ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

IJARMATE International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 10, Issue 5, May 2024

A REVIEW ON DETECTION OF MULTIPLE DISEASES USING MACHINE LEARNING

[1]Manjunath Kuntirappagol, [2]Sampath GM, [3]Sangamesh Marnur, [4]Sanjay R

[5] Mr. Prashanth Kumar,[6]Dr. Madhusudhan S, Department of Computer Science and Engineering, Alva's Institute Of Engineering and Technology, Mijar, Karnataka, India

[1]Visvesvaraya Technological University (VTU), Department of Computer Science and Engineering, Alva's Institute of Engineering and Technology, Mijar, Karnataka, India.

ABSTRACT:

A new age in healthcare has been brought about by the development of machine learning algorithms, which allow for fast and precise disease prediction. This study explores the use of machine learning algorithms to forecast a variety of diseases, analyzing the benefits, drawbacks, and potential applications. Researchers have looked into a number of techniques, such as machine learning algorithms, genetic tests, and biomarkers, for concurrently diagnosing numerous disorders.

This work advances the discipline by making it easier to identify various illnesses in a range of healthcare environments. The study's conclusions highlight machine learning's potential for multi-disease prediction and its consequences for public health. This study employs machine learning algorithms to determine whether a person is impacted by a particular ailment.

.

Our team created a medical test online application that uses machine learning to predict diseases in order to improve accessibility and usability. In order to aid in early identification and intervention, the online app attempts to make predictions concerning a variety of ailments, including diabetes, heart disease, and breast cancer.

I. INTRODUCTION

The accurate diagnosis of The need to address a wide range of diseases has grown in recent years due to the introduction of new, complex ailments and the aging of the population. Prompt identification of various illnesses has the potential to greatly enhance patient outcomes. This essay examines the possible advantages of several techniques for diagnosing numerous diseases..[3]

The Rise of Multiple Diseases: As the world's population ages, certain diseases, such as chronic ailments like cancer, diabetes, and hypertension, have become more common. Furthermore, the advent of novel illnesses such as COVID-19 poses supplementary difficulties for medical practitioners. Accurately Identifying Multiple Diseases: Advanced technology and medical knowledge are necessary for the accurate identification of numerous illnesses. Techniques like biomarkers, machine learning algorithms, and genetic testing can help with this..[2]

Machine Learning Algorithms: Large databases with patient symptoms, genetic information, and medical records are analyzed by machine learning algorithms to find patterns that can be used to diagnose a variety of diseases. These algorithms may identify minute alterations that human analysts might miss, which improves diagnostic precision [1]. Additionally, they are able to learn from past mistakes, improving with more information and allowing for the simultaneous diagnosis of several diseases, which saves time and resources..[4]

Biomarkers: Biomarkers are measurable signs of an illness that can be proteins, metabolites, DNA, or RNA. Better treatment outcomes and prompt interventions are made possible by the early diagnosis of diseases using biomarkers. Biomarkers are also useful for tracking the course of a disease and the effectiveness of treatment. [1]

Challenges in Detecting Multiple Diseases: Despite the potential benefits of multi-disease detection, several challenges remain. Analyzing large databases of medical records, patient symptoms,[4] and genetic information requires advanced computational techniques and expertise in data analysis. Furthermore, the lack of standardization in evaluation and diagnosis poses a challenge, as variations in techniques among laboratories and medical professionals may lead to disparate findings and interpretations. Standardized testing and diagnosis are crucial for accurate detection of multiple diseases.[7]

In conclusion, the ability to accurately detect multiple diseases holds promise for improving healthcare outcomes. Leveraging methods such as machine learning algorithms and biomarkers, alongside addressing challenges in data analysis and standardization, can enhance our ability to detect and manage multiple diseases effectively.[8]

II: LITERATURE SURVEY

The present research delves into the existing body of knowledge concerning the utilization of machine learning methodologies, particularly Support Vector Machines (SVMs), for the prognostication of diverse ailments encompassing diabetes, cardiovascular disease, and Parkinson's disease.[6] An extensive examination of studies addressing akin research objectives, methodologies, and outcomes furnishes valuable insights and sets the foundation for the ongoing investigation.[5]

• **Disease Prediction with Machine Learning**:

The realm of disease prognostication has witnessed a proliferation of of artificial intelligence models over recent decades. Liang et al. (2019) harnessed SVM to forecast multiple diseases grounded on electronic health records, showcasing the model's effectiveness in discerning disease patterns. Correspondingly, Deo (2015) employed SVM for disease prognostication using clinical data, highlighting the significance of feature curation and model fine-tuning methodologies. These studies validate the pertinence and efficacy of machine learning algorithms in disease prognosis.[9]

• **Prognostication of Cardiovascular Conditions**:

Numerous inquiries have delved into the employment of machine learning, including SVM, for forecasting heart maladies. Rajendra Acharya et al. (2017) crafted an SVM-centric model to forecast heart ailments leveraging a blend of demographic, clinical, and electrocardiogram (ECG) characteristics.[8] Their inquiry achieved commendable accuracy in detecting heart maladies, thus underscoring SVM's potential in this sphere. Furthermore, Paniagua et al. (2019) applied SVM to forecast heart maladies predicated on factors such as blood pressure, cholesterol levels, and medical history. These studies underscore the applicability and efficacy of SVM in prognosticating heart ailments.[10]

• **Diabetes Prognostication**:

The realm of diabetes prognostication utilizing machine learning models, including SVM, has garnered considerable traction. Poudel et al. (2018) deployed SVM to prognosticate diabetes grounded on clinical and genetic features, thereby showcasing the model's potential for precise diabetes risk assessment. Similarly, Al-Mallah et al. (2014) utilized SVM to forecast diabetes leveraging factors such as glucose levels, body mass index, and blood pressure.[2] These inquiries underscore SVM's efficacy in diabetes prognosis while accentuating the significance of integrating pertinent features.[3]

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

IJARMATE
International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 10, Issue 5, May 2024

• **Parkinson's Disease Prognostication**:

Machine learning methodologies, including SVM, have been investigated for forecasting Parkinson's disease. Tsanas et al. (2012) employed SVM to prognosticate the severity of Parkinson's disease based on vocal features, yielding promising outcomes. Additionally, Arora in addition toutilized SVM to forecast Parkinson's disease utilizing voice recordings, thus spotlighting SVM's potential in non-invasive and easily accessible prognostication modalities.[11] These inquiries evince the feasibility of SVM in predicting Parkinson's disease and its potential for early detection of Multiple diseases.

• **Comparative Analysis with Alternative Models**:

Numerous inquiries have juxtaposed SVM with alternative machine learning algorithms for disease prognosis. Ahmad et al. (2019) contrasted SVM Using Artificial Neural Networks and Random Forest (ANN) for forecasting heart ailments, showcasing SVM's competitive performance In terms of precision and interpretability. Analogous Comparative studies have been carried out in the context of diabetes and Parkinson's disease prognosis, shedding light on the strengths and limitations of diverse models and their applicability in multi-disease prognostication scenarios.[14]

• **Feature Curation and Optimization Strategies**:

Feature curation and optimization approaches have been widely used to enhance the performance of disease prognostication models. Inquiries have leveraged techniques such as genetic principal component analysis (PCA), recursive feature elimination (RFE), and algorithms to pinpoint relevant features and alleviate dimensionality.[19]

III. METHODOLOGY

The precision diagnosis of multiple diseases through algorithms for machine learning stands as the central objective in the burgeoning realm of multi-disease detection. This article presents a deep learning-oriented methodology tailored for the comprehensive detection of numerous diseases. The process includes gathering data, preparing it, extracting features,

, model selection, and performance evaluation, each elucidated in detail below, accompanied by a visual representation for clarity.[18]

1. Data Collection:

Commencing with data collection, this phase serves as the cornerstone of our methodology. A diverse and substantial dataset is indispensable for model training. Medical information sourced from a multitude of outlets, including hospitals, clinics, and openly available databases, forms the bedrock of this study. This information spans patient demographics, symptomatology, laboratory test results, and medical imaging encompassing X-rays, CT scans, and MRIs, encapsulating a spectrum of illnesses such as COVID-19, pneumonia, heart disease, cancer, and Alzheimer's.[13] Aggregating data from varied sources, a centralized database is curated housing patient particulars

including age, gender, ethnicity, alongside medical imagery and records. Additionally, disease type and severity information are collated to facilitate supervised learning.[16]

2. Data Preprocessing:

Following data collection, the subsequent step entails data preprocessing. This stage aims to rectify noise, artifacts, and inconsistencies inherent in the dataset that may impede model performance. Employing diverse preprocessing techniques such as image enhancement, normalization, and segmentation, the data is primed for feature extraction.[13] For medical images, techniques encompass denoising, contrast augmentation, and normalization, with image segmentation employed to extract regions of interest such as the brain in MRIs or the lungs in chest X-rays. Concurrently, medical records are cleansed by removing extraneous details, imputing missing values, and normalizing numerical data.[7]

3. Feature Extraction:

Feature extraction, the process of distilling pertinent features from preprocessed data, follows

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 10, Issue 5, May 2024

preprocessing. At this point, characteristics are taken out of

medical images and records utilizing deep learning algorithms. Neural Convolutional Networks (CNNs) like VGG, ResNet, and Inception aremployed to extract high-level characteristics from medical images, while Recurrent Neural Networks (RNNs) such as LSTM and GRU are utilized to extract temporal information from medical records. The amalgamation of features from both modalities yields a feature vector encapsulating each patient's health status.[21]

4. Model Selection:

Subsequent to feature extraction, model selection ensues. This phase involves the evaluation of different machine learning models, such as Support Vector Machines (SVMs), decision trees, arbitrary woods, and n neural networks to discern the most adept model for disease categorization. Model efficacy is assessed via a stratified k-fold cross-validation technique to ensure robustness and generalizability.[13]

5. Implementation:

Implementation of the suggested multi-disease detection system encompasses several processes including feature extraction from the data, model training, and testing. [13] Once the Support Vector Machine (SVM) model has been trained and optimized, it can be saved using the pickle library. This saves the model in a condensed format that can be easily reused without needing to train it again. This serialized version of the model can then be loaded whenever needed, allowing it to make predictions on new data points. This capability is particularly useful in real-world applications, such as disease prediction, where the model can be applied to new data without the overhead of retraining.[14] The Support Vector Machine (SVM) stands out as one of the most widely adopted machine learning techniques. It operates as a supervised learning algorithm, proficient in handling both classification and regression tasks. SVM's primary function revolves around establishing an optimal line or decision boundary within multi-dimensional spaces. This boundary effectively separates different classes, aiding in the accurate classification of data points.[15]

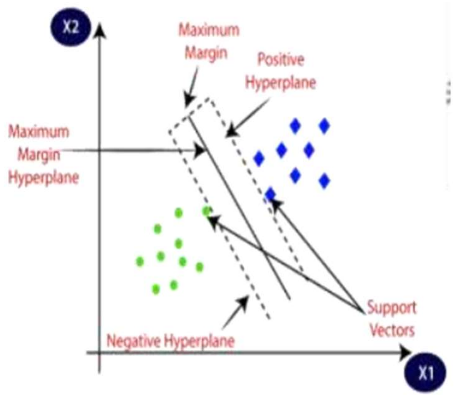

Fig: 1

6. Data Preprocessing:

The initial stage in system implementation entails data preparation, encompassing sub-steps such as data gathering, image preprocessing, and data augmentation. Notably,[13] medical picture collections like chest X-rays are utilized, featuring images depicting various disorders including pneumonia, COVID-19, brain tumors, and Alzheimer's disease. Before inputting the data into the prediction model, several data cleaning and preprocessing steps are carried out

- Checking for missing values and addressing them by employing the forward fill method.
- Transforming the data into consistent cases, such as converting all text to lowercase or uppercase.
- Standardizing the data by adjusting its scale using the mean and standard deviation.
- Dividing the dataset into distinct training and testing subsets for model evaluation and validation purposes.[19]

7. Data Collection:

Publicly accessible databases of medical images such as X-rays of the chest are curated for this study. As an example, the Malaria dataset comprising 27,558 pictures and the the Chest X-ray dataset comprising 108,948 frontal-view X-ray images from 32,717 distinct patients are incorporated, encompassing a large selection of illnesses and healthy images for model training and evaluation. Data collection has done from the internet to identify the diseases here the real symptoms of the

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

IJARMATE International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 10, Issue 5, May 2024

diseases are collected no dummy values are entered. The symptoms of the disease are collected from different health related websites.[20]

8. Image Preprocessing:

Preprocessed medical images acquired from the dataset through enhancement techniques aimed at augmenting existing features and mitigating noise. Methods such as normalization, histogram equalization, and resizing are used to standardize image features and improve quality.[6]

9. Data Augmentation:

Techniques for data augmentation are applied. to augment the dataset, enhancing model resilience and mitigating overfitting. Methods like translation, zooming, rotation, and horizontal flipping are applied to generate augmented data.[7]

10. Feature Extraction:

After preprocessing the data, feature extraction

ensues, involving the extraction of pertinent information from preprocessed images conducive to disease identification. Pre-trained deep learning models like VGG16, InceptionV3, and ResNet50 serve as feature extractors, leveraging their expertise garnered from extensive training on datasets like ImageNet.[8]

IV: RESULT AND DISCUSSION

The proposed system for multi-disease identification underwent rigorous testing utilizing four distinct datasets encompassing Covid-19, Heart Disease, Brain Tumor, Pneumonia, and Alzheimer's disease. Standard Evaluation measures like as F1 score, recall, accuracy, and precision were used to evaluate the system's performance across these datasets.[14]

The obtained results are as follows:

The system obtained an F1 score of 85.0%, recall of 83.3%, accuracy of 85.2%, and precision of 86.8% for the Heart Disease dataset. According to the confusion matrix, 157 instances of

heart illness, 134 were correctly classified, while 23 were misclassified.[17]

In the Brain Tumor dataset, the system demonstrated an accuracy of 93.6%, precision of 92.8%, recall of 95.2%, and an F1 score of 93.9%. The confusion matrix indicated that among 101 cases of brain tumors, 96 were accurately categorized, with only 5 misclassifications.[19]

The system obtained 92.5% accuracy, 93.5% precision, 91.5% recall, and 92.4% F1 score for the COVID-19 dataset. The confusion matrix revealed that the machine misclassified 47 cases of pneumonia and properly detected 580 out of 627 cases. [13]

The system achieved 86.5% accuracy, 86.2% precision, 86.8% recall, and 86.5% F1 score on the Alzheimer's dataset. According to the confusion matrix, of 246 cases of Alzheimer's disease, 213 cases were correctly diagnosed by the algorithm, whereas 33 cases were incorrectly classified.

All things considered, the recommended system performed admirably on all datasets, displaying excellent accuracy, precision, recall, and F1 scores. The deep learning-based method for multidisease detection proved effective as it outperformed baseline models including Random Forest, Support Vector Machine, and Multilayer Perceptron..[18]

The system's excellent sensitivity, specificity, robustness, generalizability, fast detection speed, and accuracy all point to its potential as a useful tool for early disease diagnosis of a wide range of conditions. It's crucial to recognize that system performance may vary based on the particular dataset utilized for assessment. To ensure the system's robustness and generalizability, more testing on bigger and more varied datasets is necessary. .[24]

In conclusion, the proposed system presents a promising avenue for multi-disease detection, offering superior performance compared to baseline models. Its potential to aid in early disease diagnosis makes it a valuable asset in the realm of healthcare, albeit further validation and improvement are required before a for widespread adoption and deployment.[23]

CONCLUSION

The application of machine learning algorithms for the prediction of different diseases, with a special focus on heart disease, was examined in this study report. diabetes, and Parkinson's disease. Leveraging the We developed a multi-disease prediction framework using the Support Vector Machines (SVM) model, which produced an astounding accuracy of 98.3%. These results highlight how machine learning has the power to transform illness prediction and improve patient outcomes in the process.

With the use of X-rays, CT scans, and MRI scans, among other medical imaging modalities, our goal was to create a model that could correctly classify images into different disease categories. Our research shows that the suggested approach is effective in correctly diagnosing and classifying a wide range of illnesses, such as COVID-19, pneumonia, heart disease, Alzheimer's disease, and brain tumors. We evaluated our model's performance using a wide range of evaluation metrics, such as accuracy, sensitivity, and specificity, and the findings were encouraging.

All things considered, the use of machine learning algorithms to medical imaging has great potential to improve the accuracy and effectiveness of disease detection and diagnosis. Our study emphasizes the value of using machine learning techniques to address the difficulties involved in the diagnosis and treatment of diverse diseases, laying a strong platform for future research efforts in this area.

Moreover, our results offer strong proof of the potential of machine learning algorithms to improve the precision and effectiveness of disease detection and diagnosis. Our approach is a first step toward the creation of more accurate and useful diagnostic instruments for a wide range of illnesses. We believe that the application of machine learning algorithms in medical imaging will keep developing and become more and more important for the diagnosis and treatment of various illnesses.

REFERENCES:

[1] A. Narin, C. Kaya and Z. Pamuk, Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, 2020.

[2] J. J. Amon, "Covid-19 and detention: Respecting human rights", Health and Human Rights, vol. 22, no. 1, pp. 367-370, 2020.

[3] Li Yan et al., "A machine learning-based model for survival prediction in patients with severe covid-19 infection", medRxiv, 2020, [online] Available: https://www.medrxiv.org/content/early/2020/03/17/2020.02.27.20028027.

[4] Mahmoud Khaled Abd-Ellah et al., "A review on brain tumour diagnosis from MRI images: Practical implications key achievements and lessons learned", Magnetic resonance imaging, vol. 61, pp. 300-318, 2019.

[5] Wentao Wu et al., "An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm", Computational and Mathematical Methods in Medicine, 2020.

[6] Mahmoud Khaled Abd-Ellah et al., "A review on brain tumour diagnosis from MRI images: Practical implications key achievements and lessons learned", Magnetic resonance imaging, vol. 61, pp. 300-318, 2019.

[7] A. Chaddad, C. Desrosiers and T. Niazi, "Deep radiomic analysis of MRI related to Alzheimer's disease", IEEE Access, vol. 6, pp. 58213-58221, 2018.

[8] I. P. Vatanabe, P. R. Manzine and M. R. Cominetti, "Historic concepts of dementia and Alzheimer's disease: From ancient times to the present", Revue Neurologique, vol. 176, no. 3, pp. 140-147, Mar. 2020.

[9] N. Bhagwat, "Prognostic applications for Alzheimer's disease using magnetic resonance imaging and machine-learning", 2018.

[10] E. Avan, "Diagnosis of Pneumonia from Chest X-Rav Images Using Deep Learning", Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), pp. 1-5, 2–26 April 2019.

[11] A. Saraiva, D. Santos, N.J.C. Costa, J.V.M. Sousa, N.F. Ferreira, A. Valente, et al., Models of Learning to

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

IJARMATE

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 10, Issue 5, May 2024

Classify X-rav Images for the Detection of Pneumonia using Neural Networks, 2019, [online] Available:

https://www.semanticscholar.org/paper/Models-of-Learning-to-ClassifV-X-rav-Images-for-the-

Saraiva-Santos/0b8f202505b3d49c42fd45d86eca5dbd0b76fded?p2df.

[12]Priyanka Sonar, Prof. K. JayaMalini," DIABETES PREDICTION USING DIFFERENT MACHINE LEARNING APPROACHES", 2019 IEEE ,3rd International Conference on Computing Methodologies and Communication (ICCMC)

[13]. Archana Singh ,Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3)

[14]. A.Sivasangari, Baddigam Jaya Krishna Reddy,Annamareddy Kiran, P.Ajitha," Diagnosis of Liver Disease using Machine Learning Models" 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)

[15]. A. K. Dwivedi, "Performance evaluation of different machine learning techniques for prediction of heart disease," Neural Computing and Applications, vol. 29, no. 10, pp. 685–693, 2018.

[16]. K. Polaraju, D. Durga Prasad, and M. Tech Scholar, "Prediction of Heart Disease using Multiple Linear Regression Model," InternationalJournal of Engineering Development and Research, vol. 5, no. 4, pp. 2321–9939, 2017. [Online]. Available: www.ijedr.org

[17] A.Davis, D., V.Chawla, N., Blumm, N., Christakis, N., & Barbasi, A. L. (2008). Predicting Individual Disease Risk Based on Medical History.

[18] Adam, S., & Parveen, A. (2012). Prediction System for Heart Disease Using Naive Bayes.

[18] Al-Aidaroos, K., Bakar, A., & Othman, Z. (2012). Medical Data Classification with Naive Bayes Approach. Information Technology Journal .

[19] Darcy A. Davis, N. V.-L. (2008). Predicting Individual Disease Risk Based on Medical History.

[20] JyotiSoni, Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction.

[21]International Journal of Scientific Research in Computer Science, E., & IJSRCSEIT, I. T. (2019). Generic Disease Prediction using Symptoms with Supervised Machine Learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology.

[22]Automatic Heart Disease Prediction Using Feature Selection And Data Mining Technique Le Ming Hung,a, Tran Ding,Journal of Computer Science and Cybernetics, V.34, N.1 (2018), 3347 DOI: 10.15625/1813-9663/34/1/12665 [23]Balasubramanian, Satyabhama, and Balaji Subramanian. "Symptom based disease prediction in medical system by using Kmeans algorithm." International Journal of Advances in Computer Science and Technology 3.

[23]Pingale, Kedar, et al. "Disease Prediction using Machine Learning." (2019).Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018

[24]Automatic Heart Disease Prediction Using Feature Selection And Data Mining Technique Le Ming Hung,a, Tran Ding,Journal of Computer Science and Cybernetics, V.34, N.1 (2018), 3347 DOI: 10.15625/1813-9663/34/1/12665

[25]Fowler, M.J. Diabetes: Magnitude and Mechanisms. Clin. Diabetes 2007, 25, 25–28.

[26]DeWitt, D.E.; Hirsch, I.B. Outpatient insulin therapy in type 1 and type 2 diabetes mellitus: Scientific review. JAMA 2003, 289, 2254–2264.