International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER SSRN

IJIEMR Transactions, online available on 25th Jun 2022. Link :http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 05

DOI: 10.48047/IJIEMR/V11/SPL ISSUE 05/21

Title **MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING**

Volume 11, SPL ISSUE 05, Pages: 136-143

Paper Authors

Mr.Polasi.Sudhakar, Sreerajamadhuveeti Keerthana, Paturi Mounika, Rowthu Vasavi Uma Maheswari, Maddireddy Praveesha

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING

Mr.Polasi.Sudhakar[1], Sreerajamadhuveeti Keerthana[2], Paturi Mounika[3], Rowthu Vasavi Uma Maheswari[4], Maddireddy Praveesha[5]

[1]Assistant Professor. Dept. of CSE, [2]18ME1A05A2, [3]18ME1A0578, [4]18ME1A0595, [5]18ME1A0568
Ramachandra College of Engineering, A.P., India
sudhakar.forall@rcee.ac.in,keerthi.satyasrm@gmail.com,mounikapathuri30@gmail.com,rowthuvasavinaidu11@gmail.com,maddireddypraveesha@gmail.com

Abstract:
Medical care is one of the most important aspects of human society. We are all susceptible to multiple diseases which cause physical ailments. Health care should be more about proactively identifying a disease and its risks then reacting to them. This project can predict a number of diseases according to the patients' listed symptoms.Many of the machine learning models for health care analysis now in use focus on just one disease at a time. This project 'multiple diseases prediction' uses machine learning and deep learning models to predict. This is based on prediction modelling that predicts disease of the users according to the symptoms provided by the users as an input. This project can predict multiple Diseases based on the user disease selection. This project is used to create intelligent systems in order to provide quick, better and more accurate outcomes. In this project, we train the model using the disease dataset and produce the output whether the person has the disease or not.

Keywords: Machine Learning, Deep Learning, Disease Prediction, Random Forest, CNN, Diabetes, Kidney, Pneumonia.

Introduction:

Earth is going through a tremendous period of technological development, with a rising desire for intelligence and accuracy. People today are more likely to be addicted to the Internet, but they are unconcerned about their own health. Humans in the twenty-first century are surrounded by technology, which is an integral part of our daily lives. With this, we are always focusing on our own health as well as our hard-earned assets. People avoid going to the hospital for minor issues that could turn into severe diseases in the future.

Machine learning has swiftly evolved among the numerous ways for studying complex data, and the deep learning component of it has emerged as the most eye-catching branch of machine learning. Deep learning outperforms prior technical approaches in terms of its ability to process complicated data, the method for extracting the primary aspects of multi-dimensional data, the efficient response to unstructured data, and the classification strategy with higher accuracy. More people will be able to understand and enter the field of deep learning as a result of this. Deep learning technology has advanced significantly in recent years. Simultaneously, the use of deep learning to forecast disease has progressively gained traction. Our main goal is to create a system that can identify diseases based on user-provided symptoms and predict their occurrence.

The The primary goal is to develop an acceptable machine learning system that is effective and precise for disease prediction. In this study, the idea of supervised machine learning is applied to the prediction of diseases. Machine Learning

will be the primary component and Deep Learning in which we will be using algorithms such as Random Forest and Convolutional Neural Nework(CNN). It will aid in precise early disease prognosis and better health management..

Related Work

Numerous studies have been conducted utilising various machine learning algorithms to develop disease prediction systems, with varying degrees of success depending on the approach.

The paper [1] "Disease Prediction System" utilised the Decision Tree, Random Forest, and Naive Bayes algorithms to predict a disease based on the systems and to enable synchronised and knowledgeable medical systems ensuring the highest level of patient satisfaction.

A CNN-MDRP technique that integrates structured and unstructured data has been proposed in the publication [2] "Disease Prediction Using Machine Learning over Big Data," and it has also been demonstrated that CNN-MDRP is more accurate than earlier prediction algorithms.

The paper [3] "Application of Machine Learning Predictive Models in the Chronic Disease " analysed the study models connected with the diagnosis of the chronic disease with an emphasis on SVM and LR algorithms. These models are quite useful for categorising and identifying CD.

The paper [4] "Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively" has put up a system for predicting different diseases that offers advice on medications and treatments for the diseases it predicts.

The paper [5] "Disease Prediction using Machine Learning" the illness prediction system was created utilising the KNN, Naive Bayes, Logistic Regression, and Decision Tree algorithms and implemented using the Grails framework.

The paper [6] "Disease Prediction using Machine Learning" created a disease prediction system with the highest level of accuracy using the Nave Bayes, Decision Tree, and Random Forest algorithms. The system also offers inspirational ideas and images.

Existing System:

The existing system with Health care analysis using machine learning focuses on a single ailment at a time.This technique can only predict specific diseases. The accuracy of the existing System for one disease per analysis is only up to 94%. Then the existing system will predict the output for one disease per analysis using machine learning models.

Proposed System:

We are putting out a system that makes use of the Flask API to forecast various diseases. In this undertaking, the system is used to analyze different disease analysis. The model will accept either structured data or an image as input. End-users, such as patients or other users, use this system. The user will enter all the symptoms they are experiencing into this system. The machine learning model and deep learning will subsequently be provided with these symptoms to anticipate the diseases like diabetes, kidney, pneumonia.

The best accuracy is then determined by applying the algorithms. TensorFlow, Flask API, and machine learning techniques were used to implement numerous illness analysis. The model behaviour is saved using Python pickling, and the pickle file is loaded when needed using Python unpickling. The final output of this system will be whether the user/patient has affected with the disease or not by the model.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Methodologies:

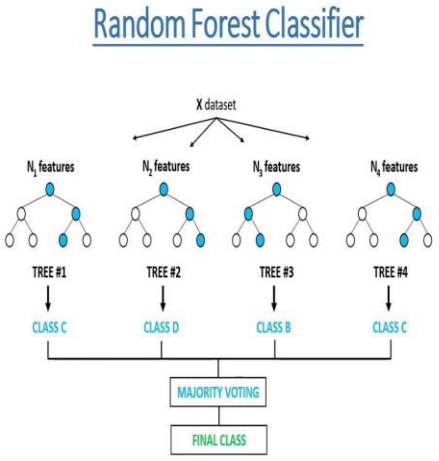
Machine Learning Models:

With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programmes can predict outcomes more accurately without having to be explicitly instructed to do so.

Our project is based on the patient-entered symptoms being used to predict multiple diseases. Identifying the problem statement is the first step. Afterward, preparing the dataset for use. In order to identify anomalies, missing values, etc. on our data and make our dataset ideal for prediction, we next conceptualise our data using scatter plot, distribution graph, etc. Finally, machine learning will be the key component, and we'll use it. Algorithm Random Forest which will accurately forecast disease to allow for early diagnosis and improved patient management. We used Python as a platform to run our System for this model..

Random Forest:

A supervised machine learning algorithm is random forest. Both classification and regression can be done with it. but it is most commonly employed for classification problems. Random Forest is really simple to set up and utilize. If we need to construct a model quickly, Random Forest is an excellent option. Random Forest is an ensemble learning technique that works by training a large number of decision trees. It uses voting to choose the best option. Simply put, a Random Forest is made up of several decision trees. It results in a forest of trees. The accuracy rate is directly proportional to the number of trees in the forest, which prevents over-fitting.Random Forest achieves strong outcomes in real-world applications due to its insensitivity to dataset noise and lack of over-fitting. It works exceptionally well and outperforms other tree-based algorithms. It mostly uses bootstrap aggregation or bagging for tree learning.



Diabetes Disease:

Diabetes is one of the world's most dangerous diseases. Diabetes is caused by obesity, excessive blood glucose levels, and other factors. It modifies the insulin hormone, resulting in abnormal crab metabolism and raising blood sugar levels. Diabetes develops when the body does not create enough insulin. Diabetes affects 422 million people globally, the majority of whom reside in low- or middle-income countries, according to the World Health Organization. This number could increase to 490 billion by the year 2030. Nevertheless, many nations, like Canada, China, and India, have high rates of diabetes. With a population of more than 100 million people, India has a total of 40 million diabetes. Diabetes is a leading cause of death in the world. Early detection of diseases such as diabetes can be controlled and human lives saved. To do so, this research looks into diabetes prediction using a variety of diabetes-related variables. We use the Diabetes Dataset for this, and we predict diabetes using Machine Learning techniques like Random forest.

Dataset:

We will go into more detail about the dataset that was utilised to train the machine learning model for this project.

We used a structured format for the dataset we used for this research. All of the names of diseases and their corresponding symptoms are included in the dataset that is being used.

Data Preprocessing:

The most crucial phase is data preprocessing. Most data pertaining to healthcare has missing values and other contaminants that can affect the data's usefulness. Data preparation is done to enhance the quality and effectiveness of the results produced through mining. This procedure is crucial for accurate results and good prediction when applying machine learning techniques to the dataset. Preprocessing for the diabetes dataset requires two steps.

1. Missing Values Removal - Remove all occurrences where the value is zero (0). It is impossible to have a value of zero. We create feature subsets by removing irrelevant features/instances, a process known as features subset selection, which decreases data dimensionality and speeds up work

2. Splitting of data- Data is standardised for the model's training and testing once it has been cleaned. After the data is spilt, we train the algorithm on the training data set while putting the test data aside. The values of the feature in the training data will be used along with logic and algorithms to generate the training model. Putting all of the qualities on the same scale is the aim of normalisation.

Algorithm:

1.The first step is to select the R features from the total features m where R<<M.

2.Among the R features, the node using the best split point.

3.Split the node into sub nodes using the best split.

4.Repeat a to c steps until l number of nodes has been reached.

5.Built forest by repeating steps a to d for a number of times to create n number of trees.

The random forest finds the best split using the Gin-Index Cost Function which is given by:

The first stage is to look at the options and utilise the foundations of each randomly built decision tree to predict the outcome and store the predicted outcome at intervals around the target location. Second, count the votes for each anticipated target and, as a result of the random forest formula's ultimate prediction, admit the predicted target with the most votes. Some of the Random Forest solutions provide accurate forecasts for a wide range of applications.

The dataset has then been split into two phases, a training dataset and a testing dataset. We used training datasets to create our models, and we then applied all of our machine learning algorithms to this data to create trained models. Finally, we gave this trained model a testing dataset to evaluate its correctness.

Kidney Disease:

We all know, that Kidney is vital organ in human body. Excretion and osmoregulation are two important functions of this organ. In layman's terms, the renal and excretion systems gather and eliminate all hazardous and useless material from the body. In India, 1 million instances of Chronic Kidney Disease (CKD) are diagnosed each year. Renal failure is another name for chronic kidney disease. It is a serious kidney condition that causes a gradual loss of renal function. CKD is a progressive loss of renal function over a long period of time. A person's kidneys will fail permanently. If CKD is not recognized and treated early, the patient may have the a decrease in immune response due to later phases when dangerous quantities of fluids, electrolytes, and wastes can build up in your blood and body, high blood pressure, anaemia, weekboans, poor nutrition, health, and nerve damage.

Machine learning does it by using CKD patient data to train and foreseeing the disease.

Model Building

This phase, which includes the creation of models for the prediction of kidney. In this we have implemented machine learning algorithm for kidney prediction.

Procedure of Proposed Methodology-

Step1: Import required libraries, Import kidney dataset.

Step2: Preprocess data to remove missing data.

Step3: Divide the dataset by 80% to create a training set, then by 20% to create a test set.

Step4: Now Select the machine learning algorithm Random Forest algorithm.

Step5: Create the classifier model based on the training set for the aforementioned machine learning technique.

Step6: Using a test set, evaluate the Classifier model for the aforementioned machine learning algorithm.

Finally, we gave this trained model a testing dataset to evaluate its correctness.

Deep learning Model:

Machine learning algorithms are stacked in a hierarchy of increasing complexity and abstraction as part of deep learning, an iterative approach to artificial intelligence (AI). Knowledge from the layer of the hierarchy before it is used to generate each deep learning level.

The most advanced AI architecture currently in use is deep learning.

Convolutional Neural Network (CNN):

When a neural network with only one fully connected layer progresses to a higher level, the number of parameters increases, increasing the memory footprint and increasing the computational cost. A solution has been steadily researched in order to address the challenges presented by some completely connected layers and encourage the growth of neural networks to a deeper level. CNN is created using the concept of local correlation and weight sharing, which not only decreases the number of parameters but also increases the training efficiency. CNN's basic framework was outlined by LeNet's proposal in 1998, which included a convolutional layer, a pooling layer, and a fully connected layer. And after convolution and pooling, the architectural alterations end get the output after fc.

Pneumonia:

PNEUMONIA is an infection that affects an affected person's lung air sacs. It is caused by bacteria, fungus, or viruses infecting the air sacs of the lungs, which fill up with discharge fluids, causing chills, fever, mucus coughing, and breathing difficulties in those who have been diagnosed with this disease. This type of sickness affects children under the age of five and older people with a weakened immune system. Pneumonia killed more than a million children worldwide in 2018, and it is still a life-threatening disease if not recognized or diagnosed early.

The most common approach for detecting pneumonia is radiography, CT-scan, or MRI. Medical personnel examine the patient's chest radiograph to identify whether or not they have pneumonia. Furthermore, the most common way for detecting pneumonia is through the patient's medical history and laboratory results. X-rays penetrate the soft tissues of the chest, which generate a dark tint, while hard tissues, such as bones, produce a bright colour. Patients with pneumonia

show symptoms of fluid filling the air sacs of the lungs in the chest cavity, making the radiological picture appear brighter. Several abnormalities, such as cancer cells, blood vessel swelling, and cardiac abnormalities, can be observed as brighter colour in the lung cavities.

Implementation:

Preprocessing, handover learning and refinement, and classification are the three stages of the proposed methodology of architectural design.

Preparing Datasets and PreProcessing

We used the Kaggle Pneumonia dataset named chest_xray dataset, which contains 26,684 image data. The data was gathered in JPEG format and was divided into two categories: pneumonia infected and normal, with maximum size of 1024 x 1024 pixels. The picture data consumes 75%, 20%, and 5% of the total data for training, testing, and validation, respectively. The major input for feature extraction is training data, and having a large amount of training data can result in strong features and a good result.

The testing data can be used to generate conclusions about the model's ability to work in real-world situations. In the occurrence of similar features from distinct classes, the responsibility of using the dataset acquires complexity. Experts will have a difficult time dealing with pneumonia's uniform pattern of symptoms.

Transfer Learning

The models are set for correct setting to the essential task for an effective classification. Transfer learning allows the model to return to its original parameter values from a previously trained model, resulting in an effective score without the use of a lot of processing effort.Also, FC or thick layers of 4096 are reduced to 512 and 256 to avoid wasting CPU resources. Hyper This research demonstrates how a Machine Learning system can quickly predict disease using various factors and models. parameters make it easier for the deep learning model to provide meaningful results throughout the training process. Selecting exact numbers can cause a significant difference at first, necessitating hyper parameter adjustment. Even yet, not every method has precise hyper parameters. As a result, practical testing is required.

Classification

Every single block use Relu activation and interconnected layers. While classification uses completely interconnected layers and predicts the outcome, feature extraction only requires an input image, convolution, and max-pooling..

The classification of the chest X-ray image as Normal or Pneumonia using the model we have developed is accurate. Our deep learning model can more accurately classify the image. Therefore, the model we developed does a good job of classifying data that it has never seen before. Therefore, with the aid of that person's chest X-ray image, we may use this model to forecast whether that person is Normal or has Pneumonia.

Conclusion:

Under the backdrop of the era of big data and the reach of the Internet of Things, the future of medical treatment has more modern potential.

The major goal of this paper is to use a Machine Learning algorithm to forecast disease based on symptoms provided by patients. In this paper, we used four Machine Learning algorithms for prediction and achieved a mean accuracy of over 95%, demonstrating remarkable rectification and high accuracy over previous work, as well as making this system more reliable than the existing one for this job and thus providing better user satisfaction than the other.

Future Scope

We believe that Disease diagnosis in deep learning and machine learning , will grow

in a more diverse manner in the future, and will almost probably expand from today's auxiliary diagnostic to decision-making diagnosis. A growing number of Deep learning models and machine learning models have emerged for various diseases. The models are linked and learn from one another, creating a more complex system network. This is in line with the complicated medical system, which will aid in the development of medical diagnosis and clinical applications, as well as the advancement of the medical profession.

References

[1] Khurana, Sarthak . , Jain, Atishay ., Kataria ,Shikhar. ,Bhasin ,Kunal . , Arora ,Sunny.,& Gupta , Dr.Akhilesh . Das. (2019). Disease Prediction System.*International Research Journal Of Engineering and Technology , 6*(5) , 5178-5184.

[2] Shirsath ,Shraddha.Subhash .,& Patil , Prof. Shubhangi . (2018).Disease Prediction Using Machine Learning over Big Data .*International Journal Of Innovative Research in Science and Technology , 7*(6), 6752-6757.

[3] Shirsath ,Shraddha.Subhash .,& Patil , Prof. Shubhangi . (2018).Disease Prediction Using Machine Learning over Big Data .*International Journal Of Innovative Research in Science and Technology , 7*(6), 6752-6757.

[4] Battineni ,Gopi. , Sagaro, Getu.Gamo. , Chinatalapudi, Nalini. , &Amenta, Francesco. (2020). Application Of Machine Learning Predictive Models in the Chronic Disease .*International of PersonalisedMedicine , 10*(21), 1- 11.

[5] Godse, Rudra A.,Gunjal, Smita S., JagtapKaran A ., Mahamuni , Neha S., & Wankhade, Prof. Suchita. (2019). Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively.*International Journal of Advance Research in Computer and Communication Engineering, 8*(12), 50-52

[6]Pingale,Kedar., Surwase, Sushant., Kulkarni,Vaibhav.,Sarage ,Saurabh., &Karve, Prof. Abhijeet .(2019). Disease Prediction using Machine Learning.*International Research Journal of Engineering and Technology, 6*(12) , 2810-2813.

[7]Pingale,Kedar., Surwase, Sushant., Kulkarni,Vaibhav.,Sarage ,Saurabh., &Karve, Prof. Abhijeet .(2019). Disease Prediction using Machine Learning.*International Research Journal of Engineering and Technology, 6*(12) , 2810-2813.

[8] Chauhan Raj H., NaikDaksh N. , Halpati, Rinal A., Patel, Sagarkumar J. , &PrajapatiMr. A.D. (2020). Disease Prediction using Machine Learning.*International Research Journal of Engineering and Technology, 7*(5) , 2000-2002.

[9] Chauhan Raj H., NaikDaksh N. , Halpati, Rinal A., Patel, Sagarkumar J. , &PrajapatiMr. A.D. (2020). Disease Prediction using Machine Learning.*International Research Journal of Engineering and Technology, 7*(5) , 2000-2002.

[10] Chauhan Raj H., NaikDaksh N. , Halpati, Rinal A., Patel, Sagarkumar J. , &PrajapatiMr. A.D. (2020). Disease Prediction using Machine Learning.*International Research Journal of Engineering and Technology, 7*(5) , 2000-2002.

[11] Chauhan Raj H., NaikDaksh N. , Halpati, Rinal A., Patel, Sagarkumar J. , &PrajapatiMr. A.D. (2020). Disease Prediction using Machine Learning.*International Research Journal of Engineering and Technology, 7*(5) , 2000-2002.

[12] Chauhan Raj H., NaikDaksh N. , Halpati, Rinal A., Patel, Sagarkumar J. , &PrajapatiMr. A.D. (2020). Disease Prediction using Machine Learning.*International Research Journal of Engineering and Technology,7*(5),2000-2002

[13] Chauhan Raj H., NaikDaksh N. , Halpati, Rinal A., Patel, Sagarkumar J. , &PrajapatiMr. A.D. (2020). Disease Prediction using Machine Learning.*International Research Journal of Engineering and Technology, 7*(5) , 2000-2002.

[14]A. Krizhevsky, I. Sutskever, & G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Neural Information Processing Systems, 25. 10.1145/3065386, 2012.

[15]Tang, S., Yuan, S., Zhu, Y. (2020). Convolutional neural network in intelligent fault diagnosis toward rotatory machinery. *IEEE Access, 8,* 86510–86519. DOI 10.1109/ACCESS.2020.2992692.

[16] Khare, S. K., Bajaj, V. (2020). Time-frequency representation and convolutional neural networkbased emotion recognition. *IEEE Transactions on Neural Networks and Learning Systems,* 1–9. DOI 10.1109/TNNLS.2020.3008938.

[17] Siddiqi, R. (2019). Automated pneumonia diagnosis using a customized sequential convolutional neural network. *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies (ICDLT 2019)*, pp. 64–70. New York, NY, USA.