Instamedi: Healthcare Consultation System

KAMRAN RASHID THOKER[1], DEEPAK SINGH[2], ABHISHEK[3]
[1, 2, 3] *Department of Computer Science and Engineering, Inderprastha Engineering College, Uttar Pradesh Ghaziabad (U.P.), India*

Abstract— ***The proliferation of computer-based technology in healthcare has led to a surge in electronic data, posing challenges for medical practitioners in accurately analyzing symptoms and diagnosing diseases early. Supervised machine learning (ML) algorithms offer a promising solution, surpassing standard systems in disease detection. This review aims to identify trends in disease detection across various supervised ML models, including Naïve Bayes, Decision Trees, and K-Nearest Neighbor. Support Vector Machine emerges as adept in detecting kidney and Parkinson's diseases, while Logistic Regression excels in heart disease prediction. Furthermore, Random Forest and Convolutional Neural Networks demonstrate precision in breast disease and common disease prediction, respectively. This analysis underscores the potential of ML in enhancing healthcare diagnostics.***

Index Terms— Health Care, Supervised Machine Learning, Disease Prediction.

I. INTRODUCTION

A. Motivation

The rise of Artificial Intelligence (AI) has empowered computerized systems to emulate human-like perception, cognition, and decision-making abilities. AI encompasses a diverse array of disciplines, including Machine Learning (ML), Computer Vision, Deep Learning, and Natural Language Processing. ML algorithms, in particular, leverage optimization, statistical, and probabilistic techniques to glean insights from past data and inform decision- making processes. This versatility has led to their widespread application across various domains, from network intrusion detection to customer behavior analysis and disease diagnosis.

Notably, many of these applications harness the power of supervised learning, wherein datasets with known labels are utilized to train predictive models, offering a potent tool for medical practitioners to enhance disease diagnosis efficiency.

The prevalence of chronic diseases presents a formidable challenge for healthcare systems globally. In the United States, Medicaid services and centers for Medicare have reported that half of Americans suffer from multiple chronic conditions, contributing to staggering healthcare expenditures totaling $3.3 trillion in 2016.

Similarly, the World Health Organization and World Economic Forum have highlighted the substantial economic losses incurred by countries like India due to fatal diseases stemming from malnutrition and unhealthy lifestyles. These statistics underscore the critical importance of early disease detection in mitigating the impact of such ailments, not only in terms of reducing mortality rates but also alleviating the financial strain on healthcare systems and fostering community well-being.

However, the efficacy of ML algorithms in disease diagnosis is contingent upon several factors. Firstly, the quality and selection of datasets are paramount in ensuring accurate and unbiased predictions. Secondly, the extraction of relevant features from these datasets poses challenges, often requiring significant computational resources. Despite these challenges, traditional methods relying solely on standard statistical techniques and the intuition of medical practitioners have been prone to biases and errors. With the exponential growth of electronic health data, the need for advanced computational methodologies such as ML algorithms has become increasingly apparent. By uncovering meaningful patterns and hidden insights from data, these algorithms facilitate critical decision-making processes, ultimately reducing the burden on medical staff and improving patient survival rates.

B. Aim

This study aims to empirically validate the hypothesis that supervised ML algorithms hold the potential to revolutionize healthcare by enabling accurate and early disease detection. To achieve this objective, we scrutinize studies employing multiple supervised ML models for disease recognition. This approach ensures comprehensive and precise evaluations, mitigating bias and yielding more reliable results. Specifically, we focus on diseases affecting the heart, kidney, breast, and brain, evaluating various ML methodologies such as K-Nearest Neighbor (KNN), Naïve Bayes (NB), Decision Trees (DT), Convolutional Neural Networks (CNN), Support Vector Machine (SVM), and Logistic Regression (LR). By identifying the best-performing ML models for each disease category, this literature review seeks to inform future research and clinical practice, paving the way for more effective disease diagnostics and interventions in healthcare settings.

II. LITERATURE REVIEW

A. Common Diseases

Dahiwade et al. [9] introduced a machine learning (ML) system for predicting common diseases, leveraging a dataset sourced from the UCI ML repository containing symptoms of various common ailments. Employing Convolutional Neural Networks (CNN) and K-Nearest Neighbor (KNN) as classification techniques, the study incorporated additional patient lifestyle information to enhance disease risk assessment. Comparison between KNN and CNN revealed CNN's superiority in terms of accuracy and processing time, echoing findings by Chen et al. [10] suggesting CNN's outperformance of traditional supervised algorithms due to its capability to discern complex nonlinear relationships in feature spaces. However, limitations include the lack of detailed neural network parameters and exclusive reliance on accuracy metrics, neglecting bias considerations and potential enhancements through feature augmentation [9].

B. Kidney Diseases

Serek et al. [12] conducted a comparative study on classifiers' performance for Chronic Kidney Disease (CKD) detection, utilizing the Kidney Function Test (KFT) dataset. Random Forest (RF) emerged as superior in terms of F-measure and accuracy, while Naïve Bayes (NB) exhibited better precision. Similarly, Vijayarani [13] focused on detecting kidney diseases using Support Vector Machine (SVM) and NB, highlighting SVM's higher accuracy albeit with longer execution times compared to NB. Other studies by Charleonnan et al. [14] and Kotturu et al. [15] also favored SVM for its adaptability to semi-structured and unstructured data. However, the lack of hyperparameter exploration limits the robustness of conclusions, as emphasized by Uddin [3], who advocates for comprehensive parameter tuning to optimize ML model performance.

C. Heart Diseases

Marimuthu et al. [16] endeavored to predict heart diseases using supervised ML techniques, reporting Logistic Regression's (LR) superiority with an accuracy of 86.89%. Dwivedi [17] expanded on this by including additional parameters such as resting blood pressure and serum cholesterol, concluding LR's high reliability in heart disease detection. Polaraju [18] and Vahid et al. [19] echoed LR's effectiveness compared to other techniques like Artificial Neural Networks (ANN) and SVM, attributing LR's success to extensive hyperparameter exploration. However, limitations arise from the small dataset sizes, constraining model performance for diseases requiring higher accuracy and precision.

D. Breast Diseases

Shubair [20] investigated breast cancer detection using ML algorithms, finding SVM's superiority in terms of recall, accuracy, and precision, while Random Forest (RF) excelled in tumor classification. Conversely, Yao [21] favored RF over SVM due to its better information gain estimates and scalability for large datasets. Nonetheless, challenges emerged from preprocessing steps impacting data quality, potentially hindering ML model performance [21].

E. Parkinson's Disease

Chen et al. [22] introduced an effective diagnosis system for Parkinson's disease (PD), showcasing Fuzzy k-Nearest Neighbor's (FKNN) superiority over SVM. Behroozi [23] proposed a classification framework enhanced by feature selection algorithms, with SVM exhibiting superior performance metrics. Eskidere [24] focused on PD progression tracking, favoring Least Square Support Vector Machine (LS-SVM) as the highest performing model.

While frameworks were detailed, lack of model calibration prior to performance evaluation poses a limitation, as calibration enhances classification for certain ML models [26].

Overall, these studies highlight the potential of ML algorithms in disease prediction, but underscore the importance of comprehensive model evaluation and parameter tuning to optimize performance and address limitations.

CONCLUSION

The Conclusion: The application of various machine learning (ML) algorithms has significantly advanced the early detection of a myriad of diseases, encompassing conditions affecting the heart, kidneys, breast, and brain. Throughout the literature, Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR) emerged as the predominant choices for disease prediction, with accuracy being the primary performance metric of evaluation. Notably, Convolutional Neural Networks (CNN) showcased exceptional aptitude in predicting common diseases, while SVM consistently demonstrated reliability in handling high-dimensional, semi-structured, and unstructured data for kidney diseases and Parkinson's Disease (PD). RF exhibited superior probability in the correct classification of breast cancer, attributed to its scalability for large datasets and adeptness in avoiding overfitting. LR emerged as the most dependable in predicting heart diseases.

In future research endeavors, the development of more intricate ML algorithms is imperative to further enhance the efficiency of disease prediction. Additionally, there is a need for more frequent calibration of learning models post-training to potentially optimize performance.

Expansion of datasets across diverse demographics is essential to mitigate overfitting and enhance model accuracy. Moreover, the utilization of more robust feature selection methods holds promise for enhancing the performance of ML models, thus contributing to more accurate and reliable disease prediction methodologies.

REFERENCES

[1] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, *"Prediction of heart disease using machine learning."* in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1275-1278.

[2] Y. Hasija, N. Garg, and S. Sourav, *"Automated detection of dermatological disorders through image-processing and machine learning."* in 2017 International Conference on Intelligent Sustainable Systems (ICISS). 2017, pp. 1047-1051.

[3] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, *"Comparing different supervised machine learning algorithms for disease prediction,"* BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1- 16, 2019.

[4] R. Katarya and P. Srinivas, *"Predicting heart disease at early stages using machine learning: A survey,"* in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 302-305.

[5] P. S. Kohli and S. Arora, *"Application of machine learning in disease prediction,"* in 2018 4th International Conference on Computing Com- munication and Automation (ICCCA), 2018, pp. 1- 4.

[6] M. Patil, V. B. Lobo, P. Puranik, A. Pawaskar, A. Pai, and R. Mishra, *"A proposed model for lifestyle disease prediction using support vector machine,"* in 2018 9th International Conference on Computing. Communication and Networking Technologies (ICCCNT), 2018, pp. 1-6.

[7] F. Q. Yuan, *"Critical issues of applying machine learning to condition monitoring for failure diagnosis,"* in 2016 IEEE International Confer- ence on Industrial Engineering and Engineering Management (IEEM), 2016, pp. 1903-1907.

[8] S. Ismaeel, A. Miri, and D. Chourishi, *"Using the extreme learning ma- chine (elm) technique for heart disease diagnosis,"* in 2015 IEEE Canada International Humanitarian Technology Conference (IHTC2015), 2015, pp. 1-3.