DOI: http://dx.doi.org/10.26483/ijarcs.v16i3.7271

Volume 16, No. 3, May-June 2025

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

A PERCEPTIVE SURVEY ON THE PREDICTION OF DISEASE DIAGNOSIS USING MACHINE LEARNING, DEEP LEARNING AND ARTIFICIAL INTELLIGENCE

Femila K. V.
Research Scholar
Department of Computer Science
SRMV College of Arts & Science, Perianaickenpalayam, India

DR. M. Jaikumar
Associate Professor
Department of Computer Science
SRMV College of Arts & Science, Perianaickenpalayam India

Abstract: Massive volumes of data are routinely handled in the medical industry. Results can be impacted when large amounts of data are handled using traditional approaches. In medical research, machine learning algorithms are particularly useful for obtaining information related to disease prediction. Early prediction of disease is essential to improve the lifestyle of the patients and prevent them from further worsening their condition. For the examination of patient medications and specialists, early disease detection is essential. A variety of diseases are identified using machine learning algorithms, such as ensemble classifier techniques, clustering models, and classification models. Using predictive machine learning techniques can result in high-accuracy and quick illness prediction. This study examines the many sorts of diseases and how they might be predicted using machine learning techniques. This study highlighted the existing research works that mostly explored the prediction of lung problems, brain-related diseases, and heart disease. While deploying conventional model of machine learning and deep learning the problem of overfitting, hyperparameter initialization and unbalanced classes are the major issues which affect the performance of the prediction models.

Keywords: Disease diagnosis, machine learning, heart disease, lung disease, brain related disease, classifiers, deep learning

INTRODUCTION

The process of identifying the condition that is responsible for a person's symptoms is known as disease diagnosis. Since some symptoms and indicators are non-specific, making the diagnosis the most difficult issue to solve [1]. The first and most important step in treating any condition is diagnosing it. Based on historical data used for training, data mining is a discipline that can assist in predicting the diagnosis of a disease. Numerous scientists have developed a wide range of machine learning techniques that are effective in diagnosing various ailments [2]. Without being specially programmed, machines may now train because of machine learning. Machine learning algorithms can be used to build a model that predicts an early disease diagnosis and offers remedies. The best strategy to reduce the death rates brought on by any disease is to receive a diagnosis early and receive effective therapy.

Classification is the most commonly employed machine learning technique in healthcare sectors since it can be applied to real-world situations. Classification algorithms work by first creating a model with the training data, which is then used to the test data to produce a prediction [3]. Various categorization techniques have been used to diagnose diseases, and the results show great promise. With these methods, errors in diagnosis can be avoided and outcomes are accessible rapidly [4].

In this paper three most tremendous significant causes of mortality namely heart disease, brain disease and lung disease are focused. Prediction of various heart disease is a critical challenge in the area of clinical data analysis. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. The application of machine learning to aid in the diagnosis and prognosis of brain illnesses holds immense potential. However, we contend that significant obstacles *still* need to be overcome by the community in order to integrate these instruments into clinical routines: more people need to adopt standard procedures for validation and reproducible research; broad generalisations studies are necessary; and comprehensible models are needed to get past the drawbacks of black-box strategies. Clinically, the most common disorders are those associated to the respiratory system, such as bronchiectasis, pulmonary thromboembolism, chronic obstructive pulmonary disease, and pulmonary tuberculosis. While cough up, phlegm anticipation, wheezing, and chest discomfort are typical symptoms of these diseases, each one has a distinct course of therapy and aftercare. Timely diagnosis can be difficult since some diseases shares similar symptoms. In primary healthcare facilities, misdiagnosis is prevalent and can result in ineffective therapy, a protracted recuperation period, and possible worsening of the condition. The insufficient experience of primary medical professionals contributes to the situation further.

The detailed survey of heart disease, brain related disorders and lung-based disease using machine learning are discussed in the following sections.

SURVEY ON HEART RELATED DISORDER DETECTION USING MACHINE LEARNING

Galla Siva et al [5] developed a feature selection model and the hybridized linear model using random forest for

prediction of cardiovascular disease. Various machine learning models with different subset of features the performance in conducted.

A unique self-attention-based transformer model that integrates transformer networks and self-attention mechanisms to predict the risk of CVD is used in the work of Rahman,et al. [6]. The self-attention layers provide representations that successfully model intricate patterns in the input by capturing contextual information. By assigning a specific attention weight to each element of the input sequence, self-attention mechanisms enable interpretability. To gather pertinent data, this entails altering the attention processes, adding new levels, and modifying the results and input layers.

Mamum et al.'s goal [7] was to find the most accurate machine learning classifiers for these kinds of diagnostic applications. A number of supervised machine-learning algorithms were used, and their accuracy and performance in predicting heart disease were compared. With the exception of K-Nearest Neighbour and Multi-layered Perceptron, all applied methods evaluated the feature significance scores for each feature. The importance score of each trait was used to rank them in order of highest likelihood of heart disease. They determined from their investigation that random forests offer the best accuracy.

Eleven machine learning (ML) classifiers were employed by Hassan et al. [8] to find critical features, hence improving the forecasting of heart disease. Several feature pairs and well-known techniques for classification were employed to construct the prediction model. Random forest algorithm produce best result compared other state of arts.

In order to decrease mortality, Bhatt et al. [9] provide a k-modes clustering approach with Huang beginning that can increase classification accuracy to forecast cardiovascular illnesses. In order to maximise the performance of the classifiers, GridSearchCV was utilised to hyper-tune the model's parameters. Based on this underlying research, it can be concluded that the accuracy of the multilayer perceptron technique with cross-validation has exceeded all other techniques.

To accurately anticipate the onset of cardiac disease, Hossain et al. [10] examined the many components in patient data. By employing the Correlation-based Feature Subset Selection Technique with Best First Search, the most important characteristics for heart disease prediction have been identified. Two distinct kinds of heart disease datasets are used to apply and compare different artificial intelligence algorithms.

SURVEY ON BRAIN RELATED DISORDER DETECTION USING MACHINE LEARNING

Fatima Yousaf et al. [11] developed an integrated model based on convolutional neural networks to concurrently identify and categorise two brain illnesses, namely tumours and ischemic stroke. Two open-source datasets are combined to generate a new dataset for this purpose. Ito conduct feature mapping the authors improved the architecture of encoder-decoder in order to preserve the grained information without overlapping the process of feature extraction.

Using an artificial vision technique, Ahmet Kursad Poyraz et al. [12] suggested an exemplar-based automatic brain disorder diagnosis model. Pre-processing was done by an exemplar deep feature generator, feature selection by iterative neighbourhood component analysis and SVM classification comprise the four stages of the automated brain disease detection model that is being presented. The most significant attributes are extracted and used for classification. The SVM classifier uses the chosen feature as its input.

Vidhya et al [13] used laptop-based clinical needs to do a precise segmentation of MRI of brain for the diagnosis of brain tumours. It is possible to determine from each image which segmentation strategy works best for the tumour by applying one of the segmentation strategies. The learning recommends a sophisticated analytical approach grounded on deep learning and machine learning to identify abnormalities. This is a simplistic and efficient model for recognising and categorising brain tumours.

High performance in the artificial intelligence sector of processing image based on categorization issues is made possible by this huge data. Classifying different forms of tumour from brain MR images is the goal of Khaliki et al [14]. For classification, Convolutional Neural Networks, and its inception, Efficient Net and transfer learning techniques were employed.

ZainEldin et al [15] anticipated an architecture based on CNN-based Brain Tumour Classification Model, it is an adaptable dynamic sine-cosine efficiency grey wolf optimisation algorithm for hyperparameter optimisation. After optimising the hyperparameters, an Inception RESENT is built to train model. To enhance brain tumour diagnosis, the model uses widely used pre-trained models, and its output is neither normal nor tumorous.

Mohammad et al [16] conducted a detail study to improve prediction accuracy through the application of a thorough methodology that encompasses feature selection, data pre-processing, and model construction. Particle swarm optimisation was used to investigate several AI methods for feature selection, other four conventional algorithms. These findings may help medical practitioners make well-informed decisions and enhance patient outcomes by having a significant impact on early disease identification, diagnosis, and customised treatment.

Ten variables were initially chosen by Xuchun Wang et al. [17] using feature selection approaches. After that, combined the ML approach with the cost-sensitive training and SMOTE resampling techniques to use trained classifiers for class data that was imbalanced. In the unbalanced datasets, the Stacked heterogeneous ensemble model performed reasonably well. In the balance data, Logistic regression models with class weighting demonstrated the greatest classification performance when these combined variables were applied as model comparison criterion.

Gunasinghe et al [18] developed machine learning model with deep learning to predict the lung diseases by integrating various types of data sources. The author’s used binary classification with input as patient's data and output classifies presence or absence of the disease. Its objective is to assist the experts in early detection of lung disease.

Through the application of artificial intelligence to microbiological metagenomes extracted from serum from patients and coded by their cumulative taxonomic hierarchy, McDowell et al. [19] devised a predicted model of diagnosis for lung cancer, asthma, and COPD. At the level of genus and phyla levels, all models showed strong predictive power with median area under the curve that included a variety of

significant features. When the clinical models were used on mice, different foods decreased the incidence of lung tumours and bronchitis linked to high-fat diets, but COPD was scarcely altered.

Mitra et al [20] introduced a deep learning model to predict the pneumonia using chest X-ray. Both machine learning and deep learning paradigm is used to predict the presence of lung disease at its early stage.

Nageswaran et al [21] demonstrate the use of technologies made possible by the use of artificial intelligence and image processing to accurately classify and predict lung cancer. Geometric mean filter is employed to enhance the quality of image. K-means is used for segmentation and ANN is used for categorization for lung cancer prediction.

Abobaker et al [22] devised a multimodal classification to determine the severity grade of lung disease. Without conduction segmentation, only using region of interest, CNN performs automatic feature extraction and pre-trained variables are used. The machine learning is used for optimizing the performance of CNN in lung disease detection.

CONCLUSION

In this paper a detailed survey of three major diseases that greatly affect mortality rate all over the world is discussed. From the study conducted on existing literatures it is observed that random forest, neural networks, support vector machine and deep learning models are used. Most of the papers utilized certainty-based algorithms and conventional method for predicting diseases. But while handling clinical datasets it is highly relies on uncertainty situations in accurate prediction of diseases at its early stages. It is necessary to handle inconsistency, incompleteness, vagueness and uncertainty while handling disease diagnose to reduce false detection rate. It is also observed that chronic obstructive pulmonary disease prediction-based works are very few, so in future we will focus on developing empowered models of COPD prediction at its early stage more effectively by considering uncertainty conditions.

REFERENCES

1. Niloy Biswas, Md Mamun Ali, Md Abdur Rahaman, Minhajul Islam, Md. Rajib Mia, Sami Azam, Kawsar Ahmed, Francis M. Bui, Fahad Ahmed Al-Zahrani, Mohammad Ali Moni, "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques", BioMed Research International, vol. 2023, Article ID 6864343, 15 pages, 2023
2. A. Solanki, A. Vardhan, A. Jharwal and P. N. K. I, "Heart Diseases Prediction Using Machine Learning," *2023 14th* International Conference on Computing Communication and Networking Technologies, 2023, pp. 1-6.
3. Jasmine Pemeena Priyadarsini, M., Kotecha, K., Rajini, G. K., Hariharan, K., Utkarsh Raj, K., Bhargav Ram, K., Indragandhi, V., Subramaniyaswamy, V., & Pandya, S. (2023). Lung Diseases Detection Using Various Deep Learning Algorithms. Journal of healthcare engineering, 2023, 3563696.
4. Kieu STH, Bade A, Hijazi MHA, Kolivand H. A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions. *Journal of Imaging*. 2020; 6(12):131
5. Galla Siva Sai Bindhika, Munaga Meghana, Manchuri Sathvika Reddy, Rajalakshmi, Heart Disease Prediction Using Machine Learning Techniques, International Research Journal of Engineering and Technology, Volume: 07 Issue: 04, 2020
6. Rahman A.U., Alsenani, Y., Zafar, A, Enhancing heart disease prediction using a self-attention-based transformer model. Sci Rep 14, 514 (2024).
7. Md Mamun Ali, Bikash Kumar Paul, Kawsar Ahmed, Francis M. Bui, Julian M.W. Quinn, Mohammad Ali Moni, Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison, Computers in Biology and Medicine, Volume 136, 2021, 104672.
8. Hassan C. A. U., Iqbal J., Irfan R., Hussain S., Algarni A. D., Bukhari S. S. H., Alturki, N., Ullah S. S., Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers. *Sensors*, *22*(19), 7227, 2022.
9. Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective Heart Disease Prediction Using Machine Learning Techniques. *Algorithms*. 2023; 16(2):88.
10. Hossain M. I., Maruf, M. H., Khan, M. A. R., Prity, F. S., Fatema, S., Ejaz, M. S., & Khan, M. A. S. (2023). Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison. *Iran Journal of Computer Science*, 1–21. Advance online publication.
11. Fatima Yousaf, Sajid Iqbal, Nosheen Fatima, Tanzeela Kousar, Mohd Shafry Mohd Rahim, Multi-class disease detection using deep learning and human brain medical imaging, Biomedical Signal Processing and Control, Volume 85, 2023, 104875.
12. Ahmet Kursad Poyraz , Sengul Dogan, Erhan Akbal, Turker Tuncer, Automated brain disease classification using exemplar deep features, Biomedical Signal Processing and Control, Volume 73, 2022, 103448.
13. Vidhya J; Rakshana R, Advanced Detection of Brain Disease using ML and DL Algorithm, International Conference on Edge Computing and Applications, pp. 1170-1174, 2022.
14. Khaliki, M.Z., Başarslan, M.S. Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. Sci Rep 14, 2664 (2024).
15. ZainEldin, H., Gamel, S. A., El-Kenawy, E. M., Alharbi, A. H., Khafaga, D. S., Ibrahim, A., Talaat, F. M. (2022). Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization. Bioengineering, 10(1), 18.
16. Mohammad Alshraideh, Najwan Alshraideh, Abedalrahman Alshraideh, Yara Alkayed, Yasmin Al Trabsheh, Bahaaldeen Alshraideh, Enhancing Heart Attack Prediction with Machine Learning: A Study at Jordan University Hospital, Applied Computational Intelligence and Soft Computing, vol. 2024, Article ID 5080332, 16 pages, 2024.
17. Xuchun Wang , Hao Ren, Jiahui Ren , Wenzhu Song, Yuchao Qiao , Zeping Ren b , Ying Zhao b , Liqin Linghu a , b , Yu Cui a , Zhiyang Zhao a , Limin Chen c , ∗, Lixia Qiu, Machine learning-enabled risk prediction of chronic obstructive pulmonary disease with unbalanced data, Computer Methods and Programs in Biomedicine 230 (2023) 107340.
18. D. Gunasinghe, A. C. Aponso and H. Thirimanna, "Early Prediction of Lung Diseases," 2019 IEEE 5th International Conference for Convergence in Technology, 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033668.
19. McDowell, A., Kang, J., Yang, J. et al. Machine-learning algorithms for asthma, COPD, and lung cancer risk assessment using circulating microbial extracellular vesicle data and their application to assess dietary effects. Exp Mol Med 54, 1586–1595 (2022).
20. Mitra D., Rakshit, P., Jha, A., Dugar, D., Iqbal K, Lung Disease Prediction Using Deep Learning. In: Dhar, S., Do, DT., Sur, S.N., Liu, H.CM., Advances in Communication, Devices and Networking. Lecture Notes in Electrical Engineering, vol 902, 2023.
21. Nageswaran, S., Arunkumar, G., Bisht, A. K., Mewada, S., Kumar, J. N. V. R. S., Jawarneh, M., & Asenso, E. (2022). Lung Cancer Classification and Prediction Using Machine

Learning and Image Processing. BioMed research international, 2022, 1755460.
22. Abobaker M.Q. Farhan, Shangming Yang, Abdulrahman Q.S. Al-Malahi, Mugahed A. Al-antari, MCLSG:Multi-modal classification of lung disease and severity grading framework using consolidated feature engineering mechanisms, Biomedical Signal Processing and Control, Volume 85, August 2023, 104916