IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

OPTIMIZATION OF HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES

Jayashree N M[1], Jeevanjyothi D K[2], Meghana L[3]

Savita B[4], Thrisha V S[5]

[1,2,3]B.E. Students, Department of CSE, Sir M Visvesvaraya Institute of Technology, VTU, Bengaluru, India
[4,5]Assistant Professor, Department of CSE, Sir M Visvesvaraya Institute of Technology, VTU, Bengaluru, India

Abstract - *The rise of non-communicable diseases (NCDs), such as heart disease and diabetes, underscores the crucial necessity for innovative predictive solutions. This study offers a framework for improving the detection, diagnosis, and management of cardiovascular diseases (CVDs) by machine learning algorithms. Recognizing the intricate relationship between modernization, commercialization, and unhealthy lifestyles, our research presents a novel application process designed specifically for medical professionals. With an emphasis on heart disease, our software helps physicians predict the onset or recurrence of non-communicable diseases (NCDs) by utilizing patient records. Thorough testing confirms the program's effectiveness and quick prediction skills, enabling well-informed choices on patient health risks. Our software prototype uses various risk characteristics to classify the risk of abrupt cardiac events in patients diagnosed with ischemic heart disease (IHD). This paper discusses different techniques that aim to raise awareness and facilitate preventive interventions to lower the occurrence of such events.*

***Key Words*: Non-Communication Diseases (NCDs), Cardio vascular diseases(CVDs), Machine learning algorithms, Predictive solutions, Heart disease, Patient records, Risk analysis, Preventive interventions.**

1. INTRODUCTION

The rapid growth, processing, and application of information in today's world have significantly influenced organizational performance in numerous industries, including healthcare. The quick distribution and applications of data are now essential to enhancing decision-making procedures. Data mining is a powerful tool that helps decision-makers in various sectors by extracting valuable insights from large databases. The World Health Organization reports that non-communicable disease (NCDs) accounts for a larger share of deaths worldwide. Diabetes, cancers, and cardiovascular disease (CVDs) are the causes of these deaths. These works intend to use inferential data mining technics and Knowledge Discovery in Databases (KDD) to forecast the occurrences of NCDs, notably Cardiovascular Diseases (CVDs) and diabetes. Millions of fatalities globally are caused by heart problems alone, especially in industrialized countries like the United States, Europe, and Japan. Prompt forecasting and analysis of lifestyles and patient data can be used to predict CVDs and allow for the implementation of preventative measures to lower the risks of this potentially fatal condition. Crucial tasks like medical diagnosis frequently call for professional knowledge that isn't always easily accessible. The lack of qualified workers emphasizes how crucial predictive algorithms are for diagnosis support. Therefore, the creation of completely automated, semi-automated, or mechanical predictive diagnostics systems has great potential to reduce the workloads for medical practicians. This study uses KDD and inferential data mining approaches to examine health data, including vital signs, to forecast NCDs. The final objective is to offer a platform for the creation of extensive databases that physicians and other decision-makers may use, assisting in the precise prediction of noncommunicable diseases in patients.

2. RELATED WORKS

Using Data Mining Techniques to Predict Diabetes and Heart Diseases, the research presents an innovative software application designed to the likelihood of non-communicable diseases (NCDs), specifically diabetes and heart disease, utilizing data mining methods. The primary aim of the tool was to develop a forecasting model by utilizing patient data from the Bahrain Defense Force Hospital. The fundamental approach involves using "mining class comparison" to differentiate between various groups based on attributes like age, cholesterol levels, and smoking habits. The program follows a series of steps, including data collection and preprocessing, analysis for relevant dimensions, simultaneous universalization for target and comparison groups, and formulation of comparison criteria. These processes help to categorize people into risk groups for NCDs, which vary from low to high risk.

Smartphone-Based Ischemic Heart Disease(Heart Attack) Risk Prediction using Clinical Data and Data Mining Approaches, a Design Prototype, The "PredictRisk" Android application, which is the subject of this study, estimates the probability of an ischemic heart attack or heart attack using clinical data and data mining algorithms (IHD). A variety of risk factors, including age, gender, diabetes, hypertension, and lifestyle choices, are entered by users. After that, these variables are examined in conjunction with clinical information from 787 IHD patients. Statistical methods like the chi-square test, p-values, and Fisher's exact test were employed to establish a correlation between these factors and the chance of IHD. A rating system classifying users into low, medium, and high-risk groups was developed as a result of this investigation. Furthermore, users reporting symptoms suggestive of IHD are advised to take precautionary measures. Apart from indicating the risk level, the app also recommends users seek consultation from a cardiologist and offer assistance in locating nearby hospitals through Google Maps. To maintain records, user data, and risk scores are synced to a server. An important discovery from examining patient data was the strong association between high-risk groups and IHD incidence: 86.7% of high-risk individuals were found to have IHD, compared to only 12.5% in the low-risk category. This research emphasizes the innovative use of clinical data analysis in creating an efficient risk-scoring system and developing a smartphone application to facilitate widespread self-assessment of IHD risks, aiming to enhance awareness and promote timely medical evaluation in the general population.

Satin Bowerbird Optimization-Based Classification Model for Heart Disease Prediction Using Deep Learning in E-Healthcare, To accurately forecast cardiovascular illnesses, a hybrid model combining deep learning techniques and the Satin Bowerbird Optimization(SBO) algorithm has been constructed in this article. Patient data, including Framingham risk factors, information from wearable sensors, and electronic medical records, are utilized. The preprocessing steps involve feature selection, data normalization, and managing missing data. Feature selection using SBO helps reduce dimensionality. Additionally, a novel deep learning classifier with cascaded kernels and Softplus activation is introduced. Through a comprehensive analysis of cardiac disease datasets, remarkable results in recall (91.3%), accuracy (90%), precision (94%), and F1-score (92.6%) are achieved. The key advancements lie in improved deep learning methodologies and SBO-influenced feature selection, facilitating prompt clinical intervention for cardiovascular issues.

Prediction Of Cardiovascular Disease By Deep Learning and Machine Learning Combined Data Science Approach, this article aims to enhance the precision of predicting cardiovascular disease (CVD) through the application of a data science approach. The Cleveland heart disease dataset is divided into training and testing sets using ten-fold cross-validation, focusing on 14 key attributes. Post deep learning feature extraction, K-means clustering is implemented. A hybrid model is introduced by merging a Convolutional Neural Network for feature extraction and a Naive Bayes classifier for probabilistic classification. The CNN utilizes the Bayes theorem to process features, whereas Naive Bayes automatically identifies features from tabular data as images. Evaluation metrics like accuracy, precision, recall, F1-score, and area under the ROC curve are considered. The hybrid model exhibits remarkable outcomes: 96% F1-score for CVD prediction, 95% precision, 94% recall, and 96% accuracy. This investigation highlights the distinctiveness of combining CNN for feature extraction and Naive Bayes for classification, suggesting that this methodology enables precise early detection of cardiovascular diseases.

MachineLearning_and_Deep_Neural_Network_Techniques_for_Heart_Disease_Prediction, To assess how well deep learning and machine learning methods predict heart disease, an examination was carried out in this study using data from the UCI machine learning repository. Various methods including random forests, decision trees, logistic regression, XGBoost, AdaBoost, LightGBM, CatBoost, KNN, SVM, Naive Bayes, and deep neural networks (DNN3 and DNN4) were examined. The results revealed that DNN3 exceeds all other machine learning models, achieving 85.9% accuracy, while XGBoost reached 81.1%. Key metrics such as accuracy, precision, recall, and F1-score were utilized for comparison, demonstrating the superior performance of DNN3. The study concluded that deep learning, especially DNN3, exhibits enhanced accuracy in predicting heart disease compared to conventional machine learning algorithms when applied to this dataset.

Ischemic_Heart_Disease_Prediction_Using_MachineLearning_and_DeepLearning_Techniques, This article describes a study that predicted cases of ischemic heart disease using cutting-edge machine learning techniques like deep learning. The researchers experimented with various models such as ANN, Deep Neural Network (DNN), Support Vector Machine, and Random Forest, analyzing a cardiovascular disease dataset from Kaggle consisting of 340 entries and 14 features. The highest accuracy achieved was 97.2% by the feature selection in the DNN model and a 20% test set size. Other notable models include DNN without feature selection scoring 94.4% accuracy, and with feature selection in the ANN model reaching 89.25% accuracy. Comparative analyses are presented in tables showing metrics like F1-score, sensitivity, specificity, and accuracy. The outcomes indicate that in predicting ischemic heart disease, feature selection in the DNN model exceeds SVM, Random Forest, and ANN. The study concludes that the DNN model, with its 97.2% accuracy, effectively anticipates cases of ischemic heart disease, providing valuable assistance in treatment decisions. Future research for exploration includes working with image data, creating more advanced applications based on deep learning, and incorporating additional evaluation criteria.

Heart_disease_diagnosis_using_data_mining_technique, This research utilizes a Kaggle dataset containing 14 attributes and 340 instances to assess The efficiency of deep learning and machine learning methods in predicting cases of ischemic heart disease. Various algorithms, such as Random Forest, SVM, Deep Neural Network (DNN), and Artificial Neural Network (ANN), were compared. Key findings reveal that the DNN model, in conjunction with feature selection and a 20% test set size, achieved the highest accuracy of 97.2%. Conversely, Without feature selection, the DNN model's accuracy was 94.4%. while the feature selection of the ANN model showed an accuracy of 89.25%. Comparative analytical metrics such as sensitivity, specificity, accuracy, and F1-score are displayed in tables. Results indicate that the feature selection of the DNN model exceeds the ANN in performance, Random Forest, and SVM models. The researchers conclude that the DNN model accurately predicts cases of ischemic heart disease with a 97.2% accuracy rate, facilitating more effective treatment strategies. Future endeavors may involve working with image data, creating advanced deep learning-based platforms, and exploring additional evaluation criteria.

Heart_Disease_Detection_with_Deep_Learning_Using_a_Combination_of_Multiple_Input_Source, A new method utilizing cutting-edge deep learning methods has been introduced in this article to identify heart conditions from audio recordings of heartbeats. It involves combining convolutional neural networks (CNNs) and long short-term memory (LSTM) recurrent neural networks in a unique hybrid model. To preprocess input data, the method utilizes short-time Fourier transform (STFT) and mel-frequency cepstral coefficients (MFCCs); feeding STFT data into the CNN and MFCC sections, respectively. The primary aim of the prototype is to distinguish healthy and diseased heartbeat sounds to detect the presence or absence of heart disease accurately. By enhancing the combined datasets of heart sounds through data augmentation methods, On common benchmark datasets, the model has proven to be more accurate than previous state-of-the-art methods. One significant advantage of this approach is its capability to recognize important patterns within raw audio data without necessitating specific features, illustrating the promising potential for future advancements. Planned future research will concentrate on broadening the model's application to identify particular types of heart ailments beyond simple binary classification tasks. Overall, this study introduces a pioneering deep-learning architecture for cardiac disease detection, employing both CNNs and LSTMs with audio heartbeat data to elevate detection accuracy.

Efficient_heart_disease_prediction_system_using_optimization_technique, The application of methodologies for particle swarm optimization (PSO) and constricted PSO techniques is explored in this research to predict the severity of cardiac disease. The information is categorized using the Cleveland Heart Disease dataset into five levels of severity, ranging from absence of disease to severe heart disease. Parameter tuning is applied to the PSO and CPSO algorithms to enhance particle movement towards optimal solutions. Performance assessment involves the examination of confusion matrices and metrics such as accuracy, precision, and recall. The results demonstrate that, as stated on the dataset, CPSO achieves slightly higher accuracy (55.1%) compared to PSO (53.1%). Both variations of PSO outperform traditional methods like neural networks and decision trees. Potential future directions include enhancing data preprocessing methods and exploring reinforcement learning for long-term model enhancements. Overall, this study highlights the effectiveness of PSO/CPSO algorithms in predicting heart disease severity, with CPSO exhibiting improved convergence and accuracy in prediction over PSO.

Coronary Artery Disease Prediction using Data Mining Techniques, The research investigates the utilization of machine learning models like Random Forest, Naive Bayes, and Support Vector Machines (SVM) in predicting coronary artery disease. The study utilizes the Cleveland heart disease dataset with 14 medical attributes, splitting the data into training and testing sets. An application in Java is developed to execute and evaluate the algorithms. The findings reveal that Random Forest exceeds SVM and Naive Bayes, accurately classifying 95.3% of cases with the least errors. Performance metrics such as root relative squared error and relative absolute error support this result. Upon analyzing the medical data, the study concludes that Random Forest stands out as the most effective algorithm for predicting ischemic heart disease. The significance of early and precise identification of heart conditions is underscored. Overall, the paper offers valuable observations on the efficacy of various classification algorithms for this predictive task.

3. CONCLUSIONS

The study emphasizes how important data mining algorithms are in revolutionizing healthcare, particularly in predicting non-communicable diseases (NCDs) like cardiac diseases and diabetes. The developed prediction systems exhibit remarkable efficacy, aiding physicians in making informed decisions while empowering patients to proactively monitor their health status. Through the utilization of diverse approaches such as cascaded kernel extreme learning machines, stacked autoencoders, and hybrid deep learning methods, significant strides have been made in disease prediction, with algorithms like XGBoost, deep neural networks (DNN), and random forest emerging as frontrunners due to their superior precision and effectiveness in early disease detection. Moreover, the integration of advanced techniques like feature selection and genetic algorithms further enhances prediction accuracy, particularly in identifying specific conditions such as coronary artery disease (CAD). As the field progresses, continued refinement of predictive models, validation across varied datasets, and the growth of user-friendly applications, including mobile platforms, will be instrumental in facilitating widespread adoption across healthcare settings, ultimately improving patient results and healthcare delivery on a global scale.

REFERENCES

[1] M.W. Groch, J. R. Domnanovich, and W. D. Erwin, “A new heart sounds gating device for medical imaging,” IEEE Transactions on Biomedical Engineering, vol. 39, no. 3, pp. 307–310, March 1992

[2] P. Wilson, R. D’Agostino, D. Levy, A. Belanger, H. Silbershatz and W. Kannel, “Prediction of Coronary Heart Disease Using Risk Factor Categories”, Circulation, vol. 97, no. 18, pp. 1837-1847, 1998.

[3] World Health Organization. Cardiovascular Diseases (CVDS). 1999. Available online: https://www.who.int/news-room/factsheets/detail/cardiovascular-diseases-(vcds).

[4] G.L. Tsirogiannis, D.S. Frossyniotis, J. Stoitsis, S. Golemati, A. Stafylopatis, and K.S. Nikita, "Classification of medical data with a robust multi-level combination scheme," 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), vol. 3, pp. 2483–2487 vol.3, 2004

[5] G. Karraz and G. Magenes, "Automatic classification of heartbeats using neural network classifier based on a Bayesian framework," in 2006 Conference Proceedings of the IEEE Engineering and Medical Biology Society, pp. 4016–4019.

[6] S.L. Strunic, F. Rios-Gutierrez, R. Alba-Flores, G. Nordehn, and S. Bums, "Detection and classification of cardiac murmurs using segmentation techniques and artificial neural networks," in 2007 IEEE Symposium on Computational Intelligence and Data Mining, March 2007, pp. 397–404.

[7] Acharya, Rajendra U. and Yu, Wenwei., Data Mining Techniques in Medical Informatics, The Open Medical Informatics Journal, PMCID: PMC2916206, 2010

[8] Srinivas, K., "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques", IEEE Transaction on Computer Science and Education (ICCSE), 2010, pp.1344 - 1349.

[9] D. S Chaitrali and A.S Sulabha, "A Data Mining Approach for Prediction of Heart Disease Using Neural Networks", International Journal of Computer Engineering and Technology, Vol. 3, pp. 30-40, 2012.

[10] E. Gomes, P. Bentley, M. Coimbra, E. Pereira, and Y. Deng, "Classifying heart sounds: Approaches to the pascal challenge," HEALTHINF 2013 - Proceedings of the International Conference on Health Informatics, pp. 337–340, 01 2013.

[11] S. Ghwanmeh, A. Mohammad, A. Ibrahim, "Innovative Artificial Neural Networks-Based Decision Support System for Heart Disease Diagnosis," Journal of Intelligent Learning Systems and Applications, Vol. 5, pp. 176-183, August, 2013.