RESEARCH ARTICLE **SPECIAL ISSUE**

International Research Journal on Advanced Science Hub
2582-4376
Vol. 05, Issue 05S May
www.rspsciencehub.com

RSP Science Hub

http://dx.doi.org/ 10.47392/IRJASH.2023.S027

International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER) 2023

Multimodal Disease Prediction using Machine Learning and Deep Learning Techniques

Akil Arsath J [1], *Suganthi S* [1], *Rakeshwaran S* [1], *Karthiga S* [2]
[1] *Department of Information Technology, Thiagarajar College of Engineering, Madurai, India*
[2] *Assistant Professor, Department of Information Technology, Thiagarajar College of Engineering, Madurai, India*

Email: akil.j.arsath@gmail.com

Article History

Received: 28 February 2023
Accepted: 16 March 2023

Keywords:
Disease prediction;
Ensembling model;
CNN;
Tensor flow;
Deep learning

Abstract

Good health is man's greatest possession but in today's world people get a lot of diseases because of several reasons. The ability to predict diseases accurately is a critical aspect of healthcare. Machine learning techniques are increasingly being used to improve disease prediction. In this paper, we present a multi-disease prediction system that uses machine learning and deep learning algorithms to predict the likelihood of several common diseases. Even Though there are a lot of algorithms and techniques to predict a disease, there is no proper system to identify multiple diseases in a single system. Hence this paper focuses on the prediction of multiple diseases using machine learning and deep learning algorithms. Our aim is to build a model which efficiently predicts diseases such as kidney, heart and diabetes, malaria using machine learning and deep learning algorithms. This helps to make a better prediction of disease. For accurate prediction we are going to use stacking and ensembling models which help to increase the accuracy of the model. We are going to implement all these models in flask web application.

1. Introduction

In the modern era, the modern computing has rapidly increased in the field of healthcare. With the advent of digital technology and the increasing availability of healthcare data, researchers are exploring ways to leverage machine learning to improve healthcare outcomes. One of the key area in the research is the prediction of multiple diseases using machine learning algorithms. In this paper, we present a novel approach to predict multiple diseases simultaneously using machine learning and deep learning techniques. Our approach utilizes a variety of patient data, including demographic information, medical history, and laboratory test results, to train a machine learning model to predict the occurrence of multiple diseases for a given patient. The ability to predict multiple diseases for a patient can provide way for health care workers, allowing them to understand about the disease. Additionally, our method helps to identify patients who has multiple diseases, enabling early intervention and potentially reducing healthcare costs.

In this paper, we describe our approch for predicting multiple diseases using machine learning, and present our experimental results on a real-world healthcare dataset. We also discuss the potential implications of our approach for improving healthcare outcomes, as well as future directions for research in this area. This paper focuses on predicting several diseases which have a huge impact on human life. Here are some statistics about diabetes-related deaths according to the World Health Organization (WHO), diabetes was the one of the dangerous disease worldwide, responsible for an estimated 1.6 million deaths and there were an estimated 229 million cases of malaria worldwide in 2019, and an estimated 409,000 deaths from malaria and also they have estimated that 17.9 million people died from cardiovascular disease in 2019, representing 32% of all global deaths. In the United States, heart disease is the leading cause of death, responsible for an estimated 655,381 deaths in 2018. Hence it is important to identify the type of the disease in the earlier stage to prevent the serious condition. We have implemented different models to predict the different kinds of diseases. For model training we have used various classification techniques such as decision tree, logistic regression, k-nearest neighbors, random forest and we have used ensembling models such as bagging to achieve higher accuracy. To predict malaria from the image, deep learning model such as CNN is used to train a image and pertained model such as VGG16 is used to get an higher accuracy. The dataset is taken from the various sources such as kaggle, heart disease ucl and it is preprocessed and then trained with a different set of algorithms. With the help of a flask application, the data is taken from the user and the different kinds of disease are predicted.

2. Related works

In contemporary times, people are susceptible to diverse diseases caused by environmental factors and lifestyle choices. Hence, early prediction of diseases has become a crucial task. In this context, Dhiraj Dahiwade (Dahiwade, Patle, and Meshram) proposed a general disease prediction model based on the patient's symptoms. This model employs KNN and CNN machine learning algorithms for precise disease prediction. The study reveals that the accuracy of general disease prediction using CNN is 84.5%, surpassing the KNN algorithm. Tanner Christensen (Christensen et al.) presented a approach that combines natural language word embedding and network modeling techniques to generate meaningful representations of medical concepts using the weighted network adjacency matrix in the GloVe algorithm, called Code2Vec. Ch Raja Shaker (Shaker et al.) suggested a heart disease prediction model that employs the PCA approach to determine the relative importance of each of the 11 fields in the dataset. The use of sample approaches improved accuracy and recall rates.

Akkem Yaganteeswarudu (Yaganteeswarudu) introduced a system designed to predict multiple diseases using the Flask API. This article specifically analyzes Diabetes, Diabetes Retinopathy, Heart disease, and breast cancer. Furthermore, the system has the potential to include other diseases such as skin

diseases, fever analysis, and many more. The implementation of multiple disease analysis is facilitated by employing machine learning algorithms, TensorFlow, and the Flask API. Ayman Mir (Mir and Dhage) developed a classifier model for diabetes prediction using WEKA tool and evaluated the performance of Naive Bayes, Support Vector Machine, Random Forest, and Simple CART algorithms. The study recommends Support Vector Machine as the most efficient algorithm for predicting diabetes disease with the highest accuracy.

Rohit Kumar (Kumar, Thakur, and Chauhan) presented a system that assesses the symptoms provided as input and produces the disease as output. The system utilizes the Naive Bayes Classifier algorithm. The primary emphasis of this research is on accuracy, as the number of symptoms provided by the individual as input directly impacts the accuracy of the disorder prediction as output. Kuldeep Vayadande (Vayadande et al.) introduced a heart disease model that employs both machine learning and deep learning algorithms. The model was implemented on a Kaggle dataset consisting of 303 rows and 14 attributes. The utilized algorithms include Logistic Regression, NB, K-NN, SVM, Multi-Layer Perceptron's, Artificial Neural Network, Decision Tree, Random Forest, XG Boost, and Cat Boost.

Octave Iradukunda (Sivaranjani et al.) proposed a malaria disease prediction model using the Extreme Learning Machine (ELM) algorithm, a single hidden layer feedforward neural network methodology. ELM was compared to other machine learning techniques, including SVM, KNN, CART, RF, CNN, VGG16, RESNET, and DENSENET.

S Sivaranjani (Roobini, Reddy, et al.) presented a diabetes prediction work that utilized Support Vector Machine (SVM) and Random Forest (RF) machine learning algorithms to identify the likelihood of developing diabetes-related diseases. The data was pre- processed, and stepwise feature selection was used to choose features that impact the prediction. Principle Component Analysis (PCA) dimensionality reduction was applied after selecting specific features, and the prediction accuracy was 83% with RF, compared to 81.4% with SVM. M.S. Roobini (Raundale, Thosar, et al.) proposes a method for Parkinson's disease detection using audio signal

classification with machine learning classifiers. Provision regression and XGBoost classifiers are utilized along with an audio feature dataset from the UCI repository.

Pooja Raundale (Narayanan et al. Sivaranjani et al. Deeba, Patil, et al.) proposed a methodology for predicting Parkinson's disease severity using deep neural networks on UCI's Parkinson's Telemonitoring Vocal Data Set. They developed a neural network and a machine learning model using Random Forest Classifier to classify the disease. Srikanth Narayanan (Deeba, Patil, et al.) proposes a system for predicting multiple diseases, including Malaria, Alzheimer's Disease, Tuberculosis, and Pancreatic Cancer. The system utilizes CNN, Random Forest, and Logistic Regression algorithms for analysis and prediction. Farha Deeba [13] proposed a disease prediction system that utilizes machine learning algorithms on a dataset of 4922 patients with 132 symptoms to predict 42 common diseases. The algorithms found for implementation are Decision Tree, Random Forest Classifier, and Naive Bayes Classifier.

Priyanka Sonar (Sonar and Jayamalini) developed a diabetes prediction model using Decision Tree, ANN, Naive Bayes, and SVM algorithms. The model achieved 85% precision for Decision Tree, 77% for Naive Bayes, and 77.3% for Support Vector Machine. The results indicate that the methods have a high level of accuracy. C.S. Manikandababu (Manikandababu et al.) used machine learning algorithms to predict type 2 diabetes in patients. The PIMA Indian Diabetic Dataset, which contains 8 parameters for 768 patients, was used in this study. Logistic regression classification algorithm was employed, and an ensemble model was created using stacking technique to improve performance and accuracy. Based on the comprehensive literature survey, it can be inferred that the majority of prior research has concentrated on addressing singular diseases using singular approaches. However, in this study, we have adopted a multifaceted approach to tackle this issue. To achieve this, various machine learning (ML) and deep learning (DL) techniques were employed, including bagging, boosting, and a stacked model comprising all existing ML models. By utilizing these diverse techniques, we aimed to enhance the accuracy and robustness of disease prediction, ultimately leading to improved clinical outcomes. Our findings provide valuable insights

into the development of effective and efficient disease prediction models for clinical applications. These results can have significant implications for healthcare professionals, policymakers, and patients alike, by enabling early detection and timely intervention, which can improve patient outcomes and reduce healthcare costs.

3. Materials and Methods

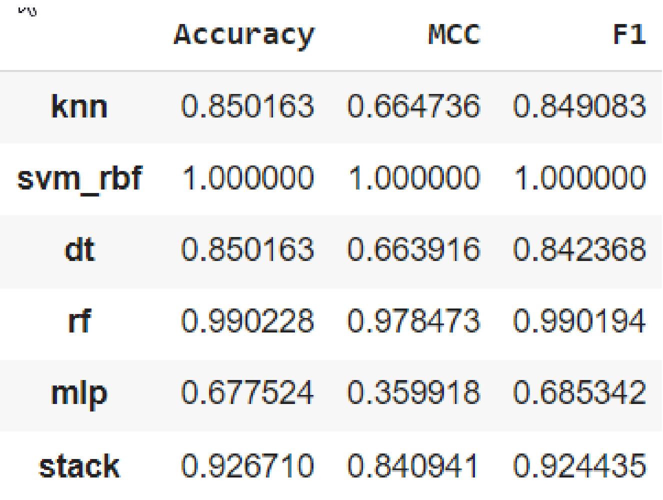
Our aim is to make a model to predict multiple diseases and also allow users to get predictions about the different diseases individually. One of the major disadvantages of the available work is, it is based on the simple machine learning algorithms which makes it difficult to fit into dataset and makes a model to perform poor on the testing set. We have implemented the ensemble model with different set of algorithm such as k-nearest neighbor, decision tree, logistic regression and the support vector machine and combined it as a stacked model which performs comparatively well on the testing dataset. Overall approach is collecting data from the kaggle and to preprocess the data such as feature selection, feature extraction and data transformation and then we transform the attribute into required types to train a model. Our paper focuses on three different disease such as diabetes, heart disease, parkinson disease prediction. Currently for diabetes prediction, many classification techniques are used that performs well on the training set but it does not performs well on the testing set. Hence we have used ensembling techniques for diabetes disease prediction we have used pima indian disease dataset and we got an accuracy of 93%. For heart disease prediction we have used Heart disease dataset from machine learning ucl repository which gives an accuracy of 99%. Similarly for parkinson disease prediction we have used dataset from kaggle and we have trained a model with support vector machine algorithm and got an overall accuracy of 90%. For malaria prediction, we have used Malaria Cell Images Dataset from kaggle which contains 27,558 images. It extracts important features in the image and trained by CNN for detecting malaria disease. The image is taken from the user and that is fed into the convolution layer. The max pooling layer and padding is used to extract the important features of the disease.

Nine percent of the data is used for the testing and the remaining images are used to train the model. The image size is reduced into the uniform size to

train a model. Flask API is used as a backend for the web application. We used pickle to load the model and from the user input, our model able to predict the disease.

4. Results and Discussion

This paper centers on forecasting multiple diseases, with particular attention given to the Diabetes Prediction task. The performance measure of various machine learning algorithms in this task is illustrated in Figure 1 and 2. Based on the findings, it is evident that the stacked model and random forest algorithms exhibit superior performance in predicting diabetes.



	Accuracy	MCC	F1
knn	0.850163	0.664736	0.849083
svm_rbf	1.000000	1.000000	1.000000
dt	0.850163	0.663916	0.842368
rf	0.990228	0.978473	0.990194
mlp	0.677524	0.359918	0.685342
stack	0.926710	0.840941	0.924435

FIGURE 1. Performance metrics

Figure 3 and 4 illustrates the performance metric of a machine learning (ML) model in predicting heart disease. Specifically, the stacked model exhibits outstanding performance, surpassing that of existing traditional models with a maximum achieved accuracy of 94%. In light of these findings, we infer that the stacked model outperforms its traditional counterparts.

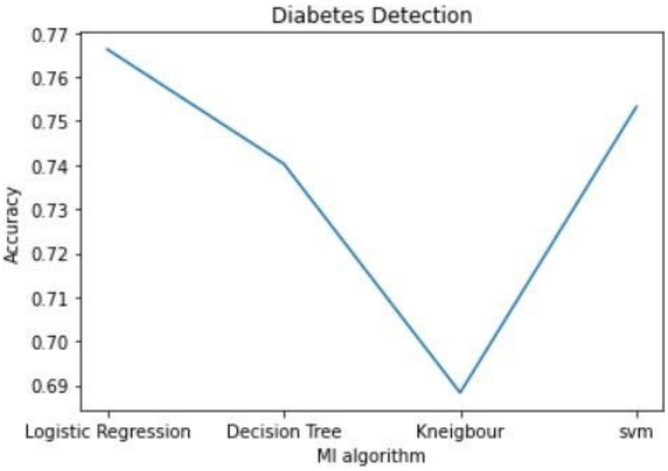
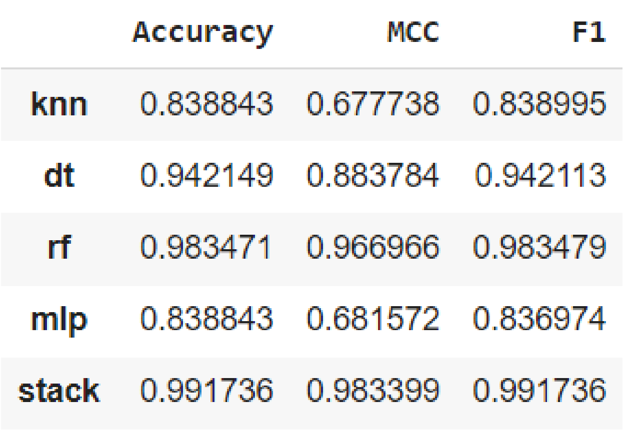


FIGURE 2. Diabetes prediction with various ml model



	Accuracy	MCC	F1
knn	0.838843	0.677738	0.838995
dt	0.942149	0.883784	0.942113
rf	0.983471	0.966966	0.983479
mlp	0.838843	0.681572	0.836974
stack	0.991736	0.983399	0.991736

FIGURE 3. Heart disease prediction - Performance metrics

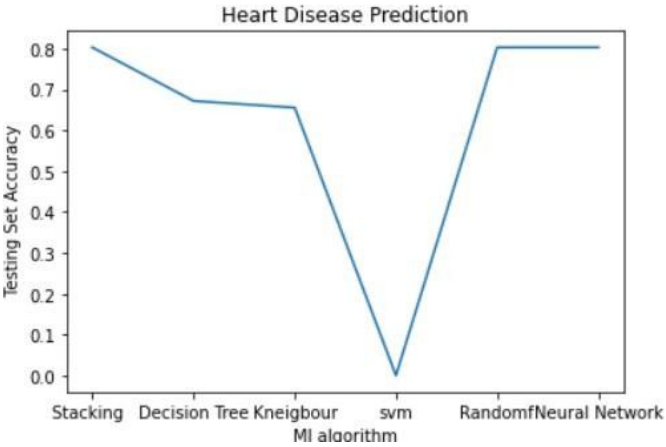


FIGURE 4. Heart disease prediction with various ML model

The results presented in Figure 5 showcase the performance metric of a support vector machine (SVM) model in predicting Parkinson's disease. To train the model, we implemented the K-fold cross-validation technique, resulting in a promising accuracy of 9.3% on the training set. However, the model's performance on the validation set did not meet our expectations.

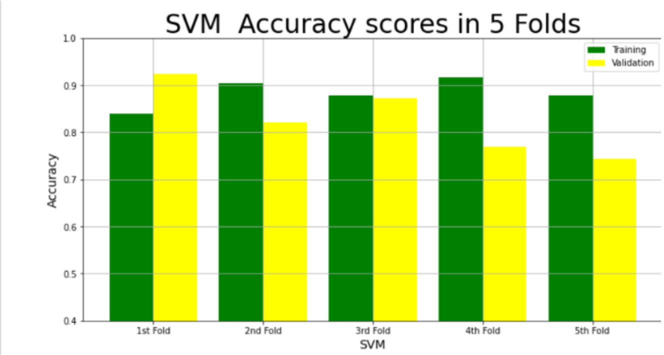


FIGURE 5. Parkinson disease prediction using SVM

Figure 6 and 7 depicts the accuracy of malaria prediction using CNN architecture. The model got accuracy metrics of 96% on training data and 95%percent on testing data.

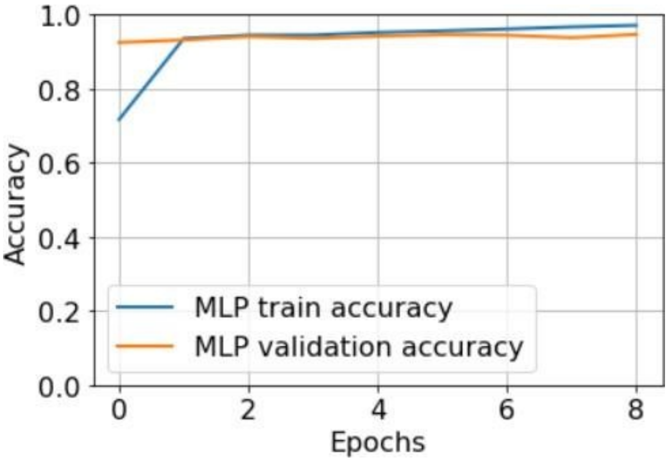


FIGURE 6. Malaria prediction using CNN architecture

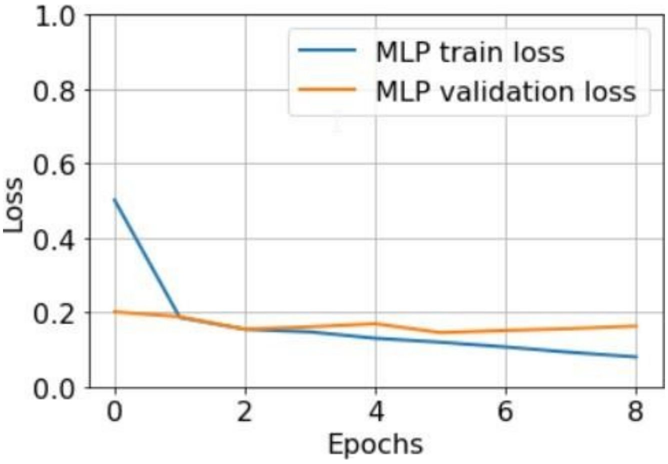


FIGURE 7. Malaria prediction using CNN architecture

FIGURE 8. Malaria prediction webpage

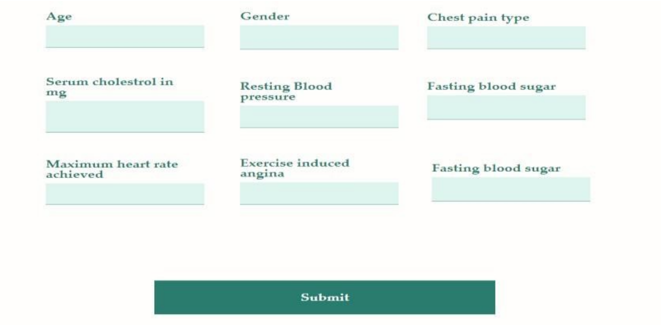


FIGURE 9. Heart disease Prediction

5. Conclusion

Multimodal disease prediction using machine learning and deep learning algorithms has the potential to revolutionize the healthcare industry by improving disease diagnosis, treatment, and patient outcomes. Our methodology demonstrates that integrating both ML and DL techniques can lead to more accurate and reliable disease predictions across multiple diseases. By leveraging large datasets and advanced algorithms, we can achieve high precision in identifying disease symptoms and risk factors, as well as predicting disease progression and treatment response. Our results highlight the potential

FIGURE 10. **Parkinson disease Prediction**

of ML and DL in improving early disease detection and personalized treatment plans, ultimately leading to better patient outcomes and reduced healthcare costs. The application of ML and DL in healthcare is still in its early stages, and there is much room for further research and development. However, our findings demonstrate the potential of these technologies to transform the healthcare industry and improve the lives of millions of patients around the world.

6. Authors' Note

We declare that there is no conflict of interest regarding the publication of this article. We confirmed that the paper is free of plagiarism.

References

Christensen, Tanner, et al. "Machine Learning Methods for Disease Prediction with Claims Data". *2018 IEEE International Conference on Healthcare Informatics (ICHI)* (2018). 10.1109/ICHI.2018.00108.

Dahiwade, Dhiraj, Gajanan Patle, and Ektaa Meshram. "Designing Disease Prediction Model Using Machine Learning Approach". *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (2019). 10.1109/ICCMC.2019.

Deeba, Farha, Siddarama R Patil, et al. "Utilization of Machine Learning Algorithms for Prediction of Diseases". *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)* (2021). 10.1109/i-%20PACT52855.2021.

Kumar, Rohit, Prince Thakur, and Sps Chauhan. "Special Disease Prediction System Using Machine Learning". *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)* (2022). 10.1109/COM-%20IT-CON54601.2022.

Manikandababu, C S, et al. "Prediction of Diabetes using Machine Learning". *2022 International Conference on Edge Computing and Applications (ICECAA)* (2022). 10.1109/ICECAA55415.2022.

Mir, Ayman and Sudhir N Dhage. "Diabetes Disease Prediction Using Machine Learning on Big Data of Healthcare". *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)* (2018). 10.1109/ICCUBEA.2018.

Narayanan, N M Srikanth, et al. "Leveraging Machine Learning Methods for Multiple Disease Prediction using Python ML Libraries and Flask API". *Journal of International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (2022). 10.1109/ICAAIC53929.2022.

Raundale, Pooja, Chetan Thosar, et al. "Prediction of Parkinson's disease and severity of the disease using Machine Learning and Deep Learning algorithm". (2021). 10.1109/INCET51464.2021.

Roobini, M S, Yaragundla Rajesh Kumar Reddy, et al. Udayagiri Sushmanth Girish, 2022.

Sivaranjani, S, et al. "Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction". (2021). 10.1109/ICACCS51430.2021.

Sonar, Priyanka and K Jayamalini. “Diabetes Prediction Using Different Machine Learning Approaches”. *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (2019). 10.1109/ICCMC.2019.

Vayadande, Kuldeep, et al. “Heart Disease Prediction using Machine Learning and Deep Learning Algorithms”. *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)* (2022). 10.1109/CISES54857.2022.

Yaganteeswarudu, Akkem. “Multi Disease Prediction Model by using Machine Learning and Flask API”. *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (2020). 10.1109/ICCES48766.2020.

Embargo period: The article has no embargo period.

To cite this Article: , Akil Arsath J, Suganthi S , Rakeshwaran S , and Karthiga S . “Multimodal Disease Prediction using Machine Learning and Deep Learning Techniques.” International Research Journal on Advanced Science Hub 05.05S May (2023): 206–212. http://dx.doi.org/10.47392/IRJASH.2023.S027