IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Literature Review for Exploring Machine Learning Approaches for Detection of Health Disorders

[a]Ibrahim Parkar, [a]Anand Tripathi, [a]Sakshi Jadhav, [a]Tanishque Prajapati, [*]Vaqar Ansari

[a]Student, [*]Assistant Professor
Department of Electronics and Telecommunication Engineering,
St. Francis Institute of Technology, Mumbai, India

Abstract: This research paper presents a comprehensive review of machine-learning approaches for detecting four critical health disorders: Arrhythmia, Sleep Apnea, Insomnia, and Stroke. Recognizing the challenge of gathering comprehensive datasets, the study advocates for separate datasets for each disease, enabling tailored-modeling approaches. Through an extensive literature review, the paper identifies the most effective machine learning models for each condition, including Logistic Regression, SVM, XGBoost, Random Forest, and Voting Classifier. The methodology proposed involves collecting diverse datasets, preprocessing data, selecting appropriate algorithms, and rigorously testing and validating predictive models. Furthermore, the integration of machine learning models with Google Maps is suggested to provide personalized recommendations and suggest nearest healthcare facilities based on predictions and user location. This initiative aims to empower individuals to proactively manage their health and revolutionize healthcare by promoting timely interventions and improving health outcomes.

IndexTerms **- Health disorders, Undetected, scalable, Website, health recommendations, data integration, advanced data analysis, early detection, proactive health management, timely interventions, health outcomes.**

INTRODUCTION

Their project aims to develop a robust health website leveraging individual health data to provide personalized recommendations. By harnessing machine learning techniques such as Logistic Regression, SVM, XGBoost, Random Forest, and Voting Classifier, they target the early detection of four critical diseases: Arrhythmia, Sleep Apnea, Insomnia, and Stroke. Recognizing the complexity of gathering comprehensive datasets, they adopt separate datasets for each disease, allowing tailored-modeling approaches. Through a thorough literature review, they explore the most effective machine learning models for each condition, integrating diverse sources of health data, smart device inputs, and user-generated information. By empowering individuals to proactively manage their health, their initiative seeks to revolutionize healthcare by promoting timely interventions and improving health outcomes.

NEED OF THE STUDY

The authors are focused on predicting four diseases: Arrhythmia, Sleep Apnea, Insomnia, and Stroke. Gathering all the data needed for one big dataset is tough, so they are using separate datasets for each disease. This means they will need different models for each one. So, they are checking out different papers to learn about the best machine-learning models for each disease.

LITERATURE SURVEY

The methodology involves an in-depth literature review to survey existing research on machine learning strategies for identifying health disorders, particularly targeting Arrhythmia, Sleep Apnea, Insomnia, and Stroke. This review encompasses an array of studies delving into diverse algorithms and methodologies employed in disease prediction.

Aishwarya Seth, Satish Babu B., S.S. Iyenger - Machine Learning Model for Predicting Insomnia Levels in Indian College Students: In this study [1], a new method is proposed for forecasting insomnia severity among college students using machine learning, specifically an artificial neural network (ANN) trained on survey data gathered from 158 college students in India. The research introduces a probabilistic model grounded in mental health indicators like depression, anxiety, stress, and social adjustment, to comprehend the connection between insomnia and mental well-being while advocating for timely intervention. Through the process of feature selection and model refinement, the ANN reliably predicts insomnia severity with minimal root mean square error (RMSE), underscoring the potential of machine learning to enhance insomnia diagnosis and intervention strategies by providing a more continuous and interpretable diagnostic approach. This advancement holds promise for both patients and healthcare providers by facilitating more effective diagnosis and intervention strategies.

Ch. Usha Kumari, R. Ankita, T.Pavani, N Arun Vignesh, N. Tarun Varma, Md Aqeel Manzar and A. Reethika - Heart Rhythm Abnormality Detection and Classification using Machine Learning Technique: This research paper [2] addresses the critical need for early detection of heart rhythm irregularities to mitigate the escalating global mortality rate attributed to heart diseases. Employing machine learning techniques, specifically Support Vector Machine (SVM) classification, the study aims to accurately classify ECG signals. Utilizing data sourced from the MIT-BIH database, the research employs Discrete Wavelet Transform (DWT) for feature extraction. The extracted features are then used to train an SVM classifier to categorize cardiac abnormalities. Notably, this approach achieves an impressive accuracy rate of 95.92%, showcasing its potential to significantly improve cardiac care by enabling the timely identification and classification of heart rhythm irregularities, thus enhancing patient outcomes in cases of heart rhythm disorders.

Tasfia Ismail Shoily, Tajul Islam, Sumaiya Jannat, Sharmin Akter Tanna, Taslima Mostafa Alif, and Romana Rahman Ema- Detection of Stroke Disease using Machine Learning Algorithms: In their project, [3] the aim is to address the significant global burden of stroke by employing machine learning algorithms to predict and classify stroke types based on physical state and medical report data. Through a literature survey, previous research in stroke prediction and classification using machine learning techniques such as Artificial Neural Networks, Support Vector Machines, Decision Trees, and ensemble methods is reviewed, providing insights into different approaches and datasets. The research methodology involves collecting and preprocessing data from multiple sources, followed by the implementation of machine learning algorithms including Naive Bayes, J48, k-NN, and Random Forest for classification using the WEKA toolkit. Evaluation metrics such as accuracy, precision, recall, and F1-score are employed to assess algorithm performance through a 10-fold cross-validation process. The study concludes that machine learning algorithms, particularly J48, k-NN, and Random Forest, demonstrate promising accuracy in stroke detection, paving the way for early disease detection and treatment through data-driven approaches in healthcare.

Daniele Padavano, Arturo Martinez-Rodrigo, Jose Manuel Pastor, Jose Joaqu'in Rieta and Raul Alcaraz -Obstructive Sleep Apnea Detection Based on Heart Rate Variability and Machine Learning Techniques: In this paper, [4] an experimental review of obstructive sleep apnea (OSA) detection using heart rate variability (HRV) and machine learning techniques is presented. OSA, a common respiratory syndrome associated with cardiovascular diseases, is traditionally diagnosed through costly and inconvenient polysomnography (PSG). The study explores alternative methods using the publicly available Apnea-ECG database, extracting traditional time-frequency domain features, complexity measures, and entropy-based measures from HRV signals. Univariate and multivariate classifiers, including support vector machines (SVM) and k-nearest neighbors (KNN), are applied, with sequential feature selection (SFS) algorithms used to reduce computational costs. Findings indicate that multivariate classifiers yield similar results to those in the literature, with frequency domain features proving effective for OSA detection. The study underscores the potential of HRV analysis and machine learning in improving OSA detection methods, providing valuable insights for researchers in the field.

A Nguyen, Sardar Ansari, Mohsen Hooshmand, Kaiwen Lin, Hamid Ghanbari, Jonathan Gryak and Kayvan Najarian. - Heart Rate Variability Analysis for Atrial Fibrillation Detection in Short Single-Lead ECG Recordings: In this project, [5] a comparative study is conducted on the detection of atrial fibrillation (AFib) using heart rate variability (HRV) analysis applied to short single-lead ECG recordings. The study aims to address the challenge of accurately classifying AFib in noisy ECG recordings from wearable devices like AliveCor. Various HRV feature extraction methods are investigated, including statistical, geometrical, frequency, entropy, Poincare plot-based, and Lorentz plot-based features. Feature selection is performed to improve classification accuracy, followed by classification using support vector machines (SVMs). The study utilizes the publicly available dataset from the 2017 PhysioNet Challenge, which includes AFib, normal, other arrhythmia, and noise classes. Results show that a combination of features from all categories achieves the highest accuracy in detecting AFib, even in short ECG recordings.

SUGGESTION

The authors have referred to multiple research papers for using various algorithms so that they can solve the current limitations in health recommendation systems. They plan to address this limitation by integrating ML models and Google Maps on a health website to suggest the nearest health centers after disease prediction.

They will begin by collecting diverse datasets to train disease prediction models. Data preprocessing will involve trimming and organizing data, transforming units, handling inconsistencies, mapping categorical variables, etc. Machine learning models will be developed for heart arrhythmia, sleep apnea, insomnia, and stroke prediction using logistic regression, random forest, and ensemble methods. A locally hosted website will provide a user-friendly interface to access and interact with the models, integrating Google Maps to recommend healthcare facilities based on predictions and user location. Rigorous testing and validation will ensure the system's accuracy and reliability. The deployment will involve setting up the website and models on a local server for optimal performance.

1. They will focus on working and choosing models developed for disease-specific machine learning models. This includes selecting appropriate algorithms for each condition, ensuring that they are well-suited to the characteristics of the data and the nature of the disease being predicted. These may include logistic regression, random forest classifiers, voting classifiers, or other suitable algorithms depending on the requirements and performance considerations.

2. Train these models using high-quality, preprocessed training data.

3. Utilize feature engineering and selection techniques to improve model accuracy and interpretability.
4. Evaluate the performance of each model using relevant evaluation metrics.

5. Integration of Models and Google Maps:
   a) Rigorous testing of the entire system will be conducted to ensure its accuracy, reliability, and security.
   b) The performance of disease prediction models will be validated with data from diverse sources.
   c) Following validation, the integration of models with Google Maps will be completed.
   d) The system will then undergo final configuration for optimal performance and secure access.

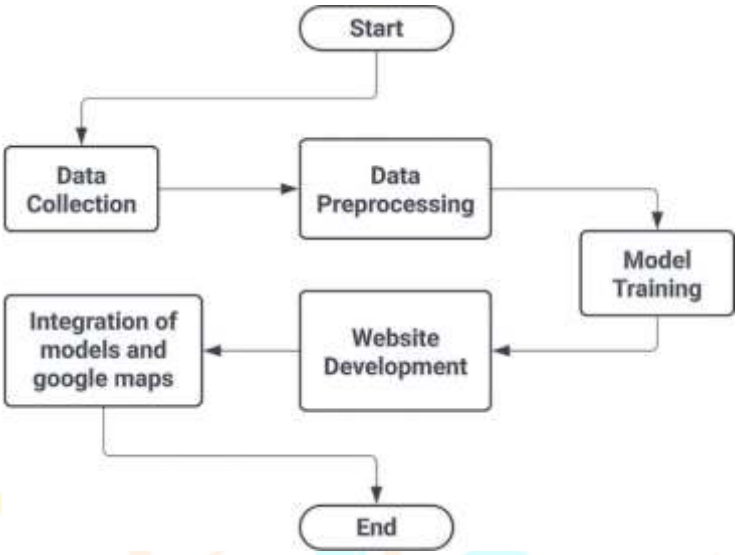

Figure 1: Block Diagram of Proposed Solution

REFERENCES

[1] A. Seth, B. S. Babu and S. S. Iyenger, "Machine Learning Model for Predicting Insomnia Levels in Indian College Students," 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), Bengaluru, India, 2019, pp. 1-6, doi: 10.1109/CSITSS47250.2019.9031041. keywords: {Sleep; Stress; Mathematical model; Probabilistic logic; Machine learning; Sociology; Statistics; Insomnia; Probabilistic Modelling; Indian College Students; Artificial Neural Network; Survey-based testing}.

[2] C. U. Kumari et al., "Heart Rhythm Abnormality Detection and Classification using Machine Learning Technique," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 580-584, doi: 10.1109/ICOEI48184.2020.9142914. keywords: {Feature extraction; Support vector machines; Electrocardiography; Heart beat; Training; Testing; ECG; ARR; NSR; CHF; Wavelet; SVM}.

[3] T. I. Shoily, T. Islam, S. Jannat, S. A. Tanna, T. M. Alif and R. R. Ema, "Detection of Stroke Disease using Machine Learning Algorithms," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-6, doi: 10.1109/ICCCNT45670.2019.8944689. keywords: {Stroke (medical condition); Classification algorithms; Decision trees; Machine learning algorithms; Hemorrhaging; Diseases; Random forests; learning; WEKA; Naive Bayes; J48; k-NN; Random Forest}.

[4] D. Padovano, A. Martinez-Rodrigo, J. M. Pastor, J. J. Rieta and R. Alcaraz, "An Experimental Review on Obstructive Sleep Apnea Detection Based on Heart Rate Variability and Machine Learning Techniques," 2020 International Conference on e-Health and Bioengineering (EHB), Iasi, Romania, 2020, pp. 1-4, doi: 10.1109/EHB50910.2020.9280302. keywords: {Heart rate variability; Feature extraction; Rail to rail inputs; Electrocardiography; Sleep apnea; Entropy; Support vector machines; Obstructive sleep apnea; Heart rate variability; Entropy; Lomb-Scargle periodogram}.

[5] A. Nguyen et al., "Comparative Study on Heart Rate Variability Analysis for Atrial Fibrillation Detection in Short Single-Lead ECG Recordings," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, pp. 526-529, doi: 10.1109/EMBC.2018.8512345. keywords: {Heart rate variability; Electrocardiography; Feature extraction; Entropy; Training; Monitoring; Standards}.