ESP IJACT
ESP International Journal of Advancements in Computational Technology
Volume 1 / Issue 1 Paper Id: IJACT-V1I1P104 / Page: 19-24

Original Article

Improve Disease Detection Performance by Reducing Risk Levels using the Classification Approach

Rutuja A Gulhane[1], Sunil R Gupta[2]

[1]*Research Scholar, Department of CSE, PRMIT&R, Badnera, Maharashtra, India.*
[2]*Assistant Professor, Department of CSE, PRMIT&R, Badnera, Maharashtra, India.*

Received Date: 09 March 2023 Revised Date: 23 March 2023 Accepted Date: 08 April 2023

Abstract: *Healthcare systems worldwide rely on health informatics to create electronic medical records (EMRs) to capture a wide range of information about patient health. Electronic medical records have the potential to transform the healthcare system by providing higher-quality care to patients. Clinical data may be more readily available if electronic medical records are available. This will aid in the identification of novel illness patterns, and the processing of this vast amount of data will aid in the delivery of individualized care to the patients. To enable quick understanding and automated processing of EMR data, the machine learning approach is used. In particular, machine learning methods obtain approval from the research community to predict disease with greater accuracy. Applying machine learning methodologies to the EMR dataset provides valuable information for health risk analysis and associated complications. This research proposes an effective system for predicting early disease and risk analysis based on machine learning algorithms to improve delivery accuracy with less risk and an auto notification approach.*

Keywords: *Accuracy, Electronic Medical Records (EMR), Healthcare Systems, Machine Learning.*

I. INTRODUCTION

The healthcare industry is a crucial aspect of modern society, directly affecting people's lives. However, the industry is fragmented and lacks adequate patient information when and where it is needed, leading to diagnostic delays and errors. The vast diversity in test-ordering for diagnostic purposes highlights the need for an appropriate and adequate test set. Researchers and practitioners analyzing medical data face the challenge of the curse of dimensionality and heterogeneity in feature sources and statistical qualities. Therefore, there is an urgent requirement for reliable and accurate methodologies that facilitate early disease identification and support physicians in decision-making.

The traditional paradigms are unable to cope with the massive and ever-growing volume of digitally stored medical data, requiring advances in data management and analysis. This developing technology is known as Big Data, which investigates intricate and ever-changing connections between data. Data mining, a subfield of Big Data, searches for and identifies functional patterns in data to make better decisions. It is a complex task to identify these patterns, requiring cutting-edge methods for characterizing data-structure patterns. Artificial Intelligence is the discipline where most of these methods were pioneered, making it crucial to the medical, computational, and statistical sciences for modeling the prognosis and diagnosis of diseases.

II. RELATED WORK

The previous AI models for disease detection were focused on identifying a single disease per examination, but the proposed system allows for multiple illnesses to be detected using machine learning classification techniques. The study focuses on the challenges inherent in automatic disease detection, such as analyzing images taken with a light microscope and detecting malaria parasites using K-means clustering. Researchers have explored various machine learning algorithms for disease detection, including SVM, LR, KNN, and ANN. CNNs have also been used for disease identification and localization in the thoracic region and for identifying pulmonary tuberculosis. Pre-processing techniques such as color-based transformation, image enhancement, noise reduction, scaling, and segmentation have been employed to improve the accuracy of disease detection. The convolution neural network (CNN) is a commonly used method for object detection and semantic segmentation, using layers of convolution, pooling, and fully connected layers. The convolution layer plays a crucial role in distributing the network's compute load by calculating a dot product between the kernel and the image, with the kernel's depth being deepened while the height and breadth are shrunk.

Zhe Li et al. [14] proposed a method for thoracic disease identification and localization with limited supervision. Another study focused on liver illness identification in patients using machine learning algorithms such as SVM, LR, KNN, and ANN. All models were evaluated based on predetermined performance metrics, and the most accurate model was found to be ANN with a success rate of 98%. The results indicated that ANN was significantly more effective than earlier studies.

Afifi et al. [15] proposed a CNN-based approach for automatic identification of diseases with limited data, and they demonstrated its effectiveness in identifying pulmonary tuberculosis. Mique et al. [16] also used CNNs for anomaly identification in chest X-rays, and they released a larger dataset of frontal chest X-rays for researchers to use in their investigations of machine learning for chest cancer detection.

One of the challenges in using machine learning for medical image analysis is the need for extensive feature extraction. Since image categorization is a tedious process that needs to be performed whenever there is a shift in the problem or data used to solve it, computer vision systems rely on custom-built functions and algorithms for enhanced images. However, deep learning and logical regression approaches can solve this problem of manual feature extraction, as they can automatically extract features from the images.

Various machine learning strategies have been employed in the context of disease diagnosis, including Artificial Neural Networks (ANNs), Decision Trees, K-Means, K-Nearest Neighbors (KNN), and Vector Support Machines (SVMs). These methods have been extensively studied for disease detection and categorization. For instance, Li et al. [14] applied SVM, LR, KNN, and ANN machine learning methods to identify liver illness in patients and showed that ANN was the most accurate model, achieving a success rate of 98%. Moreover, in the first step, RGB images and various master learning techniques are used to diagnose the illness, and this strategy yielded an 80% success rate in most cases. The use of convolutional neural networks (CNNs) in automatic identification and classification of images, particularly in the context of plant disease diagnosis. CNNs have been shown to provide superior results compared to traditional methods, and transfer learning is often used to leverage pre-trained models for domain-specific tasks. Pre-processing methods such as color-based transformations, image enhancement, noise reduction, scaling, and segmentation are used to improve segmentation precision. Extraction methods based on textures, colors, and shapes are also utilized. The convolution layer is a crucial part of any CNN system, and the appropriate filter is used to capture interdependencies between spatial and temporal dimensions. The availability of a large, labeled, and validated dataset of plant images has enabled the development of deep neural networks and machine learning models for accurate disease diagnosis. Kaggle provides such datasets for training these models.

III. PROPOSED APPROACH

A meta-heuristic method is proposed that integrates RDA (Red Deer Algorithm) and DA (Dragonfly Algorithm) methodologies. Realizing the Efficient Recurrent Neural Network (E-RNN) model is made possible by this strategy, which optimizes number of epochs and "number of hidden neurons" of RNN. Because of this, the performance can be maximized despite having detection rate that is guaranteed to be high for both breast & heart cancer [20].

A novel algorithm that is proposed here takes parts and pieces from the RDA approach as well as the DA method. RDA has been chosen for this task because of its rapid convergence to globally optimal solutions, its increased velocity of exploration and exploitation, and its capacity to strike a better balance between the two processes. RDA, on the other hand, struggles to find global solutions and to fine-tune a wide variety of "controlling parameters, fewer execution speeds." A novel algorithm that has been developed utilizes features of the DA approach to accomplish the goals of achieving greater harmony between the phases, increasing the variety of solutions, and boosting performance. Implementation is simplified when there are fewer limits placed on the controls. DA with RDA is utilized to accomplish providing superior solutions while maintaining a convergence rate that is reasonable.

In DA, the updating of the solutions is accomplished through the utilization of a uniformly distributed random vector between 0 and 1. As a result, the proposed approach is based on the sensing area. In this particular instance, the random number is determined by dividing the mean of the best fitness solutions by the best fitness solution. When this condition is met, either the DA method or the RDA algorithm is applied to the problem to update the answers and the solutions, respectively.

ML Pseudo code:

i. Dataset of Training
ii. Dataset of Testing
iii. Checking the shape/features of the input
iv. *e procedure of initiating the sequential layer
v. Adding dense layers with dropout layers and Rectified Linear Unit (ReLU) activation functions
vi. Adding a last dense layer with one output and binary activation function
vii. End repeat
viii. L (Output)

Some of the basic features are as follows:

- Predicts about different diseases
- Drag and drop images

- Drop images to predict Pneumonia X-ray
- Drag and drop csv file to predict different diseases
- Enter custom data to predict whether the person is diagnosed with any disease
- Predicts the accuracy of given dataset
- Predicts the accuracy with 4 different classifiers for each disease with proposed approach

IV. EXPERIMENTAL RESULT

This system was created to identify symptoms and forecast disease based on medical report trained on machine learning algorithms. This system supports jpg file to predict the pneumonia x-ray image and csv file to predict any other diseases as shown in Figure 1. Numerous diseases, including Heart Disease, Diabetes, Parkinson's Disease, Pneumonia Disease, and another 42 diseases can be predicted at an early stage and analyze the risk level using the system. Figure 2 shows the disease predication approach by entering proper range of data so that predication can give best outcome.

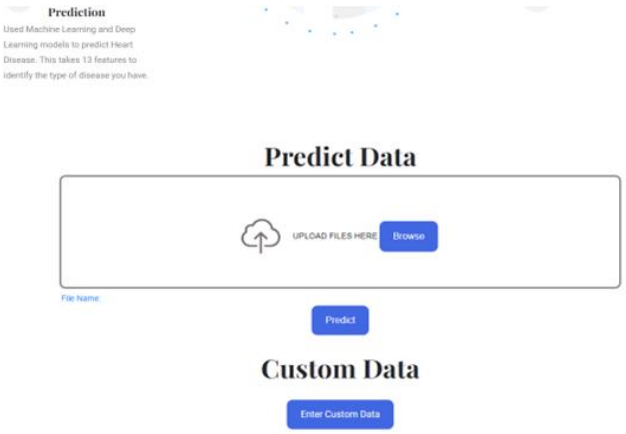

Figure 1: Entering Custom Data

The risk level for disease detection has been defined with different stages as follows:
For Positive Result:

10 - 50 % in stage 1
50 - 70 % in stage 2
70 - 100% in stage 3 (need to connect with doctor & send an auto sms on stored number)

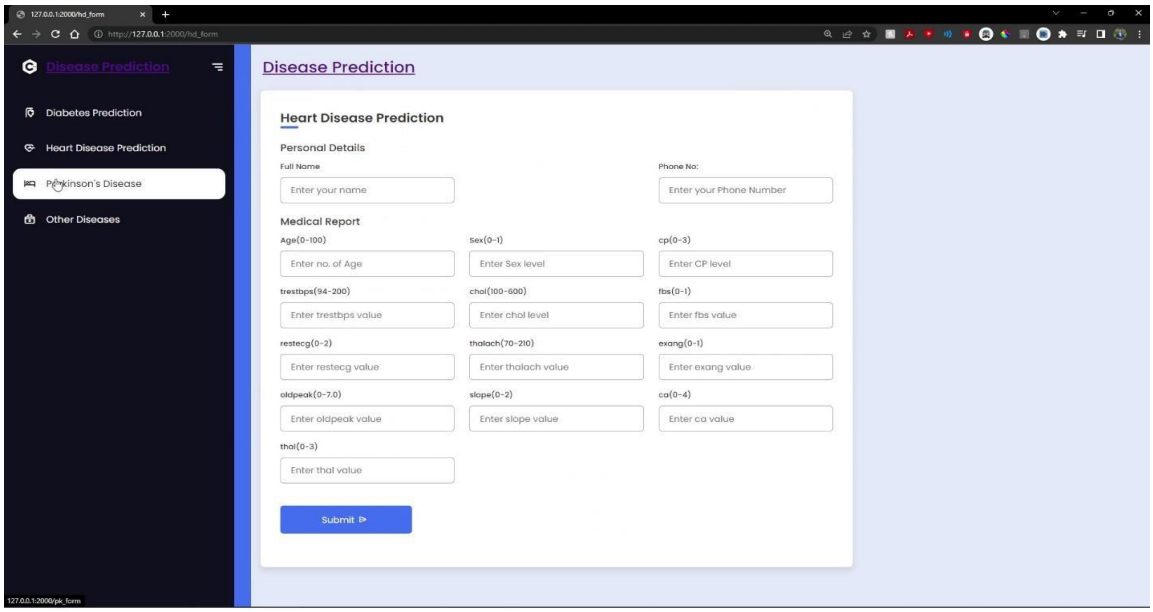

Figure 2: Disease Prediction Approach with Proper Range for Best Outcome

For Negative Result:
10 - 50 % in stage 1

50 - 70 % in stage 2
70 - 100% in stage 3 (need to connect with doctor, person may cause disease after 1.5 years if not take proper precaution by consulting to doctor on regular basis and also send an auto sms on stored number)

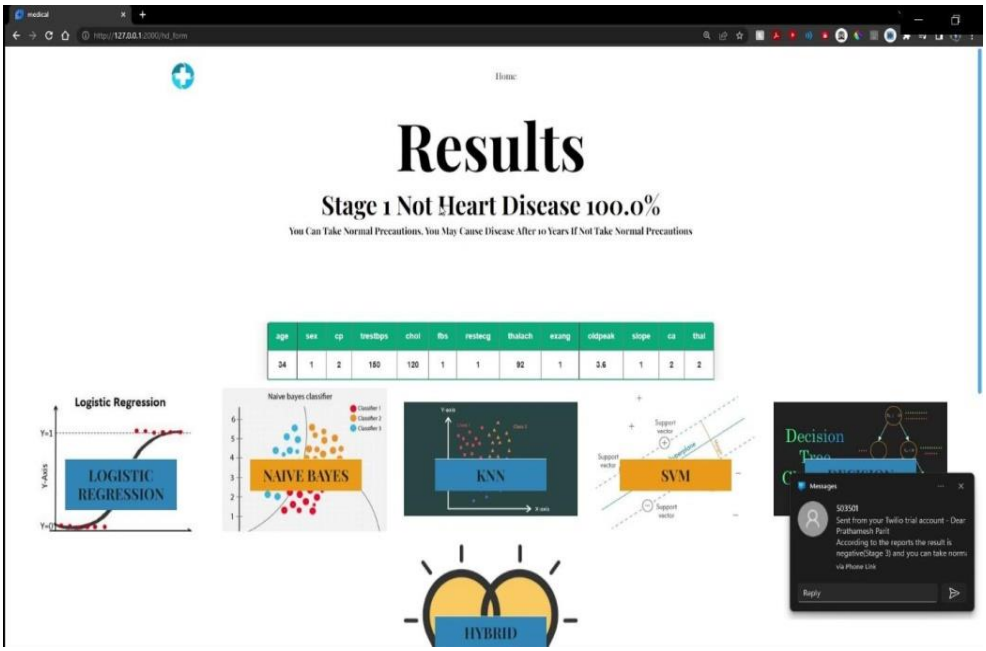

Figure 3: Stage-wise Disease Prediction with Negative Result and Auto SMS System

Figure 3 & Figure 4 shows that how the proposed method works in practice, which depicts a Stage-wise Disease Prediction with a Positive, Negative Result and an Automatic SMS System. The performance analysis suggests that this method outperforms the others in precision and automatic notification as shown in Table I.

The work is more precise by using novel hybrid meta-heuristic technique (HDO) than the existing disease detection system. The most exceptional accuracy levels for the aforementioned disorders are 85.30 percent, 79.12 percent and 95.67 percent by utilizing KNN for heart disease, LR for diabetes and breast tumor.

Figure 4: Stage-wise Disease Prediction with Positive Result and Auto SMS System

Table 1: Performance Analysis of the Proposed Approach

Disease	LR	SVM	KNN	HDD (Novel Approach)
Heart Disease	82.50	78.87	83.84	85.30
Diabetes	77.60	75.64	75.52	79.12
Breast Cancer	94.55	91.38	92.55	95.67

V. CONCLUSION

According to the analysis, most researchers only analyze structured data for the prediction system, even though unstructured texts include a significant amount of relevant medical information. This is something that may be seen. Due to the absence of a common structure, it is essential to process and evaluate unstructured information using a practical algorithm. The study that has the potential to revolutionize the healthcare business is based on the availability of a vast amount of clinical data in the form of electronic medical records. It is a significant difficulty to understand the data in electronic medical records. This is especially true when one considers the significance of the information contained in electronic medical records and their capacity to make the delivery of treatment more efficient.

Machine learning has contributed substantially to the challenge of extracting characteristics from input datasets to address predictive issues. These problems include the prediction of diseases, decision assistance, and the analysis of risk anomalies. As a consequence of this, predictive models are utilized in the process of disease diagnosis and prognosis. Therefore, a novel hybrid meta-heuristic technique (HDO) is proposed to improve the performance of disease prediction system and achieve more accurate predictions with less risk by automated notification approach.

Interest Conflicts

There are no conflicts of interest to declare. There are no relevant financial or non-financial interests to disclose that are relevant to the content of this manuscript.

Funding Statement

There are no funds, grants, or other support was received during the preparation of this manuscript.

Acknowledgement

The authors did not receive support from any organization for the submitted work.

VI. REFERENCES

[1] M. Ghassemi, T. Naumann, P. Schulam, L. A. Beam, I. Chen, and R. Ranganath, A Review of Challenges and Opportunities in Machine Learning for Health, University of Toronto and Vector Institute, Toronto, Canada, (2020) 191–200.

[2] J. Lopes, T. Guimarães, M. F. Santos, J. Lopes, T. Guimarães, and M. F. Santos, Predictive and Prescriptive Analytics in Healthcare: A Survey, Procedia Computer Science, Published by Elsevier, 170 (2020) 1029–1034. doi: 10.1016/j.procs.2020.03.078.

[3] K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, Prescriptive analytics: Literature review and research challenges, International Journal of Information Management, 50 (2020) 57–70. doi: 10.1016/j.ijinfomgt.2019.04.003.

[4] K. Hasan, A. Alam, D. Das, and E. H. Senior, Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers, IEEE Access, Special Section on Deep Learning Algorithms for Internet of Medical Things, X (2020) 1–19. doi: 10.1109/ACCESS.2020.2989857.

[5] F. Ali, S. Sappagh, S. M. R. Islam, D. Kwak, A. Ali, M. Imran, and K. S. Kwak, A Smart Healthcare Monitoring System for Heart Disease Prediction Based On Ensemble Deep Learning and Feature Fusion, Information Fusion, Published by Elsevier, (2020). doi: 10.1016/j.inffus.2020.06.008.

[6] A. Ed-Daoudy and K. Maalmi, Real-time machine learning for early detection of heart disease using big data approach, Int. Conf. Wirel. Technol. Embed. Intell.Syst., (2019) 1– 5.doi: 10.1109/WITS.2019.8723839.

[7] M. S. Satu, F. Tasnim, T. Akter, and S. Halder, Exploring Significant Heart Disease Factors based on Semi Supervised Learning Algorithms, Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2, (2018). doi: 10.1109/IC4ME2.2018.8465642.

[8] S. Mohan, C. Thirumalai, and G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access, 7 (2019) 81542–81554. doi: 10.1109/ACCESS.2019.2923707.

[9] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, R. Sun, and I. Garciá-Magarinõ, A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms, Mob. Inf. Syst., (2018). doi: 10.1155/2018/3860146.

[10] R. Sharmila and S. Chellammal, A conceptual method to enhance the prediction of heart diseases using big data Techniques, (4) (2018) 21–25.[Online]. Available: www.ijcseonline.org.

[11] M. ChalaBeyene, Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques, 118 (8) (2018) 165–174.[Online]. Available: http://www.ijpam.eu.

[12] N. A. Khan, H. Pervaz, A. Latif, and A. Musharaff, Unsupervised identification of malaria parasites using computer vision, (2017).

[13] Benjamin Antin, Joshua Kravitz and Emil Martayan, Detecting Pneumonia in Chest X-Rays with Supervised Learning, (2017).

[14] Zhe Li, Chong Wang, Mei Han, Yuan Xue, Wei Wei, Li-Jia Li, et al., Thoracic disease identification and localization with limited supervision, (2017).

[15] Afifi, Ahmed, AbdulazizAlhumam, and AmiraAbdelwahab, Convolutional Neural Network for Automatic Identification of Diseases with Limited Data, 10 (1) (2021).

[16] MiqueJr, Eusebio L., and Thelma D. Palaoag, Disease detection using convolutional neural network, Proceedings of the 2018 International Conference on Information Science and System, (2018).

[17] Bharti, R.; Khamparia, A.; Shabaz, M.; Dhiman, G.; Pande, S.; Singh, P., Prediction of heart disease using a combination of machine learning and deep learning, Comput. Intell.Neurosci., (2021).

[18] Naz, H.; Ahuja, S., Deep learning approach for diabetes prediction using PIMA Indian dataset, J. Diabetes Metab. Disord., 19 (2020) 391–403.

[19] Buvaneswari R, Vinoth R, 2022. "Kidney Stone Detection Using Hybrid Butterfly Net and Inception net Model"ESP Journal of Engineering & Technology Advancements 2(4): 45-50.

[20] Gavhale, Kiran R., and UjwallaGawande, An overview of the research on disease detection using image processing techniques, IOSR Journal of Computer Engineering (IOSR-JCE), 16 (1) (2014) 10-16.

[21] Zheng, B.; Yoon, S.W.; Lam, S.S., Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms, Expert Syst. Appl., 41 (2014) 1476–1482.

[22] Gokul Ramadoss, "Leveraging AI in ETL / ELT Designs for Enhanced Health Risk Assessment", International Journal of Science and Research (IJSR), Volume 13 Issue 8, August 2024, pp. 262-265, https://www.ijsr.net/getabstract.php?paperid=SR24802003019

[23] Gokul Ramadoss. (2021). Leveraging Affordable Care Act to Improve Global Healthcare. European Journal of Advances in Engineering and Technology, 8(5), 41–44. https://doi.org/10.5281/zenodo.13789625