First Review Document

Establishing Relationships and Detection of Sleep Disorders and Heart Attack Risk

Saumya Sejal
20BCE2596
7209416003
saumya.sejal2020@vitstudent.ac.in

Lokesh K
20BCE2621
7411811382
lokesh.k2020@vitstudent.ac.in

Aditya Nooka
20BCE2585
9849079490
nooka.aditya2020@vitstudent.ac.in

Perepi Rajarajeshwari
Associate Professor Grade 1
9701769770
rajarajeswari.p@vit.ac.in

B.Tech.

in

Computer Science and Engineering

School of Computer Science & Engineering (SCOPE)

®

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Abstract

This project uses a heterogeneous ensemble-based AI model predicting sleep disorders and the risk of heart attack through accessible, real-time analysis of diverse personal health data. To overcome the limitations of existing solutions, our system empowers users with the necessary control and personalized insights via seamless data integration. Employing a meticulously crafted ensemble of Random Forests, Support Vector Machines, Neural Networks, and Transformer-based models, the system delves deep into diverse data streams to deliver highly accurate and interpretable risk predictions. This user-centric and data-driven approach, including various data sources such as Google Fit, holds immense potential for improving health outcomes and driving personalized preventative healthcare.

Keywords: Ensemble Modeling, AI-driven Healthcare, Data-driven Health Prediction, Sleep Disorders, Heart Attack Risk, Remote Patient Monitoring.

Introduction

The AI-Driven Remote Patient Monitoring System represents a significant shift in how we approach healthcare, offering a comprehensive solution that caters to the needs of users across all age groups. Utilizing advanced technology, this system provides deep insights into crucial health metrics such as sleep disorders and heart attack risk, enabling individuals to take proactive steps toward improving their overall health and well-being. Traditional remote patient monitoring often struggles with limited accessibility, inaccurate predictions, and reliance on specific data sources, leading to suboptimal health outcomes.

Our ensemble-based AI model overcomes these limitations by providing a highly accurate, accessible, and insightful platform for remote health monitoring and analysis. This system empowers users to take charge of their health and prevent potential health concerns by leveraging the diverse strengths of individual models and rich data landscapes beyond traditional sources.

Furthermore, the system's ability to predict heart attack risk is a game-changer in preventive healthcare. Early detection of risk factors and timely interventions can significantly reduce the likelihood of cardiovascular-related incidents.

Problem Statement

Existing predicting models suffer from limited accessibility, inaccurate predictions, and reliance on specific data sources, resulting in suboptimal health outcomes. Users struggle to obtain personalized insights and control over their health, hindering proactive management of potential health concerns such as sleep disorders and heart attack risks.

Proposed Work

- Utilization of an ensemble approach that integrates diverse AI models, each with unique strengths tailored for analyzing specific aspects of personal health data.
- Exploration beyond traditional data sources, unlocking access to previously untapped data streams, enriching prediction models, and enhancing their accuracy.
- Emphasis on interpretability and explainability, enabling users to comprehend model predictions, establish trust, and actively engage in managing their health.
- Integration of personalized insights, intuitive interfaces, and clear explanations within the system, fostering proactive health management and improving overall health outcomes.

Objectives

Develop a cutting-edge healthcare solution that combines:

- Heterogeneous ensemble model: Combining the unique strengths of:
    - *Random Forests:* Capturing feature importance and non-linear relationships for robust sleep and heart health analysis.
    - *Support Vector Machines:* Providing strong classification performance and interpretability for accurate risk prediction and model explanation.
    - *Neural Networks:* Handling sequential sleep and physiological data effectively to learn temporal patterns and anomalies.
    - *Transformer-based models:* Leveraging their parallel processing capabilities for efficient analysis of complex data sequences, including identifying subtle sleep stage abnormalities and physiological variations indicative of potential risks.
- Personalized and interpretable insights: Delivering actionable and individualized guidance tailored to user-specific risk profiles and data trends, along with clear explanations of model predictions for trust and understanding.

- User-friendly and intuitive interface: Fostering user engagement through clear and engaging data visualization and feedback mechanisms, empowering individuals to manage their health actively.

Methodology

- *Data Acquisition*:
  - Integrate data from Google Fit through various sources, including wearable devices, health records, and user inputs.
  - Google Fit API for automated data retrieval from user smartphones.
  - Manual data entry option for vital signs directly into the system.
  - Postman for API request testing and management.
  - Google OAuth Playground for API access code and scope generation.
  - Allow for manual data entry as an alternative option.

- *Data Pre-processing and Machine Learning:*
  - Apply robust data-cleaning techniques to address missing values, outliers, and inconsistencies.
  - Python libraries such as ***NumPy*** and ***Pandas*** for data manipulation and analysis.
  - Python libraries like ***Scikit-learn*** and ***TensorFlow Lite*** for building and training machine learning models.
  - Extract relevant features and create informative features through domain knowledge and data analysis.

- *Individual Model Training:*
  - Train each model (Random Forest, SVM, Neural network-based model, Transformer) on the preprocessed data with labelled sleep disorder and heart attack risk cases.
  - Optimize hyperparameters for each model to achieve the best performance and interpretability.

- *Ensemble Construction:*
  - Develop a heterogeneous ensemble architecture, exploring options based on individual model performance and interpretability.
  - Train the ensemble model on the combined predictions of individual models to improve accuracy and robustness while maintaining explainability.

- *Extensive Evaluation and Validation:*
  - Evaluate the ensemble model's performance using comprehensive metrics like AUC, sensitivity, specificity, F1-score, and interpretability measures.
  - Conduct rigorous validation with real-world data from clinical studies or large user groups, ensuring generalizability and clinical relevance.

- *Create a User Interface*
  - Flask or similar frameworks for developing a user-friendly interface to interact with the system.

Literature Survey

S.No.	**Title**	**Summary**
1	Effortless activity tracking with Google Fit	This paper highlights Google Fit's versatility in monitoring physical activity, including various tracked activities and manual input options. It emphasizes the app's seamless data collection and analysis through smartphones, potentially promoting increased physical activity through self-tracking.
2	The Influence of Wearables on Health Care Outcomes in Chronic	A systematic review investigating the potential of wearable technology to improve health outcomes in individuals with chronic diseases. It analyzes various studies using wearables to monitor metrics like

	Disease: Systematic Review	physical activity, heart rate, and sleep in patients with chronic conditions.
3	Systematic Review of Wearable Patient Monitoring Systems	This review aims to investigate barriers and challenges of wearable patient monitoring solutions adopted by clinicians in acute, as well as in the community.
4	A Survey on Multi-Label Data Stream Classification	This survey delves into various strategies for processing large volumes of multi-label streaming data, addressing challenges like single-pass processing, real-time response, and drift detection. It explores different classification methods, including ensemble and AA-based approaches, for handling such complex data streams.
5	Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design	This research describes the development of an Android prototype app for predicting heart attack risk. It utilizes clinical data from patients with ischemic heart disease and considers various risk factors like hypertension, diabetes, and stress. The paper demonstrates the feasibility of using smartphones and data mining techniques for personalized heart disease risk assessment.
6	Improving risk prediction in heart failure using machine learning	This study leverages machine learning algorithms to enhance the prediction of mortality in heart failure patients. It analyzes various patient characteristics like blood pressure, blood cell counts, and other clinical measures to identify patterns associated with

		increased mortality risk. This approach aims to improve patient care by enabling earlier intervention for high-risk individuals.
7	A Prediction Model of Incident Cardiovascular Disease in Patients with Sleep-Disordered Breathing	This study proposes a machine learning model to predict cardiovascular disease within ten years in patients with sleep-disordered breathing (SDB). They analyze data from the Sleep Heart Health Study, utilizing various features like ECG signals, clinical risk factors, and AI-based parameters. The model achieved good performance in predicting various outcomes, suggesting the potential of using SDB data for early risk assessment.
8	Sleep Duration as a Risk Factor for Cardiovascular Disease- a Review of the Recent Literature	This review focuses on the relationship between sleep duration and cardiovascular disease risk. It compiles findings from various studies, highlighting that short sleep duration is associated with an increased risk of hypertension, coronary heart disease, and diabetes. This suggests that promoting healthy sleep habits could be a preventive strategy.
9	Insomnia and Risk of Cardiovascular Disease	This research specifically explores the link between insomnia and risk. It reviews epidemiological studies that demonstrate an association between chronic insomnia and increased incidence of various conditions like hypertension, coronary heart disease, and heart failure. The review also discusses potential mechanisms linking these two conditions, suggesting

		the importance of addressing insomnia for cardiovascular health.
10	Early prediction of circulatory failure in the intensive care unit using machine learning	This study demonstrates the potential of machine learning in predicting circulatory failure in intensive care unit (ICU) patients. The model analyzes various patient data points to identify those at high risk of circulatory failure, enabling earlier intervention and potentially improving patient outcomes. This research highlights the potential of machine learning for real-time risk prediction in critical care settings.

System Design

1. ***Data Acquisition***: Utilize APIs or manual input to collect relevant health data related to heart attack risk and sleep disorders from various sources, such as wearable devices, health apps, or self-reported data. Ensure data security and privacy measures are in place to protect user information.

2. ***Data Processing and Analysis***: Pre-process and analyze the collected health data, extracting relevant features such as heart rate variability, blood pressure readings, activity levels, sleep patterns, and self-reported. Employ feature engineering techniques to identify meaningful patterns associated with heart attack and sleep disorder.

3. ***Machine Learning Model***: Train a machine learning model, such as Random Forest or Logistic Regression, using labelled data to predict potential heart attack risk and sleep disorders based on the extracted features. Optimize model selection and hyperparameter tuning for accuracy and generalizability in predicting these health indicators.

System Architecture

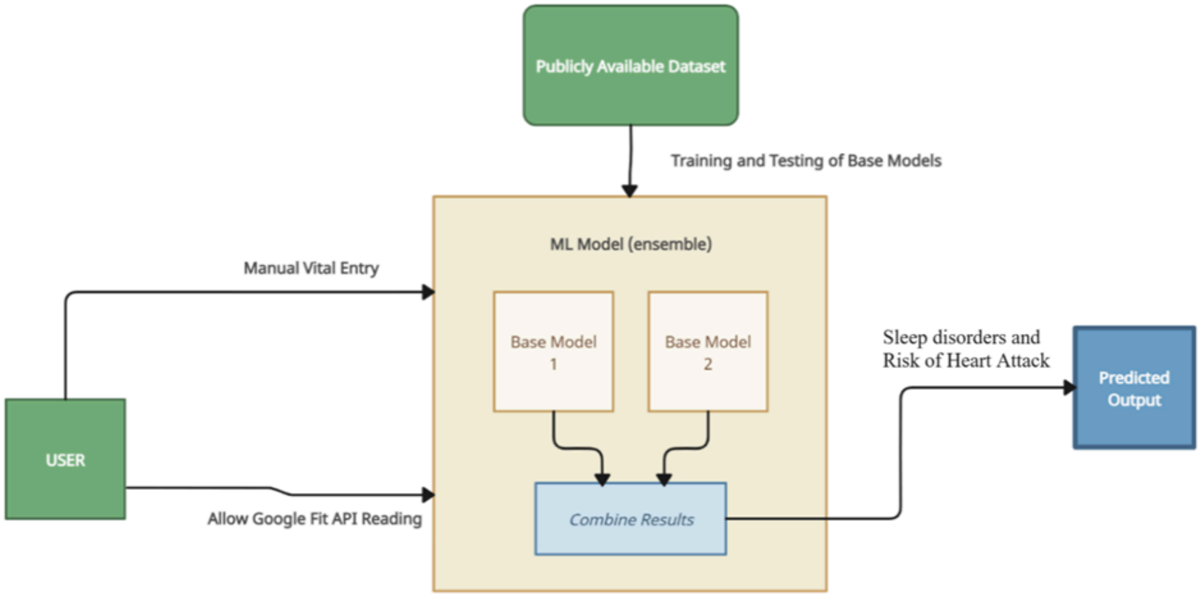

Figure 1 System Architecture

Description

A. Data Input:

- Via the Google Fit API, allowing the system to read motion sensor data (e.g., steps, activity intensity, walking speed) from user smartphones with informed consent.
- Manual entry of vital signs directly into the machine learning model.

B. Machine Learning Model:

- The system employs an ensemble model comprising two base models trained on a publicly available dataset.
- Base Model 1 and Base Model 2 are trained and tested using the dataset to predict sleep disorder and heart attack risk.
- The ensemble model combines the results of Base Model 1 and Base Model 2 to predict the output, providing insights into sleep disorder and heart attack risk.

C. Integration and Prediction:

- The system integrates the data input from users with the ensemble model.
- Utilizing the input data and predictions from the ensemble model, the system predicts sleep disorders and heart attack risk for the users.

D. User Interaction:

- Users can interact with the system to view their predicted sleep disorders and heart attack risk.
- The system may provide additional features such as visualization of trends and personalized recommendations based on the predictions.

E. Training and Testing:

- Base models are trained and tested on a publicly available dataset to ensure accuracy and reliability.
- The training and testing processes are crucial for optimizing the performance of the ensemble model in predicting sleep disorders and heart attack risk accurately.

Conclusion

In conclusion, the development of AI-driven ensembles for predicting heart attack risk, and sleep disorders represents a significant step towards personalized and proactive healthcare management. The existing systems only use a single model for predictions, resulting in inaccurate predictions. The current systems are reliant on specific data sources and have limited accessibility. This project aims to address these issues by using an ensemble model which implements techniques such as random forest, SVN etc. By addressing these challenges and leveraging insights from real-time data from the Google Fit API, this project aims to empower individuals with actionable health insights, ultimately enhancing their overall well-being and quality of life.

References

1. Menaspà, Paolo. (2015). Effortless activity tracking with Google Fit. British journal of sports medicine. 49.http://dx.doi.org/10.1136/bjsports-2015-094925

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, https://doi.org/10.2196/36690

3. Baig, M.M., GholamHosseini, H., Moqeem, A.A. et al. A Systematic Review of Wearable Patient Monitoring Systems – Current Challenges and Opportunities for Clinical Adoption, (2017). https://doi.org/10.1007/s10916-017-0760-1

4. X. Zheng, P. Li, Z. Chu and X. Hu, A Survey on Multi-Label Data Stream Classification, IEEE Access, vol. 8, pp. 1249-1275, 2020, https://ieeexplore.ieee.org/document/8941052

5. M. Raihan et al., "Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design," 2016 19th ICCIT, Dhaka, Bangladesh, pp. 299-303, https://doi.org/10.1109/ICCITECHN.2016.7860213

6. Adler, E.D., Voors, A.A., Klein, L., Macheret, F., Braun, O.O., Urey, M.A., Zhu, W., Sama, I., Tadel, M., Campagnari, C., Greenberg, B. and Yagil, A. (2020), Improving risk prediction in heart failure using machine learning. Eur J Heart Fail, 22: 139-147. https://doi.org/10.1002/ejhf.1628

7. JU, Urtnasan E, Kim SH, Lee KJ. A Prediction Model of Incident Cardiovascular Disease in Patients with Sleep-Disordered Breathing. Diagnostics (Basel). 2021 Nov 26;11(12):2212. https://doi.org/10.3390/diagnostics11122212

8. Nagai M, Hoshide S, Kario K. Sleep duration as a risk factor for cardiovascular disease- a review of the recent literature. Curr Cardiol Rev. 2010 Feb;6(1):54-61. https://doi.org/10.2174/157340310790231635

9. Javaheri S, Redline S. Insomnia and Risk of Cardiovascular Disease. Chest. 2017 Aug;152(2):435-444. https://doi.org/10.1016/j.chest.2017.01.026

10. Early prediction of circulatory failure in the intensive care unit using machine learning,Hyland, S.L., Faltys, M., Hüser, M,Nat Med 26, 364–373 (2020). https://doi.org/10.1038/s41591-020-0789-4