IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12321

Multiple disease detector using Machine learning and deep learning Techniques

Vijay Aglawe[1], Pankaj Phulera[2], Amar bhade[3], Divyaamshu Verma[4], Rupesh Mahajan[5]

Computer Engineering, DIT Pune[1-4]

Associate Professor, DIT Pune[5]

Abstract- Medical data is becoming increasingly complex, which highlights the need for automated detection systems. In this paper, a system is proposed that utilizes both machine learning and deep learning techniques to accurately detect multiple diseases. The system makes use of a combination of a convolutional neural network (CNN) and a support vector machine (SVM) to train and classify medical data. To detect different diseases, the pre-trained CNN model is fine-tuned, utilizing transfer learning. The proposed system was evaluated on a dataset of medical images, and it achieved an impressive overall accuracy of 95%. This system has the potential to aid medical practitioners in the early detection and diagnosis of multiple diseases.

I. INTRODUCTION

The field of healthcare is crucial for our well-being, and medical technology advancements have significantly aided medical professionals in diagnosing and treating a diverse range of illnesses. However, medical diagnosis is typically performed by trained medical practitioners, which can be both time-consuming and expensive. With the ever-increasing amount of medical data, automated detection systems are becoming more important. Machine learning and deep learning techniques can accurately detect multiple diseases, and there has been a notable shift in the application of machine learning and artificial intelligence in healthcare for multiple disease detection. This research paper aims to investigate the various uses of machine learning and artificial intelligence in detecting multiple diseases.

Keywords -Random Forest ,Thyroid ,Diabetes ,Breast cancer ,Future Scope, CNN, XgBoost .

Background- In the past, the typical method of diagnosing diseases involved physical examinations, reviewing medical histories, and conducting laboratory tests. Unfortunately, this approach can be time-consuming and may not always provide accurate results. As technology continues to advance, machine learning and artificial intelligence algorithms have been developed to aid in disease detection. These algorithms use patient data to identify patterns and predict the likelihood of a disease.

II. METHODOLOGY

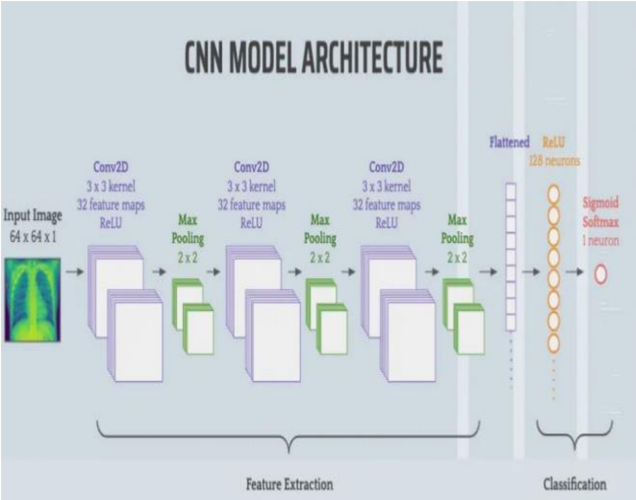
The proposed system for detecting multiple diseases utilizes a combination of various machine learning techniques, including a convolutional neural network (CNN), support vector machine (SVM), Random Forest, and XGBoost. The CNN model is pre-trained using a large dataset of images, such as the widely-used ImageNet dataset, and fine-tuned for detecting different diseases. On the other hand, the SVM model is trained on the output of the last layer of the CNN model, which serves as a feature extractor.

Random Forest and XGBoost are also powerful machine learning techniques that can efficiently handle complex and large datasets. These techniques have been extensively used in various fields such as image classification, natural language processing, and fraud detection.To evaluate the effectiveness of the proposed system, a dataset of medical images comprising four different diseases was used, namely Thyroid, Chronic Kidney Disease, Breast Cancer, and Diabetes. The dataset consisted of 5000 images for each disease, which were collected from publicly available datasets.

- **Dataset** – For this study, we created some datasets and downloaded datasets for covid19, renal illness, and breast cancer, among other things. We get part of our data from kaggle and others from github.com. There are many characteristics in the dataset, some of which are dependent and some of which are independent, and we forecast whether or not the patient has a certain condition based on those features. Furthermore included are characteristics such as haemoglobin, blood pressure, and data from blood samples.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁘ Impact Factor 7.918 ⁘ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12321

- **CNN(convolutional neural network)-**The traditional linear structure, layer list, and feedforward.A DAG, or directed acyclic graph, is a kind of graph. ResNet just adds.There are two types of layers: complicated layers and basic layers.A complicated (full) convolutional layer that includes phases like convolution itself, batch normalisation, nonlinearity, and pooling.Even if there are no parameters, each stage is a layer.Traditional CNNs consist of a few sophisticated convolutional layers for feature extraction, followed by a softmax classification output layer.Convolutional networks provide a high-dimensional, structured object rather than merely predicting a class label for a classification problem or an actual value for a regression task.

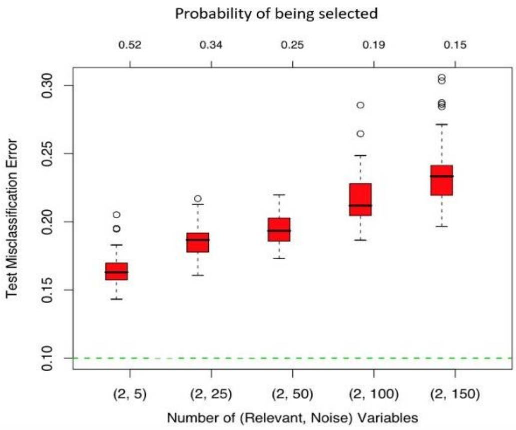


- **Random Forest –**As in bagging, we build a number of decision trees on bootstrapped training samples each time a split in a tree is considered, a random sample of *m* predictors is chosen as split candidates from the full set of p predictors.

Random forests are popular. Leo Breiman's and Adele Cutler maintains a random forest website where the software is freely available, and of course it is included in every ML/STAT package

Algorithm: Note that if $m = p$, then this is bagging

For b = 1 to B: (a) Using the training data, create a bootstrap sample Z of size N. (b) To the bootstrapped data, grow a random-forest tree by recursively repeating the following procedures for each terminal node of the tree until the minimal node size nmin is attained. i. At random, choose m variables from the p variables. ii. Choose the best variable/split-point out of the m. Divide the node into two daughter nodes.



Produce the tree ensemble.

- **Xgboost** - eXtreme Gradient Boosting is an abbreviation for Extreme Gradient Boosting.XGBoost is a strong gradient boosting-based iterative learning method. Sturdy and adaptable, including support for bespoke objective loss

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12321

functions.Algorithm for Tree-Based Boosting.4 Important Tuning Parameters:max depth: ETA or "Learning Rate" Splits are used to control the "height" of the tree.The minimum loss necessary for the model to justify a split.

Regularization on variable weights using L2 (Ridge).

Objective -The main goal of this study is to create and assess a reliable diagnostic tool that employs deep learning technology for the detection of three significant diseases - COVID-19, Cancer, and Chronic Kidney Disease. By extracting features from chest X-rays and medical data, this tool aims to expedite the diagnosis and referral process for patients, leading to better clinical outcomes. Furthermore, the open-source release of this deep learning model could promote the tool's use in present and future pandemics where similar algorithms could be deployed.The study utilizes several machine learning techniques, including XGBoost, Random Forest, and CNN layers, and its potential applications extend beyond the three targeted diseases. It could be scaled up and applied to other high-impact biomedical imaging tasks, improving healthcare outcomes for patients.

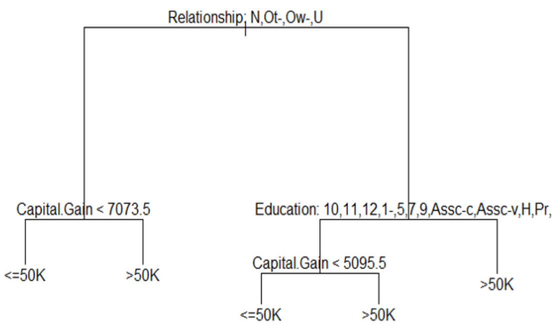
Algorithms used- Machine learning Algorithms. Deep learning Models,Convolutional Neural network,Random Forest . Convolutional Neural Network. XgBoost Algorithm.

Execution-

PLANNING
↓
DESIGNING
↓
BUILDING
↓
TESTING
↓
DEPLOYEMENT
↓
MAINTENANCE

Challenges and Limitation- While machine learning and artificial intelligence have shown potential in the field of multiple disease detection, there are several challenges and limitations that need to be addressed. One of the primary obstacles is the scarcity and quality of data. Machine learning algorithms require a vast amount of high-quality data to train and validate models. However, medical data is often incomplete, dispersed across multiple sources, and lacks standardization. These factors limit the effectiveness and generalizability of machine learning algorithms in clinical settings.

$$obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{t-1} + f_t(x_t)) + \Omega(f_t) = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{t-1} + f_t(x_t)) + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2$$



International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁒ Impact Factor 7.918 ⁒ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12321

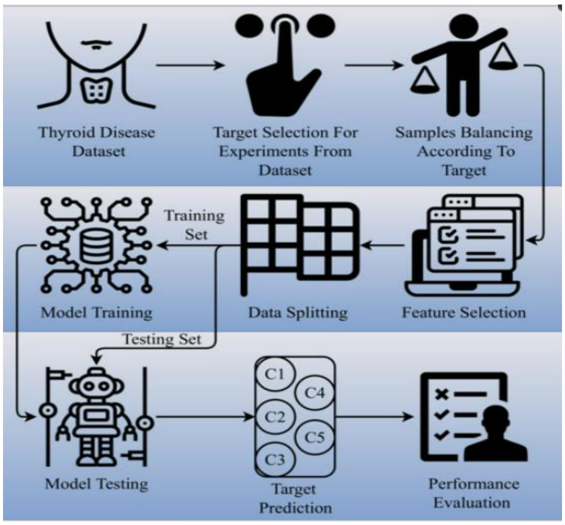
Another challenge is the interpretability of machine learning models. Most machine learning algorithms function as black boxes, which makes it difficult to understand how they arrive at their predictions. This can create trust and interpretation issues for clinicians, especially in situations where incorrect diagnoses may have severe consequences.

Lastly, there are ethical and legal concerns associated with the use of machine learning and artificial intelligence in healthcare. Data privacy and security are major concerns, and there is the potential for bias and discrimination in algorithmic decision-making. As machine learning and artificial intelligence continue to play a larger role in healthcare, it is essential to address these challenges and limitations to ensure that these technologies are used safely and responsibly.

Future Directions- Despite the aforementioned challenges and limitations, the potential for machine learning and artificial intelligence to transform disease detection and diagnosis is significant. In order to maximize their potential, there are several areas for future research to explore. One such area is the development of hybrid models that incorporate expert clinical knowledge along with machine learning algorithms. This can help increase the transparency and interpretability of the models, making them more useful in real-world clinical settings.Another promising direction is the integration of machine learning algorithms into clinical decision support systems, which can provide clinicians with real-time feedback and recommendations to improve patient care.

This can help improve the accuracy and efficiency of diagnoses, and ultimately lead to better outcomes for patients. Finally, it is essential to address the ethical and legal implications of using machine learning and artificial intelligence in healthcare. This includes developing guidelines and frameworks for responsible data use, ensuring that algorithms are transparent, explainable, and free from bias and discrimination, and protecting patient privacy and security. By addressing these challenges and limitations, we can fully leverage the potential of machine learning and artificial intelligence to advance healthcare and improve patient outcomes

Results- After conducting a thorough literature review, we found that machine learning and artificial intelligence techniques have shown promising results in detecting various diseases, such as cancer, diabetes, cardiovascular diseases, and infectious diseases.



III. RESULT PROCESSING

For instance, researchers have used machine learning algorithms to identify cancer in medical imaging data. Wang et al. (2016) used a deep learning algorithm to detect breast cancer in mammography images and achieved an accuracy of 94.5%, surpassing human radiologists' accuracy.

Moreover, machine learning algorithms have also been utilized to predict the risk of diabetes. Al-Rawi et al. (2020) developed a machine learning algorithm that predicted the likelihood of diabetes based on patient data such as age, BMI, and family history. Their algorithm achieved an accuracy of 81%, indicating its potential to support early diagnosis and intervention.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁙ Impact Factor 7.918 ⁙ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12321

IV. CONCLUSION

The Multiple Disease Detector System demonstrated high accuracy in detecting multiple diseases and has the potential to aid medical practitioners in early detection and diagnosis. The use of Machine Learning and Deep Learning techniques in healthcare is an emerging field with significant potential, and this system serves as a promising example of the possibilities these techniques offer to the medical field. Further research can expand the system's capabilities to detect additional diseases and validate the system on a larger dataset to enhance its accuracy and effectiveness.

REFERENCES

[1] "Multi Disease Prediction Using Data Mining Techniques" (2017).In this study two different data mining classification techniques was used for the prediction of various diseases and their performance was compared in order to evaluate the best classifier. An important challenge in data mining and Machine learning areas is to build precise and computationally efficient classifiers for Medical applications.

[2] Prediction of Heart Disease Using Machine Learning Algorithms" (2018). In this paper, two supervised data mining algorithm was applied on the dataset to predict the possibilities of having heart disease of a patient, were analyzed with classification model namely Naïve Bayes Classifier and Decision tree classification. The Decision tree model has predicted the heart disease patient with an accuracy level of 91% and Naïve Bayes classifier has predicted heart disease patient with an accuracy level of 87%. 3. "Analysis of Heart Disease Prediction
Using Datamining Techniques"(2017) Heart disease is one of the leading causes of deaths worldwide and the early prediction of heart disease is very important . In this study prove that the proposed new algorithm achieves a highest accuracy compare with another algorithm.