques. The accuracy and correctness of the diagnosis are the most important aspects of the treatment procedure when it comes to sickness. AI has demonstrated significant accuracy in the diagnosis of diseases based on images as well as in the forecasting of treatment results in terms of survival rate and response to treatment. The knn algorithm, the xgboost algorithm, and the random forest algorithm were utilized in the system's diabetic disease prediction model since they each provided the highest levels of accuracy. When the patient adds the disease-specific parameter, it will indicate whether or not the patient has the ailment in question. The parameters will display the necessary value range, and if the value is outside of that range, is invalid, or is empty, a warning sign advising the user to input a valid value will be displayed.

IV. CONCLUSION

The primary goal of this study was to develop a system that could accurately predict several diseases. The user doesn't have to navigate via several websites thanks to this project, which also saves time. Early diagnosis of diseases can both lengthen your life and spare you from financial hardship. To achieve the highest level of accuracy, we have used a variety of machine learning algorithms, including Random Forest, XGBoost, and K nearest neighbor (KNN).

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my college, Kings College of Engineering, for providing a platform for me to create a project on the theme of "Multiple Disease Prediction System". I greatly appreciate Prof. Dr. S. Kannan for serving as my mentor. Furthermore, without the cooperation, advice, and assistance of my friends and family, the complexity of this study article would not have been achievable.

V. REFERENCES

- [1] T Xia C, Li X, Wang X, Kong B, Chen Y, Yin Y, et al. A multi-modality network for cardiomyopathy death risk prediction with CMR images and clinical information in lecture notes in computer science. Cham: Springer; 2019.
- [2] Wu X, Liu X, Zhou Y. Proceedings of 2021 Chinese intelligent systems conference: a review of unsupervised learning techniques in lecture notes in electrical engineering. Singapore: Springer; 2022.
- [3] Dhal P, Azad C. A comprehensive survey on feature selection in the various fields of machine learning. Appl Intell. 2022;52(4):4543.

- [4] Xi X, Meng X, Yang L, Nie X, Yang G, Chen H, Fan X, Yin Y, Chen X. Automated segmentation of choroidal neovascularization in optical coherence tomography images using multi-scale convolutional neural networks with structure prior. *Multimed Syst.* 2019;25(2):95.
- [5] Rizk Y, Hajj N, Mitri N, Awad M. Deep belief networks and cortical algorithms: a comparative study for supervised classification. *App Comput Inform.* 2019;15(2):81.
- [6] Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A. The state of the art of deep learning models in medical science and their challenges. *Multimedia Syst.* 2021;27(4):599.
- [7] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large-scale visual recognition challenge. *Int J Comput Vision.* 2015;115(3):211.
- [8] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM.* 2017;60(6):84.
- [9] Owais M, Arsalan M, Choi J, Mahmood T, Park KR. Artificial Intelligence-based classification of multiple gastrointestinal diseases using endoscopy videos for clinical diagnosis. *J Clin Med.* 2019;8(7):986.
- [10] Nithya A, Appathurai A, Venkatadri N, Ramji DR, Palagan CA. Kidney disease detection and segmentation using artificial neural networks and multi-kernel k-means clustering for ultrasound images. *Measurement.* 2020;149: 106952.
- [11] Abedi V, Khan A, Chaudhary D, Misra D, Avula V, Mathrawala D, et al. Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework. *Therap Adv Neurol Disorders.* 2020;13:1756286420938962.
- [12] Gorris M, Hoogenboom SA, Wallace MB, van Hooft JE. Artificial intelligence for the management of pancreatic diseases. *Digestive Endoscopy.* 2021;33(2):231.
- [13] Sinagra E, Badalamenti M, Maida M, Spadaccini M, Maselli R, Rossi F, et al. Use of artificial intelligence in improving adenoma detection rate during colonoscopy: might both endoscopists and pathologists be further helped. *World J Gastroenterol.* 2020;26(39):5911.
- [14] Xu L, Gao J, Wang Q, Yin J, Yu P, Bai B, et al. Computer-aided diagnosis systems in diagnosing malignant thyroid nodules on ultrasonography: a systematic review and meta-analysis. *European Thyroid J.* 2020;9(4):186.
- [15] Bharti R, Khamparia A, Shabaz M, Dhiman G, Pande S, Singh P. Prediction of heart disease using a combination of machine learning and deep learning. *Comput Intell Neurosci.* 2021;2021:8387680.
- [16] Suresha PB, Wang Y, Xiao C, Glass L, Yuan Y, Clifford GD. A deep learning approach for classifying nonalcoholic steatohepatitis patients from nonalcoholic fatty liver disease patients using electronic medical records. In: *Explainable AI in Healthcare and Medicine.* Switzerland: Springer; 2021.
- [17] Goenka N, Tiwari S. Deep learning for Alzheimer prediction using brain biomarkers. *Artif Intell Rev.* 2021;54(7):4827.
- [18] Ramani V, Shendure J. Smash and DASH with Cas9. *Genome Biol.* 2016;17:42.
- [19] Rajaraman S, Candemir S, Kim I, Thoma G, Antani S. Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs. *Appl Sci.* 2018;8(10):1715.
- [20] Depeursinge A, Chin AS, Leung AN, Terrone D, Bristow M, Rosen G, et al. Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution computed tomography. *Invest Radiol.* 2015;50(4):261.
- [21] Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imag.* 2016;35(5):1207.
- [22] Dabowska NI, Amaitik NM, Maatuk AM, Aljawarneh SA. A hybrid intelligent system for skin disease diagnosis in the international conference on engineering and technology (ICET), Antalya Turkey, 2017.
- [23] Jader R, Aminifar S. Fast and accurate artificial neural network model for diabetes recognition. *NeuroQuantology.* 2022;20(10):2187.
- [24] Momin MA, Bhagwat NS, Dhiwar AV, Chavhate SB, Devekar NS. Smart body monitoring system using IoT and machine learning. *Int J Adv.*