https://africanjournalofbiomedicalresearch.com/index.php/AJBR
Afr. J. Biomed. Res. Vol. 27 (September 2024); 679-684

Research Article

Early - Prediction of Critical Diseases Using Machine Learning and Image Processing

Prof. Varsha Pimprale[1*], Prof. Nutan Deshmukh[2], Dr. Sandhya Arora[3]

[1,2]Assistant Professor,[3]Professor, Computer Engineering Department, Cummins, College of Engineering for Women, Pune, varsha.pimprale@cumminscollege.in, nutan.deshmukh@cumminscollege.in, sandhya.arora@cumminscollege.in

Abstract:
The health care sector experiences a significant transformation. Post-Covid-19, the world undergoes dynamic changes. Many hospitals now offer online consultations. Early intervention is crucial for treating numerous major and minor diseases. The early prediction of a disease, based on image reports, diagnosis reports, or symptoms, allows for prompt treatment by doctors and patients. This system will diagnose various diseases based on given symptoms, including major diseases like multiple types of cancer. The proposed system can identify diseases based on diagnosis report parameters. Machine learning is widely used in the healthcare system for predictions. Three algorithms of machine learning used are to predict the critical diseases - Naive Bayes Decision Tree and Random Forest for more accurate results. Using machine learning algorithms and image processing techniques multiple diseases can be predicted with best accuracy in less time saving in a more safe way.

Keywords: Disease, Cancer, Image Processing, Prediction, Machine Learning.

***Author for correspondence: Email:** *varsha.pimprale@cumminscollege.in*

Received: 10/07/2024 Accepted:16/08/2024

DOI: https://doi.org/10.53555/AJBR.v27i3.2369

Introduction

25% of people die due to symptom neglect. And even 25% of the people succumb to death because of ignoring symptoms. Early detection and diagnosis of the disease may lead to a full recovery for the patient. Why early stage detection - As per the WHO more than 10 million of deaths are because of cancer but it is also analyzed that if cancer is detected at the early stage then probably more than 91% is survival rate of patients. Predicting diseases in their early stages using image and pathology reports is very crucial. Various diseases, such as breast cancer and pneumonia, can be identified from respective diagnostic imaging tests. Image processing and machine learning methods yield more precise outcomes than human observation alone. The majority of individuals overlook their health concerns. More people are affected by diseases, including cancer, that do not exhibit noticeable symptoms. Also the presence of a tumor or lump in the human body makes it impossible for doctors to definitively diagnose cancer. Determining the cancerous nature of a tumor is another crucial task. Predicting the cancerous status of a tumor holds significant importance and treating the patients becomes more challenging for doctors. Thus , Image processing and Machine Learning can play vital roles in determining the likelihood of cancer in the human body. According to the predicted probability, the patient should consult doctors for an early decision for faster recovery and less pain. In healthcare, technology enhances the convenience of detecting and treating diseases. With advanced data technology, medical records can now be collected and stored more easily, enabling earlier detection. Machine Learning algorithms combined with image processing can effectively detect various diseases, including cancer, pneumonia, and diabetes. Image processing significantly enhances doctors and patients' ability to visualize internal diseases. Early detection of diseases relies on accurate

prediction of both major and minor health threats. These diseases significantly impact overall health. Timely detection of diseases such as cancer and pneumonia is crucial for both the patient and doctors to initiate appropriate treatments promptly. The latest technology trends of image processing and machine learning combination helps to simplify this task. This technology enables the identification of diseases. Machine learning and image processing produce results that are more accurate and faster than human perception. This web application predicts diseases based on symptoms, image reports, and diagnosis report parameter values, and suggests nearby specialists accordingly.

Literature review

1. In this paper , the author is predicting critical disease with Chest X-rays which helps in diagnosing asthma, lung cancer, pneumonia, and COVID-19. AI identifies multiple diseases by employing machine learning and deep learning techniques to address complex image detection challenges and improve its accuracy. Deep CNN's effectiveness in Chest X-rays (CXR) analysis is bolstered by merging it with large input features. By employing a new multi disease diagnosis model using deep learning techniques on chest X-ray images, this study reduces computational cost. The images undergo pre-processing for cleaning and contrast enhancement. Further, the image segmentation is carried out. Multi-disease classification is done using Optimized Ensemble Transfer Learning (OETL). The main objective of the author is enhanced sensitivity, precision, and specificity, resulting in correct classification and detection accuracy. **[Kavitha B., Srinivas K., 2024]** Cardiovascular disease is the leading global cause of death. Predicting heart disease is a complex undertaking requiring extensive knowledge and expertise. Prediction accuracy in heart disease research remains suboptimal despite numerous studies addressing the issue. An accurate Heart Disease Prediction (HDP) can save a life, while an inaccurate one can be lethal. In this paper the author has discussed various Deep Learning, Machine Learning methods and optimization based HDP techniques to address the associated challenges. Currently, various DL and ML algorithms are being employed by researchers to aid professionals and the healthcare industry in forecasting heart disease. The text also explores optimization-based algorithms and evaluates their performance. The optimization-based HDP algorithm in this review paper may aid doctors in predicting heart disease and providing timely treatment.**[**Girish B, Agam D, 2024] In this paper , the author demonstrated Breast Cancer prediction. Predicting breast cancer recurrence with precision and in a timely manner is essential for effective medical intervention. Machine learning holds potential in this field. Effectiveness of ML models depends on proper hyperparameter setting. They had implemented the effect of this process on ML models' final performance by predicting the five-year recurrence of breast cancer patients in a real-life scenario. They optimized hyperparameters for five ML algorithms (Logistic Regression, Decision Tree, Gradient Boosting, eXtreme Gradient Boost, and Deep Neural Network) then compared their performance. Using default hyperparameters, simpler algorithms give superior results. After optimization, the complex algorithms outperformed the simpler ones. The significance of hyperparameter selection in ML algorithms for cancer recurrence prediction is emphasized in the study's findings.[Lorena G., Marcela C, 2024]

Pneumonia, an infectious disease with potential to be fatal, is diagnosed primarily via physical exams and diagnostic techniques including chest X-rays, ultrasounds, and lung biopsies. A misdiagnosis can have lethal outcomes for patients. Deep learning advancements have significantly improved medical experts' pneumonia diagnosis abilities. In this paper the author has focused on the deep learning concept and its algorithms. Utilizing deep learning models, healthcare professionals can improve both diagnostic precision and effective patients who are suspected of pneumonia. Six various deep learning models have been implemented and assessed in this study. It is observed that the deep learning models help to detect and predict pneumonia from chest X-ray images, supporting clinical decision-making and enhancing patient treatment more accurately . **[Mudasir A,Mobeen Shahroz,** 2024] Cloud computing and other advanced wireless technologies have revolutionized our lifestyle, particularly in healthcare. With the rising number of cancer cases, a thorough investigation is necessary. Although proper use or implementation of cloud technology in cancer-care services is still in process. This paper provides an extensive review of cancer treatment methods within a cloud-based healthcare system. In hospitals, cloud computing facilitates physicians' usage of sophisticated applications and tools while enabling nurses to promptly retrieve patient information via wireless technologies. Understanding the practicalities of cloud computing is essential for scientists to effectively manage and analyze large data sets in cancer research. This study, through the comparison of analyzed articles' benefits and drawbacks, offers a current and thorough account of the advancements in cloud-based cancer research. **[Binghui X., Fengcheng Z.,** 2022] **Brain tumor detection is a common concern in healthcare nowadays.**

A brain tumor is characterized by uncontrollable, abnormal cell growth. Abnormal brain tumor regions are extracted using image segmentation. Identifying brain tumors relies on brain tissue segmentation in MRI scans. Accurate data mining classification methods allow for effective early disease prediction. In the medical field, machine learning and data mining techniques enjoy substantial importance. The study explores brain tumor surveillance systems' identification of risk factors. The proposed method guarantees both efficiency and precision for brain tumor detection, classification, and segmentation. Precise results necessitate the use of automatic or semi-automatic methods. The study suggests an automatic segmentation approach using CNNs. These processes include average filtering, segmentation, and feature extraction. Machine learning and data mining techniques are instrumental in early stage brain tumor detection and prevention. The techniques of machine learning (ML) and Data mining combinely are being effectively useful for brain tumor detection and prevention at an early stage.[G. Hemanth**; M. Janardhan, 2019**] In this paper the author has focused on hybrid machine learning application in predicting brain tumors A precise brain tumor diagnosis depends on early and accurate prediction. Manual analysis of medical images via traditional methods is time-consuming. Deep learning has demonstrated

potential for enhancing brain tumor prediction in machine learning applications. Individual approaches can have their own limitations. Hybrid models hold multiple techniques for improved prediction accuracy, stability and clarity. This survey examines recent research on hybrid models, evaluates their techniques, and compares their performance to conventional methods. Hybrid machine learning shows promise in advancing brain tumor prediction for better patient outcomes. [P. Saravanan; S Saravanakumar, 2023]

Methodology

As per shown in Fig.1 based on symptoms/ image reports diseases can be predicted based on its parameters.Three machine learning algorithms were applied for best accuracy. The disease will be predicted using three machine learning algorithms based on symptoms or diagnosis report parameters. These three machine learning algorithms include: Decision tree, Random Forest, and Naive Bayes. The algorithms with identical, accurate results will be displayed as the predicted disease. To enhance the precision, the outcomes of these 3 algorithms are compared. Based on the given dataset, machine learning algorithms are employed to predict diseases from the provided parameters/symptoms. Image processing techniques are employed for predicting diseases based on reports. The image will first be converted to grayscale, followed by thresholding, erosion, and dilation for noise removal. The TensorFlow model is used to identify tumor regions. This model is trained on multiple image datasets. Using marker-based watershed algorithms, the affected region image segmentation can be viewed.

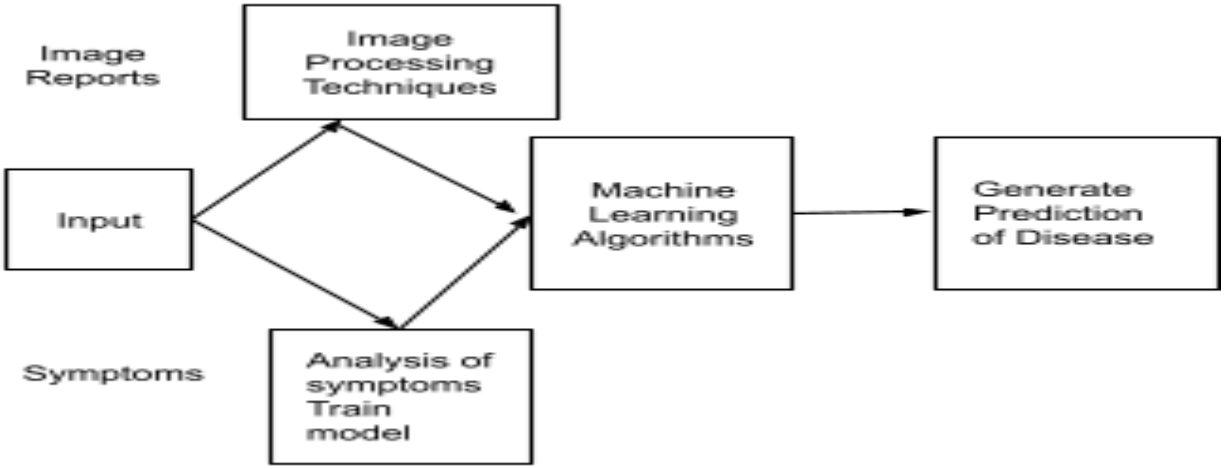

Figure 1: System Architecture

Reports can manifest as imagery (e.g., MRIs, x rays, mammographies) or as numerical input data, accompanying their results. For a particular disease, the system accepts user input. Figure 1 depicts the structure of the presented system and figure 2 will apply image processing techniques for extracting results from the image. According to our literature review, a significant need exists. During the pandemic, the system can predict major diseases like Cancer, Pneumonia, and diabetes. In Module 1, predicting diseases based on symptoms is the primary focus. Users can choose symptoms from the provided list. Machine learning algorithms are employed for disease prediction based on symptoms. These algorithms use the provided dataset for training and testing. The CSV data, as presented, needs to be revised. This condition manifests symptoms and the disease it's linked to. This CSV file contains a list of symptoms in combination. Assign 1 to the presence and 0 to the absence of symptoms. Decision Tree, Random Forest, and Naïve Bayes were employed as the selected Machine Learning algorithms. Using their own techniques, individuals can predict diseases, and the results from these 3 algorithms will be given. The final detected disease is displayed as the best-predicted or suitable result. This test not only identifies diseases in their initial stage but also increases the likelihood of successful treatment. In Module 2, the prediction of disease is done based on image reports. Based on the selected disease, users need to upload the corresponding images for accurate disease prediction. The report on disease images will be detailed. Once uploaded, the image undergoes various image processing techniques. The image will undergo grayscale conversion followed by erosion and dilation processes to eliminate noise. After following these steps, the image is given to a pre-trained machine learning model, which has been trained on a large image dataset. A pre trained machine learning model determines whether a selected disease is present or not. This system utilizes image processing and machine learning within the COVID-19 learning model to determine whether a specific disease is present or not. The user can observe the affected area marked by the Marker-based Watershed Algorithm in the image report.

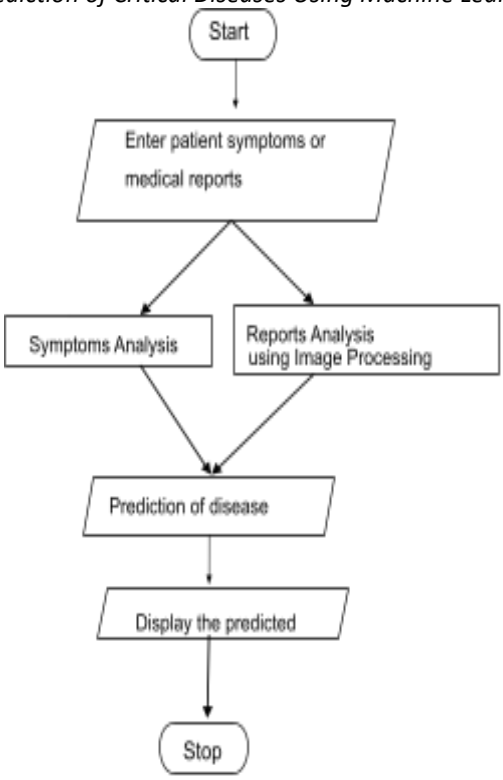

Fig 2: Data Flow Diagram

In Fig.2 , users enter their symptoms. The given list can be used for symptoms-based disease prediction or it can serve to input parameter values derived from diagnostic medical reports. The symptoms/values will be analyzed accordingly. Using machine algorithms, the model generates a predicted value after processing the input. Based on the prediction, the result will be displayed. In Module 3 where prediction of disease is based on parameter values, Users should input the extracted parameters from a diagnosis report for accurate disease prediction. The given diagnosis report provides values such as radius, diameter and skewness extracted from a brain tumor. The mentioned machine learning algorithms will analyze these values. These values will be analyzed by the three machine learning algorithms mentioned above. The dataset, which includes parameter values, is supplied for both training and testing purposes for the same.The best suitable prediction for the presence or absence of a disease will be displayed after comparing the individual predictions from various Machine Learning algorithms.

Results and Discussion :

The given dataset is used for training and testing image-based disease prediction models. The dataset encompasses a range of images, some diseased and others normal, depicting various orientations, blurriness, sizes, and degrees of affected regions. The proposed system enhances the input image by removing noise, sharpening, and segmenting for improved results. The table below indicates the prediction accuracy data.

Table 1. Accuracy of Machine Learning Algorithms

Learning Algorithm	Accuracy (%) Score
Random Forest	92.20
Naïve Bayes	88.70
Decision Tree	84.60

Table 2 : Accuracy of Prediction of disease

Year	Cancer Accuracy %	Pneumonia Accuracy%	Brain Tumor Accuracy %
2020	85	80	75
2021	87	82	78
2022	90	85	80

2023	92	88	83
2024	94	90	85

Here is the combined graph showing the prediction accuracy for cancer, pneumonia, and brain tumors from 2020 to 2024. The graph illustrates that the prediction accuracy for all three conditions has been improving over the years, with cancer prediction having the highest accuracy, followed by pneumonia, and then brain tumors.

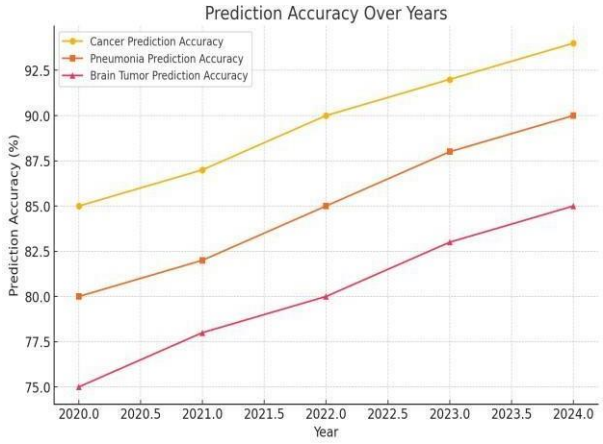

Fig. 3: Prediction of critical diseases

The proposed system aims to deliver accurate predictions for diseases. Our system aims to predict both cancer and general diseases. Our system, proposed here, can predict multiple diseases. The text provides details about the diagnosed disease. This system enables prediction of diseases in the early stage for better recovery. Image processing can detect major diseases with 95% accuracy, surpassing human capabilities.

Conclusion :

The system accurately forecasts diseases. This system enables severe predictions. Diseases can be diagnosed promptly. The system can predict multiple diseases. Currently, various conditions including pneumonia, breast cancer, and skin disease are being treated. The text provides data on detected diseases. Obtain the contact details for a specialist doctor in this field. Invaluable during the pandemic crisis. To ensure safety. This system enables the swift prediction of disease and immediate corrective measures. Timely treatment is crucial. At reception or laboratories, this system is useful for hospitals. Based on the patient's reports, I will assign the appropriate doctor. It seems that Image processing techniques can improve the accuracy of detecting major diseases like cancer by up to 95%. The naked eye can't discern details as finely as the revised sentence. The system generates a doctor suggestion list for users. Detecting diseases early permits risk reduction and effective treatment. The treatment of the patient can be explored further in the future. It can also be extended. Also, it can be extended. We aim to construct a predictive system for some more complex diseases.

Future Scope :

Advanced HCI technologies enable the detection of various diseases with the same system and methods. In this system , we can add new diseases and train the machine learning model. 3D images and video datasets such as sonography can be utilized similarly for disease detection. Adding roles for users, doctors, and hospitals enhances the system's convenience for communication between patients and healthcare professionals and improves the prediction functionality.

References

Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare. Elsevier; 2020. pp. 25–60.

Binghui X, Fengcheng Z. et.al, “The Roles of Cloud-Based Systems on Cancer-Related Studies: A Systematic Literature Review”, Research article IEEE access, doi 10.1109/ACCESS.2022.3181147, June 2022.

Chiranjeev Sagar and Lalit Mohan Saini, “Color Channel Based Segmentation of Skin Lesion from Clinical Images for the Detection of Melanoma”, 2016 IEEE 1st International Conference on Power Electronics,Intelligent Control and Energy Systems (ICPEICES), 2016.

Chugh G, Kumar S, Singh N. Survey on machine learning and deep learning applications in breast cancer diagnosis. Cognitive Computation 2021; 13(6): 1451–1470.doi: 10.1007/s12559-020-09813-6.

C K Gomathy, et.al., “THE PREDICTION OF DISEASE USING MACHINE LEARNING”, International Journal of Scientific Research in Engineering and Management (IJSREM), Volume: 05 Issue: 10 Oct - 2021.

Daghottra A, Jain D. From humans to robots: Machine learning for healthcare. International Journal of Scientific Research in Computer Science Engineering and Information Technology 2021; 705–714. doi: 10.32628/CSEIT2173152.

Frunza, O., Inkpen, D., & Tran, T. (2011). A Machine Learning Approach for Identifying Disease Treatment Relations in Short Texts. IEEE Transactions on Knowledge and Data Engineering, 23(6), 801–814.

Girish B, Agam D, et.al., “Heart disease prediction using machine learning, deep Learning and optimization techniques-

A semantic review", Springer article in Multimedia Tools and Applications, July 2024.

Hemanth; M. Janardhan, "Design and Implementing Brain Tumor Detection Using Machine Learning Approach", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), April 2019.

Han, Zhongyi, Benzheng Wei, Yuanjie Zheng, Yilong Yin, Kejian Li and Shuo Li. "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model." Scientific Reports, 2017.

Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology 2017; 2(4): 230–243. doi: 10.1136/svn-2017-000101.

K. Gaurav, A. Kumar,et.al., "Human Disease Prediction using Machine Learning Techniques and Real-life Parameters " , International Journal of Engineering, IJE TRANSACTIONS C: Aspects Vol. 36 No. 06, (June 2023) 1092-1098.

Kavitha B, Srinivas K,et.al, 'Automated and reliable detection of multi-diseases on chest X-ray images using optimized ensemble transfer learning" , ACM-Digital Library, Expert Systems with Applications: An International Journal, Volume 246, Issue C, July 2024. 3.

Lorena G., Marcela C. et.al, "Impact of Hyperparameter Optimization to Enhance Machine Learning Performance: A Case Study on Breast Cancer Recurrence Prediction", MDPI Special Issue Artificial Intelligence for Healthcare , 2024.

Mudasir A,Mobeen Shahroz,et.al., "Pneumonia Detection Using Chest Radiographs With Novel EfficientNetV2L Model", Research article IEEE access, doi 10.1109/ACCESS.2024.3372588 ,vol. 12, February, 2024.

P. Saravanan; S Saravanakumar, "A Survey on Brain Tumor Prediction with Various Machine Learning Approaches", 2023 International Conference on Computer Science and Emerging Technologies (CSET), October 2023.

Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health and Technology 2017; 7(4): 351–367. doi: 10.1007/s12553-017-0179-1

Prannoy Giri And Saravanakumar " Breast Cancer detection using Image processing Techniques" , Oriental Journal of Computer Science & Technology", Vol. 10(2), June 2017.

https://www.who.int/cancer/prevention/diagnosis-screening/breast- cancer/en.

Sneha Grampurohit, et.al., "Disease Prediction using Machine Learning Algorithms", 2020 International Conference for Emerging Technology (INCET), 05-07 June 2020.

Sonal Shilimkar, Varsha Pimprale, et.al., "Disease Prediction System in New Normal", International Journal of Scientific Research & Engineering Trends, Volume 7, Issue 1, JanFeb-2021, ISSN (Online): 2395-566X.

Smitha P. , Shaji L. et.al," A Review of Medical Image Classification Techniques", International Conference on VLSI, Communication & Instrumentation (ICVCI) 2011 Proceedings published by International Journal of Computer Applications® (IJCA). [23]. Sunarti S, Rahman FF, Naufal M, et al. Artificial intelligence in healthcare: Opportunities and risk for future. Gaceta Sanitaria 2021; 35: S67–S70. doi: 10.1016/j.gaceta.2020.12.019

R. Kasmi, Oana Frunza, "A Machine Learning Approach for Identifying Disease Treatment Relations in Short Texts" , IEEE transactions on knowledge and data engineering, vol. 23, Issue no. 6, June 2011.

Stewart J, Sprivulis P, Dwivedi G. Artificial intelligence and machine learning in emergency medicine Emergency Medicine Australasia: EMA 2018; 30(6): 870–874. doi: 10.1111/1742-6723.13145. https://www.kaggle.com/datasets?search=disease