International Journal of Research in Engineering, Science and Management
Volume 5, Issue 6, June 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

Online Intelligent Smart Health Prediction Using Machine Learning

Pola Aashray[1*], Maryala Manish[2], Bingi Abhishek[3], Mohan Dholvan[4]
[1,2,3]*Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, India*
[4]*Head of the Department, Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, India*

Abstract*: There are a number of single illness prediction ML initiatives that are primarily focused on prediction, making it difficult for users to use. To address this issue, we developed this sophisticated system that can forecast multiple diseases and has a user-friendly interface. During a pandemic, most people would prefer to be treated at home as much as possible to avoid polluted locations such as hospitals and clinics. The purpose of our project is to establish a user-friendly platform for cross-validating results while on the road, as well as to promote general awareness and provide preventative measures [11]. Many deaths are caused by diseases such as diabetes, lung cancer, and heart-related ailments around the world; nevertheless, the bulk of deaths occur due to a lack of regular health exams [12]. Because of a lack of medical infrastructure and a low doctor-to-population ratio, the following problem occurs. According to WHO guidelines, the doctor-to-patient ratio should be 1:1000, however India's doctor-to-population ratio is 1:1457, indicating a doctor deficit. The purpose of this research is to apply artificial intelligence (AI) to forecast serious ailments such kidney infection, diabetes, liver disease, pneumonia, Parkinson's disease, Covid-19, malaria, and cellular breakdown in the lungs. To make this work more consistent and accessible to the general public, our group designed an online application that uses AI to generate disease forecasts. The major purpose of this project is to create an online application that uses artificial intelligence to anticipate the diseases stated. There would be a significant danger to our humanity. If diseases such as heart disease, kidney disease, pneumonia, diabetes, malaria, and other maladies are not recognised and prevented at an early stage, they might lead to death. Many lives can be saved by early detection and prevention of these disorders [10]. The Random Forest Classifier and Convolution Neural Network algorithms are used in the Online Intelligent Health Care system to predict illness risk levels.

***Keywords*: Neural Network, Convolution, Parkinson's, WHO, Classifier.**

1. Introduction

Online Intelligent Smart Health Prediction is a website that predicts the user's ailment based on the symptoms provided by the user. The Intelligent Health Care system contains data sets gathered from several health-related websites. The user will be able to determine the likelihood of a condition based on the symptoms provided by Intelligent Health Care. This initiative will aid in the prediction of nine different types of sickness on a single platform.

As civilization progresses, people's quality of life patterns and environmental situations are gradually changing, increasing people's hidden hazards from a variety of ailments. Liver health, cardiovascular disease, vision impairment, diabetes, cancer, and other major diseases have a global impact. Diabetes affects 421 million people globally, with Type 2 diabetes accounting for over 90% of all cases. [13] Due to cardiac senescence and loss of function, the risk of heart disease rises with age. [8] Heart disease is responsible for more than 32% of all fatalities worldwide. These major diseases have a significant impact on people's health and productivity. Disease prediction's main purpose is to predict the likelihood of an individual having a given disease in the future. There are numerous influencing elements that must be considered for various diseases in different populations.[14] Including a set of features with exceptionally broad dimensions that must be detected, as well as complex and varied individual and disease differences. Simply wishing to complete these activities manually is not only difficult, but it also takes a significant amount of human and financial resources.

Deep learning concepts such as probability, backpropagation, overfitting, multi-layered perceptron, gradient descent, python syntax, various libraries such as Tensorflow, as well as Machine Learning concepts like Preprocessing Techniques, Exploratory Data Analysis, Handling Imbalanced Data Techniques, HyperParameter Tuning Techniques, and so on, were used to develop the project.

In our project, we use two algorithms to predict diseases: CNN (Convolutional Neural Network) and RFC (Random Forest Classifier). HTML, Bootstrap, Cascading Style Sheets, and JavaScript are used to create the web application's front-end. Flask and Python programming are used to build the system's back-end.

2. Literature Survey

Prediction of Chronic Kidney Disease Using Machine Learning Algorithm [1]

Authors: Siddheshwar Tekale, Pranjal Shingavi, Sukanya Wandhekar, Ankit Chatorikar

Summary: Everyone is aiming to be health-conscious in today's environment, but due to work and a hectic schedule, one only pays attention to one's health when symptoms appear.

*Corresponding author: aashraypolaa@gmail.com

However, because CKD is a disease with no symptoms or, in some cases, no illness-specific indicators, it can be difficult to predict, detect, and prevent, which can result in long-term health consequences. Machine learning, on the other hand, may be a solution to this issue because it excels at prediction and analysis. Data from CKD patients with 14 features and 400 records was analyzed using several machine learning techniques such as Decision Tree, SVM, and others.[1] To create a model that can properly predict whether or not someone has CKD and, if so, how serious their condition is.

Prediction of Heart Disease Using Machine Learning [2]

Authors: Aditi Gavhane, Gouthami Kokkula, Isha Pandya, Kailas Devadkar

Summary: Heart disease is one of the most frequent diseases worldwide, and health care providers must collaborate to protect their patients' health and lives. In order to effectively categorize the heart disease dataset and/or anticipate heart disease occurrences with minimal features, the performance of various classifiers was investigated in this work. The healthcare industry has amassed a vast amount of personal information. This data collection supports smart decision-making. To achieve trustworthy results and develop acceptable data analyses, modern data mining techniques are required. To forecast the risk level of heart disease, a Heart Disease Prediction System is built using KNN, Random Forest Classifier, Decision Tree, Support Vector Machine and Logistic Regression techniques.

The Random Forest Classifier (RFC) and Support Vector Machine (SVM) have the highest accuracy of 90.30 percent, while logistic regression, K-Nearest Neighbor (K-NN), and decision tree have 87.10 percent, 83.86 percent, and 70.95 percent, respectively.[2]

Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm [3]

Authors: Samrat Kumar Dey, Ashraf Hossain, Md. Mahbubur Rahman

Summary: Diabetes is caused by the accumulation of too much sugar in the blood. It is now considered one of the world's most deadly diseases. This deadly disease affects people all around the world, whether they are aware of it or not. Diabetes can also cause heart attacks, paralysis, kidney failure, blindness, and other disorders. [3]

Several computer-based detection techniques for predicting and evaluating diabetes have been developed. The standard approach of discovering diabetic patients takes longer and costs more money. However, because of advances in machine learning, we now have the ability to create a solution to this difficult challenge. As a result, we developed a system that can predict if a patient has diabetes or not. The major purpose of this investigation is to develop a web application that improves forecast accuracy by utilizing a sophisticated machine learning algorithm. The Pima Indian benchmark dataset was used, which can predict the onset of diabetes based on diagnostics. With an accuracy rate of 82.35 percent, the Artificial Neural Network (ANN) displays a significant improvement in accuracy, prompting us to develop an Interactive Web Application for Diabetes Prediction. [3]

Designing Disease Prediction Model Using Machine Learning Approach [4]

Authors: Dhiraj Dahiwade, Gajanan Patle, Ektaa Meshram

Summary: People nowadays are susceptible to a wide range of diseases as a result of the environment and their lifestyle choices. As a result, anticipating disorder at an early stage is critical. Nonetheless, experts find it difficult to set precise expectations in light of negative effects. The most difficult test is predicting infection accurately. Information mining plays a key role in disease forecasting to address this problem. Every year, clinical science generates a massive amount of data. Because of the extended measure of information production in the clinical and medical services areas, the suitable analysis of clinical information has benefited from early understanding. In order to find stowed example data in vast amounts of clinical data, information mining is used. We proposed an approach for forecasting illness in general, based on the patient's side effects. We use AI computations such as K-Nearest Neighbor (KNN) and Convolutional Neural Network (CNN) for convincing sickness prediction. The collection of illness side effects is critical for infection prediction. In this broad sickness expectation, the individual's lifestyle proclivities and exam results are taken into account for a precise guess. The accuracy of CNN in general disease prediction is 84.5 percent, which is higher than the KNN technique. Furthermore, KNN takes longer and uses more memory than CNN.[4] This technique can assess the risk associated with general disease, which can range from low to high, based on general illness prediction

Development of machine learning model for diagnostic disease prediction based on laboratory tests [5]

Authors: Dong Jin Park, Min Woo Park, Homin Lee, Young-Jin Kim, Yeongsic Kim & Young Hoon Park

Summary: In medical science, deep learning and machine learning are rapidly being applied, notably in the worlds of visual, audio, and linguistic data. We sought to combine a DNN model with two ML models to construct a new optimized ensemble model for disease prediction using laboratory test data. 86 attributes (laboratory tests) were picked from datasets based on value counts, clinical importance-related factors, and missing values. We collected information on 5145 patients, including 326,686 laboratory test results. We looked into a total of 39 disorders using the International Classification of Disorders, 10th edition codes. TensorFlow-based light gradient boosting machines and extreme gradient boosting ML models, as well as a DNN model, were created using these datasets. The optimized ensemble model had an F1-score of 81 percent and a prediction accuracy of 92 percent for the five most common diseases. Deep learning and machine learning models had different prediction power and disease categorization patterns. Using a confusion matrix, the SHAP value approach was employed to analyze feature importance. Our unique ML model obtained great disease prediction efficiency by disease classification. This study will aid in disease detection and prediction. [5]

Efficient Automated Disease Diagnosis Using Machine Learning Models [6]

Authors: Naresh Kumar, Nripendra Narayan Das, Deepali

Gupta, Kamali Gupta, and Jatin Bindra

Summary: Using supervised learning models, a number of academics have recently created automated diagnosis methods. If certain disorders are detected early, the death rate from them can be lowered. This study develops an effective automated illness diagnostic model using machine learning techniques.[6] In this research, we've picked three key disorders: Diabetes, Heart Disease and Covid-19. The data is entered into an android app, the analysis is done in a real-time database with a pre-trained machine learning model trained on the same dataset and deployed in firebase, and the sickness detection result is displayed in the android app. Logistic regression is used to calculate the prediction. Early detection can help avoid diseases including Diabetes, Heart Disease and COVID-19. According to a comparative analysis, the proposed methodology can assist doctors in providing timely treatment.[6]

3. Problem Statement

The main goal is to develop a prediction engine that will allow users to diagnose covid-19, lung cancer, malaria, pneumonia, heart disease, diabetes, and kidney disease, to mention a few, from the comfort of their own homes. Unless the user has one of the ailments listed above, he does not need to consult a doctor for further treatment. The prediction engine needs a large dataset and effective machine learning techniques to forecast the presence of the disease. Pre-processing the dataset before using it to train machine learning models, such as removing redundant, null, or invalid data to improve prediction engine performance.

Doctors rely on common knowledge to treat patients. Studies are summarized after a given number of instances have been analyzed when there is a lack of common knowledge. This technique, however, takes time, whereas machine learning can find patterns more quickly. To apply machine learning, a large amount of data is required. There is very little data available depending on the illness.[9] Furthermore, the number of samples that are free of disease is significantly greater than the number of samples that have the disease. This research entails completing two case studies to assess the feasibility of several machine learning algorithms in assisting in the detection of such patterns, as well as establishing a platform to facilitate data exchange and collaboration.

4. Objective

The primary goal of this research is to analyze data from prestigious sources such as the University of California Irvine, Indian Liver Patient Records, National Institute of Health for malaria disease, lung cancer dataset, to name a few, using the Kaggle open-source platform. Predictions will be made using the Random Forest Classifier ML technique and Convolutional Neural Networks.

The secondary goal is to turn our model into a web application that uses Flask and Python for the backend and HTML, CSS, and JavaScript for the frontend, allowing users to predict diseases using the prediction engine.

5. System Analysis

A. *Existing System*

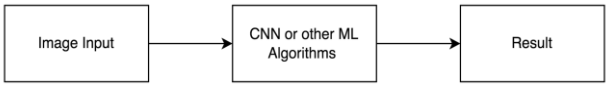

Fig. 1. Existing system flowchart

The model forecasts chronic diseases for a given region and population. Prediction of disease is only applicable to particular diseases. Big Data and the Convolution Neural Networks Algorithm are utilized in this strategy to forecast disease risk.[15] The disadvantage of this strategy is that it is limited to a single ailment or input type. Overflow might occur due to the inputs provided and the accuracy is not up to the mark for users to rely on this existing system. [14]

B. *Proposed System*

A new multimodal illness risk prediction algorithm is introduced for Convolution Neural Networks to address the shortcomings of the previous model. Data sets were gathered from a variety of health-related websites. Predicts a user's disease based on the user's or patient's symptoms. On the basis of the symptoms displayed in the web application, the consumer will be able to determine the possibility of a condition. Users can search for diseases based on a variety of symptoms and probability estimates and circumstances.

The WSGI has become a widely used platform in between the web servers and applications through the internet. Our model is first loaded on the local system using the Heroku cloud platform. Making it a user-friendly platform and reaching a wide audience across the globe through an online web application which acts intelligently. Incorporating Machine Learning into healthcare using the algorithms such as Convolution Neural Network and Random Forest Classifier is a breakthrough for our model.

Making use of features and various attributes of the diseases we will be able to achieve the expected accuracy.

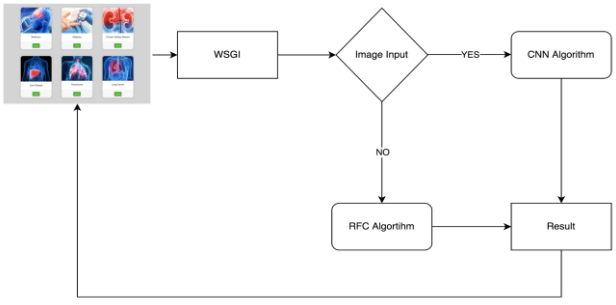

Fig. 2. System architecture

C. *System Requirements*

1) *Software Requirements*

- Python ide 3.7 version (or)
- Anaconda 3.7
- Flaski

2) Hardware Requirements

- Operating System: Windows, Linux, iMacOs
- Processor: minimum Intel i3
- RAM: minimum 4GB

6. Algorithm

A. Convolution Neural Network Algorithm

A Deep Learning model that can accept an image of the disease to be predicted as an input and give importance as learnable weights and biases to different aspects/objects in the image, allowing them to be distinguished from one another. Other classification systems require substantially more pre-processing than Convolutional neural networks. Convolutional networks may acquire these features, whereas primitive approaches' filters are hand-engineered.

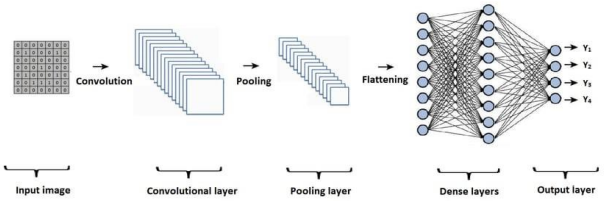

Fig. 3. CNN architecture

The layout of the Visual Cortex inspired the design of a Convolutional network, which is analogous to the human brain's linking network of neurons. Individual neurons respond solely to stimuli in the Receptive Field, which is a limited portion of the visual field [15]. A group of fields that overlap to fill the full visual field.

The input layer is where we supply data to the model. The total number of properties in our data is equal to the number of neurons in this layer; in the event of an image, the number of pixels is considered.

The data is sent from the input layer to the hidden layer, which processes it. Depending on our model and the quantity of data we have, there could be a lot of hidden layers. The number of neurons in each hidden layer varies, although it is almost always more than the number of features. The output of each layer is formed by multiplying the output of the previous layer by that layer's learnable weights, adding learnable biases, and finally applying the activation function, which makes the network nonlinear [17].

The output of the hidden layer is then fed into a logistic function like sigmoid or softmax, which turns the output of each class into a probability score for each class.

As a neural network with only one fully connected layer progresses in level, the number of parameters rises, increasing the memory footprint and increasing the computational cost. A solution has been steadily studied to overcome the challenges provided by some completely linked layers and encourage the growth of neural networks to a deeper degree.[4] A convolutional neural network (CNN) is developed using the ideas of local correlation and weight sharing, which not only minimizes the number of parameters but also increases training efficiency.

Convolutional neural network algorithms are used to predict Covid 19, Pneumonia, Malaria and Lung cancer which have image datasets. Using the right filters, a Convolution network may successfully capture relationships, Because of the reduced number of parameters and the reusability of weights, the design achieves a better fit to the image dataset.[17] The network may be trained to recognise the image's sophistication level.

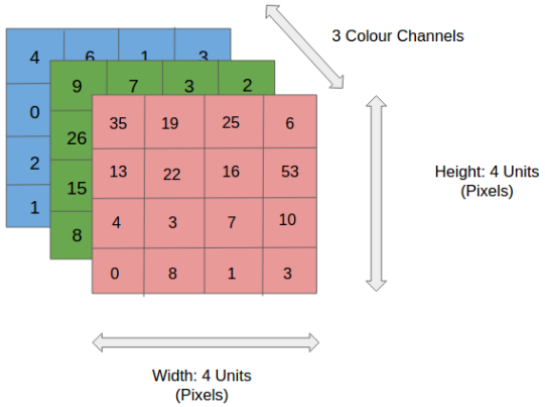

Fig. 4. RGB planes

The RGB image has been separated into three colour planes (Red, Green, and Blue) as shown in figure 3. Colour spaces such as Grayscale, CMYK, RGB and more can be used to represent images.

It shows how computationally intensive things will become as image dimensions get exceeded. The task of the Convolutional Network is to compress the images into a format that is easier to handle while keeping crucial features for a good prediction. This is necessary for creating an architecture that can learn features while also being scalable to huge datasets.

B. Random Forest Classifier

The Decision tree algorithm is improved by this algorithm. It's commonly used to develop prediction models in classification and regression problems. It builds decision trees at random in predefined subspaces. The main idea is to make small decision trees out of random data subsets, each with a biased classifier and various data trends. Splitting input variables are used to generate each tree at random [16]. The ideal split is then calculated using the training set features. It takes less time to train than other algorithms, accurately predicts outcomes, and operates smoothly even on big datasets. Even when a large amount of data is missing, it may maintain accuracy.

In general, a variety of learning models are used, but in a random forest, a model for the entire forest is constructed to find the best responses. The random forest method is used to investigate a medical instance of diseases having statistical data [9]. The random forest approach can include many eigenvalue features, and the combined outputs of multiple decision trees led to increased prediction accuracy of our research. The diseases which consists of numeric value input to be given according to the symptoms have been predicted using the Random Forest approach.

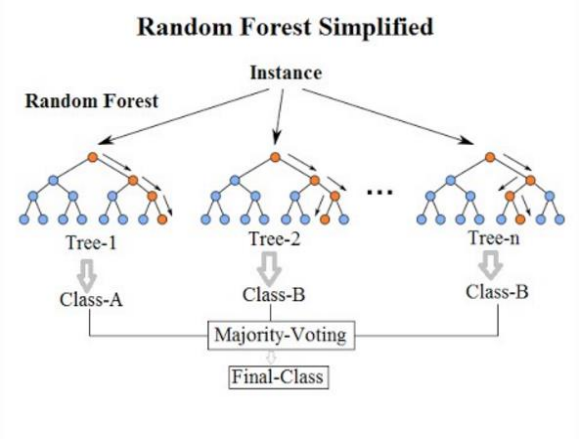

Fig. 5. RFC algorithm

7. Results

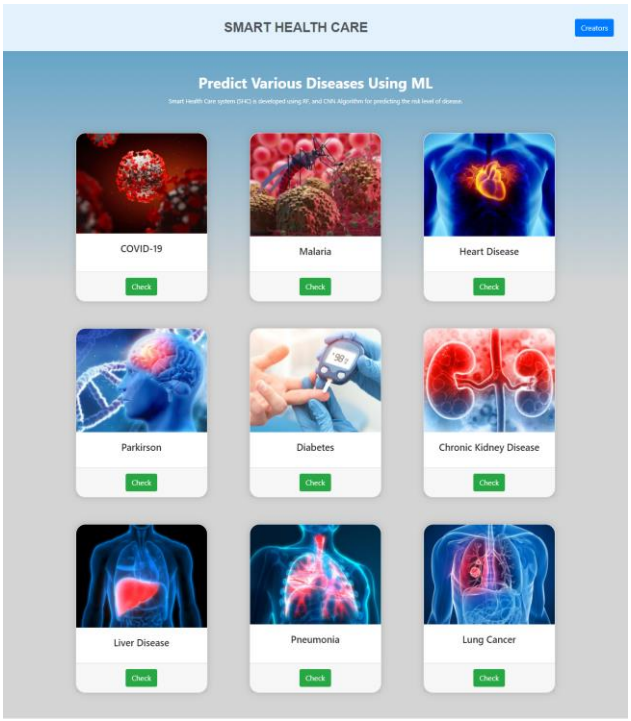

Fig. 6. UI of the system

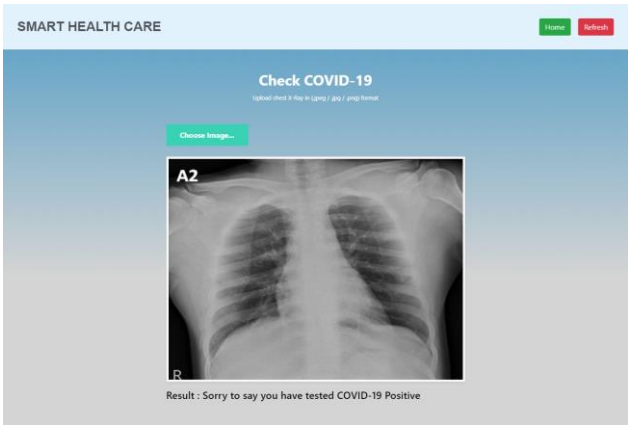

Fig. 7. Providing image input for COVID-19 prediction

Fig. 8. Providing image input for Malaria prediction

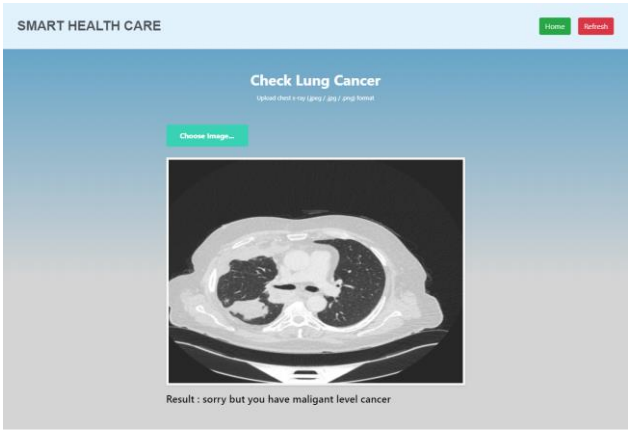

Fig. 9. Providing image input for lung cancer prediction

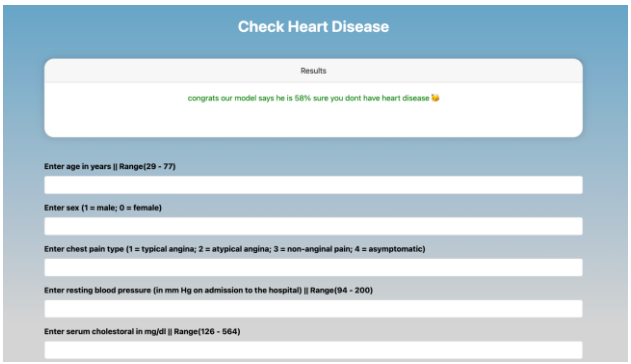

Fig. 10. Providing statistical input for heart disease prediction

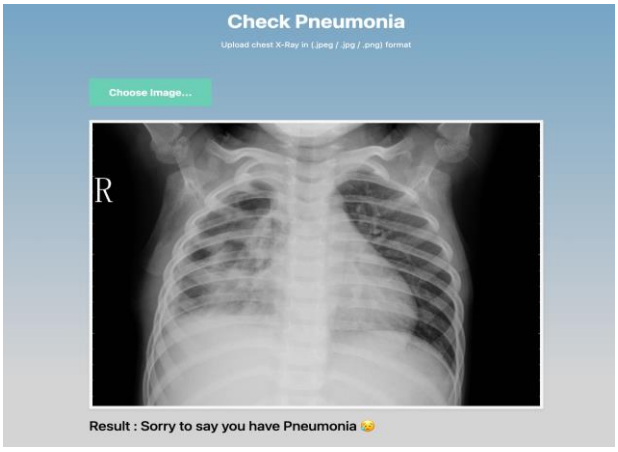

Fig. 11. Providing image input for Pneumonia prediction

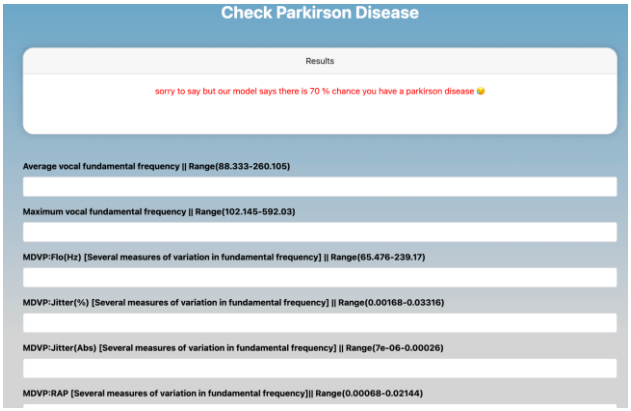

Fig. 12. Providing statistical input for Parkinson's disease prediction

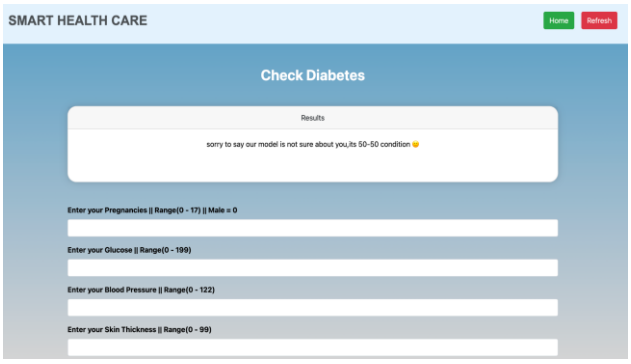

Fig. 13. Providing statistical input for Diabetes prediction

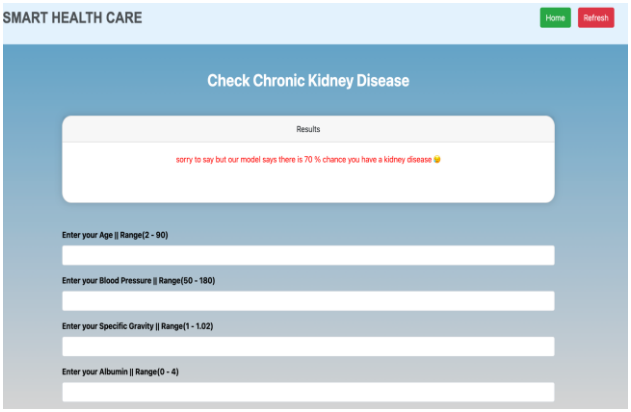

Fig. 14. Providing statistical input for chronic kidney disease prediction

8. Conclusion and Future Enhancement

The major purpose of this research is to use machine learning algorithms to improve disease diagnosis. After receiving the user's symptoms, the system forecasts and generates a disease prognosis. The main focus is on predicting diseases such as Covid-19, Malaria, Diabetes, Pneumonia, and heart disease using approaches such as Random Forest Classification and Convolution Neural Networks algorithms. The user interacts with the Prediction system by entering a few values or an image, which includes the parameter set that the trained models utilize as input. The average accuracy probability of prediction is determined to be 88.3 percent. Against the backdrop of the data science era, the future of medical treatment has more modern potential. Because of its unique feature processing method, deep learning has emerged as a crucial driving force for future development in the face of the high dimension and instability of medical data [15]. This research will be further improved in order to automate and improve the accuracy of many different disease predictions. To bring this into the real world, this idea can be expanded by using real-time hospital data and automating and deploying the mobile application to assist consumers. Many other ailments can be inculcated, and enhanced accuracy benefits all sectors of the users.

References

[1] Siddheshwar T, Pranjal S, Sukanya W, Wandhekar, Ankit C., "Prediction of Chronic Kidney Disease," Hindawi, 2018.
[2] Aditi Gavhane, Gouthami K, Isha P, Kailas D, "Prediction of Heart Disease Using Machine Learning," IEEE Xplore, 2018.
[3] Samrat Kumar Dey, Ashraf H, Mahbubur R Md, "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm," IEEE Xplore, 2019.
[4] Dhiraj Dahiwade, Gajanan P, Ektaa M, "Designing Disease Prediction Model Using Machine Learning Approach," IEEE Xplore, 2019.
[5] Dong Jin Park, Min Woo Park, Homin L, Young-Jin K, Yeongsic K, Young Hoon Park, "Development of machine learning model for diagnostic disease prediction based on laboratory tests," Scientific Reports, 2021.
[6] Naresh Kumar, Nripendra N D, Deepali Gupta, Kamali Gupta, Jatin B, "Efficient Automated Disease Diagnosis Using Machine Learning Models," Hindawi, 2021.
[7] Singh M, Bhatia V, Bhatia R, "Big data analytics solution to healthcare."
[8] Santhana Krishnan J, Geetha S (2019) Prediction of heart disease using machine learning algorithms. In: 2019 1st international conference on innovations in information and communication technology (ICIICT).
[9] R. Sreevalli, Prediction of diseases using random forest classification algorithm.
[10] Thirunavukkarasu K, Singh AS, Irfan M, Chowdhury A (2018) Prediction of liver disease using classification algorithms. In: 2018 4th international conference on computing communication and automation (ICCCA).
[11] Allam, Z., Dey, G., & Jones, D. S. (2020). Artificial intelligence (AI) provided early detection of the coronavirus (COVID-19) in China and will influence future Urban health policy internationally. *Ai*, *1*(2), 156-165.
[12] Jagtap, A., Malewadkar, P., Baswat, O., & Rambade, H. (2019). Heart disease prediction using machine learning. *International Journal of Research in Engineering, Science and Management*, *2*(2), 352-355.
[13] Fang, L., Karakiulakis, G., & Roth, M. (2020). Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?. *The lancet respiratory medicine*, *8*(4), e21.
[14] Md R. Hoque and M. Sajedur Rahman, "Predictive modelling for chronic disease: machine learning approach," in *Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis*, pp. 97–101, Silicon Valley, CA, USA, March 2020.
[15] S. Programming, "Retracted: enhancing health risk prediction with deep learning on big data and revised fusion node paradigm," *Hindawi*, vol. 2019, Article ID 9757658, 1 page, 2019.
[16] M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," *BMC Medical Informatics and Decision Making*, vol. 11, no. 1, pp. 1–13, 2011.
[17] R. Ge, R. Zhang, and P Wang, "Prediction of chronic diseases with multi-label neural network," *IEEE Access*, vol. 8, pp. 138210–138216, 2020.