engineering proceedings

MDPI

Proceeding Paper

Microcontroller-Based EdgeML: Health Monitoring for Stress and Sleep via HRV †

Priyanshu Srivastava ‡, Namita Shah ‡ and Kavita Jaiswal *

Department of Computer Science and Engineering, International Institute of Information Technology, Naya Raipur 493661, India; priyanshu22101@iiitnr.edu.in (P.S.); namita22101@iiitnr.edu.in (N.S.)

* Correspondence: kavita@iiitnr.edu.in

† Presented at the 1st International Conference on AI Sensors & the 10th International Symposium on Sensor Science, Singapore, 1–4 August 2024.

‡ These authors contributed equally to this work.

Abstract: The healthcare sector is undergoing a transformation with the integration of cutting-edge technologies such as machine learning (ML), the Internet-of-Things (IoT), and Cyber–Physical Systems (CPS). However, traditional ML systems often face challenges in real-time processing and resource efficiency, limiting their application in life-critical scenarios. This research explores the potential of edge ML, particularly TinyML with TensorFlow Lite, implemented on microcontroller-based AI sensors for real-time health monitoring. By leveraging model quantization, the system analyzes heart rate variability (HRV) data to deliver continuous and personalized insights into stress levels and sleep quality. Trained on SWELL and ISRUC datasets, the system is highly energy-efficient, consuming 33 mW in idle mode, 66 mW during data collection, and 99 mW during real-time inference, making it suitable for resource-constrained environments. Performance analysis reveals significant demographic variations: younger individuals (18–25) achieved 90% accuracy due to higher HRV and lower baseline stress, while middle-aged (26–50) and older adults (50+) demonstrated declining HRV, reducing accuracy to 82% for the latter. Gender differences were also observed, with males exhibiting greater stress response sensitivity and better accuracy (89%) compared to females. This study underscores the transformative potential of TinyML for real-time, energy-efficient health monitoring and emphasizes the need for demographic-specific optimizations to enhance system reliability and accessibility.

Keywords: TinyML; EdgeML; AI sensors; microcontrollers; healthcare

check for updates

Citation: Srivastava, P.; Shah, N.; Jaiswal, K. Microcontroller-Based EdgeML: Health Monitoring for Stress and Sleep via HRV. *Eng. Proc.* **2024**, *78*, 3. https://doi.org/10.3390/engproc2024078003

Academic Editor: Po-Liang Liu

Published: 4 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The healthcare sector is experiencing a profound transformation driven by the integration of cutting-edge technologies such as machine learning (ML), the Internet-of-Things (IoT), and Cyber–Physical Systems (CPS). While these innovations have spurred the development of solutions across various health sectors, many rely heavily on traditional ML infrastructure [1], which can be resource-intensive and lacks real-time capabilities. Moreover, cloud-based solutions often fail to meet the stringent requirements of life-critical healthcare applications due to latency, bandwidth, and privacy concerns. Traditional ML systems also suffer from the need for continuous cloud connectivity, which can result in data privacy risks and inefficiencies in bandwidth usage. In response to these challenges, edge computing has emerged as a compelling alternative, offering the capability to perform computations closer to the source of data generation, thereby reducing dependence on continuous cloud connectivity. By leveraging edge computing, healthcare applications [2,3] can benefit from improved latency, bandwidth efficiency, and data privacy. Several studies have explored the potential of edge computing in real-time health monitoring tasks, such as electrocardiogram (ECG) analysis and fall detection. For example, Khaled et al. (2022)

examined the use of TinyML for continuous blood pressure estimation on the extreme edge [4]. However, existing solutions still face challenges such as high power consumption, lack of scalability, and limited real-time capabilities.

The focus of our project lies in harnessing the potential of embedded ML, specifically TinyML, to address the growing demand for real-time health monitoring solutions. TinyML refers to the deployment of machine learning models on resource-constrained embedded devices [5,6], enabling them to perform inference tasks locally without relying on cloud infrastructure. Our project aims to leverage TinyML technology to create a real-time platform for stress and sleep monitoring using heart rate variability(HRV), with the overarching goal of enhancing personalized health interventions.

While machine learning techniques have been widely applied to health monitoring tasks, including HRV analysis [7], stress monitoring [8], and disease prediction, existing solutions often rely on cloud-based architectures. These systems face challenges such as high power consumption, delayed data processing, and privacy concerns due to cloud dependency. Moreover, many existing designs focus on general population models and fail to personalize their health interventions. For instance, Li et al. (2013) used wavelet analysis and a support vector machine classifier to detect driver drowsiness based on heart rate variability [9], while Radha et al. (2019) performed sleep stage classification from HRV using long short-term memory neural networks [10]. These systems, however, did not address the real-time processing challenges associated with resource-constrained environments. By integrating TinyML with HRV metrics, our system seeks to overcome these limitations by providing a low-power, real-time solution that ensures data privacy and personalized health monitoring. The comparison table provided in Table 1 highlights the novelty of our proposed approach compared with published works. The table provides a detailed comparison based on model architecture, quantization techniques, system latency, power consumption, and cloud dependency. Our TinyML-based system offers unique advantages, such as real-time inference on embedded devices, lower power consumption, and improved latency, which are crucial for wearable health-monitoring applications.

Table 1. Comparisonwith existing approaches.

Study	Model Used	HRV Metrics	Quantization Technique	Cloud Dependency	Latency	Power Consumption	Contribution
[9] Li et al. (2013)	SVM (Support Vector Machine)	Time-domain (SDNN, RMSSD)	None	Yes	High	High	Drowsiness detection using HRV for drivers, cloud-based, no real-time capabilities.
[10] Radha et al. (2019)	LSTM (Long Short-Term Memory)	Time-domain and Frequency-domain (LF/HF)	None	Yes	Medium	High	Sleep stage classification from HRV, limited real-time functionality, requires cloud processing.
[4] Khaled et al. (2022)	TinyML (Regression)	Continuous Blood Pressure Estimation	Yes	No	Low	Low	Edge-based BP monitoring using TinyML, reduced cloud reliance but limited to blood pressure estimation.
Our Approach	TensorFlow Lite (TFLite) and custom feedforward neural network	Time-domain, Frequency-domain, Non-linear	Yes (Post-Training)	No	Low	Very Low	Real-time stress and sleep analysis using HRV metrics, personalized interventions, energy-efficient and fully cloud-independent.

2. Materials and Methods

This section describes the methodology followed in this research, outlining data collection, model development, deployment, and evaluation. The process flow of the entire methodology is depicted in Figure 1.

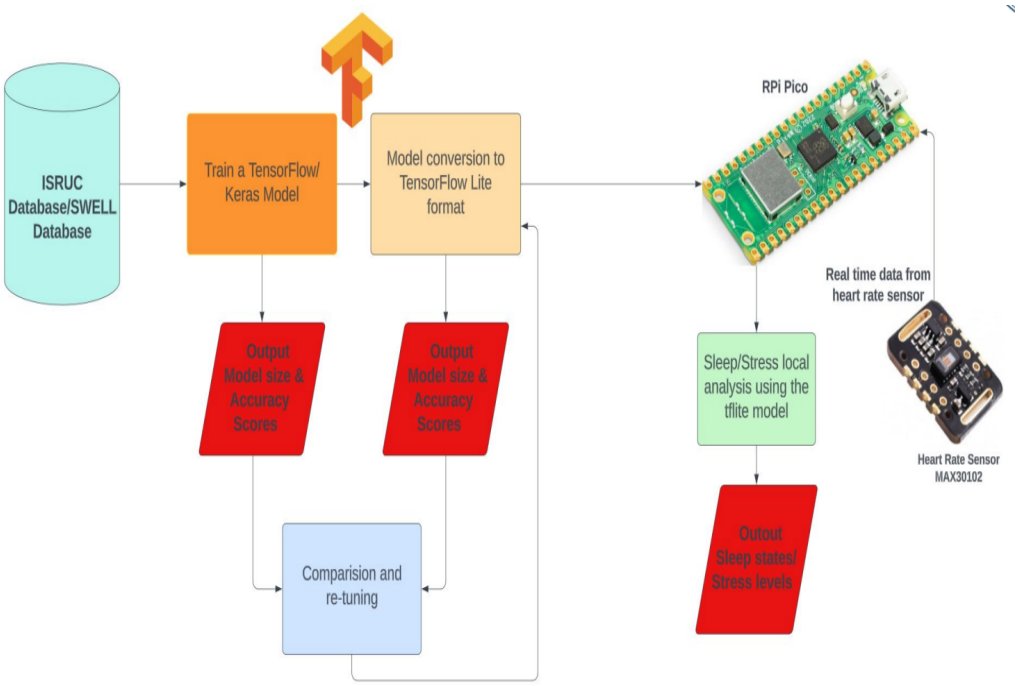


Figure 1. Flowchart of the proposed methodology.

2.1. Data Collection and Preprocessing

The HRV data utilized in this study were obtained from the ISRUC and SWELL datasets, both widely used for sleep and stress prediction. These datasets provide over 30 HRV metrics, including time-domain features such as Standard Deviation of Normal-to-Normal Intervals (SDNN), Root Mean Square of Successive Differences (RMSSD), and Percentage of successive RR intervals differing by more than 50 ms (pNN50), as well as frequency-domain features like Low Frequency (LF), High Frequency (HF), and LF-HF ratios. Preprocessing involved removing noise and artifacts using signal filtering methods, normalizing the data to account for sensor variations, and segmenting it into epochs corresponding to activities such as sleep and stress for targeted analysis.

2.2. Model Development and Optimization

The machine learning model was a custom feedforward neural network designed for stress classification using 34 HRV features (e.g., RR intervals, RMSSD, pNN50, LF/HF ratios). The architecture included an input layer with 34 nodes, 2 hidden layers (20 and 10 nodes with ReLU activation), and a softmax output layer for classifying stress levels into low, moderate, and high categories. The model was trained using categorical cross-entropy loss and the Adam optimizer (learning rate 0.0001). Techniques such as data augmentation and cross-validation improved generalization and robustness. Quantization reduced the model size by converting weights from 64-bit floats to 32-bit floats, enabling efficient deployment on edge devices.

2.3. Implementation on Microcontroller

The trained model was deployed on the Raspberry Pi Pico microcontroller (Manufactured by Raspberry Pi Ltd, based in Cambridge, UK), which features 256 kB SRAM and 2 MB storage. The microcontroller was set up using the Pico SDK (versionv1.4.0) , and a compatible heart rate sensor was integrated to provide real-time HRV data. Optimization techniques such as quantization and pruning reduced the computational complexity and memory footprint, ensuring efficient inference within the resource constraints of the hardware. To protect user privacy, all health data processing occurred locally on the microcontroller. Encryption ensured the confidentiality of sensitive information during storage and transmission.

2.4. Real-Time Health Insights and Validation

The testing data were collected using a heart rate sensor connected to the Raspberry Pi Pico microcontroller, which captures real-time heart rate measurements. These measurements were used to compute HRV metrics such as RR intervals, SDNN, RMSSD, pNN50, LF, and HF. The HRV metrics were then fed into the TinyML model for stress and sleep state prediction. The system's performance was evaluated through extensive testing, including assessments of accuracy, responsiveness, and energy efficiency. Comparisons with existing health monitoring systems confirmed its viability and competitiveness. Pilot studies gathered user feedback on system usability and effectiveness, which informed iterative improvements. Field deployment in real-world settings, such as homes and workplaces, validated the system's scalability and adaptability.

2.5. Deployment and Scalability

The HRV analysis system was designed to be scalable and adaptable, allowing for future enhancements. Data visualization tools were developed to present HRV analysis results in an accessible format, providing users with feedback on their health trends over time. Regular updates and maintenance are planned to ensure the long-term usability and success of the system.

3. Results and Discussion

This section presents the key findings from our study on stress prediction and sleep stage classification using TinyML on resource-constrained edge devices, along with a discussion of the implications of these results.

- **Stress Prediction Model Optimization:** The stress prediction model was successfully optimized for deployment on the Raspberry Pi Pico W microcontroller platform. The TensorFlow Lite (TFLite) model was compressed to one-sixth of its original size while maintaining an accuracy of approximately 90% (Figure 2). This model size reduction ensures efficient performance on edge devices with constrained resources. The high accuracy achieved despite the reduced model size demonstrates the feasibility of running advanced machine learning models on low-power hardware, highlighting the practicality of TinyML in real-world health monitoring applications.
- **HRV Metrics Computation:** The system processes real-time heart rate inputs to compute key HRV metrics such as RR intervals, SDRR, RMSSD, pNN25 and pNN50 and LF and HF components. HRV metrics are derived from heart rate by analyzing RR intervals—the time between consecutive heartbeats. As shown in (Figure 3), the heart rate samples are used to calculate the HRV metrics such as MEAN_RR, RMSSD, pNN25, and pNN50 using the appropriate formula. These HRV metrics are critical for accurately predicting stress levels, as they reflect the autonomic nervous system's response to stress. Leveraging these metrics in the model allows for a robust assessment of stress conditions.
- **Real-time Stress Prediction:** The TFLite model predicts stress conditions by analyzing the HRV metrics and estimating the probability of stress in real time (Figure 4). The ability to provide real-time stress assessment is a key advantage of this approach, as it allows continuous monitoring without the need for external cloud infrastructure. This not only reduces latency but also ensures that sensitive health data remain local to the device, thereby protecting user privacy.
- **HRV and Sleep Stage Classification:** We also investigated how HRV metrics vary across different sleep stages, including wake, N1, N2, N3, and REM. This analysis revealed distinctive HRV patterns associated with each stage, which can aid in accurately classifying sleep states. The box plot (Figure 5) visualizes the variations in heart rate during these stages, providing a deeper understanding of how sleep quality and disruptions can be detected using HRV data. This finding demonstrates the potential of the model to extend beyond stress prediction and contribute to more comprehensive health monitoring, including sleep analysis.

- **Model Performance Based on Demographics:** Our study found that age and gender significantly affect HRV and stress prediction accuracy (see Table 2). Younger individuals (18–25) had higher HRV and lower baseline stress, resulting in high model accuracy. Middle-aged groups (26–50) showed a gradual HRV decline, with slightly increased false negatives due to subtler changes. Older adults (50+) had lower HRV, reducing prediction accuracy, suggesting that age-specific tuning could help. Gender differences also impacted results: males had lower HRV but higher stress response sensitivity, while females had subtler changes, leading to slightly more false negatives. Future adjustments will further improve model sensitivity across demographics.
- **Power Consumption Evaluation:** To ensure the feasibility of deploying our TinyML-based system in wearable health-monitoring devices, we evaluated its power consumption across different operational stages. Power measurements were conducted during the idle state, data collection, and inference stages using a power monitoring tool, such as the Monsoon Power Monitor or INA219. The power consumption was calculated using the following formula:

$$\text{Power (mW)} = \text{Voltage (V)} \times \text{Current (mA)}$$

The system consumes approximately 33 mW in the idle state when powered on but not performing active computations. During data collection, which involves acquiring heart rate data, power consumption increases to 66 mW due to the activation of sensors and data processing. Power consumption peaks at 99 mW during real-time stress and sleep analysis using TinyML, which is still within the acceptable range for wearable devices. These measurements demonstrate the system's energy efficiency, which is attributed to the optimized deployment of the TinyML model on a microcontroller (e.g., ARM Cortex-M series). The results highlight the suitability of the proposed system for continuous health monitoring in resource-constrained environments, operating well within the power budgets of typical wearable and embedded systems.

Table 2. Model performance based on demographics.

Demographic Group	Accuracy	Precision	Recall	F1-Score
Male	89%	0.90	0.88	0.89
Female	86%	0.87	0.84	0.85
18–25 Age Group	90%	0.91	0.89	0.90
26–35 Age Group	87%	0.88	0.86	0.87
36–50 Age Group	85%	0.86	0.83	0.84
50+ Age Group	82%	0.83	0.80	0.81

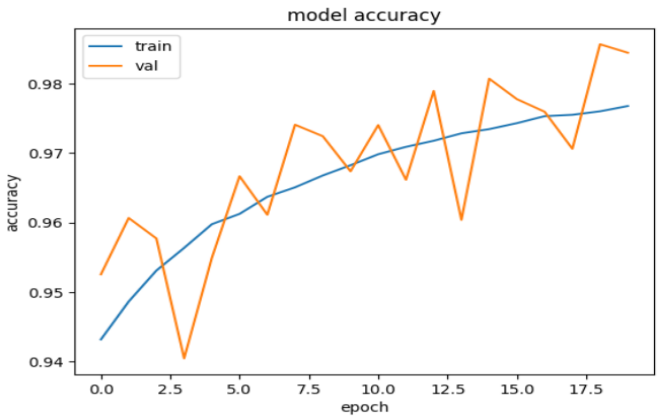


Figure 2. Model accuracy score.

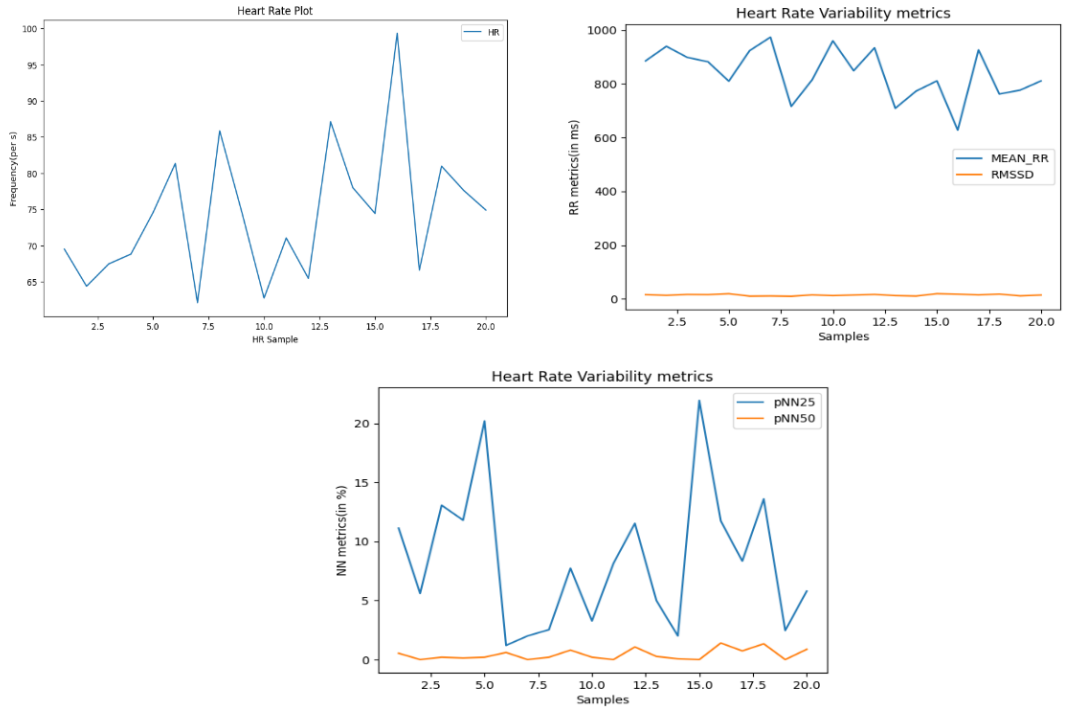


Figure 3. HRV metrics calculated from heart rate.

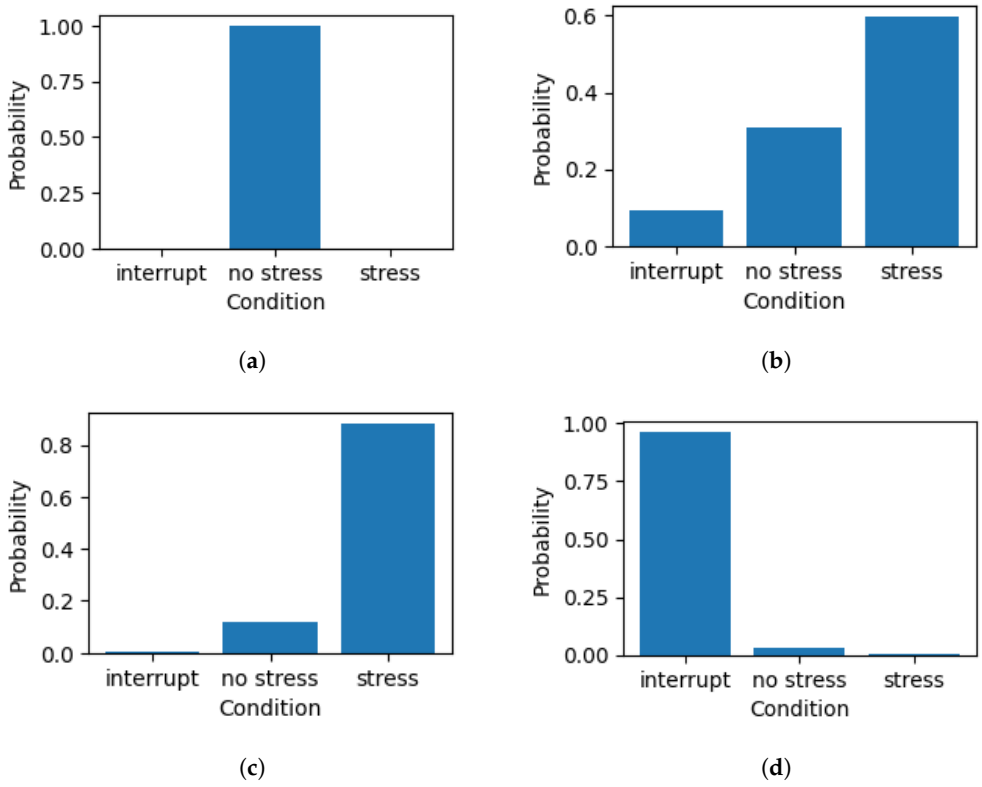


Figure 4. Different levels of stress prediction. (**a**) No stress. (**b**) Medium stress. (**c**) High stress. (**d**) Interruption during measurement.

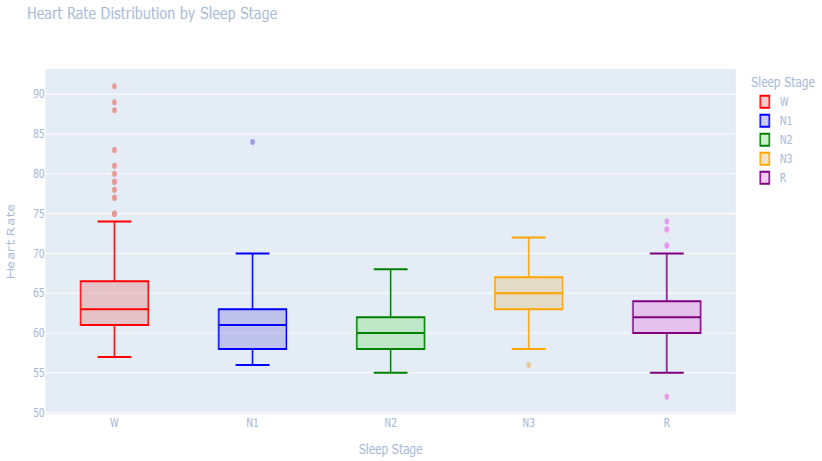


Figure 5. Heart rate variations for different stages.

The results of this study align with prior research on using edge devices for health monitoring, such as TinyCare, which deployed TinyML for blood pressure estimation on constrained devices. HRV metrics have proven effective for stress and fatigue detection, such as in driver drowsiness studies.

This study underscores the potential of TinyML for personalized health monitoring, particularly in applications requiring privacy and low latency. By keeping all computations on the edge device, the model eliminates the need for cloud infrastructure, thereby securing sensitive heart rate data from third-party access. Although this research focused on stress prediction, similar methods could be applied to sleep monitoring. Running two models simultaneously in the same environment was beyond the scope of this study, but this remains a promising area for future research. The aim of this study to optimize health-related models for deployment on edge devices was successfully realized.

4. Conclusions and Future Work

This paper demonstrates the successful deployment of a stress prediction model on a resource-constrained microcontroller, showing that advanced machine learning models can be effectively run on low-powered hardware without compromising accuracy. By operating entirely on the edge, this approach eliminates the need for cloud infrastructure, thereby reducing latency and enhancing privacy by keeping sensitive data local. The results reinforce the potential of TinyML for personalized health monitoring, specifically in stress and sleep analysis. The use of a microcontroller such as the RPi Pico further validates the feasibility of running health-related models on devices with limited memory and processing power, making this an important step towards more efficient and private health monitoring systems.

Future research could focus on optimizing the current model to handle additional health parameters, such as sleep stage classification, in the same environment. Energy efficiency improvements are also essential for ensuring long-term deployment, particularly on wearable devices. Expanding the system to integrate multiple sensors would enhance the model's capabilities for more comprehensive health monitoring, covering a broader range of metrics like physical activity and environmental factors. Additionally, investigating adaptive learning techniques for personalized health models on edge devices could further improve accuracy and user experience.

Author Contributions: Conceptualization, P.S. and N.S.; methodology, P.S.; writing—original draft preparation, N.S.; writing—review and editing, K.J. and N.S.; supervision, K.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors wish to acknowledge that this research was conducted without any external funding or support. The research work was carried out at the International Institute of Information Technology (IIIT), Naya Raipur, Chhattisgarh, India. We extend our gratitude to the institute for providing the necessary infrastructure and environment for this work.

Conflicts of Interest: The authors declare that they have no conflicts of interest to disclose in relation to this research. All authors have adhered to ethical guidelines and have no financial or personal relationships that could be perceived as influencing the research.

References

1. Du, G. Vision-Based Fatigue driving recognition method integrating heart rate and facial features. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 3089–3100. [CrossRef]
2. Farida, S.; Tamer, E.; Wadha, L.; Khawla, A.; Qutaibah, M. Machine Learning for Healthcare Wearable Devices: The Big Picture. *J. Healthc. Eng.* **2022**, *2022*, 4653923.
3. Mamta, B.; Pradnya, V.K.; Rashmi, R.; Sarika, B.; Ruhi, P. TinyML applications and use cases for healthcare. In *TinyML for Edge Intelligence in IoT and LPWAN Networks*; Academic Press: Cambridge, MA, USA, 2024.
4. Ahmed, K. tinyCare: A tinyML-based Low-Cost Continuous Blood Pressure Estimation on the Extreme Edge. In Proceedings of the 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), Rochester, MN, USA, 11–14 June 2022; IEEE: lPiscataway, NJ, USA, 2022.
5. Element 14. Available online: https://community.element14.com/challenges-projects/design-challenges/pi-fest/b/blog/posts/songspire---running-tensorflow-lite-in-a-microcontroller-such-as-pico (accessed on 13 February 2024).
6. MJRoBot. Available online: https://mjrobot.org/2021/03/12/tinyml-motion-recognition-using-raspberry-pi-pico/ (accessed on 25 March 2024).
7. Sajjadieh, A.; Shahsavari, A.; Safaei, A.; Penzel, T.; Schoebel, C.; Fietze, I.; Mozafarian, N.; Amra, B.; Kelishadi, R. The Association of Sleep Duration and Quality with Heart Rate Variability and Blood Pressure. *Tanaffos* **2020**, *19*, 135. [PubMed]
8. Kim, H.G.; Cheon, E.J.; Bai, D.S.; Lee, Y.H.; Koo, B.H. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Tanaffos* **2018**, *15*, 235. [CrossRef] [PubMed]
9. Li, G.; Chung, W. Detection of driver drowsiness using Wavelet analysis of heart rate variability and a support vector machine classifier. *Sensors* **2013**, *13*, 16494–16511. [CrossRef] [PubMed]
10. Radha, M.; Fonseca, P.; Moreau, A.; Ross, M.; Cerny, A.; Anderer, P.; Long, X.; Aarts, R.M. Sleep stage classification from heart-rate variability using long short-term memory neural networks. *Sci. Rep.* **2019**, *9*, 14149. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.