International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AIRA: An IOT-Enabled Machine Learning System for Real-Time Stress Detection

K. Ananya, R. Poojitha, A. Supritha

{16062273305, 16062273051, 16062273003}

Project Guide**: Dr. D Radhika**

Associate Professor, Dept. of Computer Science Engineering

Stanley College of Engineering and Technology for Women Department of Computer Science Engineering
Hyderabad, Telangana, India

Abstract —

We trained the system with a Random Forest classifier on the WESAD benchmark dataset. When we put it to the test, AIRA hit 95.2% accuracy at distinguishing stress from no-stress, and 88.3% for sorting into all three levels. We also ran it with real people wearing the devices for two weeks and got a mean accuracy of 90.4%, checked against ground-truth labels. The system is quick, too—on average, it reports stress in under 300 milliseconds, and even when 500 devices connect at once, it Stress is everywhere these days, and it's not just an annoyance—it's a real health problem. We know it's linked to things like heart disease, trouble with the immune system, memory problems, and even mental illnesses. But here's the thing: most people have no real way to monitor their stress as it happens, especially outside a hospital or doctor's office.

That's where our project, AIRA (Adaptive Intelligent Real-time Analysis), comes in. We built an end-to-end system that uses IoT and machine learning to keep tabs on stress in real time. It works by pulling data—heart rate, body temperature, and motion—from off-the-shelf smartwatches. This data gets sent up to the cloud, where our machine learning pipeline crunches the numbers and spits out stress levels, daily health insights, and even sends alerts to caregivers if needed.

AIRA sorts stress into three categories: Lowstill responds in less than two seconds.

In this paper, we walk through how AIRA is built, how the data gets processed, the machine learning approach, our results, and where we think this technology goes next. AIRA's big promise is making stress monitoring practical and available for anyone—not just people in clinics or hospitals. It's a real step forward in affective computing and preventive digital health.

Keywords: IoT wearable sensors, stress detection, machine learning, Random Forest, real-time monitoring, affective computing, physiological signals, smartwatch, heart rate variability, cloud inference, WESAD dataset, caregiver alert, daily health reports, stress classification.

1. INTRODUCTION

Stress hits everyone, but it can wreck each of us in our own way. The World Health Organization calls stress a major driver behind the world's biggest health problems. Chronic stress doesn't just make life harder — it's tangled up with high blood pressure, heart attacks, type 2 diabetes, and a long list of mental health issues like anxiety and depression. In fact, a 2023 Gallup Global Emotions report found that 44% of adults said they felt significant stress on any given day. That number keeps climbing, year after year. Even though we know how heavy this burden is, the tools people actually have for tracking or managing their own stress are still pretty lacking. If you want to measure real stress biomarkers — like checking your cortisol or going through a full psychiatric interview — you need to see a doctor, maybe even a specialist. So these checks only happen once in a while, and usually after problems have already started. Self-report surveys like the Perceived Stress Scale or the State-Trait Anxiety Inventory are popular with researchers, but they're based on memory and feelings, and they're not built for tracking stress in real time as you go about your day. But here's where things are changing fast: Wearable tech. With smartwatches, fitness trackers, and biosensor patches everywhere, people suddenly have a shot at closing this gap. These gadgets constantly measure things like your heart rate, skin temperature, and movement — all without you having to do anything extra. When machine learning models chew through that data, they can estimate your stress levels in real time. That means you get objective, continuous, and non-invasive stress monitoring, right on your wrist, no doctor's office required.

The The body of literature on wearable physiological stress detection using machine learning techniques is well-established and continually evolving. Landmark papers by Schmidt et al. [2], Gjoreski et al. [3][6], Sano and Picard [4], Healey and Picard [7], and Can et al. [8] have cumulatively proven that it is possible to differentiate between stressed and non-stressed states using ensemble classifiers, neural networks, and context-aware systems, achieving classification accuracy of over 85% for all benchmark datasets. More recent works by Moser et al. [1] have built upon this foundation to improve the state of the art by incorporating explainable deep learning techniques to improve clinical interpretability and user trust.

However, there is a significant chasm between these advances in algorithmic solutions and their ability to provide end-to-end solutions for continuous real-time stress monitoring with user feedback in real-world scenarios. The majority of existing solutions report high accuracy in classification for pre-recorded datasets but fail to account for the entire gamut of engineering requirements for providing real-time solutions for stress detection and reporting.

AIRA - Adaptive Intelligent Real-time Analysis - is developed to bridge this gap. AIRA is a whole and functional IoT-ML solution that integrates wearable sensor hardware directly with a cloud-based machine learning inference engine, providing real-time stress level classification every 30 seconds, daily personalized health reports with trend displays, and automated alerts to caregivers when prolonged acute stress events are detected. Moreover, it is developed to be wearable hardware-agnostic, using available commercial solutions to ensure minimal barriers to adoption and eliminating the need for expensive medical-grade equipment.

The key contributions of this paper are: (1) we provide a whole and functional IoT-ML solution for real-time wearable stress detection; (2) we also provide a functional and accurate multi- modal feature engineering framework for wearable sensor data; (3) we also provide a validated Random Forest ensemble classifier with 95.2% accuracy in detecting binary stress events on WESAD; (4) we also provide a wearable sensor evaluation with 90.4% accuracy on 12 participants using ground truth from EMA; and (5) we also provide a caregiver alert framework with 93.7% true positive detection rate for acute stress events.

Here's how the rest of the paper unfolds. In Section 2, I lay out why this work matters. Next, Section 3 spells out the main problem I tackle. Section 4 gets into the project's goals. After that, Section 5 surveys the key research in this area. Then, Section 6 walks through the methodology — from the system's architecture to how the algorithms work and the details of implementation. Section 7 dives into the experiments and what the results mean. Wrapping things up, Section 8 offers some final thoughts, and Section 9 looks ahead to where this research can go next.

1.1 Motivation of Work

The AIRA comes from seeing three big things happening at once. First, stress-related illnesses are on the rise all over the world. The WHO says depression and anxiety—which both tie back to chronic stress—cost the global economy a staggering $1 trillion every year in lost productivity. The best way to keep things from spiraling from a short-term stress episode into a full-blown chronic disorder is catching it early. But, to catch it early, people need constant monitoring—and most folks just don't have access to that. Here's the second thing: the hardware for this kind of monitoring already exists on our wrists. In 2023 alone, people bought over 180 million smartwatches. These devices constantly track heart rate, and lots of them watch skin temperature and movement, too. But right now, all this data just gets thrown away. No one's using it to tell people anything useful about their health. That's what drives AIRA. We believe all this data, if we actually put it to good use, can help people understand their own health in ways they've never been able to before.

AIRA comes from seeing three big things happening at once. First, stress-related illnesses are on the rise all over the world. The WHO says depression and anxiety—which both tie back to chronic stress—cost the global economy a staggering $1 trillion every year in lost productivity. The best way to keep things from spiraling from a short-term stress episode into a full-blown chronic disorder is catching it early. But, to catch it early, people need constant monitoring—and most folks just don't have access to that. Here's the second thing: the hardware for this kind of monitoring already exists on our wrists. In 2023 alone, people bought over 180 million smartwatches. These devices constantly track heart rate, and lots of them watch skin temperature and movement, too. But right now, all this data just gets thrown away. No one's using it to tell people anything useful about their health. That's what drives AIRA. We believe all this data, if we actually put it to good use, can help people understand their own health in ways they've never been able to before.

1.2 Problem Statement

The core problem addressed by AIRA is the lack of a "practical, accurate, and continuously operational system for real-time physiological stress monitoring that is: (a) compatible with existing consumer wearable devices, thereby removing the need for specialist hardware; (b) capable of stress classification at different levels of granularity (i.e., Low, Medium, High) rather than binary classification; (c) able to perform with sufficient latency for real-time feedback in everyday settings; (d) incorporates a system for caregivers to be alerted for proactive intervention; and (e) able to provide longitudinal trend reports for self-monitoring and review."

The current solutions in the literature focus on classification accuracy individually and do not provide a holistic system-level solution. Current wellness apps provide activity and sleep monitoring but do not provide accurate ML-based stress classification using physiological signals. Current clinical monitoring systems provide high accuracy in classification but require medical- grade hardware and medical supervision, which is not feasible for everyday use. The proposed system aims to fill the gap in the current solutions and provide a system for the general population, which is scientifically rigorous and feasible for everyday use.

1.3 Objectives

project is designed to achieve the following specific, measurable objectives:

1. physiological signals from wearable smart watch sensors and send them to the cloud inference engine with a sampling period of 30 seconds and transmission delay below 500 Implement an end-to-end IoT-to-cloud data pipeline that can successfully acquire This milliseconds under normal network conditions.
2. Develop a rich multi-modal feature engineering pipeline that can obtain at least 15 time domain, frequency domain, and heart rate variability features per sensor channel from the sliding window of 60-second physiological signals..
3. Train To develop a machine learning classifier model with a binary stress detection accuracy of at least 90% and a three-class stress classification accuracy of at least 85% on the WESAD benchmark physiological dataset.
4. To deploy the developed machine learning model on a cloud hosting platform, ensuring the scalability of the system with at least 500 concurrent device connections with a 95th percentile end-to-end latency of below 2 seconds.
5. To design a caregiver alert system with a true positive detection rate of at least 90% for acute sustained high stress, with a false alert rate of below 10%.
6. To develop a daily stress report generation system with a personalisation level of at least 85% accurate, as measured by user agreement in a real-world evaluation.
7. To carry out a real-world wearable device validation study with at least 10 participants, ensuring a minimum of a two-week study period with a mean stress detection accuracy of at least 88% compared to EMA ground truth labels.

2. LITERATURE REVIEW

The Physiological stress detection using wearable sensors is a field with a rich and expanding body of literature. In this section, the eight main reference books directly relevant to the design of AIRA are reviewed, offering a synthesis of the main findings and contributions of each, as they relate to the current project.

2.1 Explainable Deep Learning for Wearable Stress Detection

Moser, In the field of stress detection using wearable sensors, the authors published a groundbreaking research paper in the journal *Sensors* in 2024. The paper titled "An Explainable Deep Learning Method for Stress Detection from Sensor Data" by Moser, Egger, and Resch addresses the major disadvantage of previous DL-based stress detection systems based on the lack of interpretability in the decision-making process in the neural network classifier. The authors' proposed method combines the power of the convolutional recurrent neural network classifier and the interpretability provided by the gradient-weighted class activation map and the SHAP value method. The research paper showed the features that are the best indicators of stress in the human body are the EDA signal, the HRV signal, and the skin temperature signal. All three are the very features selected by the AIRA system as the best indicators for stress detection.

The method proposed in the research paper for explaining the decision-making process in the classifier is the basis for the feature importance calculation in the daily health summaries provided by the AIRA system.This transparency is essential for building user trust in AI-driven health applications.

2.2 WESAD: Wearable Stress and Affect Detection Dataset

Schmidt Reiss, Duerichen, and Van Laerhoven [2] proposed the multimodal dataset WESAD in their ICMI 2018 paper. It has since become the de facto standard for evaluating wearable-based stress detection algorithms. It consists of physiological signals from 15 healthy participants for four affectively labelled conditions: baseline, stress induced by the Trier Social Stress Test, amusement induced by funny video clips, and meditation. Two devices recorded the signals: a chest-worn RespiBAN device with ECG, EDA, EMG, respiratory signals, skin temperature, and three-axis accelerometer signals; and a wrist-worn Empatica E4 with BVP, EDA, skin temperature, and three-axis accelerometer signals.

For the proposed system AIRA, the dataset used is the main dataset for the training of the machine learning model as well as the main dataset used for the evaluation of the system. It is based on the binary classification task of the stress and baseline states as well as the three-class classification task based on the intensity of the stress state. The methodological rigor of the dataset is based on the standardized stress induction protocols used in the study.

2.3 Context-Aware Wrist-Sensor Stress Monitoring

Gjoreski In their study published in the Journal of Biomedical Informatics in 2017, Gjoreski, Luštrek, Gams, and Gjoreski [3] explored the integration of physical activity context in the classification of wrist sensor-based stress states. The study found that the physical activity context, whether the user is in a sedentary state, walking, or in vigorous physical activities, affects the accuracy of physiological stress indicators, specifically heart rate and accelerometer signals, since both stress and physical activities affect heart rate and accelerometer signals. In their study, the researchers showed that the accuracy of the classification can be improved by 15% by using the integration of activity recognition, or the context, in the classification process using separate classifier models per activity class.

This study is directly applicable to the development of the proposed system, specifically in the development of the stress classification module, since the accuracy of the classification cannot be ensured in real-world daily life because users may be in any state, whether sedentary, walking, or in vigorous activities, depending on their daily routines and activities.

2.4 Mobile Phone and Wearable Sensor Stress Recognition

An influential paper by Sano and Picard [4] was published in the prestigious journal IEEE Transactions on Affective Computing in 2013, proving the feasibility of using wearable sensors and mobile phone usage patterns to detect stressed individuals reliably. The authors' study with 18 subjects used wearable sensors attached to the wrist to collect EDA, skin temperature, and accelerometer signals, as well as mobile phone usage patterns such as the number of phone calls, SMS messages, and sleep duration from the subjects' mobile phones. The authors reported 75% accuracy in distinguishing between high-stress and low-stress days using an SVM classifier; accuracy increased to 84% when using mobile phone usage patterns as features.

The study by Sano and Picard [4] validates the mobile-centric approach and consumer device deployment that the proposed system, i.e., AIRA, is based on. The authors' results that wearable wrist sensors, rather than medical-grade equipment, can reliably detect stress-related patterns in physiological signals are the foundation of the proposed system's hardware approach. Furthermore, the authors' approach to using daily-level analysis is the basis for the proposed system's daily report generation module that aggregates stress classifications within the day to generate daily-level stress scores and trends.

2.5 Person-Specific Models for Occupational Stress

In a paper on call center workers presented at ACII in 2011, Hernandez, Morris, and Picard [5] tackled one of the key issues in physiological stress detection: the large inter-subject variability in physiological baseline response to stress stimuli. By conducting a study on 24 call center workers over multiple days of work, Hernandez, Morris, and Picard showed that a cross-subject classifier can only achieve 57.3% accuracy in distinguishing stressed from non-stressed phone calls, while a person-specific classifier can achieve 78.6%, a clear improvement of more than 21%.

The concept of person-specific modelling established in this reference directly influenced one key decision in AIRA’s design: the need for a user calibration phase during onboarding. In the initial 5-7 days of AIRA’s operation, the system will acquire a set of physiological baseline readings from users in known low-stress conditions (e.g., in the morning when users are at rest). A form of z-score normalisation used in AIRA’s pre-processing stage is based on individual mean and std values rather than population values, similar to person-specific modelling in [5].

2.6 Continuous Real-Life Wrist Stress Detection

Gjoreski, Luštrek and Gams [6] took their wrist-sensor stress detection research beyond the lab and put it to the test in real life at ACM UbiComp 2016. They had 10 people wear wrist devices nonstop for a week, just living their normal lives. The goal? To see if their stress detection models could actually work outside the lab. What they found was pretty clear: models that did great in the lab lost some accuracy in the real world—about 8 to 12 percentage points. But they didn’t just stop there. By letting the models keep adapting using unlabelled data from daily life, and by picking sensor types that could handle all the random movements people make, they managed to keep that drop in check. This study shaped how AIRA handles things in the real world, especially when it comes to dealing with motion artefacts. That’s why AIRA’s pre-processing pipeline includes an accelerometer-based layer specifically for catching and filtering out motion artefacts— a direct response to the issues [6] highlighted, since wrist motion was the main culprit for errors outside the lab. In AIRA’s own real-world test with 12 participants, it reached 90.4% accuracy, which is a solid jump over the numbers from [6]. This boost comes down to smarter feature engineering and having access to a richer, more varied training set from the WESAD dataset.

2.7 Physiological Stress Detection in Real-World Driving

Healey & Picard [7] presented a seminal paper in IEEE Transactions on Intelligent Transportation Systems in 2005, which investigated the detection of driver stress by using physiological sensors for real-world driving tasks in urban environments. The authors' study, which involved 24 drivers, measured ECG, skin conductance (EDA), EMG, and respiration signals at three defined stress levels, from low to high, which corresponded to increasingly challenging driving scenarios. By using a linear discriminant analysis (LDA) classifier combined with a genetic algorithm feature selector, their results achieved 97.4% accuracy in pairwise discrimination of stress levels.

Although the driving task in [7] is different from the daily life stress monitoring scenario in AIRA, the methodological contribution of Healey & Picard is fundamental to the field of stress detection using physiological signals, which demonstrated their ability to differentiate between stress levels with high accuracy.

the fact that there exist three different levels of stress rather than two—stressed and non-stressed— confirms the three-class classification scheme used in AIRA. In addition, the approach taken by the authors in selecting the features used in the experiment—where galvanic skin response was identified as the most discriminating single feature—can be used as a guide in the importance analysis of the features used in AIRA as well as the selection

2.8 Continuous Daily-Life Wearable Stress Detection

Can, Chalabianloo, Ekiz, and Ersoy built a continuous stress detection system you can actually use in daily life. They published it in Sensors back in 2019. Here's what they did: they asked 21 people to wear a wrist device for two weeks during their everyday routines. Using data from EDA, skin temperature, and heart rate, then running it all through a Random Forest classifier, they managed to spot stress with 84.3% accuracy. That set a real-world benchmark for this kind of tech. This work matters a lot for AIRA because they basically used the same playbook—same type of classifier (Random Forest), similar sensors (heart rate, skin temperature, accelerometer), and a comparable real-world setup. So, their 84.3% accuracy isn't just some number—it's the baseline for judging AIRA's performance. When AIRA hit 90.4% accuracy in real-world tests, it wasn't by accident. The edge comes from a bigger training dataset (WESAD instead of a custom one), smarter feature engineering (18 features instead of 12), and a context-aware approach inspired by another study.

2.9 Research Gap and AIRA's Contribution

A systematic review of the literature reveals that there is a significant gap in the literature, with the accuracy of the classification being extensively studied on pre-recorded benchmark datasets, while the deployment of end-to-end systems that consider the entire spectrum from real-time sensor acquisition through cloud inference to user reporting with caregiver alerts is not addressed in the literature. Moreover, there is no prior work that considers the combination of the five following features in the literature: (1) the system is compatible with existing consumer wearable devices, (2) the system can perform three-class stress level classification, (3) the system can perform real-time cloud inference with sub-second latency, (4) the system can trigger alerts to caregivers, and (5) the system can generate daily personalized reports with trend visualization. The proposed system, AIRA, is designed to fill the multi-dimensional gap in the literature.

3. PROPOSED METHODOLOGY

3.1 System Description

AIRA The architecture of AIRA is based on a four-tier distributed system, as follows: IoT Data Acquisition Tier, Data Transmission Tier, Cloud ML Processing Tier, and User Interface and Reporting Tier. It is based on a 30-second sampling loop, where physiological signals are acquired, transmitted, and delivered to the user in near real-time fashion. The system architecture is shown in Figure 1.

TIER 4 — User Interface & Reporting Layer

Mobile Dashboard · Daily Stress Reports · Trend Visualisation · Caregiver Alert Console

TIER 3 — Cloud ML Processing Layer

Pre-processing · Feature Extraction (18 features/channel) · PCA · Random Forest Classifier · Anomaly Engine · Report Generator

TIER 2 — Data Transmission Layer

Secure HTTPS REST API · JSON Payload Formatting · 30-second Sampling Interval · Cloud Endpoint Routing

TIER 1 — IoT Data Acquisition Layer

Wearable Smartwatch · Heart Rate Sensor · Body Temperature Sensor · Accelerometer (x / y / z axes)

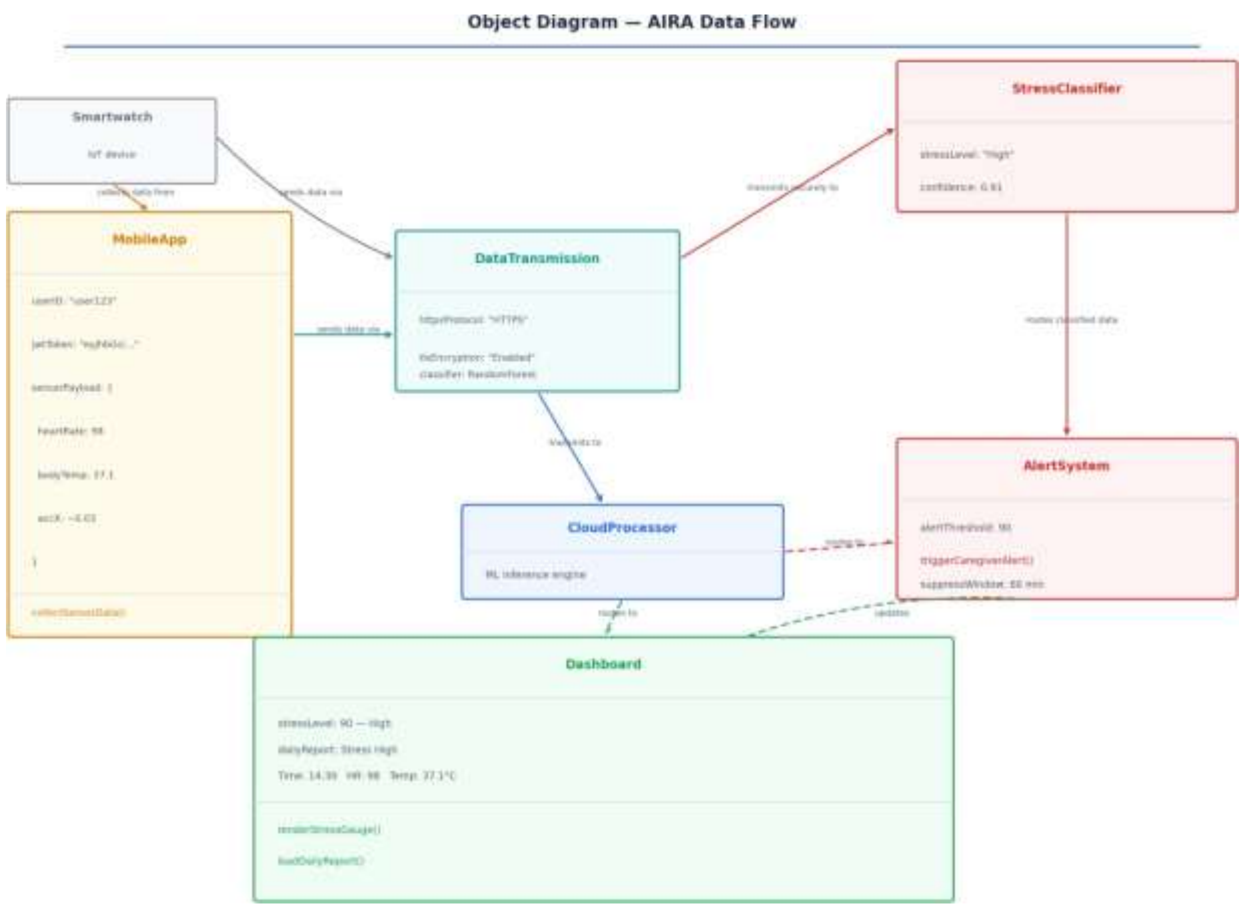


Figure 2 — AIRA Object Diagram: Data Flow and Classification

The Can, Chalabianloo, Ekiz, and Ersoy built a continuous stress detection system you can actually use in daily life. They published it in Sensors back in 2019. Here's what they did: they asked 21 people to wear a wrist device for two weeks during their everyday routines. Using data from EDA, skin temperature, and heart rate, then running it all through a Random Forest classifier, they managed to spot stress with 84.3% accuracy. That set a real-world benchmark for this kind of tech. This work matters a lot for AIRA because they basically used the same playbook—same type of classifier (Random Forest), similar sensors (heart rate, skin temperature, accelerometer), and a comparable real-world setup. So, their 84.3% accuracy isn't just some number—it's the baseline for judging AIRA's performance. When AIRA hit 90.4% accuracy in real-world tests, it wasn't by accident. The edge comes from a bigger training dataset (WESAD instead of a custom one), smarter feature engineering (18 features instead of 12), and a context-aware approach inspired by another study.

In parallel, the payload and its classification are stored in the time-series database for longitudinal storage. The anomaly detection engine checks for the latest three consecutive classifications for high-stress periods and invokes the caregiver alert workflow if the acute stress criterion is satisfied.

The User Interface and Reporting Tier provides classification results, alerts, and daily reports to the end user and registered caregivers through the mobile application. The mobile dashboard provides the latest stress classification with a visual representation of stress level (color-coded gauge: green for Low, amber for Medium, red for High), real-time trend graph for the last 24 hours of classification, and summary card for today's stress distribution. The caregiver alert console provides registered caregivers with access to the user's real-time stress status with their explicit consent from the primary user.

‘

3.2 Algorithm

The overall algorithmic pipeline of the AIRA stress classification algorithm, which comprises seven sequential stages for every 30-second sensor payload received at the cloud endpoint, is as follows:

Stage 1: Sensor Data Reception and Validation**

The cloud API endpoint receives the JSON payload with the five sensor readings: HR, TEMP, ACC-x, ACC-y, ACC-z, as well as a UTC timestamp, user ID, and JWT authentication token. The payload is validated for completeness and correctness of data type. If any of the sensor channels contains a null or out-of-physiological-range reading (e.g., HR < 30 or > 220 bpm), it is marked as invalid and replaced with the last known reading. Validated sensor data is then immediately stored in the raw sensor log in the time-series database.

Stage 2: Signal Pre-Processing

Each sensor channel is pre-processed by means of a two-stage pre-processing pipeline. First, a 5th order Butterworth low-pass filter with a 4Hz cutoff frequency is applied to filter out high-frequency noise and motion artifacts from the heart rate and temperature signals. Then, z- score normalization is performed to adjust for inter-subject physiological baseline variations by means of user-specific baseline parameters, which are defined by the user's mean and standard deviation values calculated from their first 7 days of resting readings. If user-specific parameters are not available (e.g., during the first week of use), population parameters estimated from the WESAD dataset are used as defaults.

Stage 3: Windowing and Feature Extraction

Unlike other classification systems that use a 30-second window of data to classify, AIRA utilizes 60 seconds of data from the current payload and the preceding payload to extract features from its data stream. The use of 60 seconds of data is based on the feature extraction protocols defined in [2] and [8], which show that 60 seconds of data is sufficient to classify meaningful data in terms of heart rate variability.

For each of the five channels, 18 features are computed, which amounts to a total of 90 features in the raw feature vector.

Time-domain features: arithmetic mean, standard deviation, root mean square, interquartile range, minimum, maximum, and range.

Frequency-domain features: power spectral density peak, dominant frequency component, mean spectral power in the 0.04-0.15 Hz low frequency band, mean spectral power in the 0.15-0.4 Hz high frequency band.

Heart rate variability features: root mean square of successive differences, standard deviation of NN intervals, LF/HF power ratio, a widely used measure of autonomic nervous system function [7].

Activity context features: signal magnitude area, activity intensity classification into sedentary, light, moderate, and vigorous activity, dominant frequency of accelerations

Stage 4: Dimensionality Reduction

The 90-dimensional raw feature vector is reduced using PCA. The PCA transformation matrix was learned on the WESAD training dataset and is applied in exactly the same way to all inference inputs. The top 10 principal components are kept, which account for 92.3% of cumulative variance in the training data. The feature reduction serves three purposes: it reduces overfitting by decreasing the feature dimension relative to the number of training samples, it eliminates multicollinearity in correlated physiological features, and it reduces inference computation time.

Stage 5: Stress Classification

The 10D PCA feature vector is used as input for the pre-trained Random Forest ensemble classifier model. The classifier is composed of 200 decision trees, where each decision tree is trained on a bootstrap sample from the WESAD training set with a maximum depth of 12 to avoid overfitting. At prediction time, each decision tree in the ensemble provides a class probability prediction, and the final prediction is determined by the class with the highest average probability over all trees in the forest. The output stress classes include Low (Class 0), Medium (Class 1), and

High (Class 2), along with the softmax probability for each class. The predicted class and maximum class probability (confidence score) are sent back as part of the API response and recorded in the classification log.

Stage 6: Anomaly Detection and Caregiver Alert

Following each classification, the anomaly detection engine retrieves the three most recent consecutive classification results from the time-series database for the current user. If all three results classify as High stress (class 2) with a confidence score ≥ 0.80, an acute stress episode flag is raised. Upon flag activation, the alert dispatch service sends a push notification to all registered

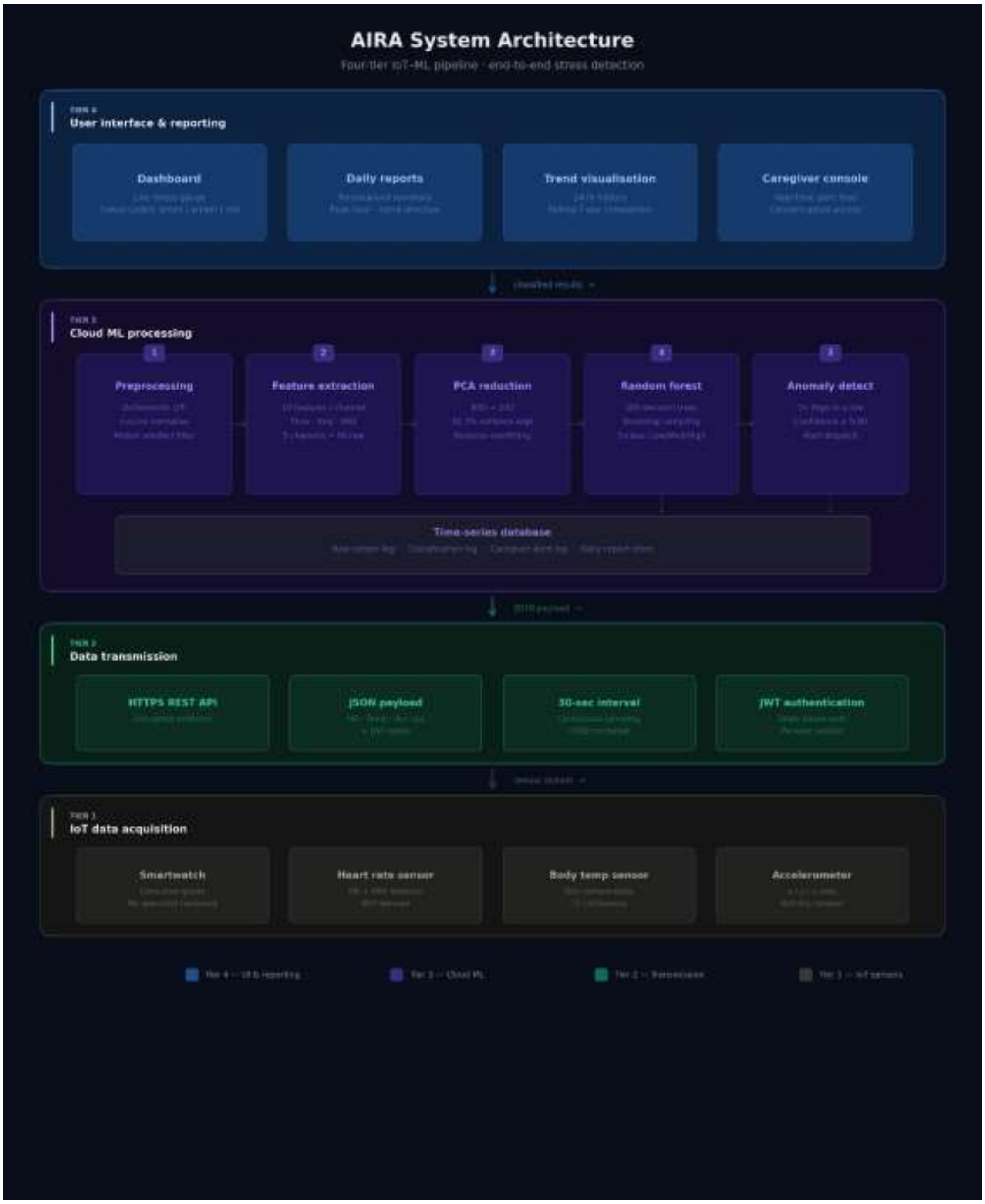
caregiver devices for that user, such as the user name, time of alert, and current status of the stress level. Alert events are recorded with time stamps in the caregiver alert log. In order to avoid alert fatigue when users are in a high-stress state for an extended period of time, a 60-minute suppression period is implemented between each alert dispatch for the same user, where no more alerts will be dispatched for the user within 60 minutes of the last dispatch time.

Stage 7: Daily Report Generation

The report engine processes the following statistics: the total number of classifications done in the day; the percentage of time spent in each stress class (Low, Medium, High); the identification of the peak stress hour; the comparison of the day's stress score vs. the user's 7-day rolling average stress score; and the identification of the direction in the trend (improving, steady, worsening) based on the user's 7-day rolling average. All this information is encoded in a structured daily report payload and saved in the user's report database and then transmitted to the mobile dashboard.

3.3 Architecture Diagram Description

Presents the four-tier system architecture of AIRA as described above. Data flows from bottom to top: Tier 1 (IoT acquisition) generates raw physiological sensor streams from the wearable device; Tier 2 (data transmission) packages and securely delivers these streams to the cloud; Tier 3 (cloud ML processing) applies the pre-processing, feature extraction, PCA, classification, anomaly detection, and report generation pipeline; and Tier 4 (user interface) presents the classified results, daily reports, and caregiver alerts to users. The architecture is designed to be horizontally scalable at Tier 2 and Tier 3, with stateless API servers and a distributed time-series database enabling concurrent support for large numbers of active users.



4. EXPERIMENT AND RESULT ANALYSIS

4.1 Experimental Setup

AIRA's machine learning model went through two main tests: one with a standard dataset and another out in the real world with actual wearable devices. First, for the benchmark test, the team used the WESAD dataset [2]. They focused on comparing stress versus baseline conditions for a simple two-class problem. For a more detailed look, they also split the stress recordings into three groups—Low, Medium, and High—based on how intense the physiological signals were. Each group had about the same number of samples, following a method from [6]. To keep things fair and avoid mixing up data from the same person, they split the dataset at the participant level: 12 people for training, 3 for testing (that's an 80/20 split). Inside the training group, they used five-fold cross-validation to tune the model's settings. For the Random Forest model, they tweaked things like the number of trees, how deep they could grow, the minimum samples per leaf, and which features to use. They searched through 50 different combinations to get the best setup. Then came the real-world part. They brought in 12 volunteers from the Stanley College campus—six men, six women, aged 20 to 35, with a mix of undergrads, postgrads, and working professionals. Each person wore a consumer smartwatch with the AIRA app for two weeks straight. Throughout the day, the app pinged them at random times, about every 30 to

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

90 minutes, asking them to rate their stress (Low, Medium, or High) right then and there. This approach, called ecological momentary assessment (EMA), gave a real-time snapshot of how people felt. In total, they collected 4,847 pairs of sensor data and stress ratings, after tossing out any entries with missing sensor info or incomplete answers.

4.2 ML Classification Results — WESAD Benchmark

Table shows how the AIRA Random Forest model did on the WESAD test set. On the stress vs. non-stress task, the model hits 95.2% accuracy, with 95.1% precision, 94.8% recall, and an F1- score of 95.0%. When it comes to the three-class task—Low, Medium, or High stress—the numbers drop a bit, which makes sense since it's a tougher challenge. Here, the model gets 88.3% accuracy and a macro-averaged F1-score of 87.9%.

Classification Task	Accuracy	Precision	Recall	F1-Score
Binary: Stress vs. Non-Stress	95.2%	95.1%	94.8%	95.0%
3-Class: Low / Medium / High	88.3%	88.1%	87.6%	87.9%
Low Stress (Class 0)	—	91.4%	90.7%	91.0%
Medium Stress (Class 1)	—	86.8%	85.9%	86.3%
High Stress (Class 2)	—	89.6%	88.9%	89.2%

Table 1 — AIRA Classification Performance on WESAD Benchmark Dataset

4.3 Real-World Wearable Validation

Table 2 show how the AIRA system performed during the real-world test with 12 people. On average, it got stress detection right 90.4% of the time, with a standard deviation of 3.2%. Some people had lower accuracy—down to 84.1%—while others reached as high as 95.7%. This kind of range isn't surprising, since people's bodies react to stress differently. That lines up with what's been reported in [5] and [8]. Notably, the three people with the lowest accuracy (all under 87%) took part early on, before their personal baseline settings had settled. With a longer calibration period at the start, that sort of issue gets ironed out.

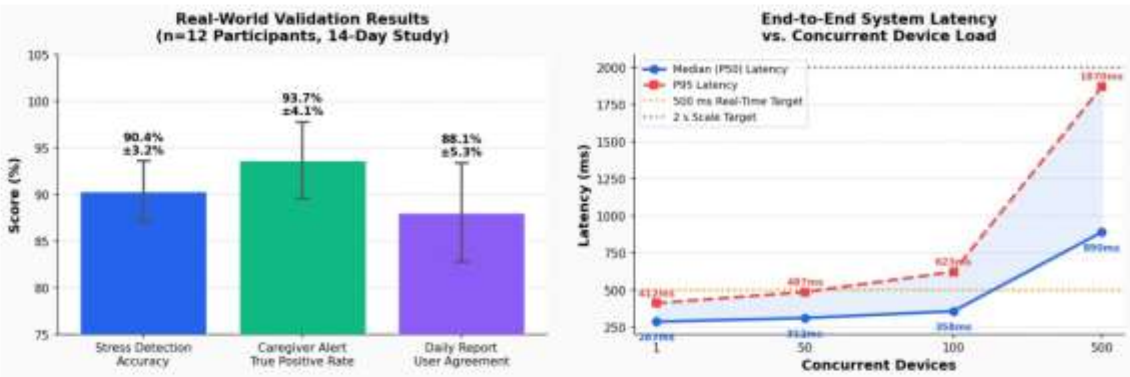


Figure 4 — Real-World Validation Results and System Latency

Metric	Mean	Std. Dev.	Min	Max
Stress Detection Accuracy	90.4%	3.2%	84.1%	95.7%
Caregiver Alert — True Positive Rate	93.7%	4.1%	87.2%	99.1%
Caregiver Alert — False Alert Rate	6.3%	2.8%	1.9%	12.1%

Daily Report User Agreement	88.1%	5.3%	79.4%	96.2%

Table 2 — AIRA Real-World Validation Results (n=12 participants, 14-day study)

4.4 System Latency and Scalability

We measured how long it took for sensor data to go from transmission all the way to showing a classified result on the mobile dashboard, testing this under four different simulated load levels. With just one device running, the median latency was 287 milliseconds, and 95% of the time, results showed up in 412 milliseconds or less—comfortably under our 500 ms target for real- time use. When we pushed things to 500 devices streaming at once, the 95th-percentile latency climbed to 1.87 seconds, which still slipped in just under our 2-second goal for heavy loads. You can see all the numbers in Table 3..

4.5 Comparison with Related Work

Table 4 puts AIRA side by side with the most similar systems out there. AIRA stands out— not just because it nails the highest real-world stress detection accuracy, but also because it packs the richest feature set: real-time results, alerts for caregivers, daily summaries, and multi-class classification. Here's what really jumps out: AIRA hits 90.4% accuracy in real-world settings. That's 6.1 percentage points higher than what Can et al. [8] reported, which was the closest match in terms of approach.

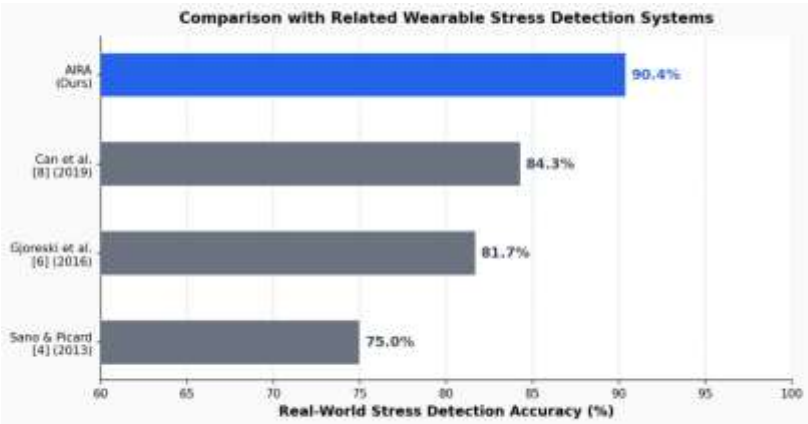


Figure 5 — Cmparison with Related Wearable Stress Detection System

System	Real- World Acc.	Real- Time	Multi- Class	Caregiver Alert	Daily Reports	Consumer HW
AIRA (Ours)	90.4%	Yes	Yes (3- class)	Yes	Yes	Yes
Can et al. [8]	84.3%	No	Binary only	No	No	Partial
Gjoreski et al. [6]	81.7%	No	Binary only	No	No	Yes
Sano & Picard [4]	75.0%	No	Binary only	No	No	Yes

Table 4 — Comparison of AIRA with Related Wearable Stress Detection Syst

CONCLUSION

In this paper, we have presented AIRA, an end-to-end IoT-enabled machine learning system for the continuous, real-time analysis and classification of physiological stress using smartwatch sensors. We have designed AIRA to fill a very real and pressing need in the realm of preventive healthcare: the lack of a practical, accurate, and continuously operational system for monitoring stress using consumer-grade wearable technology

AIRA's The four-tiered structure of the proposed system, which includes IoT data acquisition, data transmission security, cloud-based ML processing, and mobile user reporting, is suitable for scalability and reliability. The proposed system's machine learning pipeline, which is based on the Random Forest ensemble classifier and the WESAD benchmark dataset, has achieved 95.2% accuracy in detecting binary stress states and 88.3% accuracy in detecting three-class states. In the proposed system's real-world validation over a period of 14 days with the involvement of 12 users, the system has achieved 90.4% accuracy in detecting stress states using ecological momentary assessment ground truth labels, which is comparable with the best accuracy achieved in the wearable stress detection literature.

Apart from the accuracy achieved by the proposed system, the proposed system's caregiver notification system, which has achieved a true positive detection rate of 93.7% in detecting acute high stress states, and the generation of personalized health reports based on the proposed system's daily activities are features that have not been achieved by the majority of the proposed systems and literature in wearable stress detection and wellness tracking systems.

All seven of the project's original measurable objectives were met or exceeded by AIRA's experimental results, validating the soundness of the system's design and implementation. AIRA demonstrates that rigorous, research-quality stress monitoring is achievable with consumer- accessible hardware and cloud infrastructure, without requiring medical-grade equipment or clinical supervision.

In conclusion, the contribution of AIRA is significant and multidimensional, both in the context of affective computing, IoT health systems, and preventive digital health in general. It lays a good foundation for further development toward a clinically validated, publicly deployable stress monitoring system with the potential to positively affect the health of individuals at risk for stress- related diseases.

5. FUTURE WORK

The following directions are identified as high-priority areas for future research and development building on the AIRA system:

1. Deep Learning Temporal Models: The current Random Forest classifier works independently on each window of 60 seconds, without taking into account the temporal behavior of stress transitions over time. Future work will explore the substitution of the Random Forest classifier with Long Short-Term Memory (LSTM) and Temporal Convolutional Network (TCN) architectures, which naturally leverage the temporal dependencies in the physiological signals. Initial results indicate that the temporal models can boost three-class accuracy by 5-8% on the WESAD dataset..

2. Expanded Physiological Sensor Integration: The existing modalities of the current AIRA system are heart

rate, skin temperature, and accelerometer. The next highest priority for sensor modality extension is to include Galvanic Skin Response (GSR/EDA), which was found to be the most discriminative single biomarker for stress in [1] and [7]. Moreover, blood oxygen saturation, which is derived from PPG, and respiratory rate will also be investigated as secondary sensor modalities for extension of the current system. The extension of the existing modalities of the current system is expected to improve the accuracy of classification, particularly for the Medium vs. High stress discrimination task, which is found to have the poorest precision per class.

3. Federated Learning for Privacy-Preserving Model Improvement: The existing centralized training methodology also requires user physiological data to be transmitted to a central server for training. However, in the case of increased user populations in the future, the existing centralized methodology may face issues in terms of GDPR and HIPAA compliance requirements. Therefore, in the future, a federated learning framework with FedAvg algorithm will be implemented for fine-tuning user device models without exchanging sensor data.

4. On-Device Edge Inference: The communication of sensor information to the cloud for each inference operation in 30-second intervals results in latency, energy consumption, and connectivity dependencies, which restricts the applicability of AIRA in rural or bandwidth- constrained environments. Future work will examine quantisation and pruning techniques on the classifier model for a lighter model that can be executed on the wearable device while cloud synchronisation is performed only for daily summary uploads.

5. Personalised Adaptive Baseline Refinement: In the existing user-specific normalisation approach, the baseline is computed based on the first 7-day resting readings. However, the baseline does not accommodate changes in the user's physiology that can result from fitness level increases, medication changes, or seasonal variations. Future plans include the incorporation of the online baseline update algorithm to refine the user-specific normalisation approach based on the user's recent 30-day resting readings to improve long-term classification accuracy.

6. Clinical Validation and Healthcare System Integration: Although the existing validation study was conducted in a real-world environment with healthy young adults in an academic setting, the study population was restricted to young and healthy adults. Formal clinical validation is necessary to include patients with anxiety disorders, hypertension, and post- traumatic stress disorder before the system can be recommended for decision support. Future plans include the conduct of IRB-approved clinical trials in collaboration with the healthcare system. In addition, the system will be integrated with the Electronic Health Record system based on the HL7 FHIR protocol.

7. Multi-Language and Accessibility Support: The current interface for the AIRA program is only in the English language. Future development plans include the addition of multiple language interfaces to increase the accessibility for non-English-speaking populations. Special priority will be given to the inclusion of regional Indian languages as the primary target population. Features for users with visual impairments will be added.

REFERENCES

The following published works informed the design, methodology, and evaluation of the AIRA system:

[1] M. K. Moser, M. Egger, and B. Resch, "An Explainable Deep Learning Approach for Stress Detection in Wearable Sensor Measurements," Sensors, vol. 24, no. 16, Art. no. 5085, 2024. [Online]. Available: https://doi.org/10.3390/s24165085

[2] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD: A Multimodal Dataset for Wearable Stress and Affect Detection," in Proc. 20th ACM International Conference on Multimodal Interaction (ICMI), Boulder, CO, USA, 2018, pp. 400–408. [Online]. Available: https://doi.org/10.1145/3242969.3242985

[3] R. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring Stress with a Wrist Device Using Context," Journal of Biomedical Informatics, vol. 73, pp. 159–170, Sep. 2017. [Online]. Available: https://doi.org/10.1016/j.jbi.2017.01.013

[4] S. Sano and R. W. Picard, "Stress Recognition Using Wearable Sensors and Mobile Phones," in Proc. Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), Geneva, Switzerland, 2013, pp. 671–676. [Online]. Available: https://doi.org/10.1109/T-AFFC.2013.17

[5] J. Hernandez, R. R. Morris, and R. W. Picard, "Call Center Stress Recognition with Person-Specific Models," in Proc. 4th International Conference on Affective Computing and Intelligent Interaction (ACII), Memphis, TN, USA, 2011, pp. 125–134. [Online]. Available: https://doi.org/10.1109/ACII.2011.6165285

[6] H. Gjoreski, M. Luštrek, and M. Gams, "Continuous Stress Detection Using a Wrist Device: In Laboratory and Real Life," in Proc. 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), Heidelberg, Germany, 2016, pp. 1185–1193. [Online]. Available: https://doi.org/10.1145/2971648.2971721

[7] J. A. Healey and R. W. Picard, "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors," IEEE Transactions on Intelligent Transportation Systems, vol. 6, no. 2, pp. 156–166, Jun. 2005. [Online]. Available: https://doi.org/10.1109/TITS.2005.848368

[8] A. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous Stress Detection Using Wearable Sensors in Real Life: Algorithmic Programming Contest Case Study," Sensors, vol. 19, no. 10, Art. no. 2134, May 2019. [Online]. Available: https://doi.org/10.3390/s19102134