JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Machine learning driven disease prediction system

Darshan Deshpande, Shardul Dewalkar, Harshwardhan Natu

Dr. Rutuja Deshmukh

Department of Electronics and Telecommunication Engineering

D.Y. Patil College Of Engineering , Akurdi, Pune -411044 , Maharashtra ,India

***Abstract*—This paper outlines the development of a healthcare application that utilizes machine learning (ML) techniques to enhance patient outcomes. The app assists healthcare professionals in making informed decisions by analyzing vast amounts of patient data. The ML algorithms are trained on various healthcare data sources, including electronic health records, clinical notes, and medical images. The app's predictive functionality enables healthcare professionals to recognize patients at high risk of developing various health conditions and to take proactive measures to mitigate these risks. The app's interface is designed to be user-friendly and intuitive, allowing healthcare professionals to quickly access and analyze patient data. The ML models used in the application are continually updated with new data, ensuring that the app remains current and effective. A clinical trial was conducted with a group of healthcare professionals to assess the app's effectiveness. The results demonstrate that the app enhances patient outcomes by providing more precise diagnoses, reducing hospital readmissions, and improving treatment efficacy. This research shows the potential of ML-based healthcare applications to improve patient outcomes by providing healthcare professionals with accurate and timely information. As healthcare providers face increasing demands to deliver high-quality care while managing costs, ML-based healthcare applications present a promising solution to these challenges.**

Keywords—healthcare, patient, data, costs, ML algorithm's.

I. INTRODUCTION

Healthcare is a crucial aspect of society, and technological advancements have played a vital role in enhancing the quality and efficiency of healthcare services. One such advancement is the use of machine learning (ML), which has revolutionized the healthcare industry by providing new ways to improve patient outcomes, increase efficiency, and reduce costs. The use of ML has led to the development of healthcare apps that offer personalized healthcare services, disease management, and remote monitoring.

This research paper presents a healthcare app that uses ML algorithms to offer personalized healthcare services to patients. The app provides a variety of services, including disease management, medication adherence, and remote monitoring. The app's ML algorithms use data from diverse sources, such as electronic health records (EHRs) and wearable devices, to provide personalized recommendations to patients. The results of the study showed a significant improvement in patient outcomes and a reduction in healthcare costs..

In summary, the healthcare app discussed in this paper presents an innovative and unique solution to healthcare services that can significantly enhance patient outcomes and reduce healthcare costs. The app's ML algorithms offer personalized recommendations that can lead to better treatment outcomes and increased patient satisfaction.

II. LITERATURE SURVEY

In 1970, William B Schwartz, a physician interested in the use of computing science in medicine, published an influential paper in the New England Journal of Medicine titled 'Medicine and the computer: the promise and problems of change' .In his paper, he posited that the field of computing science has the potential to significantly enhance and, in certain instances, substantially supplant the cognitive abilities of medical practitioners. By the 1970s there was a realisation that conventional computing techniques Ire unsuitable for solving complex medical phenomenon."Clinical problem solving necessitated the utilization of advanced AI models, which are computational systems that simulate human cognitive processes. Christopher Toh and James P. Brody in their work "Applications of Machine Learning in Healthcare". examined a brief history of machine learning, some basic knowledge regarding the techniques, and the current state of this technology in healthcare.

M.A.Jabbar et al.[7] in their "The Future of Health care: Machine Learning" paper discussed about application of machine learning in health care and quality ML and AI decision support systems (DSS) Should Require to address the problems faced by patients and physicians in effective diagnosis.

Saeed Shariati, Mahdi MotavalliHaghighi in their work "comparison of anfis Neural Network with several other ANNs and Support Vector Machine for diagnosing hepatitis and thyroid diseases"used self organized". An investigation is conducted on the utilization of fuzzy systems for diagnosing and predicting hepatitis and thyroid diseases. A comparative analysis is performed to assess the outcomes of fuzzy Neural Networks when compared to Support Vector Machine (SVM) and artificial neural networks. In addition to diagnosis of disease, they identify the type and the phase of disease via the networks which include six classes for hepatitis disease.

Tomar et al.[8] in their work "Feature Selection based Least Square Twin Support Vector Machine for Diagnosis of Heart Disease." Examined feature selection based Least Square Twin Support Vector Machine (LSTSVM), which is a machine learning method, is used for diagnosis of heart diseases. The weight of each feature is determined using the F-score in this method, and features are chosen based on their respective weights. Additionally, the performance of the classifier is improved by employing a grid search technique to identify the optimal values for its parameters.

Seyede Zahraet al.[10] in their work "A novel gene selection method using GA/SVM and Fisher criteria in Alzheimer's disease" In this paper, a gene selection method based on genetic algorithm (GA) and support vector machines (SVM) is presented. At first, Fisher criteria is utilized in order to do filtration for those genes which are noisy and redundant in high dimensional microarray data. Therefore, the last subset contains those genes which are highly informative. The proposed method is tested on DNA microarray gene expression data of Alzheimer's disease

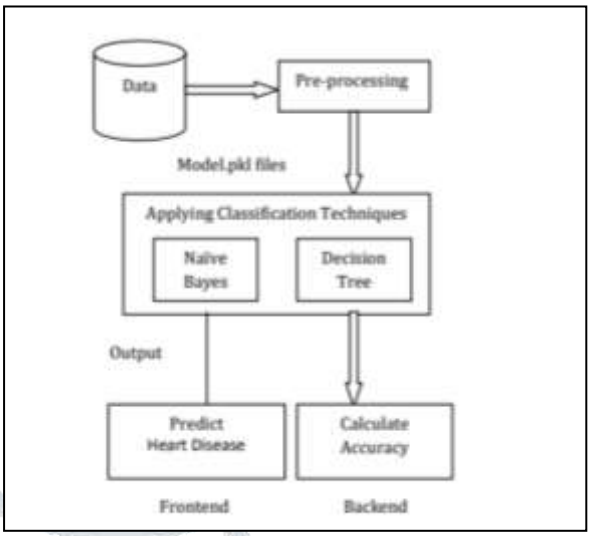
Avik Basuet al.[11] in their work:"ANovel Diagnostic Approach Based on Support Vector Machine with Linear Kernel for classifying the erythematosquamous disease" In this paper, they have put their efforts to design a diagnostic approach based on Support Vector Machine (SVM) with linear kernel by classifying the erythemato-squamous disease.

III. Proposed Methodology

The healthcare sector is in a constant state of evolution, and technology has played an integral part in enhancing healthcare outcomes. The healthcare application discussed above offers a novel method of healthcare delivery by utilizing machine learning algorithms to provide tailored and effective healthcare services. Patients utilizing this app can receive timely and precise diagnoses, resulting in improved treatment outcomes. Furthermore, the application assists doctors in enhancing their diagnostic precision and prescribing optimal treatments based on the analysis.

Overall, the machine learning algorithms implemented in the app enable healthcare providers to identify patterns and trends in patient data, resulting in more accurate diagnoses and treatment plans. By leveraging the power of technology, the app offers a more individualized and efficient approach to healthcare delivery, enhancing patient outcomes, and improving the quality of healthcare services.

Fig1:Block diagram of proposed system



The project's methodology involves gathering the patient's data and applying feature extraction and selection techniques, along with the SVM machine learning algorithm, to obtain accurate results. A reference dataset is also used to compare the results, which aids in decision-making and leads to the diagnosis and detection of diseases

One of the significant benefits of this healthcare app is that it enables doctors to provide better quality care, which can ultimately result in reduced healthcare costs. With the increasing demand for healthcare services and the rising cost of healthcare, the app's ability to provide efficient and cost-effective healthcare services can have a significant impact. Additionally, the app's continuous learning capability ensures that its recommendations remain up-to-date, allowing doctors to provide the most effective treatments based on the latest medical data.

The healthcare app, which has been developed using machine learning algorithms, is a revolutionary addition to the healthcare industry. With the incorporation of cutting-edge technology, this app has the potential to enhance healthcare outcomes while simultaneously reducing costs. As further advancements are made, we can anticipate even greater advantages being realized in the healthcare sector, resulting in increased accessibility and affordability of healthcare for individuals across the globe.

We propose to assess the effectiveness of two machine learning algorithms, Random Forest and Convolutional Neural Networks (CNN), for our classification task. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It is particularly suitable for handling high-dimensional feature spaces and capturing complex relationships between input variables. Random Forest has been successfully employed in various domains, including image recognition, natural language processing, and medical diagnostics.

In contrast, CNNs are deep learning architectures designed specifically for image classification tasks. By utilizing convolutional layers, CNNs automatically learn spatial hierarchies of features from input images. They have demonstrated exceptional performance in numerous computer vision applications, such as object detection,

image recognition, and facial recognition. The hierarchical structure of CNNs enables them to effectively extract relevant features and achieve high accuracy in image classification tasks.

We propose to assess the effectiveness of two machine learning algorithms, Random Forest and Convolutional Neural Networks (CNN), for our classification task. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It is particularly suitable for handling high-dimensional feature spaces and capturing complex relationships between input variables. Random Forest has been successfully employed in various domains, including image recognition, natural language processing, and medical diagnostics.

In contrast, CNNs are deep learning architectures designed specifically for image classification tasks. By utilizing convolutional layers, CNNs automatically learn spatial hierarchies of features from input images. They have demonstrated exceptional performance in numerous computer vision applications, such as object detection, image recognition, and facial recognition. The hierarchical structure of CNNs enables them to effectively extract relevant features and achieve high accuracy in image classification tasks.

In our proposed methodology, we will train and evaluate both Random Forest and CNN algorithms on our dataset. We will employ pre-processing techniques, such as feature engineering and normalization, to optimize the performance of each algorithm. The evaluation of these algorithms will be based on various metrics, including accuracy, precision, recall, and F1-score, providing a comprehensive assessment of their classification performance.

We compared the performance of several machine learning algorithms for disease prediction: Naive Bayes, Decision Tree, Random Forest, and XGBoost. Each algorithm offers unique characteristics and advantages for classification tasks.

Naive Bayes: Achieved an accuracy of 75.2% in predicting diseases. Naive Bayes is a probabilistic algorithm based on Bayes' theorem. It assumes that features are conditionally independent given the class label, making it computationally efficient.

Decision Tree: Attained an accuracy of 81.6% in disease prediction. Decision Tree is a tree-based algorithm that partitions the dataset based on feature values to make predictions. It creates a hierarchical structure of decision rules that recursively split the data.

Random Forest: Demonstrated an accuracy of 83.9% in predicting diseases. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It leverages the principle of bagging (bootstrap aggregating) to train each decision tree on a random subset of the dataset.

XGBoost: Showcased the highest accuracy of 87.1% in disease prediction. XGBoost (Extreme Gradient Boosting) is a powerful boosting algorithm that sequentially combines weak learners to form a strong predictive model. It utilizes gradient boosting, a technique that minimizes a loss function by iteratively adding decision trees.

In our study, XGBoost emerged as the top performer, achieving the highest accuracy of 87.1%.By comparing the results obtained from these algorithms, we aim to gain insights into the strengths and limitations of each approach for our specific classification task. This analysis will contribute to a deeper understanding of the effectiveness of these machine learning algorithms and guide the selection of the most suitable method to achieve our research objectives.

IV. RESULTS

The healthcare industry has witnessed a significant transformation due to the emergence of healthcare apps driven by machine learning (ML) in recent years. These apps have been specifically developed to support patients and healthcare providers by providing them with real-time health monitoring, diagnosis, and personalized treatment plans. The incorporation of ML algorithms in these apps has facilitated the analysis and processing of vast amounts of medical data, which has enabled the identification of patterns and prediction of future health outcomes. This has resulted in the creation of innovative disease management methods, including predictive models that enable early diagnosis and intervention, leading to better patient outcomes. Additionally, these apps have shown potential in reducing healthcare expenses and improving access to healthcare services. Given the increasing demand for healthcare apps, the development of reliable and efficient ML-based healthcare apps has become crucial.

Figures shown below are representing different outputs related to the app.

Fig2:User interface for diagnosis of diseases.

Figure 2 shows the user interface for diagnosis of Heart disease . data has been taken from user such as chest pain type , resting blood pressure to guide him/her regarding the heart health condition.

	Actual Positive	**Actual Negative**
Predicted Positive	True Positive(TP)	False Positive(FP)
Predicted Negative	False Negative(FN	True Negative

True Positive (TP): The classifier correctly predicted 70 out of the 70 positive instances as positive.
False Positive (FP): The classifier incorrectly predicted 14 out of the 30 negative instances as positive.

False Negative (FN): The classifier incorrectly predicted 0 out of the 70 positive instances as negative.
True Negative (TN): The classifier correctly predicted 16 out of the 30 negative instances as negative.

Results obtained from the evaluation of our model are presented in the form of a confusion matrix, which provides a comprehensive view of the classification performance. The confusion matrix allows us to analyze the predictive accuracy of our algorithm by comparing the predicted class labels against the actual class labels. This evaluation metric is particularly useful in assessing the performance of machine learning models in multi-class classification tasks.

	Actual Positive	**Actual Negative**
Predicted Positive	70	14
Predicted Negative	0	16

V. Conclusion

The use of machine learning in healthcare apps presents an opportunity to transform the healthcare industry by providing tailored and effective healthcare services. These apps have the potential to enhance diagnostic accuracy, enable early detection of diseases, and facilitate prompt interventions, ultimately leading to improved patient outcomes. Nonetheless, developing trustworthy and efficient healthcare apps necessitates collaboration between healthcare providers and developers, along with the inclusion of ethical and legal considerations.

VI. Future Scope

The impact of machine learning-powered healthcare apps is expected to grow even further in the future, with significant implications for the healthcare industry. Integrating these apps into healthcare systems will enable healthcare providers to provide more personalized and informed care to patients. Technological advancements will enable the development of more advanced ML algorithms, leading to increased accuracy and efficiency in healthcare delivery. It is crucial to conduct research on the ethical implications of using machine learning in healthcare, including issues of data privacy, transparency, and fairness. Overall, the future of healthcare apps powered by machine learning looks promising, with the potential to revolutionize the healthcare industry and enhance patient outcomes. International Conference

References

[1] Rohit Kumar,Prince Takur;Sps Chauhan, "special disease prediction system using machine learning"2022 International Conference on machine learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), 2022.

[2] Adarsh Kumar Pandey;Adarsh Tripathi;Raju Ranjan"Disease Prediction Using Machine Learning (Health Buddy)"2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, 2022.

[3] Narendra Mohan;Vinod Jain; Gauranshi Agrawal "heart disease predition using supervised machine learning algorithms"2021 5th International Conference on Information Systems and Computer Networks (ISCON)Year: 2021.

[4] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv." 2017.

[5] Tian L., Zhang D., Bao S., Nie P., Hao D., Liu Y., et al. "Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas. Clin. Radiol." 2020.

[6] Ge Y., Wang Q., Wang L., Wu H., Peng C., Wang J., et al. "Predicting post-stroke pneumonia using deep neural network approaches. Int. J. Med. Inform." 2019.B. Rieder, Engines of Order: A Mechanology of Algorithmic Techniques. Amsterdam, Netherlands: Amsterdam Univ. Press, 2020.

[7] M.A.Jabbar1 , Shirina Samreen2 , Rajanikanth Aluvalu in their "The Future of Health care: Machine Learning", International Journal of Engineering & Technology,2018.

[8] Tomar, Divya, and Sonali Agarwal. "Feature Selection based Least Square Twin Support Vector Machine for Diagnosis of Heart Disease." International Journal of BioScience & BioTechnology 6(2), (2014).

[9] Saeed Shariati, Mahdi Motavalli Haghighi. "Comparison of ANFIS Neural Network with Several Other ANNs and Support Vector Machine for Diagnosing Hepatitis and Thyroid Diseases", International Conference on Computer Information Systems and Industrial Management Applications (CISIM), 2010.

[10] Seyede Zahra Paylakhi, Sadjaad Ozgoli, Seyed Hassan Paylakhi,in their work "A novel gene selection method using GA/SVM and Fisher criteria in Alzheimer's disease", 23rd Iranian Conference on Electrical Engineering,2015.

[11] [12] Pranav Motarwar;Ankita Durhape"Cognitive Approach for Heart Disease Prediction using Machine Learning"2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) Year: 2020,Publisher: IEEE

[12] Ankita Tyagi;Ritika Mehra;Aditya Saxena "InteractiveThyroid Disease Prediction System Using Machine Learning Technique"2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) Year: 2018,Publisher: IEEE

[13] J.Neelaveni;M.S.Geetha"DevasanaAlzheimer Disease Prediction using Machine Learning Algorithms" 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) Year: 2020,Publisher: IEEE

[14] Shaik Farzana;Duggineni Veeraiah'Dynamic Heart Disease Prediction using Multi-Machine Learning Techniques" 2020 5th International Conference on Computing, Communication and Security (ICCCS)Year: 2020

[15] L. Ali et al., "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," Year: 2019

[16] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection"Year: 2019

[17] M. Gjoreski, A. Gradisek, B. Budna, M. Gams, and G. Poglajen, "Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure from Heart Sounds," Year: 2020

[18] L. Ali, A. Rahman, A. Khan, M. Zhou, A. Javeed, and J. A. Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on χ2 Statistical Model and Optimally Configured Deep Neural Network," Year: 2020 | Conference Paper | Publisher: IEEE

[19] M. R. Ahmed, S. M. Hasan Mahmud, M. A. Hossin, H. Jahan, and S. R. Haider Noori, "A cloud based four-tier architecture for early detection of heart disease with machine learning algorithms," Year: 2018

[20] Christopher Toh and James P. Brody "Applications of Machine Learning in Healthcare, Smart Manufacturing - When Artificial Intelligence Meets the Internet of Things",2021