iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 12 **Issue:** IV **Month of publication:** April 2024

**DOI:** https://doi.org/10.22214/ijraset.2024.61379

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 12 Issue IV Apr 2024- Available at www.ijraset.com*

# Personalized Health Recommendation System using Machine Learning

Ibrahim Parkar[1], Sakshi Jadhav[2], Anand Tripathi[3], Tanishque Prajapati[4], Dr. Vaqar Ansari[5]
*Electronics and Telecommunication Engineering, Mumbai University*

***Abstract:*** ***The core objective of the project is to develop a health website capable of harnessing individual health data to generate personalized health recommendations, fostering early detection and more effective management of health issues. Utilizing Machine Learning (ML), the team aims to predict diseases like arrhythmia, Sleep Apnea, Insomnia, and Stroke, pivotal for early intervention and adapting diagnosis strategies. In their approach, ML algorithms, including Logistic Regression, Random Forest, and Voting classifier, analyze diverse health data sources to build a comprehensive recommendation system. It is noteworthy that all models exhibit an accuracy exceeding 90%, which underscores the reliability and effectiveness of the system. By seamlessly integrating various health data sources and emphasizing proactive health management, this initiative holds transformative potential, empowering individuals to make informed decisions about their well-being and promoting improved health outcomes.***

***Keywords: health website, individual data, personalized health recommendations, early detection, Machine Learning, diverse health data sources, recommendation system, accuracy, reliability, proactive health management, informed decisions, improved health outcomes.***

## I. INTRODUCTION

The primary motivation of the authors is to revolutionize healthcare by empowering individuals to make informed decisions about their health, thereby enhancing healthcare outcomes through early disease prediction and personalized advice. By developing a user-friendly and scalable health website, they aim to harness health data from diverse sources. Integrating data analysis algorithms such as Logistic Regression, Random Forest, and Voting classifier, their website serves as the interface for predicting diseases like arrhythmia, Sleep Apnea, Insomnia, and Stroke, enabling timely intervention and management. This approach, which considers individual medical history, lifestyle, and preferences, facilitates early detection and proactive health management. By seamlessly integrating various health data sources and encouraging individuals to take an active role in their well-being, their initiative has the potential to transform healthcare, promoting improved health outcomes for all.

## II. LITERATURE SURVEY

The authors aim to predict four diseases: Arrhythmia, Sleep Apnea, Insomnia, and Stroke, as mentioned in the introduction. Obtaining a single master dataset for training models encompassing all these diseases proves challenging. Thus, they intend to train models using distinct datasets. Given this approach, employing multiple models is inevitable. Consequently, they are consulting various papers to gain insights into different ML models.

Aishwarya Seth, Satish Babu B., S.S. Iyenger - Machine Learning Model for Predicting Insomnia Levels in Indian College Students: In this work [1], a novel approach is proposed to predict insomnia levels in college students using machine learning, specifically an artificial neural network (ANN) model trained on survey data collected from 158 college students in India. The study introduces a probabilistic model based on mental health parameters such as depression, anxiety, stress, and social adjustment, aiming to understand the relationship between insomnia and mental health and advocating for early intervention. Through feature selection and model development, the ANN effectively predicts insomnia levels with low root mean square error (RMSE), highlighting the potential of machine learning in improving insomnia diagnosis and intervention strategies by offering a more continuous and understandable approach to diagnosis, thus benefiting both patients and healthcare providers.

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: The research paper [2] tackles the urgent necessity for early detection of heart rhythm irregularities, aiming to curb the increasing global mortality rate associated with heart diseases. Through the utilization of machine learning techniques, specifically Support Vector Machine (SVM) classification, the study seeks to effectively categorize ECG signals.

©IJRASET: All Rights are Reserved 

Drawing upon data obtained from the MIT-BIH database, the research employs Discrete Wavelet Transform (DWT) to extract features. These extracted features are then utilized to train an SVM classifier for the classification of cardiac abnormalities. Noteworthy is the achievement of an impressive accuracy rate of 95.92%, underscoring the potential of this approach to significantly enhance cardiac care by facilitating timely identification and classification of heart rhythm irregularities, thereby improving patient outcomes in instances 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 the paper [3] the objective is to combat the substantial global impact of stroke by utilizing machine learning methods to forecast and categorize stroke occurrences based on both physical condition and medical records. A comprehensive review of existing literature examines prior studies on stroke prediction and categorization employing machine learning techniques like Artificial Neural Networks, Support Vector Machines, Decision Trees, and ensemble methods, offering insights into various methodologies and datasets. The research methodology entails gathering and preprocessing data from diverse sources, followed by the application of machine learning algorithms such as Naive Bayes, J48, k-NN, and Random Forest for classification tasks using the WEKA toolkit. The performance of these algorithms is evaluated using metrics like accuracy, precision, recall, and F1-score, assessed through a 10-fold cross-validation procedure. The findings suggest that machine learning algorithms, especially J48, k-NN, and Random Forest, exhibit promising accuracy in stroke detection, indicating potential for early disease identification and treatment via 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: The paper [4] presents an experimental examination of detecting Obstructive Sleep Apnea (OSA) using heart rate variability (HRV) and machine learning techniques. OSA, a common respiratory condition linked to cardiovascular ailments, is traditionally diagnosed through expensive and inconvenient polysomnography (PSG). The research investigates alternative approaches using the publicly accessible Apnea-ECG database. It extracts conventional time-frequency domain features, complexity metrics, and entropy-based measures from HRV signals. Both univariate and multivariate classifiers, such as support vector machines (SVM) and k-nearest neighbours (KNN), are deployed, with sequential feature selection (SFS) algorithms utilized to reduce computational overhead. The results indicate that multivariate classifiers produce comparable outcomes to those reported in existing literature, particularly highlighting the efficacy of frequency domain features for OSA detection. This study emphasizes the promise of HRV analysis coupled with machine learning in advancing OSA detection.

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: The study [5] delves into comparing different methods for detecting atrial fibrillation (AFib) by analyzing heart rate variability (HRV) in short single-lead ECG recordings. The primary objective is to tackle the challenge of accurately identifying AFib amidst noisy ECG data from wearable devices like AliveCor. The research explores various techniques for extracting HRV features, encompassing statistical, geometrical, frequency, entropy, Poincare plot-based, and Lorentz plot-based approaches. To enhance classification accuracy, the study also incorporates feature selection, followed by classification using support vector machines (SVMs). Leveraging the publicly available dataset from the 2017 PhysioNet Challenge, which covers AFib, normal, other arrhythmias, and noise classes, the findings reveal that a blend of features from all categories yields the highest accuracy in AFib detection, even in brief ECG recordings.

## III. METHODOLOGY

### A. *Training and Development*

In the training and development phase, a meticulous approach was undertaken to ensure the accuracy and reliability of the disease prediction system. The process encompassed several key stages, including data collection, preprocessing, model training, website development, and integration of Google Maps.

1) *Data Collection:* Diverse datasets were sourced from reputable sources such as Kaggle, encompassing Fitbit fitness tracker data, sleep health and lifestyle data, and stroke data. These datasets provided a robust foundation for training accurate prediction models.
2) *Data Preprocessing:* Rigorous preprocessing techniques were employed to cleanse and prepare the datasets for model training. This included handling missing values, converting categorical variables to numerical ones, and engineering new features tailored to each disease being predicted.
3) *Model Training:* Disease-specific machine learning models were trained using the pre-processed data. During the model training phase, we evaluated the performance of each classifier to ensure robustness and accuracy.

We employed logistic regression to predict Arrhythmia, achieving an accuracy score of 0.935 with an F1 score of 0.94. For predicting sleep disorders, a Random Forest Classifier was utilized, yielding an accuracy score of 0.929 with an F1 score of 0.93. Our Voting Classifier, comprising Decision Trees and Random Forest models, was applied to predict stroke, achieving an accuracy score of 0.957 with an F1 score of 0.96. These comprehensive evaluation metrics provide insight into the performance of each model.

4) *Website Development:* A user-friendly website interface was developed using Flask, HTML, CSS, and JavaScript. The website provided seamless interaction with the prediction models, offering features such as data forms, diagnosis, results display, and integration with Google Maps.
5) *Integration of Models and Google Maps:* The trained models were seamlessly integrated into the website, allowing users to receive accurate predictions for various diseases. Integration with Google Maps enabled the recommendation of nearby healthcare facilities based on health predictions and user location, enhancing the utility of the system.

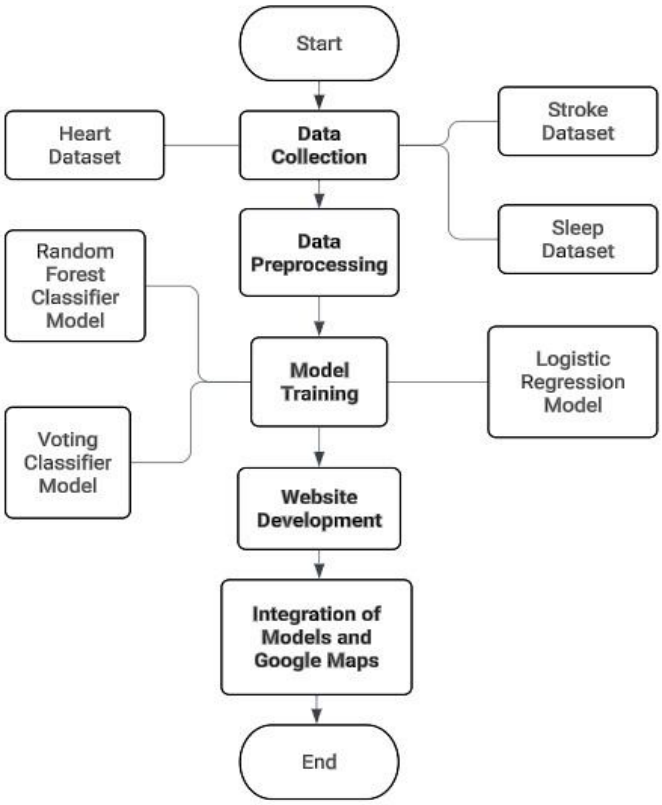



Fig. 1 Training and Development

*B. Testing*

The testing phase involves users accessing a locally hosted website and creating a detailed health profile. They upload data from smart devices capturing vital metrics. Using predictive analytics, the system anticipates potential health conditions. Users are then directed to the nearest healthcare facilities for tailored evaluation and treatment.

1) *Access Website:* Users access the locally hosted website. The index page serves as the starting point.
2) *Create Health Profile:* Users fill out a form with personal health information, including demographics, medical history, and lifestyle factors. This information is saved for further analysis.
3) *Upload Data with Smart Device:* Users must upload health-related data obtained from a smart device. This data typically includes physiological measurements like heart rate, blood pressure, activity levels, etc.
4) *Prediction:* Based on the submitted health profile and uploaded data, the system predicts potential health conditions such as Arrhythmia, Stroke, Sleep Apnea, or Insomnia. Users are presented with the predicted condition.
5) *Recommend Nearest Healthcare Facility:* After the prediction of a disease users can choose to receive recommendations for the nearest healthcare facilities where they can seek further evaluation and treatment based on the predicted condition.

©IJRASET: All Rights are Reserved 

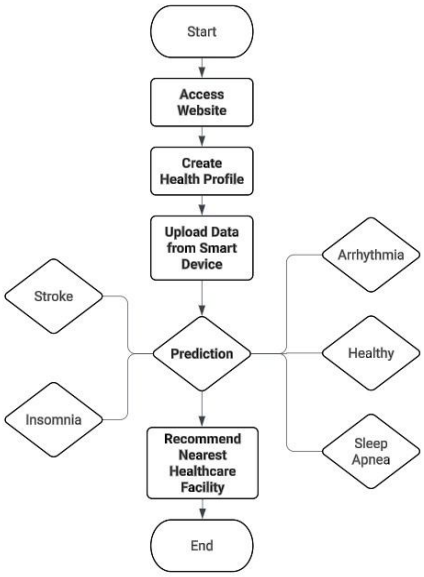



Fig. 2 Testing

## IV. PERFORMANCE METRICS AND RESULTS

### A. *Performance Metrics*

1) *Precision:* Precision measures the proportion of true positive predictions among all positive predictions made by the model. A high precision value indicates that the model is making accurate positive predictions, while a low precision value suggests that the model is incorrectly labeling negative instances as positive.
2) *Recall:* Recall measures the proportion of true positive instances that were correctly identified by the model out of all actual positive instances in the dataset. A high recall value indicates that the model is effectively capturing most of the positive instances in the dataset, while a low recall value suggests that the model is missing a significant number of positive instances.
3) *F1 Score:* The F1 Score provides a single score that combines both precision and recall into one measure, offering a more comprehensive evaluation of a model's performance. The F1 score rewards models that have both high precision and recall, striking a balance between them. It is particularly useful when the class distribution is imbalanced, meaning one class significantly outweighs the other. A high F1 score indicates that the model has both high precision and high recall, meaning it is making accurate positive predictions while capturing most of the positive instances in the dataset. Conversely, a low F1 score suggests that the model is either lacking in precision, recall, or both.
4) *Accuracy:* It measures the proportion of correctly classified instances out of the total instances in the dataset. A high accuracy value indicates that the model is making accurate predictions overall, while a low accuracy value suggests that the model is making a significant number of incorrect predictions.
5) *Cross-validation:* Cross-validation involves partitioning the dataset into multiple subsets, called folds, and training the model on a combination of these folds while evaluating its performance on the remaining fold. This process is repeated multiple times, each time using a different fold as the validation set and the remaining folds as the training set.

Here's how cross-validation works:

a) *Partitioning the Dataset:* The dataset is divided into k equal-sized folds. Common choices for k are 5 or 10, but other values can also be used. We have used 5 folds.
b) *Training and Validation:* The model is trained k times. In each iteration, one of the k folds is held out as the validation set, and the model is trained on the remaining k-1 folds.
c) *Performance Evaluation:* After training the model on each fold, its performance is evaluated on the validation set (the fold that was held out). Metrics such as accuracy, precision, recall, or F1 score are calculated.
d) *Aggregation of Results:* The performance metrics obtained from each iteration are averaged to obtain a single performance estimate for the model.

TABLE I
PERFORMANCE METRICS OF ALL MODELS

| Disease Detected | Model Name | Precision | Recall | F1 score | Accuracy | Cross Validation |
|---|---|---|---|---|---|---|
| Heart Arrhythmia | Logistic Regression | 90.0 | 94.5 | 91.5 | 93.5 | 94.7 |
| Insomnia Sleep Apnea | Random Forest Classifier | 90.33 | 83.67 | 85.67 | 91.07 | 91..4 |
| Stroke | Random Forest Classifier | 92.18 | 95.9 | 94 | 93.83 | - |
| | Decision Tree | 95.75 | 94.67 | 95.2 | 95.19 | - |
| | Voting Classifier | 96 | 96 | 96 | 95.68 | 95.96 |

*B. Results*

As mentioned in the testing process, users first access a locally hosted website to create a detailed health profile and upload data from smart devices. This data serves as the foundation for subsequent predictive analysis.

In line with this process, the image below illustrates a crucial aspect of the system's functionality: the presentation of predicted health conditions. The image showcases the system's prediction of Insomnia, one of the potential health conditions identified through predictive analytics. This visual representation shows the system's capability to generate personalized health insights based on the user's profile and uploaded data.

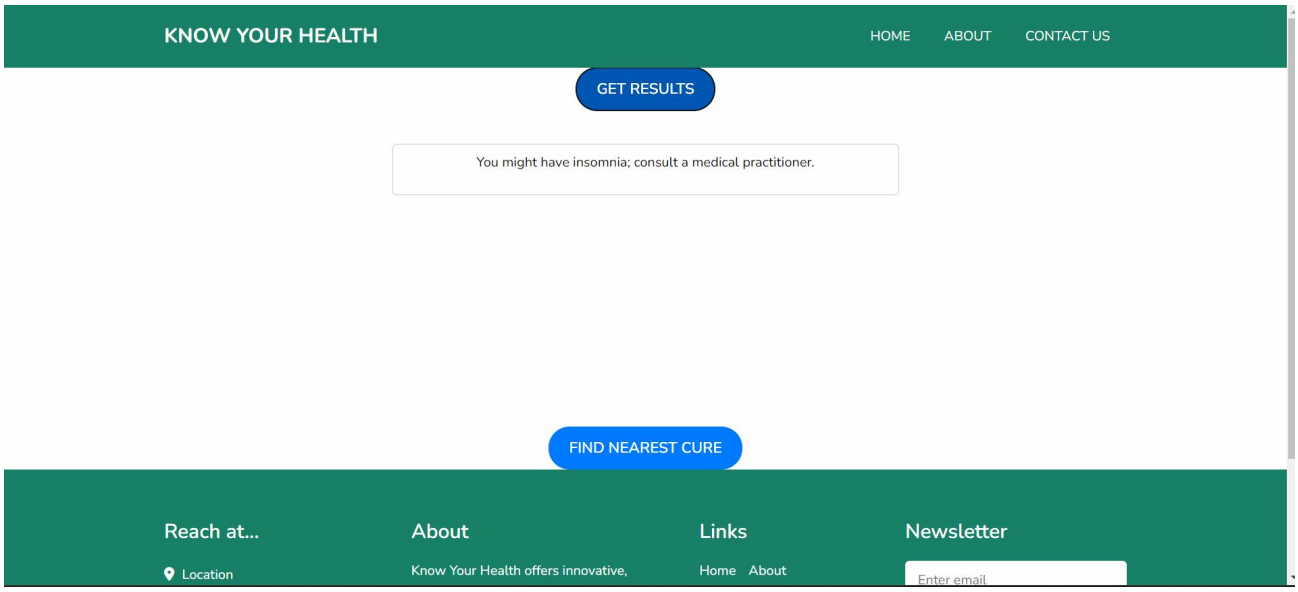



Fig. 3 Prediction of Health Disorder

The following image showcases a vital feature of the system: recommending the nearest healthcare facility for the predicted disease, Insomnia, as shown in the previous image. Users are directed to the closest facility via Google Maps, ensuring timely medical evaluation and treatment. This highlights the system's personalized approach to facilitating timely medical evaluation and treatment.

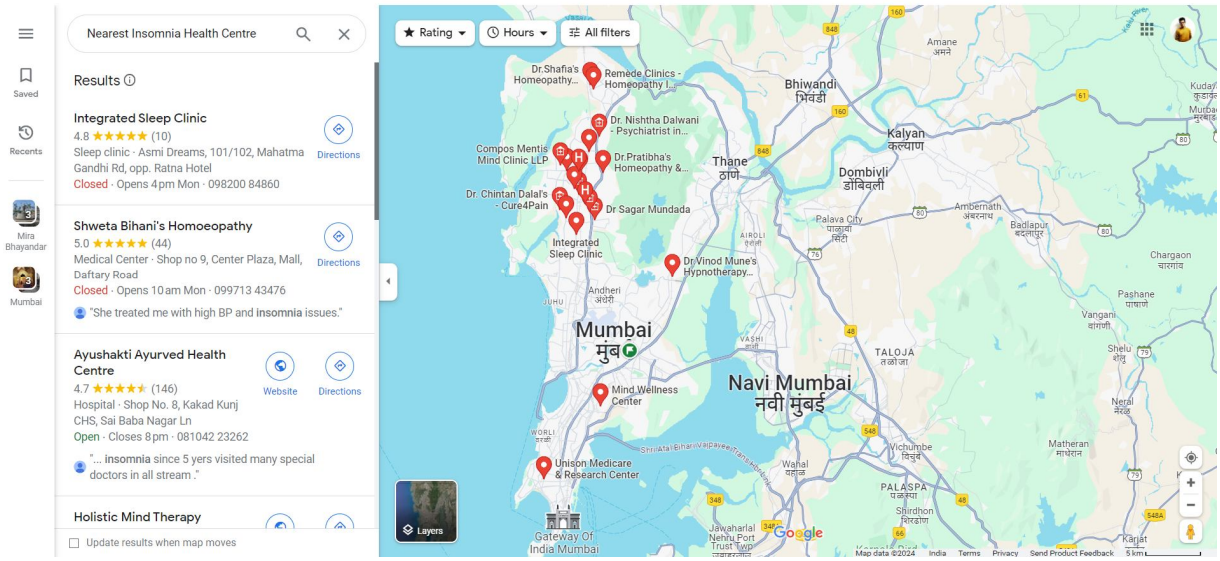



Fig. 4 The user is Recommended to the Nearest Healthcare Facility for Insomnia

## V. CONCLUSION

The concluding remarks provided underscore the significant promise of the research project for the future of healthcare. It emphasizes the user-friendly platform's focus on early detection and proactive health management, suggesting its potential to revolutionize the industry. The effectiveness of disease-specific machine learning models, each boasting an accuracy of above 90%, underscores the success of the approach.

Looking ahead, there is an opportunity for further refinement and expansion, including the integration of AI-driven health coaching and telehealth services to enhance the platform's impact. Additionally, the future scope involves integrating a personalized health recommendation system directly into smart devices, eliminating the need for manual data extraction and upload.

In summary, the project signifies a meaningful step towards a more patient-centered and efficient healthcare system. With ongoing dedication to innovation, the vision of preventive, personalized, and empowering healthcare for all can be realized..

## REFERENCES

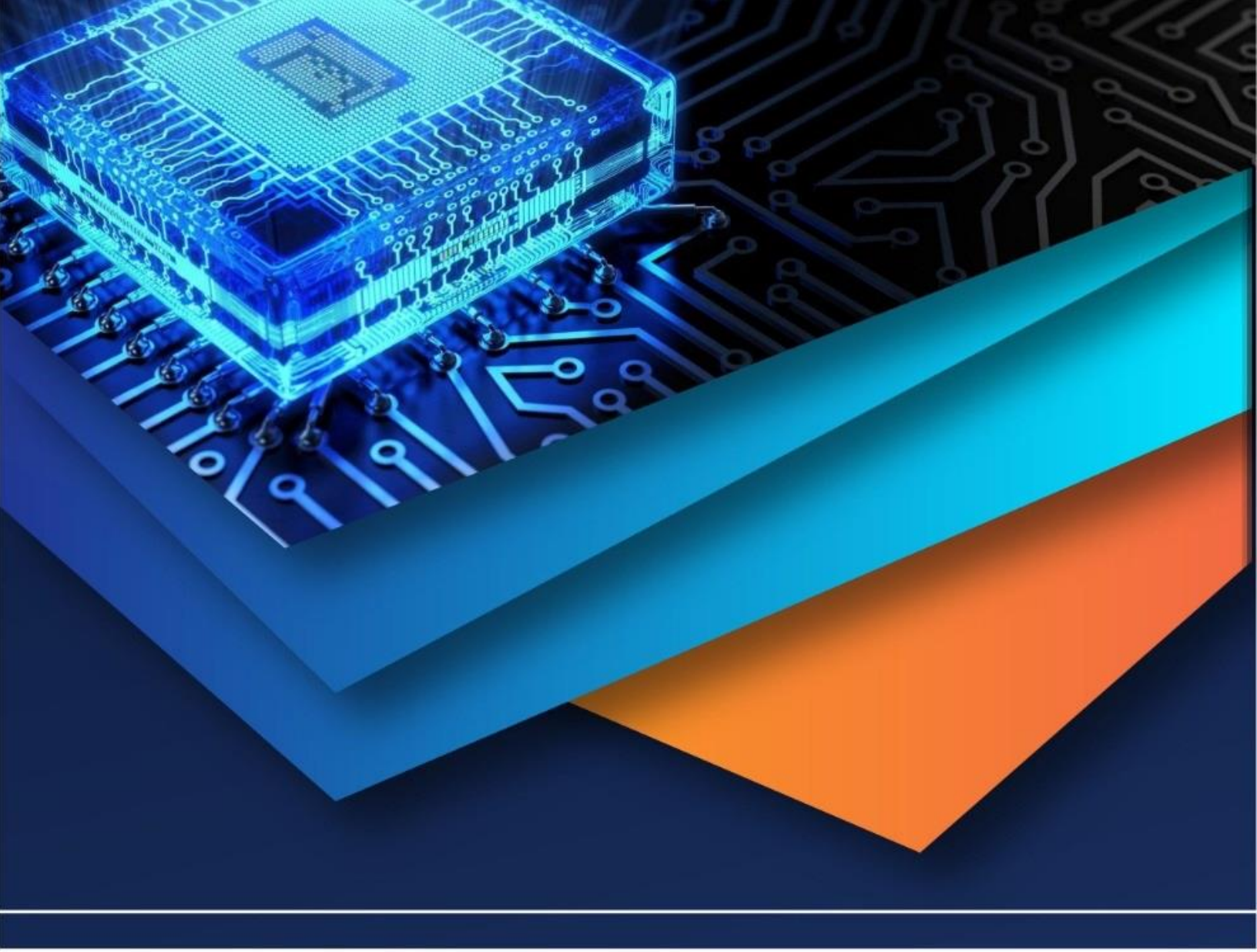
[1] C. U. Kumari et al., "Heart Rhythm Abnormality Detection and Classification using Machine Learning Technique," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 580-584, doi: 10.1109/ICOEI48184.2020.9142914. keywords: {Feature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

[2] 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}.

[3] 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}.

[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}.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)