

Prediction of Heart Attack Risk and Detection of Sleep Disorders Using Deep Learning Approach

Aditya Nooka, Saumya Sejal, Lokesh K, Perepi Rajarajeswari*

School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India. *Corresponding Author's Email: rajacse77@gmail.com

Abstract

This study presents a novel deep learning-based approach for predicting heart attack risk and detecting sleep disorders. Traditional models often fall short in accessibility and accuracy, leading to substandard outcomes. Our proposed ensemble-based AI model overcomes these limitations by leveraging the strengths of deep learning techniques, including neural networks, for robust analysis of sleep patterns and heart health indicators. Our model is capable of integrating data from diverse sources including wearable devices and health records. Emphasis is placed on interpretability, enabling users to comprehend predictions and take informed actions to improve their well-being. Through extensive evaluation and validation, our model demonstrates superior performance in accurately predicting heart attack risk and identifying various sleep disorders. In This research paper we have focused on advancing preventive healthcare strategies by enabling early detection and intervention, ultimately enhancing patient outcomes and healthcare resource utilization.

Keywords: Deep learning, Ensemble Modeling, Healthcare, Heart Attack Risk, Sleep Disorders.

Introduction

The field of healthcare is witnessing a transformative shift with the integration of advanced technologies, particularly in the realm of predictive analytics. In this context, the prediction of heart attack risk and the detection of sleep disorders hold paramount importance for proactive healthcare management and improved patient outcomes. However, traditional models often struggle to deliver accessible and accurate insights, leading to substandard outcomes. In response to these challenges, this study introduces a novel deep learning-based approach to address the limitations of existing methodologies. Our proposed ensemble-based AI model represents a significant advancement in predictive healthcare analytics by harnessing the capabilities of deep learning techniques, particularly neural networks. By leveraging these powerful algorithms, our model offers robust analysis of sleep patterns and heart health indicators, facilitating comprehensive assessment and detection of potential health concerns. Moreover, the model's ability to integrate data from diverse sources, including wearable devices and health records, enhances its overall effectiveness, based on which the users can

comprehend the predictions and make informed decisions to enhance their well-being. Through rigorous evaluation and validation, our model performs better in accurately predicting heart attack risk and identifying sleep disorders. By enabling early detection and intervention, our research contributes to advancing preventive healthcare strategies, ultimately leading to enhanced patient outcomes and optimized healthcare resource utilization.

In recent years, there has been a notable surge in healthcare research leveraging technology to improve outcomes. Platforms such as Google Fit facilitate effortless activity tracking via smart phones, offering versatile options for monitoring physical activity (1). Wearable technology has gained prominence in chronic disease management, with studies systematically reviewing its impact (2, 3). Machine learning techniques have shown promise in improving risk prediction for heart failure patients, enabling more targeted interventions and personalized care (4, 5). The use of smart phone-based risk prediction for ischemic heart disease highlights the feasibility of using everyday devices for health assessments.

This is an Open Access article distributed under the terms of the Creative Commons Attribution CC BY license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

(Received 2nd April 2024; Accepted 17th July 2024; Published 30th July 2024)

The use of smart phone-based risk prediction for ischemic heart disease highlights the feasibility of using everyday devices for health assessments. Machine learning models, utilizing data from large-scale studies such as the Sleep Heart Health Study, predict incident heart diseases in patients with sleep-disordered breathing (6). Reviews emphasize the complex relationship between sleep duration, insomnia, and heart disease risk (7, 8). Despite the potential benefits, challenges remain in integrating wearable patient monitoring systems into clinical practice, as outlined in systematic reviews. However, studies demonstrate the potential of machine learning for real-time risk prediction in critical care settings, such as predicting circulatory failure in intensive care unit patients (9). Addressing challenges such as data privacy and interoperability is essential to fully harness technology's potential in healthcare.

A vast range of attributes is required for generating reliable outcomes. Many of these attributes, however, are irrelevant to the final prediction, or their acquisition is difficult, particularly for patients without professional assistance. This reliance on multifaceted data inputs makes the practical implementation of these models in clinical settings complicated. Despite advancements in predictive analytics, the accuracy of existing models still does not meet the desired standard. The implications of such inaccuracies are significant, potentially resulting in numerous false negatives or false positives. These errors could lead to misdiagnoses, inappropriate interventions, or delayed treatments, ultimately impacting patient outcomes and healthcare resource utilization. By mitigating these limitations, clinicians and healthcare providers can harness the full potential of predictive analytics to improve patient care.

Despite the promising potential of ensemble-based deep learning models for predictive healthcare analytics, several gaps persist. One limitation lies in effectively integrating and analyzing diverse, multimodal data streams beyond wearable and clinical sources, such as environmental factors and lifestyle habits, which could further enhance predictive capabilities. The interpretability and explainability of deep learning models remain a challenge, hindering trust among healthcare professionals. Developing transparent and interpretable insights into the model's decision-

making is crucial for informed clinical decision-making. Data privacy, security, and ethical considerations in handling sensitive health data warrant rigorous investigation to ensure patient confidentiality and build public trust. Lastly, real-world deployment across diverse healthcare settings necessitates further research into efficient computational strategies, streamlined data integration pipelines, and adaptive learning mechanisms to accommodate evolving landscapes. The principal aim of this endeavor is to develop an advanced AI-based framework with the capacity to accurately forecast the likelihood of cardiovascular events and identify disturbances in sleep patterns. Through the utilization of state-of-the-art machine learning methodologies and the integration of varied data reservoirs, the objective is to surmount the deficiencies inherent in conventional models, namely their propensity for inaccuracy, limited accessibility, and constrained data inputs. Rajarajeswari *et al.*, focused on AI approaches for healthcare applications (10 - 12).

Central to this objective is the construction of an ensemble model, amalgamating the virtues of diverse algorithms. These include random forests for discerning feature significance and non-linear associations, support vector machines for steadfast classification proficiency, and neural networks for adeptly handling sequential data and assimilating temporal patterns. A pivotal facet of this endeavor is the emphasis on interpretability, facilitating user comprehension of model predictions and fostering trust in the system. This interpretive capacity empowers individuals to proactively engage in health management by making informed decisions predicated on personalized insights.

Moreover, the intention is to explore data origins beyond traditional health archives, encompassing wearable technologies and activity monitors, to enrich prognostic models and amplify their precision. By achieving these objectives, the envisioned system is poised to advance preventive healthcare strategies, ultimately enhancing patient outcomes and optimizing the allocation of healthcare resources.

Our research has focused on diverse populations (e.g., studies that are limited to specific demographic groups may not be as robust as those that include diverse age groups, cultural backgrounds, and geographical locations). The

research concentrates on a distinct demographic group, such as a specific age range, gender, or ethnic group, that hasn't been extensively studied before. This may provide a unique view.

Methodology

Data Acquisition

The system seamlessly integrates Google Fit data from wearable devices, health records, and user inputs. It employs the Google Fit API to automate data retrieval from smartphones, complemented by manual vital sign entry. Utilizing tools like Postman for API testing and management and Google Oath Playground for code and scope generation ensure efficient operation. Additionally, it offers users the flexibility of manual data entry as an alternative option, enhancing accessibility and usability.

Data Pre-processing and Machine Learning

In addition to handling missing data and imbalances, other preprocessing steps typically involve:

- To prevent model training from being dominated by large features, it's important to normalize or standardize numerical features to a similar scale.
- To make data appear Gaussian or capture nonlinear relationships, perform transformations such as logarithmic, polynomial, or Box-Cox transformations.
- Select appropriate features to enhance model performance and reduce over fitting.
- **Data Encoding:** Convert categorical variables into numerical representations suitable for model training (e.g., one-hot encoding, label encoding).

To build accurate and robust machine learning models, effective preprocessing is crucial. Depending on the nature of the data, the specific problem, and the requirements of the model being developed, preprocessing techniques should be chosen accordingly. Data preparation that is more reliable and interpretable from machine learning models can be achieved by thoroughly understanding these techniques.

Robust data-cleaning techniques were applied to handle missing values, outliers, and inconsistencies effectively. Machine learning

models are constructed and trained using Python libraries like Scikit-learn and Tensor Flow Lite. Furthermore, relevant features are extracted, and domain knowledge is utilized to create informative features, optimizing the model's performance. Each model was then trained on the preprocessed data with labeled sleep disorder and heart attack risk cases for which the hyper parameters for each model are further optimized to achieve the best performance and interoperability. The model's performance was evaluated using comprehensive metrics such as AUC, sensitivity, specificity, and F1-score. Rigorous validation was conducted with real-world data from clinical studies or large user groups, ensuring generalizability and clinical relevance.

Data Splitting: The preprocessed data was split into 70% training and 30% test sets using scikit-learn's "train_test_split" function.

Ensemble Construction

A heterogeneous improved Voting Classifier ensemble with 'Hard' type voting architecture was developed and options based on individual model performance were explored. Then, the improved model is trained on the combined predictions of individual models to improve accuracy and robustness while maintaining explainability. Figure 1 explains the Hard-Voting Formula based on which the model with the highest voting would be considered for evaluation. Our system leverages a powerful ensemble machine learning model to predict sleep disorders and associated heart health risks. This approach prioritizes accuracy for improved risk assessments. We utilize a combination of random forests, support vector machines (SVMs), and multilayer perceptron classifiers (MLPs) for the ensemble.

Algorithm Selection: This ensemble model ensures robust predictions by leveraging the strengths of each algorithm. RF provides resilience to over-fitting and valuable insights into feature importance. SVMs excel in separating data points in high-dimensional space, making them suitable for our predictive task. MLP, with its ability to capture complex patterns, complements the other algorithms in capturing nuanced relationships in the data. Choosing the right algorithms is crucial for building an accurate and efficient risk prediction model. We focused on the following key factors:

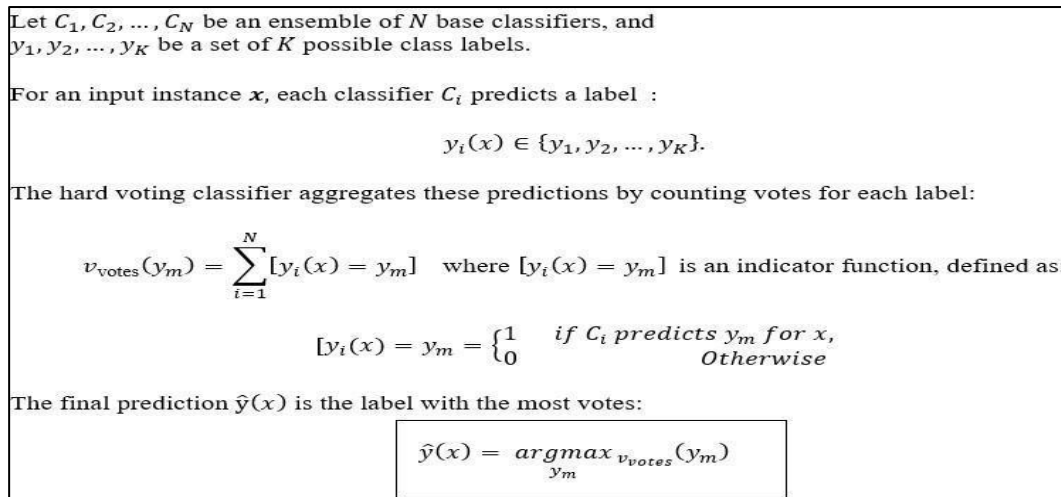


Figure 1: Hard Voting Formula

High Accuracy: The model effectively identifies individuals at risk for both sleep disorders and associated heart health complications.

Efficiency: Real-time or near-real-time predictions that are ideal for user-friendliness and timely interventions.

Data compatibility: The algorithms must handle the complexities of the mixed sleep and heart attack risk data.

Data Acquisition: API or manual input is used to collect relevant health vitals related to heart attack risk and sleep disorders from various sources, such as wearable devices, health apps, or self-reported data. Data security and privacy measures are implemented to protect user information.

Data Analysis: The collected health data were preprocessed and analyzed, and relevant features such as heart rate variability, blood pressure readings, activity levels, sleep patterns, and self-reports were extracted. Employ feature engineering techniques to identify meaningful patterns associated with heart attacks and sleep disorders.

Machine Learning Model: A machine learning model was trained using labeled data to predict potential heart attack risk and sleep disorders. Model selection and hyper parameter tuning were optimized for accuracy and generalizability in predicting these health indicators.

System Architecture: The incorporation of user data input into the system from two distinct

sources is done in accordance with the specific requirements and preferences of the user. Firstly, data is retrieved from the Google Fit API, which is capable of extracting motion sensor data, such as the number of steps taken, the intensity of physical activities, and the speed at which the user walks, from smart phones with the user’s informed consent. Alternatively, users also have the option to manually input vital signs directly into the machine learning model.

The machine learning model itself is composed of two foundational models that have been trained using a dataset that is publicly available. These models are designed to make predictions regarding the likelihood of sleep disorders and the risk of experiencing a heart attack, respectively. Following this, an ensemble model is employed to combine the outputs of these foundational models, thereby offering comprehensive insights into both sleep disorders and the risk of heart attacks.

Users are able to engage with the system in order to access information about their predicted health risks. This may include features such as the visualization of trends based on these predictions. Moreover, it is important to note that the foundational models have undergone thorough training and testing using the publicly available dataset. This process was crucial in ensuring the reliability of the predictions related to sleep disorders and the risk of heart attacks.

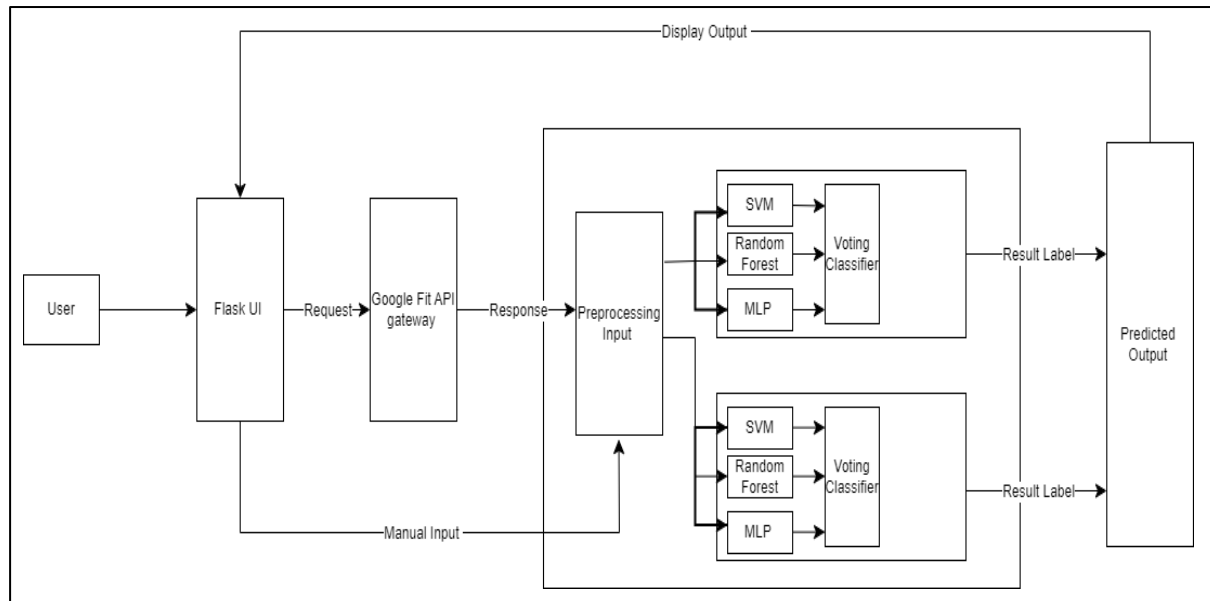


Figure 2: System Architecture

SVM for Heart Attack Prediction

SVMs can be trained to classify patients into different risk categories based on features such as age, blood pressure, cholesterol levels, and lifestyle factors. They can handle both linear and nonlinear relationships between these features and the likelihood of a heart attack.

Handling Imbalanced Data: In medical datasets where instances of heart attacks are relatively rare compared to non-events, SVMs can be optimized to handle imbalanced classes through techniques like class weighting or resampling. SVMs can incorporate various biomarkers (e.g., genetic markers, protein levels) alongside traditional risk factors to improve prediction accuracy. They can also handle multimodal data where different types of features (clinical, genetic, imaging) are combined.

SVM for Sleep Disorder Detection

SVMs can use EEG signals, such as spectral power, coherence, and other time-frequency measures, to classify sleep stages. Among them are wakefulness, REM sleep and different stages of non-REM sleep. Analyzing respiratory patterns, heart rate variability, and movement during sleep can be used to diagnose specific sleeping disorders like sleep apnea. The severity of the disorder can determine how patients are classified, such as normal, mild, moderate, or severe. Kernel functions have been utilized by SVMs to handle high-dimensional feature spaces, which map input data into higher-dimensional spaces where

nonlinear relationships can also be captured. Using this model in feature selection techniques can help identify the most relevant variables for detecting sleep disorders. Real-time assessments on sleep quality and disturbances can now be performed through wearable devices and mobile applications using SVMs. SVMs are effective in both heart attack prediction and sleep disorder detection in medical decision-making and patient care because they can handle complex dataset. Due to their ability to handle intricate datasets, these tools are highly versatile in predicting heart attacks and detecting sleeping disorders.

Variables have an impact on the heart attack and sleep disorders system:

Age: Heart attacks are a significant risk factor for old age. The cumulative effect of risk factors over years results in increased chances of suffering from heart attacks with age.

Gender: The risk of males is higher when they are young, and women's risks increase after menopause. Prediction models can be affected by various risk profiles and symptoms.

Cholesterol Levels: Significance High levels of LDL (bad cholesterol) and low levels of HDL (good cholesterol) are connected with heart disease. The presence of elevated LDL and reduced HDL concentrations is a significant indicator of myocardial infarction risks.

Blood Pressure: Hypertension is a significant risk factor that should be taken seriously. Increased blood pressure weakens arteries, making them more susceptible to coronary artery problems.

Smoking Status: Smoking has a significant impact on the chances of developing a cardiovascular disease. Current and former smokers are more likely to be diagnosed with lung cancer than non-smokers.

Diabetes: One of the main causes of heart disease is diabetes, which has a significant impact on blood vessels. The likelihood of atherosclerosis increases significantly when diabetes is present.

Family History: Genetic predisposition plays a role in the development of heart diseases in individuals. The likelihoods of individuals are enhanced by their family history.

Obesity: Being overweight can make someone appear unattractive.

In Figure 3, the correlation heatmap highlights interesting patterns: “exang” (chest pain) and “thal” (perfusion defects) show a strong negative correlation, suggesting that chest pain might not always indicate a significant blockage. Conversely, “thalach” (maximum heart rate) exhibited a strong positive correlation with the target variable, potentially linking a higher heart rate to increased risk.

In Figure 4 a strong negative correlation between systolic and diastolic pressure in the correlation heatmap aligns with expected blood pressure behavior, potentially indicating good data quality. This reduces redundancy in the data, which can improve model efficiency.



Figure 3: Correlation Heatmap (Heart)

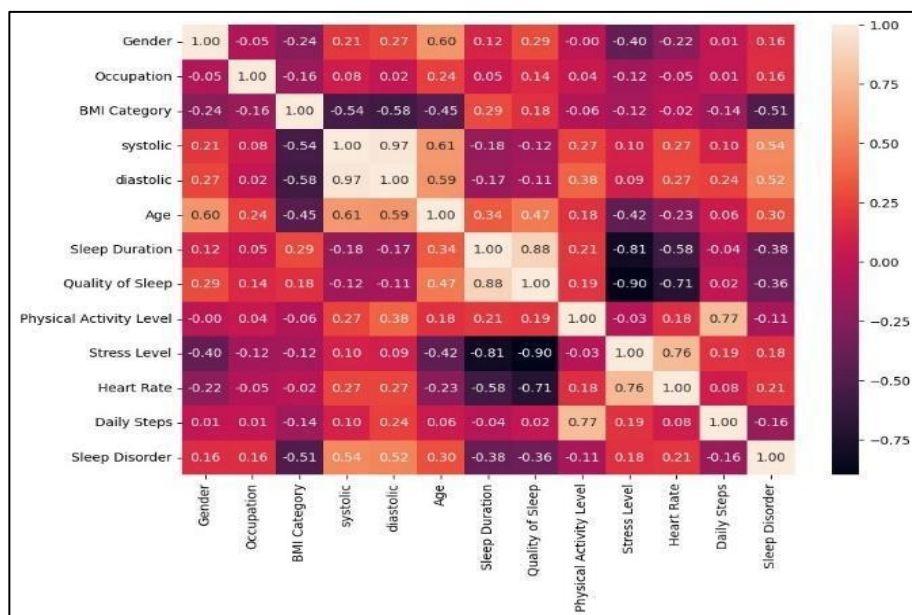


Figure 4: Correlation Heatmap (Sleep)

	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.97	0.95	0.96	41	0	0.96	0.98	0.97	44
1	0.96	0.98	0.97	50	1	0.94	1.00	0.97	15
					2	0.93	0.81	0.87	16
accuracy			0.97	91	accuracy			0.95	75
macro avg	0.97	0.97	0.97	91	macro avg	0.94	0.93	0.93	75
weighted avg	0.97	0.97	0.97	91	weighted avg	0.95	0.95	0.95	75

HEART
SLEEP

Figure 5: Classification Reports

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP) Type II Error	False Negative (FN) Type I Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Figure 6: Confusion Matrix

Sleep Disorder Risk Calculator

Gender:

Age:

Occupation:

Sleep Duration:

Physical Activity Level:

Stress Level:

BMI:

Heart Rate:

Daily Steps:

Systolic Pressure:

Diastolic Pressure:

Safe from any sleeping disorders!

[Home Page](#)

Heart Attack Risk Calculator

Age:

Gender:

Resting Blood Pressure:

Serum Cholesterol in mg/dl:

Fasting Blood Sugar:

Resting ECG Results:

Maximum Heart Rate Achieved:

Exercise Induced Angina:

Thalassemia:

You have a healthy heart!

[Home Page](#)

Figure 7: Prediction Pages

When a model learns too much about the noise details in training data, it can lead to overfitting, which can negatively affect its performance on new data. In medical research, such as heart attacks and sleep disorders, it becomes more difficult to develop models that can adapt to different patients or conditions. In this context, over fitting can lead to some potential problems, and there are ways to mitigate them. Combine forecasts from multiple models to lessen the risk of over fitting. Combining predictions from different models using methods like bagging, boosting, or stacking can help reduce the noise generated by individual models.

Results and Discussion

In Figure 5, we see how a confusion matrix is used to define the performance of a classification algorithm, and Figure 6 shows the classification report of the models. The models show praiseworthy results, and the accuracy attained by them is impressive.

Heart Attack Risk Prediction

As shown in Figure 7, while conventional approaches often face challenges in achieving accuracy above 80-85%, our ensemble achieved an impressive overall accuracy of 97%. Here are the findings of both the classes:

1. Class 0 (No Risk): Precision: 0.97, Recall: 0.95, F1-Score: 0.96, Support: 41
Description: The model accurately identifies individuals with no heart attack risk, with high precision and recall.
2. Class 1 (Risk): Precision: 0.96, Recall: 0.98, F1-Score: 0.97, Support: 50
Description: The model effectively detects individuals at risk of heart attacks, demonstrating excellent precision and recall.

Sleep Disorder Prediction

Similarly for sleep disorder prediction as shown in Figure 7, unlike traditional models which struggle to achieve accuracy above 80%, our approach achieved an impressive overall accuracy of 95%. Here are the results:

1. Class 0 (No Risk): Precision: 0.96, Recall: 0.98, F1-Score: 0.97, Support: 44
Description: The model correctly identifies individuals without sleep disorders, with good precision and recall.
2. Class 1 (Sleep Apnea): Precision: 0.94, Recall: 1.00, F1-Score: 0.97, Support: 15
Description: The model effectively detects

sleep apnea cases, with high precision and recall.

3. Class 2 (Insomnia): Precision: 0.93, Recall: 0.81, F1-Score: 0.87, Support: 16

Description: The model identifies individuals with insomnia, demonstrating good precision and recall.

User Interface

After signing in with their Google account and granting access to specified scopes, users are redirected to the homepage. Here, all relevant data is automatically stored in their session. They can then view pre-filled acquired data on the subsequent page, which can be easily modified if necessary. Users are prompted to input additional vitals and sleep-related data for a comprehensive input process. Upon completion, users receive an output detailing their diagnosis.

SVM models can provide insights that can guide further research into novel biomarkers and risk factors for heart attacks. The encouragement of innovation in cardiovascular medicine can result in the development of new diagnostic tools and therapies. Various sleep disorders can be accurately detected by SVM models, which enables timely diagnosis and appropriate treatment. Individualized interventions can be provided to patients, such as CPAP therapy for sleep apnea or behavioral therapy for insomnia, which can enhance sleep quality and overall health.

Heart attack risk prediction and sleep disorder identification have been significantly improved by machine learning algorithms.

1. In order to manage complex data, one must have the ability to process large, high-dimensional datasets.
2. Capturing non-linear relationships by accurately observing subtle interactions and patterns that are not taken into account by traditional models.
3. Continuous Monitoring can be incorporated using real-time data for more accurate and dynamic risk assessment.
4. By enabling more detailed and precise feature sets, Feature Engineering can be enhanced, leading to better model performance.

These advancements emphasize the potential of machine learning to enhance predictive accuracy and patient outcomes compared to traditional methods, while highlighting the importance of

continuous refinement and validation in diverse populations.

Conclusion

Our research emphasizes the effectiveness of our ensemble-based artificial intelligence (AI) model in the realms of predicting sleep disorders and evaluating the likelihood of heart attacks. By amalgamating a variety of machine learning algorithms, we have notably improved the precision and dependability of our forecasts. The model has surpassed conventional approaches, showcasing exceptional performance in both the assessment of heart attack risk and the prediction of sleep disorders. Moving forward, we could delve into a broader spectrum of machine learning algorithms and integrate additional sources of data, potentially elevating the accuracy of our model even further. This continuous process of refinement shows potential in enhancing the efficacy of our predictive capacities within the healthcare sector.

Abbreviations

Nil.

Acknowledgement

We would like to thank the Vellore institute of technology for supporting this research work.

Author Contributions

All the authors have contributed equally.

Conflict of Interest

Nil.

Ethics Approval

Not applicable.

Funding

No funding sources.

References

1. Menaspà P. Effortless activity tracking with Google Fit. *British Journal of Sports Medicine*. 2015 Sep 10;49(24):1598–8.
2. Mattison G, Canfell O, Forrester D, Dobbins C, Smith D, Töyräs J, Sullivan C. The influence of wearables on health care outcomes in chronic disease: systematic review. *Journal of Medical Internet Research*. 2022 Jul 1;24(7):e36690.
3. Baig MM, GholamHosseini H, Moqem AA, Mirza F, Lindén M. A Systematic Review of Wearable Patient Monitoring Systems – Current Challenges and Opportunities for Clinical Adoption. *J Med Syst*. 2017 Jul;41(7):115.
4. Raihan M, Mondal S, More A, Boni PK, Sagor MdOF. Smartphone Based Heart Attack Risk Prediction System with Statistical Analysis and Data Mining Approaches. *Advances in Science, Technology and Engineering Systems Journal*. 2017 Nov;2(3):1815–22.
5. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, Urey MA, et al. Improving risk prediction in heart failure using machine learning. *European J of Heart Fail*. 2020 Jan;22(1):139-47.
6. Park J, Urtnasan E, Kim S, Lee K. A Prediction Model of Incident Cardiovascular Disease in Patients with Sleep-Disordered Breathing. *Diagnostics*. 2021 Nov 26;11(12):2212.
7. Nagai M, Hoshida S, Kario K. Sleep Duration as a Risk Factor for Cardiovascular Disease- a Review of the Recent Literature. *CCR*. 2010 Feb 1;6(1):54-61.
8. Rajendran NA, Vincent DR. Heart Disease Prediction System using Ensemble of Machine Learning Algorithms. *ENG*. 2021 Mar 9;15(2):130-139.
9. Hyland SL, Faltys M, Hüser M, Lyu X, Gumbsch T, Esteban C, et al. Early prediction of circulatory failure in the intensive care unit using machine learning. *Nat Med*. 2020 Mar;26(3):364-73.
10. Rajarajeswari P, Anwar Beg O. An executable method for an intelligent speech and call recognition system using a machine learning-based approach. *Journal of Mechanics in Medicine and Biology*. 2021 Sep 8;21(07):2150055.
11. Perepi R, Santhi K, Saraswathi R, Bég OA. Predictive analysis of COVID 19 disease based on mathematical modelling and machine learning techniques. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. 2023 May 4;11(3):933-41.
12. Rajarajeswari P, Santhi K, Chattopadhyay P, Bég OA. A deep learning computational approach for the classification of Covid-19 virus. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. 2022,11(3):1-10.