International Journal of Innovative Research and Scientific Studies, 8(6) 2025, pages: 191-208

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Assessing early cardiovascular risk: Heart rate variability as a predictor of air pollution's impact in young adults

Zhanel Baigarayeva[1*], Assiya Boltaboyeva[1,3], Sarsenbek Zhussupbekov[2], Mergul Kozhamberdiyeva[1], Gulshat Amirkhanova[1]

[1]*Faculty of Information Technologies, Al-Farabi Kazakh National University, Almaty 050040, Kazakhstan.*
[2]*Department of Automation and Control, Energo University, Almaty 050013, Kazakhstan.*
[3]*LLP « Kazakhstan R&D Solutions», Almaty 050056, Kazakhstan.*

Corresponding author: Zhanel Baigarayeva (*Email: zhanel.baigarayeva@gmai.com*)

Abstract

Stress is associated with significant behavioral and physiological changes, including decreased heart rate variability (HRV) at rest. Environmental factors such as air pollution are increasingly recognized as potential triggers of physiological stress responses, especially in highly polluted cities such as Almaty, Kazakhstan. However, the relationship between air quality and HRV as a physiological stress marker has not been sufficiently studied. This study explores the development of an IoT system for assessing physiological stress levels based on HRV under various environmental conditions, with a particular focus on air pollution. The study was conducted in three contrasting locations in Almaty, Kazakhstan: a green vegetation area (Botanical Garden), a busy urban area (Al-Farabi Avenue), and an enclosed space with regulated conditions (laboratory). HRV data were synchronously recorded from 10 healthy volunteers using both an optical photoplethysmography (PPG) sensor and an electrocardiographic (ECG) sensor, while air quality parameters ($PM_{2.5}$, PM_{10}, CO_2) were measured simultaneously. The results showed that sympathetic nervous system activation was most pronounced in the Botanical Garden, where elevated levels of particulate matter ($PM_{2.5}$ and PM_{10}) were detected. Fine $PM_{2.5}$ particles had the most significant impact on HRV, followed by PM_{10} and CO_2, leading to a reduction in overall HRV and an increase in the low-frequency to high-frequency (LF/HF) ratio, indicating heightened physiological stress. Machine learning models, including DNN, XGBoost, Random Forest, and TabNet, were developed and trained to assess stress levels based on air quality data. Among them, the XGBoost model achieved the highest classification accuracy of 91.92%. This research provides valuable insights for evaluating disease risks and analyzing the potential impact of long-term exposure to polluted air on the cardiovascular system.

Keywords: Air pollution, Autonomic nervous system, Electrocardiography (ECG), Heart rate variability (HRV), Machine learning. Photoplethysmography (PPG), Stress assessment.

DOI: 10.53894/ijirss.v8i6.9526
Funding: This work is supported by the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant number: AP23488586).
History: Received: 7 July 2025 / **Revised:** 6 August 2025 / **Accepted:** 13 August 2025 / **Published:** 27 August 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Air pollution is a widespread global issue with profound health implications, including the development of stress and cardiovascular diseases [1]. Stress is a natural physiological response of the body to negative environmental influences, which, when prolonged, transforms into chronic stress [2] associated with an increased risk of various pathologies. One of the key factors contributing to such stress is air pollution, particularly fine particulate matter ($PM_{2.5}$), which is generated by industrial activities and other anthropogenic sources [3]. Epidemiological and toxicological studies define $PM_{2.5}$ as particles with an aerodynamic diameter of less than 2.5 μm, posing a significant health hazard [4]. Once inhaled, these particles penetrate the lung alveoli and enter the bloodstream, where they stimulate the production of reactive oxygen species (ROS), induce oxidative stress, and trigger the release of inflammatory mediators, leading to nearly 4 million deaths worldwide due to cardiopulmonary diseases [4].

According to WHO estimates [5], approximately 99% of the world's population is exposed to air pollution levels exceeding the safe threshold for $PM_{2.5}$, with around 4.2 million deaths linked to this factor. Prolonged exposure to polluted air also negatively affects the brain, as it weakens the integrity of the blood-brain barrier and causes neuronal damage [6, 7]. This leads to increased levels of anxiety, depression, and cognitive dysfunction among residents of polluted areas [3]. Pregnant women are particularly vulnerable, as exposure to $PM_{2.5}$ significantly elevates stress levels in this group [6].

Air pollution exacerbates inflammatory responses, disrupts endothelial function, and intensifies oxidative stress [8], which forms a pathophysiological basis for the development of cardiovascular diseases [9]. At the same time, chronic stress factors activate the sympathetic nervous system and contribute to endothelial dysfunction [10], which causes an imbalance between antioxidant mechanisms and reactive oxygen species, a decrease in nitric oxide levels, impaired vascular tone, and an increased risk of thrombosis [11]. The resulting inflammatory responses further damage the endothelium and accelerate the progression of atherosclerosis, while associated metabolic disorders, including obesity and insulin resistance, aggravate pathological processes in the cardiovascular system [12]. A key indicator of the degree of stress exposure on the body is heart rate variability, recorded using electrocardiography and photoplethysmography, which reflects the balance between the sympathetic and parasympathetic nervous systems [13] and provides an objective assessment of stress levels.

1.1. Heart Rate Variability as an Indicator of Autonomic Regulation and Stress

Heart rate variability (HRV) analysis is widely used for the objective assessment of stress levels. HRV reflects the heart's adaptive capacity to changing conditions and is a reliable indicator of stress and overall health status [14]. HRV is primarily measured using electrocardiography (ECG), which is considered the gold standard due to its high measurement accuracy, and photoplethysmography (PPG), which is employed in portable devices and allows continuous monitoring [15]. These technological solutions enable real-time tracking of HRV dynamics and help assess the body's response to various stressors, including air pollutants.

HRV is directly linked to the functional state of the autonomic nervous system (ANS), which regulates the activity of internal organs and consists of two main components: the sympathetic (SNS) and parasympathetic (PNS) divisions. The sympathetic nervous system is activated in response to stress, physical exertion, and danger, triggering tachycardia, increased blood pressure, and the mobilization of energy resources known as the "fight-or-flight" response [16]. In contrast, the parasympathetic system promotes recovery during rest, reducing heart rate, lowering blood pressure, and stimulating digestive processes, referred to as the "rest-and-digest" response [17]. Under normal conditions, a balance between these systems ensures an adequate physiological response to various stimuli. However, chronic stress induced by air pollution leads to SNS hyperactivation and simultaneous PNS suppression, resulting in an increased heart rate, elevated blood pressure, and impaired recovery mechanisms. This imbalance can reduce the body's adaptive potential and increase the risk of cardiovascular diseases [16]. High HRV values indicate an optimal balance between SNS and PNS activity, reflecting effective adaptation, whereas low HRV suggests predominant sympathetic activity and a heightened level of chronic stress.

Individuals exposed to $PM_{2.5}$ particles exhibit suppression of parasympathetic activity, which manifests as decreased HRV, increased anxiety, and irritability [18]. This effect is observed in natural conditions and experimental studies utilizing virtual reality (VR) technologies. In an experimental setting, it has been established that stress-inducing stimuli significantly reduce HRV parameters, indicating SNS activation [19].

A study by Tonacci et al. [20] observed that different types of stress have varying effects on the nervous system. For instance, mental strain, such as solving complex mathematical problems, substantially impacts the autonomic nervous system more than the effects of odors and other sensory stimuli. These observations suggest that the mechanisms of stress influence may vary depending on its origin, but ultimately, they lead to a decrease in HRV and dysfunction of ANS regulatory mechanisms.

Advances in machine learning technologies have improved the accuracy of stress state assessment based on HRV analysis. The application of depthwise separable convolutional neural networks (DSCNN), support vector machines (SVM), and random forest (RF) algorithms enables the automatic classification of stress states and assessment of stress levels based on HRV parameters [21]. The integration of these algorithmic solutions into monitoring systems paves the way for the development of intelligent platforms that assess the impact of air pollution on stress and overall health indicators [22]. Combining data on air quality characteristics with physiological parameters allows for the development of new approaches to early stress detection and individualized adaptation strategies. In the long term, the implementation of environmental monitoring technologies alongside HRV analysis could significantly mitigate the negative effects of chronic stress caused by adverse environmental conditions and contribute to improving the population's quality of life [23].

A study Lim et al. [24] demonstrates that the most influential microclimatic factors affecting biometric indicators include noise, carbon dioxide, dust concentrations, light, temperature, humidity, and odors. The subjective perception of these factors correlates with insomnia symptoms, sleep quality, and perceived stress [25] highlighting the comprehensive impact of the environment on physiological well-being, this influence is particularly crucial when organizing rest and sleep conditions, which are key factors in recovery from stress, as the quality of microclimatic conditions directly determines the effectiveness of parasympathetic restorative processes [26].

A comparison of the effects of natural and urban environments on physiological indicators reveals significant differences. Being in a natural environment contributes to increased HRV and reduced salivary cortisol levels, indicating decreased physiological stress [27]. Walking in natural areas significantly lowers cortisol levels compared to walking on urban streets, demonstrating the stress-reducing effect of interaction with nature. At the same time, the presence of green spaces and the intensity, duration, and frequency of interaction with them play a crucial role [27, 28]. Enhancing visual and auditory experiences with tactile and olfactory stimuli in urban green spaces reduces stress and improves overall well-being [29]. This underscores the importance of creating a balanced environment to maximize psychological benefits.

1.2. Advanced Approaches for Assessing Environmental Impact on Stress Levels

Machine learning and artificial intelligence represent key approaches to researching the impact of the environment on stress. These technologies, combined with wearable devices, enable real-time collection of physiological data and allow continuous monitoring of psychological and physiological responses to various stressors. Campanella et al. [30] investigated the use of a wristband for recording photoplethysmographic and electrodermal signals, which were subsequently processed using Random Forest, SVM, and Logistic Regression algorithms. Among these, Random Forest achieved the highest accuracy (76.5%) by utilizing 27 extracted features for the binary classification of stress states [30]. A similar methodology was applied by Georgas et al. [31], where a galvanic skin response (GSR) sensor was used to assess anxiety in patients during COVID screening, revealing a correlation between environmental stress factors and psychological reactions in 51 subjects through automatic classification of physiological signals. Significant progress is demonstrated by Moser et al. [32] who proposed an LSTM network with a deep generative ensemble GAN, outperforming traditional algorithms by 7.18% in accuracy when using integrated gradients to identify key stress features and process sparsely labeled data. The innovative "Stress-Track" system [33] based on Internet of Things (IoT) technologies, the system achieved an impressive 99.5% accuracy in monitoring body temperature, sweating, and motor activity, which is particularly crucial for the early detection of stress responses to air pollution and other environmental factors that trigger oxidative stress.

Expanding this research direction, scientists have developed other effective methods for stress assessment. Bin Heyat et al. [34] used smart shirts to collect ECG signals from two groups of 10 participants each, one group experiencing stress after a 12-hour work shift and the other in a normal state [34]. Analysis of these data using various classifiers showed that the decision tree algorithm achieved the highest performance with a recall of 93.30% and an accuracy of 94.40%, which confirms the link between psychological stress and physiological disruptions, including mitochondrial dysfunction, oxidative stress, and elevated blood pressure. A study Almadhor et al. [27] observed that an innovative federated learning method with deep neural networks was applied, achieving 86.82% accuracy in classifying electrodermal activity from the WESAD dataset into five different states: transitional, baseline, stress, amusement, and meditation. A key feature of this method was ensuring patient data privacy, which is essential for large-scale monitoring of environmental stressors. A large-scale study Abd Al-Alim et al. [35] compared the effectiveness of five different machine learning models (KNN, SVC, DT, RF, and XGBoost) in conditions simulating real environmental exposure, using SWEET data from 240 participants. The analysis included electrocardiography, skin temperature, and skin conductance data, with the Random Forest algorithm achieving the best results in binary classification without SMOTE, with an accuracy of 98.29% and an F1-score of 97.89%. XGBoost outperformed other models in three-level classification with SMOTE, reaching an accuracy and F1-score of 98.98%. In research [36], researchers demonstrated the effectiveness of the k-Nearest Neighbors algorithm, which achieved an accuracy of 83.3% when analyzing GSR and PPG data from 37 participants exposed to air raid sirens under regulated conditions. This broadens the range of studied environmental factors and confirms the feasibility of using machine learning methods for their assessment.

With the deterioration of air quality and changing climatic conditions, physiological research is becoming increasingly important for developing reliable models of environmental impacts on health and for creating scientifically grounded recommendations for population protection [37].

Air pollution is a significant stress factor affecting physiological indicators; however, most available data on its effects have been obtained under controlled laboratory conditions [38]. This raises concerns about the applicability of this data for assessing real-world scenarios where air pollution levels fluctuate depending on climatic conditions, terrain type, and other factors [39].

Field deployment at a medium-scale plant showed 15% energy consumption, 18% peak demand, 30% CO_2 emissions, and 15% downtime with approximately 90% forecasting accuracy, confirming the effectiveness of integrating SCADA, machine learning, and digital twins [1,1]. On synchronized continuous-monitoring data, XGBoost achieved 91.25% accuracy, with passenger occupancy emerging as the key determinant of CO_2/$PM_{2.5}$/PM_{10} levels, which justifies prioritizing ventilation upgrades on overloaded lines [2,1]. Experimental validation of the clean-room IoT system at the KazNU IT Faculty (SCADA Genesis64, Modbus TCP, OPC UA, Google Coral USB) demonstrated approximately 99% sensor accuracy and control-loop reliability for automated temperature control and air conditioning [3,1]. To verify heating and cooling loops, a robotic test framework was built with control via a Google Coral USB Accelerator and an ADC, providing reproducible assessment of temperature/leak/switch sensors and their cyclic behavior, with logging and supervision in SCADA Genesis64 over OPC UA/Modbus TCP [40].

In this regard, this study conducted a review and analysis to assess the impact of air quality on stress levels in different environmental conditions. The main objective was to compare physiological responses recorded in three different locations under exposure to various air pollutants with responses observed in real-world environments with different pollution levels. The study analyzed changes in cardiovascular stress markers in response to polluted air exposure.

The complexity of field research, which requires control over various parameters such as weather conditions, physical activity levels, and individual participant characteristics, makes direct experimental investigation of this issue challenging. However, the analysis of collected data allows us to evaluate differences between laboratory and field conditions while considering significant factors influencing stress levels. Additionally, the use of machine learning algorithms enables us to obtain predictive estimates of physiological parameter changes in response to air pollution across various climatic and environmental conditions.

2. Method

2.1. System Architecture

This section presents the architecture of the developed system, as shown in Figure 1. The system consists of three subsystems: physiological parameter monitoring, environmental control, and machine learning algorithms.

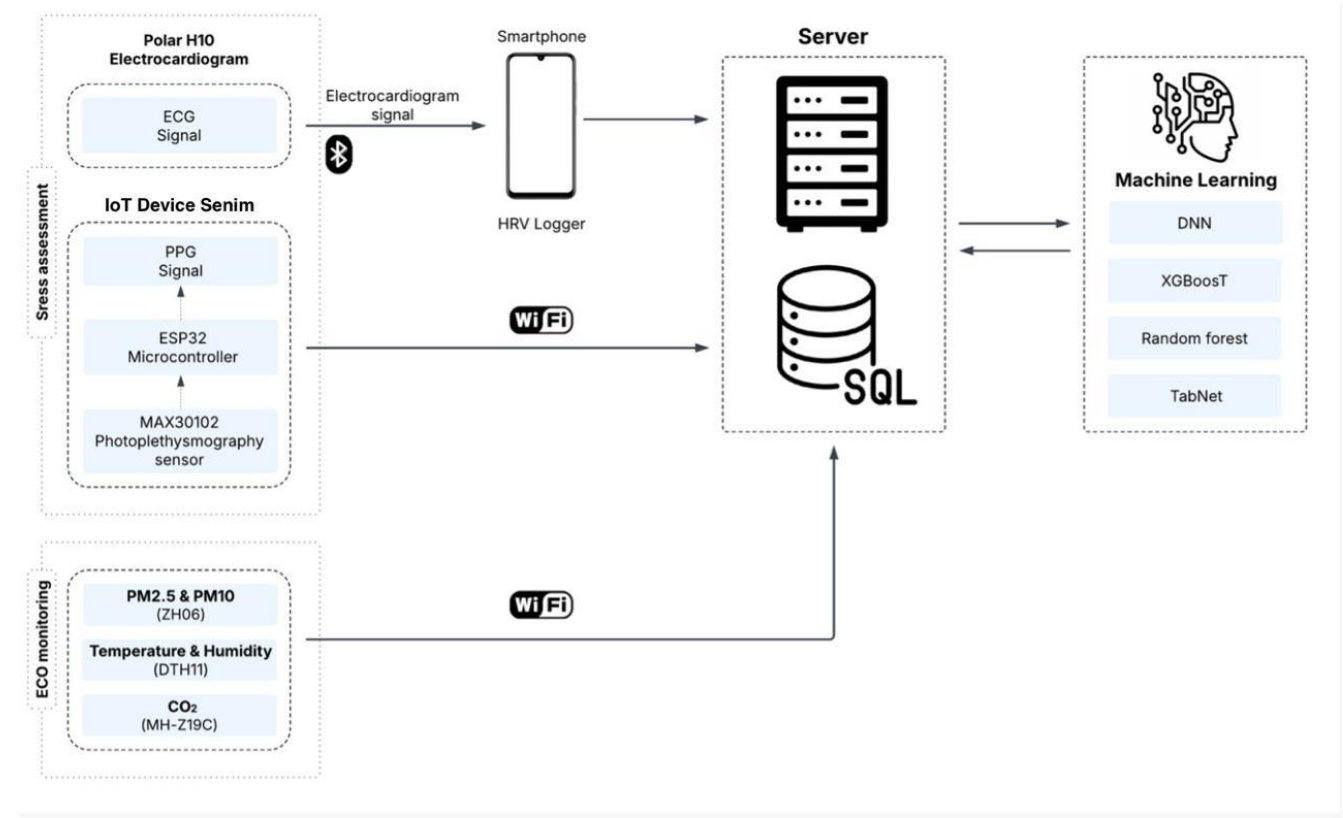


Figure 1.
System Architecture.

The system architecture is designed to monitor physiological parameters and environmental context, followed by machine learning analysis. The system comprises three key components: a data acquisition module, a server backend, and an intelligent processing unit. To assess stress levels, two data sources are utilized: the Polar H10 chest strap device, which records electrocardiographic (ECG) signals, and the custom-developed IoT device Senim, equipped with a MAX30102 photoplethysmographic (PPG) sensor and an ESP32 microcontroller.

ECG signals are transmitted via Bluetooth to the mobile application HRV Logger, which extracts heart rate variability (HRV) metrics and sends them to the server. Simultaneously, IoT device Senim records PPG signals and transmits them to the server via Wi-Fi. An essential aspect of the methodology is the use of data from the Polar H10 as a reference for validating the measurements obtained from the IoT device Senim to ensure their reliability and accuracy. In parallel, environmental parameters are monitored using a set of sensors: ZH06 (for $PM_{2.5}$ and PM_{10}), DHT11 (for temperature and humidity), and MH-Z19C (for carbon dioxide). These data are also sent to the server via Wi-Fi. The server backend provides centralized data storage in an SQL relational database and prepares the information for analysis.

After preliminary processing, the data is passed to the machine learning stage, where it is analyzed using deep neural networks (DNN), gradient boosting (XGBoost), Random Forest, and TabNet. These algorithms help identify patterns between environmental parameters and the body's physiological responses.

Synchronizing physiological data with environmental parameters enables a comprehensive analysis of the impact of external factors on the human body.

2.2. Experiment Description

The experiment was conducted in three contrasting environments: an open natural setting, a noisy urban environment, and an enclosed indoor space. To analyze the impact of environmental factors on stress levels, experimental HRV data were collected from a group of 10 participants who were sequentially exposed to three different conditions: a natural outdoor environment in a botanical garden, a noisy urban environment on Al-Farabi Avenue, and the laboratory with regulated conditions. HRV recordings were conducted for five minutes under each condition with continuous monitoring of physiological parameters, while environmental conditions (temperature, humidity, and air pollutants) were recorded in real-time. To ensure consistency across measurements, all sessions at the three locations were conducted on the same day, under the same weather conditions, and during the same time frame. This helped control for external environmental variability that could have otherwise influenced HRV outcomes. To minimize the influence of random factors on HRV, participants were instructed to sit and breathe naturally, avoiding controlled breathing or movement.

The experimental conditions were selected to create clear contrasts between different environmental settings. In the first phase of the study, participants were in a botanical garden, where noise levels were low, and natural greenery was present. The second phase took place in an urban environment on Al-Farabi Avenue, where participants were exposed to moderate environmental stressors such as air pollution, traffic noise, and high pedestrian density. The final phase of the experiment was conducted indoors under regulated conditions, including the use of humidifiers and air purifiers, which allowed for the assessment of HRV changes in an environment with minimal variability in external factors.

In previous research, strict participant selection criteria were commonly applied to control for confounding factors and ensure valid physiological data. This approach is reflected in a clinical health intervention study that required participants to be predominantly healthy and, in cases of known medical conditions, to present a certificate from a general practitioner verifying their ability to participate. Individuals whose health issues could interfere with the intervention were excluded [41]. Similarly, studies assessing cognitive performance in older adults excluded individuals with neurological, psychiatric, or chronic illnesses to maintain a homogeneous and reliable sample [42]. In physical health studies, participants with cardiovascular disease or temporary injuries were also excluded to reduce external influences on outcomes [43]. In line with these approaches, the present study established strict inclusion and exclusion criteria to ensure the reliability of HRV measurements. It included individuals aged 18 to 22 years without cardiovascular diseases, not taking medications affecting HRV, and abstaining from alcohol and caffeine for 24 hours before testing as shown in Table 1. Participants were excluded if they had insufficient sleep (less than six hours), experienced significant psychological or physiological stress on the test day, or if technical artifacts were detected in HRV data during preprocessing.

Table 1.
The Participant Demographics and Inclusion Criteria for the Experiment.

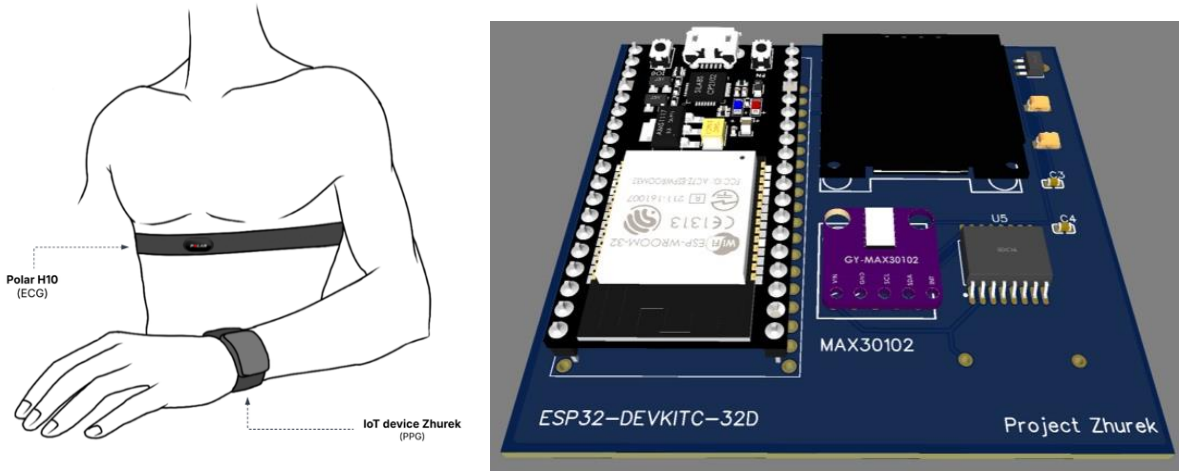
Location	**Male**	**Female**	**Total**
Participants	6	4	10
Age Range	19-22	18-20	18-22
Average Range	21	19	20
Alcohol Consumers	None	None	None
Caffeine Consumers	None	None	None
Smokers	None	None	None
CVD Cases	None	None	None

2.3. Data Collection

Data collection was conducted systematically during each experimental session, ensuring consistency in recording physiological parameters and environmental factors. Participants avoided unnecessary movements that could introduce

artifacts into the data. HRV and environmental factors were continuously measured for five minutes in each test condition. This time interval provided sufficient data for further analysis of the impact of environmental conditions on autonomic nervous system function.

The foundation of physiological stress monitoring was the registration of HRV using the IoT device Senim, which includes the MAX30102 photoplethysmographic sensor. This integrated biomedical sensor combines the functions of a pulse oximeter and PPG sensor, operating on the principle of differential light absorption at red (660 nm) and infrared (880 nm) wavelengths [44]. This configuration allows for the simultaneous recording of pulse waves and blood oxygen saturation (SpO_2) levels, which is particularly important for assessing the respiratory effects of air pollution. The sensor provided valuable data on heart rate, RR intervals, and key HRV parameters (SDNN, RMSSD). It was attached to the participant's wrist as shown in Figure 2, and recorded changes in blood volume associated with pulse waves, reflecting the cardiovascular system's response to stress caused by polluted air. To minimize delays in real-time physiological monitoring, the collected data were transmitted to the ESP-32 microcontroller. The SMD board of the IoT device Senim and its design are shown in Figure 2.



(a) (b)

Figure 2.
IoT device Senim for measuring heart rate variability (HRV) parameters: (a) Operation mechanism of PPG sensor on the finger for pulse wave measurement. (b) ESP-32 microcontroller with PPG sensor.

To ensure high measurement accuracy and validate PPG data, a Polar H10 ECG sensor was attached to the participant's chest following standard methodology, as shown in Figure 2. ECG sensors record the heart's electrical activity, providing reference HRV data, including time-domain parameters such as SDNN and RMSSD, as well as frequency-domain parameters, such as low-frequency (LF) and high-frequency (HF) oscillations and their LF/HF ratio. Electrocardiographic measurement is highly accurate and is considered the gold standard for HRV assessment, making it an essential tool for validating data obtained from the PPG sensor [45]. Polar H10 sensor employs advanced noise suppression technologies and eliminates interference caused by movement or changes in electrical contact with the skin. Both sensors, Polar H10 and PPG GY MAX30102, were worn simultaneously, ensuring concurrent data recording and synchronization. The data collected from both PPG and ECG sensors allowed for an assessment of measurement accuracy. The simultaneous use of both devices ensured data comparability and provided reliable validation of heart rate parameters, which enabled accurate interpretation of participants' physiological responses under different environmental conditions.

To assess the accuracy of measurements and validate the data obtained using the MAX30102 PPG sensor, a comparison was conducted with the Polar H10 ECG sensor. As shown in Table 2, the heart rate values and HRV features from both sensors exhibit a similar dynamic pattern, which indicates a relatively high reliability of the IoT device Senim. Data analysis revealed that the average correlation between values obtained with the Polar H10 Electrocardiographic Sensor and the IoT device Senim was $r = 0.87$ ($p < 0.001$). However, some discrepancies in absolute values were observed, particularly during the period 10:34:11–10:34:15, which may be attributed to motion artifacts or changes in sensor contact with the skin. The mean difference between RR intervals recorded by both sensors was 1.2 ± 3.4 ms, which falls within the clinically acceptable margin of error. This confirms that the IoT device Senim can be a reliable alternative to traditional electrocardiographic methods for assessing heart rate variability in environmental monitoring conditions. The results of validating the IoT device Senim with a Polar H10 ECG sensor are shown in Table 2.

Table 2.
Validation of the Polar H10 Electrocardiographic Sensor and IoT device Senim for Measurement.

Timestamp	Polar H10 ECG Sensor				IoT device Senim			
	HR	HR	RMSSD	SDNN	HR	RR	RMSSD	SDNN
07.03.2025 10:34:11	90	666.67	45	35	90.73	666.67	43	34
07.03.2025 10:34:12	89	674.16	45	38	86.01	697.67	47	39
07.03.2025 10:34:13	88	681.81	50	40	88.3	681.81	45	38
07.03.2025 10:34:14	88	681.81	48	39	87.40	689.65	48	40
07.03.2025 10:34:15	89	674.16	44	37	87.66	689.65	46	39

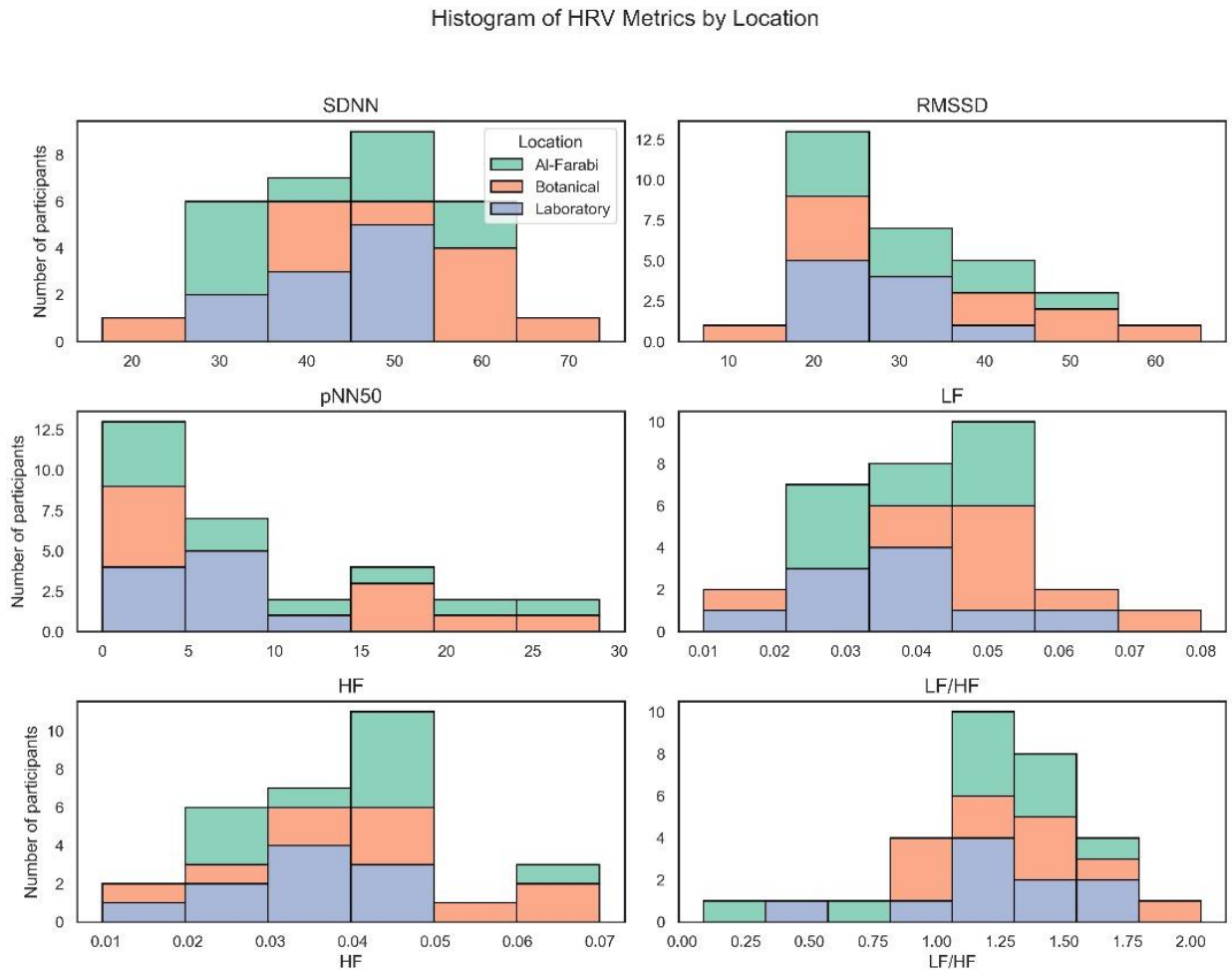


Figure 3.
Histograms of HRV Metrics by Location.

The histograms in Figure 3 display the distribution of six heart rate variability metrics across the three experiment locations, providing insights into how the autonomic nervous system (ANS) responds to varying factors such as location or stress. Heart rate variability (HRV) is a key indicator of the balance between the sympathetic and parasympathetic branches of the autonomic nervous system (ANS). Lower HRV values are generally associated with increased sympathetic activity or stress, whereas higher HRV values reflect stronger parasympathetic influence and relaxation [46]. By comparing HRV data across different locations, this analysis highlights how environmental factors can influence autonomic regulation and potentially impact overall health. During the study, signal processing was performed on the low-frequency component (LF Signal) using the discrete wavelet transform (DWT). The main goal was to determine the optimal wavelet function order and decomposition level for accurately extracting the respiratory rhythm. A comparison of different orders of Daubechies wavelets (db1, db2, db3, db4) showed that the most stable results were achieved with the fourth-order wavelet (db4), which provided the best balance between noise reduction and preservation of physiologically significant information.

In Figure 4, the signal decomposition stages are presented: the original LF signal and its decomposition at the second (Resp_A2), third (Resp_A3), and fourth (Resp_A4) approximation levels. At the fourth decomposition level (Resp_A4), a peak detection method was applied, enabling highly accurate identification of the respiratory rhythm. The automatically detected peaks, corresponding to respiratory cycles, are marked with red indicators. To verify the accuracy of the method, a comparison between automatic and manual counting of respiratory cycles over 5 minutes was conducted, showing an error of less than 4%, confirming the reliability of the detection.

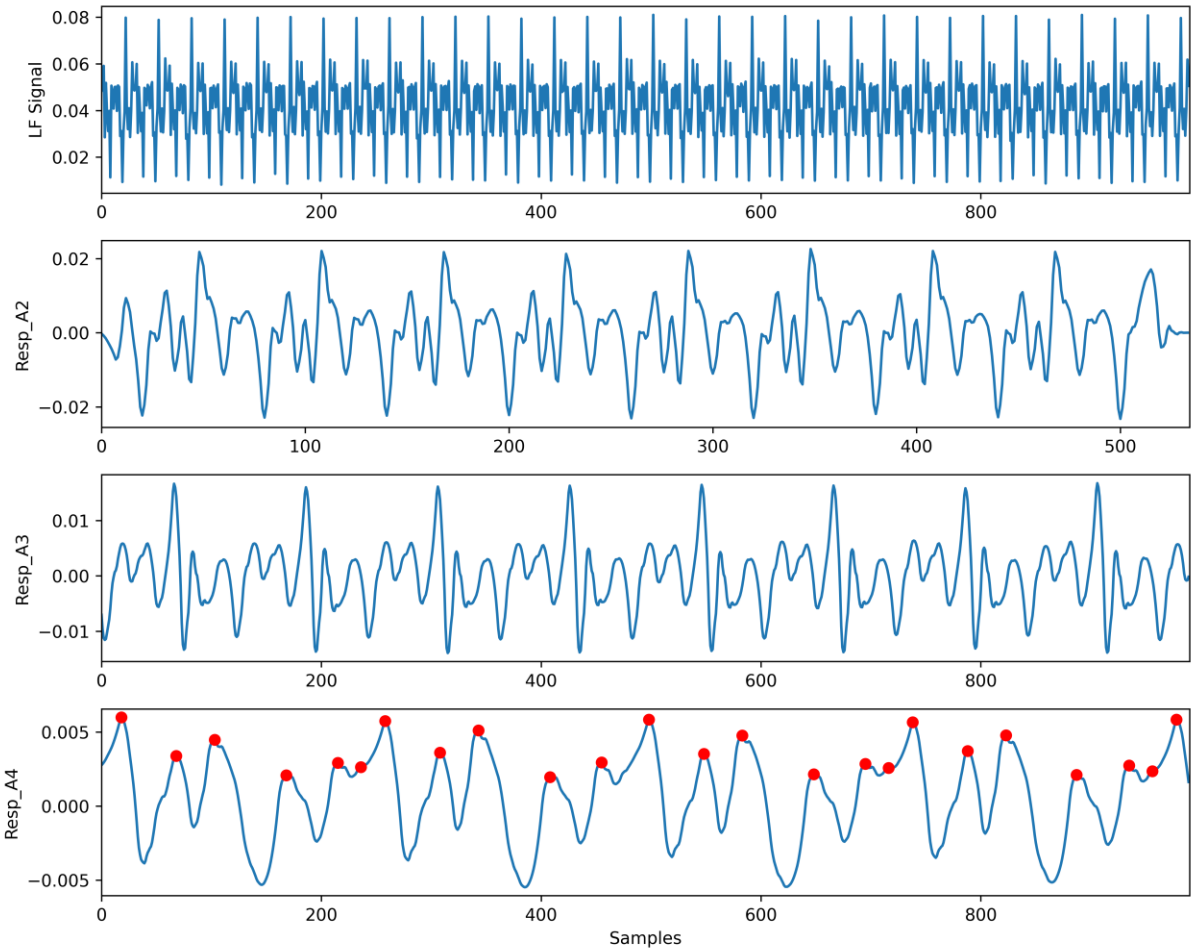


Figure 4.
Decomposition of the LF signal using DWT (db4) and comparison of approximation levels.

To understand the impact of external factors on the cardiovascular system, environmental parameters were monitored alongside physiological measurements using the air monitoring system. This compact device provided a comprehensive assessment of environmental conditions by measuring the concentration of fine particulate matter ($PM_{2.5}$), larger particles (PM_{10}), and carbon dioxide (CO_2), which were visually displayed on the device screen as shown in Figure 5. The monitor uses a laser sensor based on the light-scattering principle to detect suspended particles ranging from 0.3 to 10 µm in diameter, corresponding to $PM_{2.5}$ and PM_{10} fractions, which are most relevant for evaluating the health effects of air pollution. The measurement accuracy for particulate concentration is ±10% within a range of 0 to 500 µg/m³, allowing for the reliable detection of even slight pollution fluctuations [43]. For carbon dioxide measurement, the device employs a nondispersive infrared (NDIR) sensor with an accuracy of ±50 ppm within a range of 400 to 5000 ppm [47]. The device's high sensitivity enabled the detection of even minor changes in air composition under different test conditions.

Figure 5.
Real-time monitoring of environmental data.

In Figure 5 the data from sensors were displayed on a dedicated website, eco.com.kz, which provides real-time updates every 5 minutes. The key parameters shown include CO_2 levels, humidity, PM_{10}, and $PM_{2.5}$, enabling efficient tracking and analysis of air quality data. Users can select a specific date at the top right corner of the website to view the recorded parameters for that day. Graphs for each parameter are available, which can be navigated through to observe variations over time. These parameters are critically important for assessing the environmental impact on participants' physiological states, as air pollution and climatic factors can significantly influence the autonomic nervous system through mechanisms of neuroinflammation, oxidative stress, and direct chemoreceptor stimulation [48]. Special attention was given to $PM_{2.5}$ concentrations, as these particles can penetrate the alveolar-capillary barrier directly into the bloodstream, triggering systemic inflammation and endothelial dysfunction processes closely linked to sympathetic nervous system activation and stress responses [48]. The use of this sensor enabled the identification of correlations between air quality and changes in cardiovascular system parameters across different environmental conditions.

2.4. Data Preprocessing

Stress levels were classified using Baevsky's Stress Index (SI), which was calculated based on RR intervals. SI was divided into three categories: low (0), medium (1), and high (2) stress levels, according to the following thresholds: a low-stress level (0) corresponds to scores between 0 and 50, a medium stress level (1) covers scores from 51 to 150, and a high stress level (2) includes scores above 150. This classification enabled the determination of stress levels based on physiological data, which was then combined with environmental data to form a unified dataset. This dataset included both physiological parameters and environmental factors such as $PM_{2.5}$, PM_{10}, and SO_2. These environmental factors were recorded alongside the Stress Index to reflect the influence of external stressors on the autonomic nervous system.

The combined dataset served as the foundation for building machine learning models. To further enhance the dataset and ensure good generalizability of the models across various conditions, synthetic data generation was employed. Gaussian noise distribution was added to the cleaned dataset, introducing controlled variations in environmental parameters ($PM_{2.5}$, PM_{10}, CO_2) and the Stress Index, which increased the original 30 data points to 990 instances.

Gaussian noise was selected as a data augmentation technique due to its demonstrated effectiveness in enhancing machine learning performance on small datasets, particularly in clinical and environmental applications. It has been shown to perform comparably to, and in some cases outperform, other augmentation methods, such as SMOTE and ADASYN, while preserving data variability and structure. The use of Gaussian noise also supports improved model generalization and reduces overfitting, especially when training data is limited [49, 50]. It was used to introduce controlled variations in environmental parameters ($PM_{2.5}$, PM_{10}, CO_2) and the Stress Index, increasing the original 30 data points to 990 instances. The noise was generated with a mean of 0 and a standard deviation selected based on the real data distribution to ensure realistic augmentation. This synthetic data generation process was designed to preserve the characteristics of the real dataset, ensuring that the added noise reflected natural variability.

Gaussian noise can be mathematically represented as:

$$x = \mu + \sigma \cdot z \quad (1)$$

where x represents the noisy data, μ is the mean of the data (which is 0 in this case), σ is the standard deviation determined by the distribution of real data, and z is a random variable drawn from a standard normal distribution N (0,1).

The final dataset consisted of 990 records, including both real and synthetic data points, with key environmental features such as PM_{10}, $PM_{2.5}$, CO_2, and Stress Index. The introduction of synthetic data helped balance the dataset, making it more representative of different stress levels and improving model generalization. These synthetic records preserved realistic correlations with real data, particularly between the environmental features and the stress index, ensuring that the models could generalize well across diverse conditions. This approach enabled improved assessment quality in the subsequent machine learning models.

2.5. Data Analysis

This research section employed supervised machine learning classification algorithms to develop a model capable of assessing a person's stress level based on various air quality parameters. During this phase, air parameters (CO_2, $PM_{2.5}$, and PM_{10}) were extracted from real-time environmental monitoring data and used as input features. This preliminary process enabled the creation of a larger number of air parameter samples for use in machine-learning classification. The model was trained and evaluated using labeled stress index data derived from RR intervals to establish correlations between air pollution levels and physiological stress responses. We applied grid search for hyperparameter tuning to identify the optimal configuration for our model. This approach systematically explored a predefined parameter space to enhance overall performance.

The model's performance was assessed using precision, recall, and F1-score metrics, which validated its effectiveness in stress assessment. Precision represents the ratio of correctly classified positive instances to the total predicted positive instances. Recall measures the proportion of actual positive instances that were correctly identified. The F1-score provides a balanced measure by calculating the weighted average of precision and recall. Feature importance analysis was indicated using the SHAP method [51].

2.5.1. Statistical Method

The stress level of individuals was determined using RR intervals of heartbeats, which were recorded at three different locations. To quantify stress, Baevsky's Stress Index (SI) was computed according to the formula (2): preferred.

$$SI = \frac{AM_o \times 100\%}{2M_o \times M_x DM_n} \quad (2)$$

where the mode (M_o) is the most frequent RR interval expressed in seconds [49]. The amplitude of the mode (AM_o) was calculated, using 50 ms bin width, as the number of the RR intervals in the bin containing the M_o, expressed as a percentage of the total number of intervals measured. The variability is reflected in $M_x DM_n$ as the difference between the longest (M_x) and shortest (M_n) RR interval values, expressed in seconds. The SI is expressed as s^{-2}.

2.5.2. Machine Learning Classification

Using an expanded dataset, DNN, XGBoost, Random Forest, and TabNet were used to assess the subject's stress level based on air parameters. As a three-class classification problem, DNN, XGBoost, TabNet, and Random Forest were employed due to their ability to handle complex patterns, non-linearity, and structured data effectively. The pipeline of machine learning algorithms is presented in Figure 6.

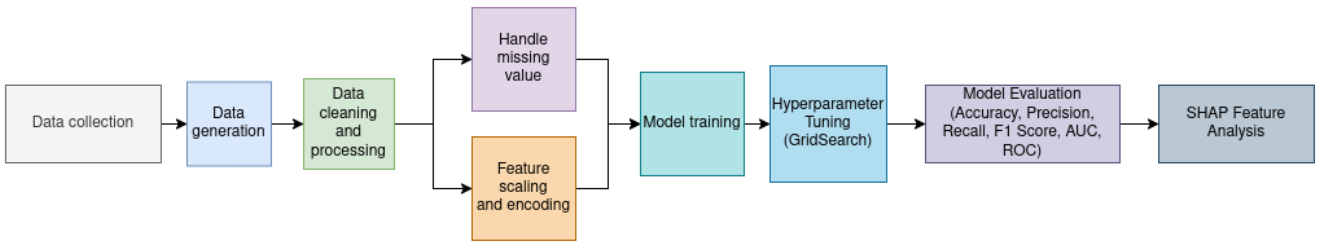


Figure 6.
Machine Learning Pipeline.

DNN uncovers hidden interactions and complex dynamics in systems purely from observational data, allowing analysis without relying on predefined models [52]. XGBoost and RF utilize ensemble learning to enhance the accuracy and reliability of assessment [53]. Moreover, TabNet was incorporated due to its unique ability to perform interpretable deep learning using a sequential attention mechanism [54].

Preprocessing operations included the normalization of CO_2, $PM_{2.5}$, and PM_{10} values to ensure all variables were computed on the same scale, along with label encoding of the prediction target. As a result, categorical variables representing stress levels were encoded into numerical values using a label encoder: '0' for low stress, '1' for medium stress,

and '2' for high stress. The dataset consists of three features (CO_2, $PM_{2.5}$, and PM_{10}) and a target variable representing stress levels. The dataset was based on these three features and a target variable indicating stress levels (low, medium, and high).

The classification models were developed using Python and implemented with Scikit-learn (for RF), XGBoost (for gradient boosting), PyTorch (for DNN), and PyTorch TabNet (for TabNet).

The performance of the machine learning models was assessed using evaluation metrics such as classification accuracy, precision, recall, and the F1-score.

3. Results and Discussion

3.1. Measurements in the Experimental Environment

Environmental conditions were continuously monitored in real-time across the three experimental settings. Temperature, humidity, and air pollutant levels were recorded during each session to assess potential external influences on physiological responses. These measurements provided a standardized assessment of environmental variability and guaranteed that each condition was characterized objectively. Differences in these parameters across environments contributed to the varying physiological effects observed in HRV analysis. The recorded air parameters from the three locations are presented in Table 3. Figure 7 illustrates Air Quality Index (AQI) values. The AQI range spans from 0 to 500 [55], where lower values represent good air quality with minimal health impact, and higher values indicate unhealthy to hazardous air quality, posing significant health risks, especially for sensitive individuals. Monitoring AQI is crucial to assess health risks, adopt preventive measures during high pollution periods, and evaluate the effectiveness of environmental policies to improve air quality.

Table 3.
Air quality parameters were measured across the three locations.

Location	Mean CO2, ppm	Mean $PM_{2.5}$, µg/m³	Mean PM_{10}, µg/m³
Al-Farabi avenue	463.821	40.200	31.871
Botanical Garden	413.300	57.500	46.500
Laboratory	1478.975	31.325	24.324

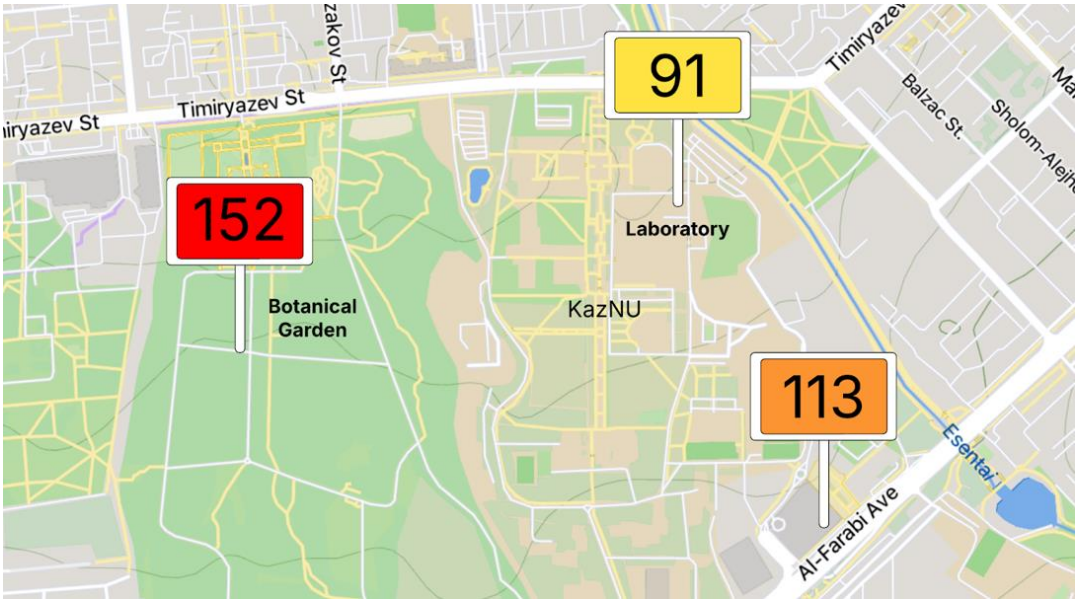


Figure 7.
Air Quality Index (AQI) values across the three different locations, based on sensor data collected during the experiment.

The higher AQI in the botanical garden, as shown in Figure 7, was primarily due to increased humidity in Table 4, which played a significant role in trapping particulate matter ($PM_{2.5}$ and PM_{10}). Vegetation contributes to higher humidity through transpiration and reduced air circulation, leading to moisture accumulation in the air [56]. This excess moisture causes fine particulate matter to absorb water, increasing its mass and concentration in the atmosphere. As a result, the botanical garden exhibited higher levels of $PM_{2.5}$ and PM_{10}, which directly impacted the AQI.

Table 4.
Mean Humidity Levels Across Different Locations.

Location	Mean Humidity, %
Al-Farabi avenue	48.33 ± 5.08
Botanical Garden	83.54 ± 0.97
Laboratory	32.61 ± 4.11

3.2. Influence of Environmental Exposure on HRV

HRV data were collected for 5 minutes in each environment to track fluctuations in RR intervals as indicators of autonomic nervous system activity. Participants' RR intervals varied depending on the surrounding environment, which demonstrated the body's physiological adaptation to external stimuli. The observed differences in RR interval variability suggest that external environmental factors influence autonomic activity, even when participants maintain a stable seated position and a natural breathing pattern. The results support the hypothesis that environmental conditions affect nervous system regulation. The calculations for each participant were detected, as shown in Figure 8.

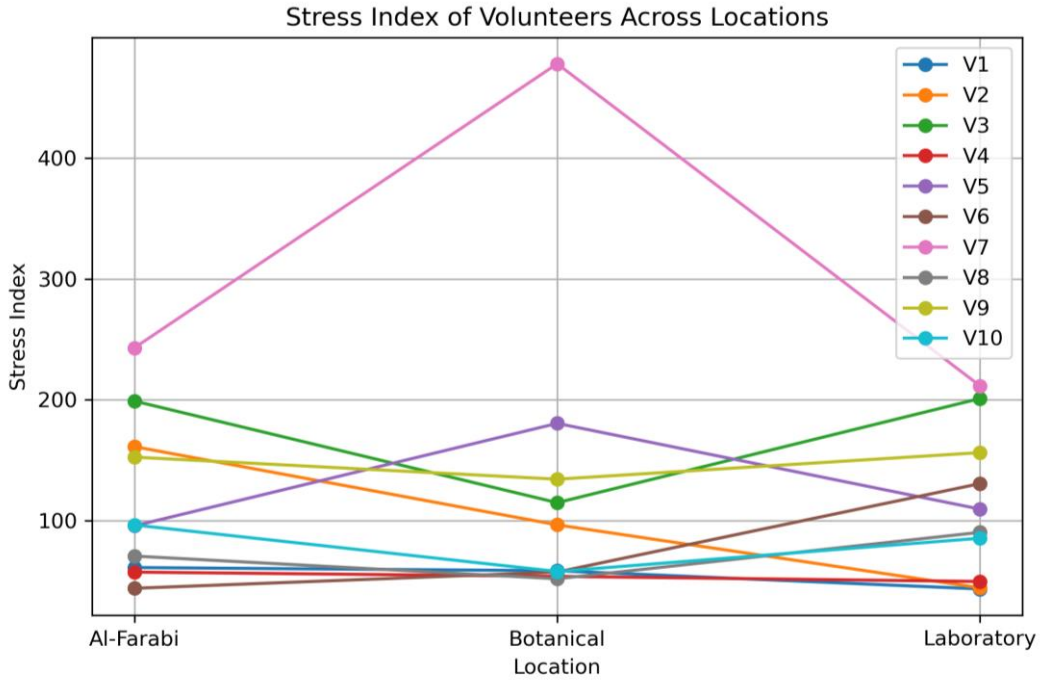
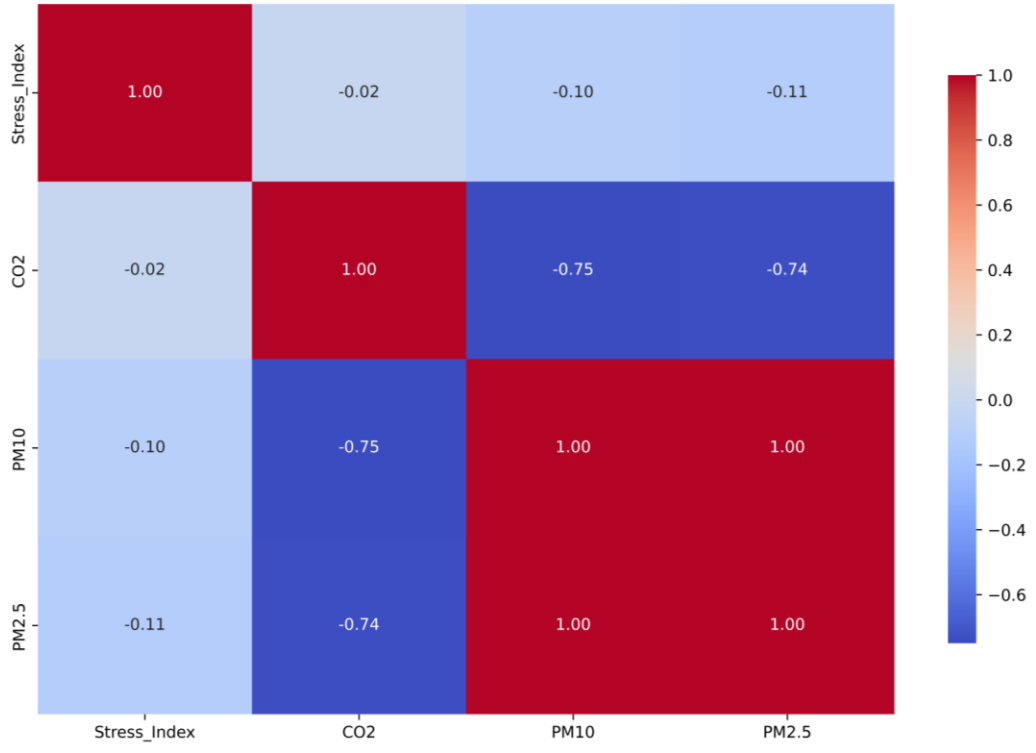
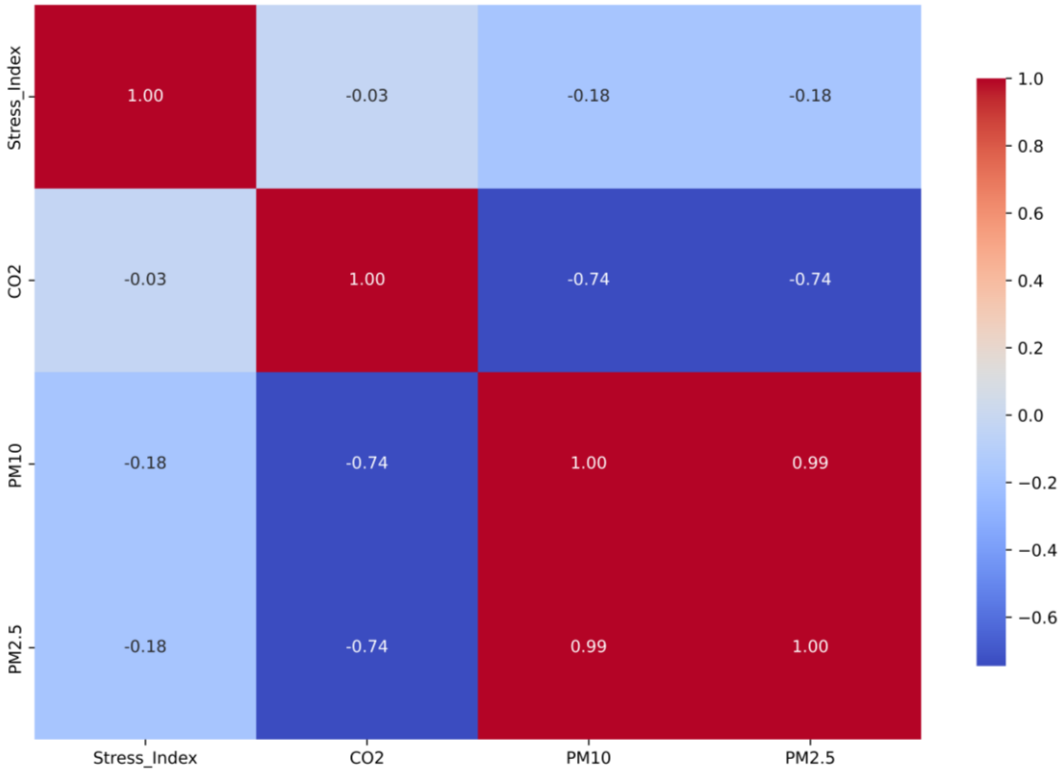


Figure 8.
Stress index of all volunteers across locations.

The comparison of mutual information between features in the original and generated air parameters datasets is illustrated in Figure 9. Mutual information quantifies the dependency between variables and helps assess how well the generated data preserves relationships present in the original dataset. By comparing these correlation matrices, we evaluate the similarity in feature interactions and determine the effectiveness of data generation in maintaining underlying patterns.



(a)



(b)

Figure 9.
Correlation matrix between Stress index and environmental parameters for: (a) Original data. (b) Synthetic data.

3.3. Classification of Stress Level Based on Machine Learning

3.3.1. Data Augmentation

The dataset was expanded by generating synthetic data using Gaussian noise. Figure 10 presents a comparison of the original dataset and the generated dataset, visualized using Kernel Density Estimation distribution. This analysis used CO2, $PM_{2.5}$, and PM_{10} as features. The data generation process effectively preserved these air quality parameters' underlying patterns and relationships. By the final training stage, the synthetic data closely mirrored the distribution of the original dataset across all features, as depicted in Figure 10. However, synthetic data generation may not perfectly replicate the original dataset or fully capture intricate feature relationships. This variation is not necessarily a drawback, as some level of divergence is expected. Nevertheless, it is crucial that the generated data retains the statistical properties and dependencies between features to ensure its reliability and applicability.

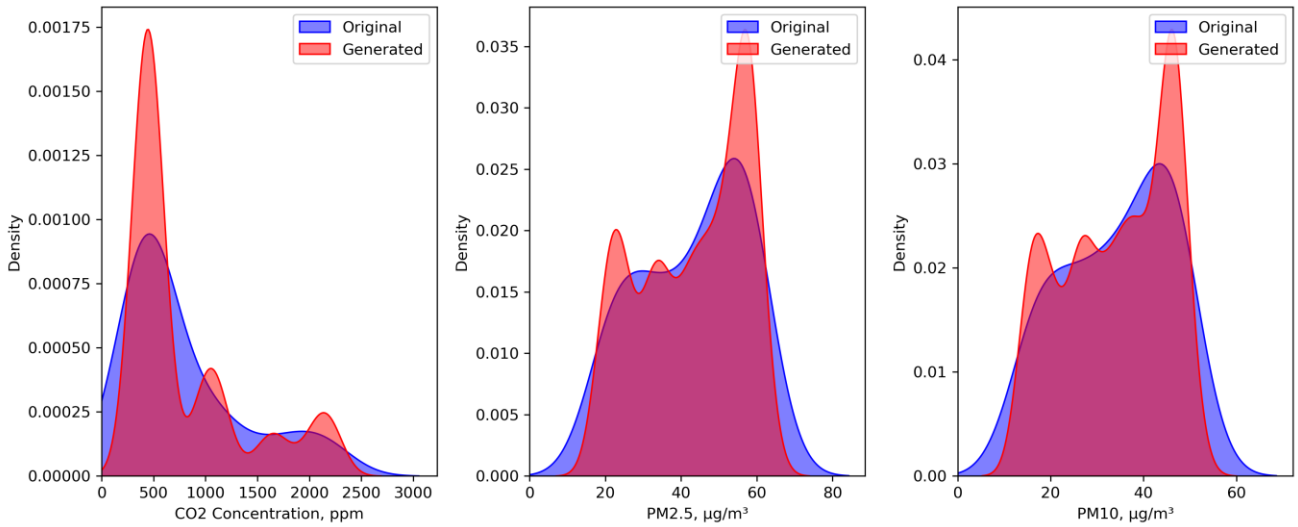


Figure 10.
Comparison of the features in the original and generated datasets.

3.3.2. Machine Learning Classification

The synthetic data samples were merged with the original training data to enhance balance and expand the training dataset for the machine learning classifiers. Machine learning models are used to assess subjects' stress levels (low, medium, or high) based on air parameters. Specifically, stress levels are classified as low for values between 0-50, medium for values between 51-150, and high for values of 151 and above. The performance results of the DNN, XGBoost, RF, and TabNet algorithms are summarized in Table 5. After training the ML models, the average evaluation metric values for each classifier were computed. All classifiers demonstrated strong performance, with the XGBoost model achieving the highest accuracy of 91.92%.

The confusion matrices of the assessment process, using the air parameters, are shown in Figure 11. The XGBoost model achieved a precision of 91.82% in assessing stress levels, with an F1-score of 90.42% and a recall of 89.28%, indicating strong overall classification accuracy.

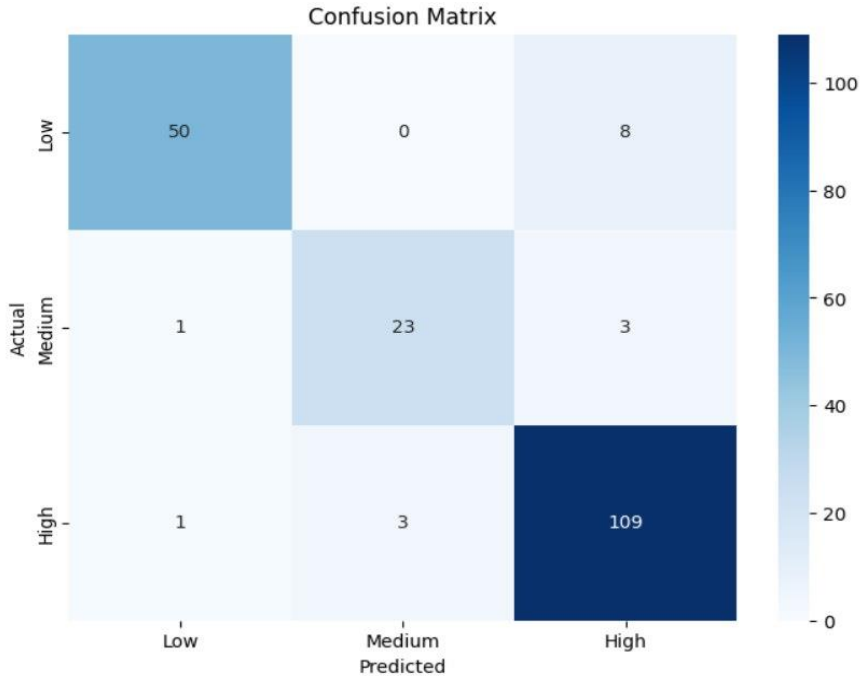
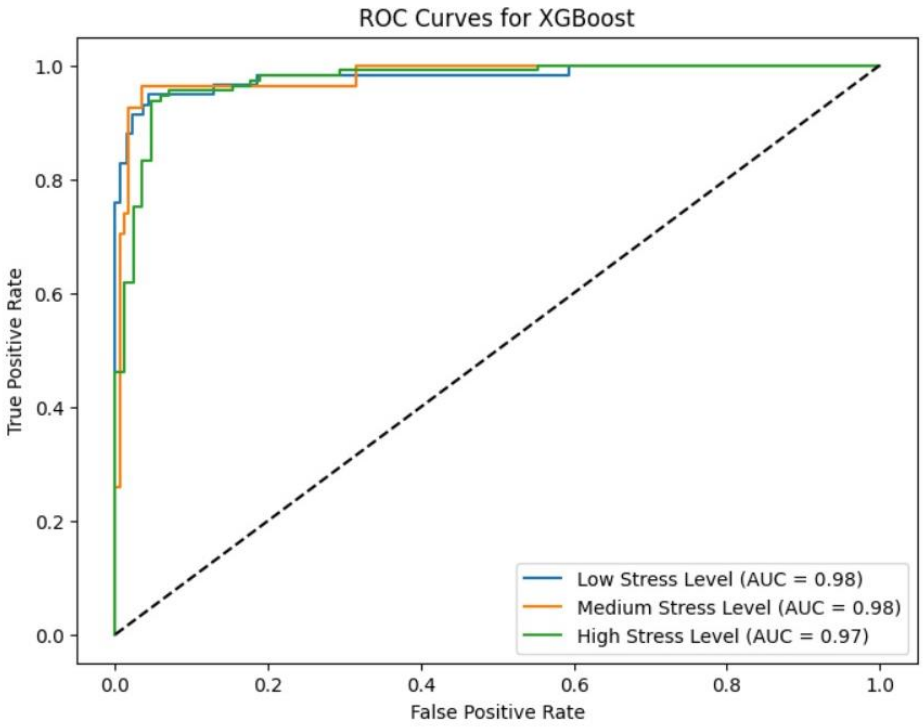
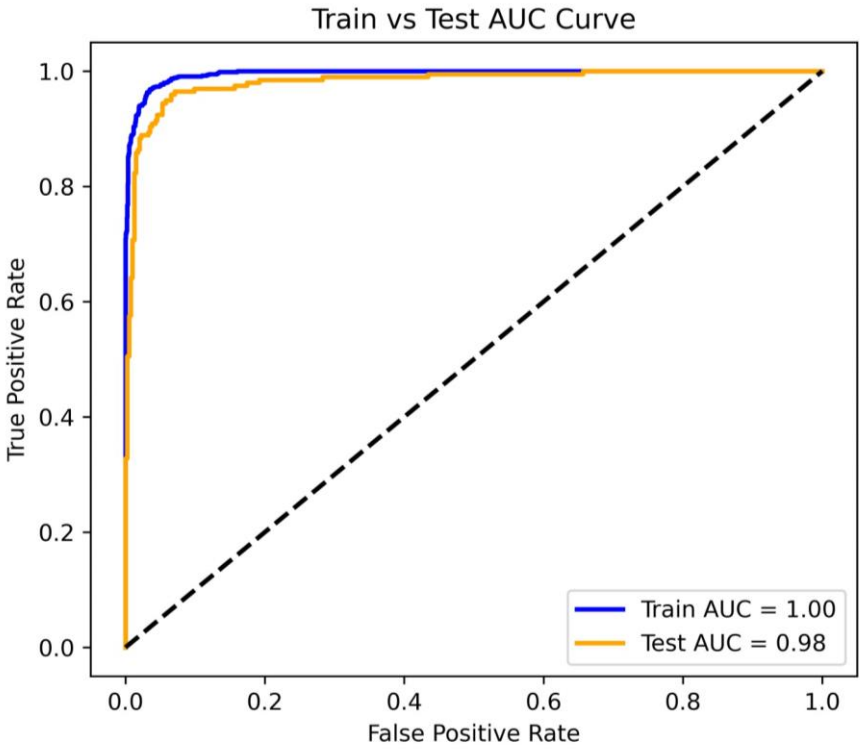


Figure 11.
Confusion matrix using the air parameters for XGBoost.



(a)



(b)

Figure 12.
Model performance evaluation using the air parameter: (a) Receiver operating characteristic (ROC) curve for classification assessment. (b) Area under the curve (AUC) comparison during training and testing for XGBoost.

The ROC curves, along with the AUC values, are shown in Figure 12, respectively. The Area Under the Curve (AUC) values obtained during training and testing provide a robust indication of the model's discriminative ability. The training AUC of 0.9945 suggests that the model has learned the underlying patterns in the data exceptionally well, achieving near-perfect classification. Meanwhile, the test AUC of 0.9786 demonstrates that the XGBoost model generalizes well to unseen data, with only a slight decrease in performance. The minimal gap between training and test AUC values suggests that overfitting is not a significant concern, indicating a well-regularized model that maintains high predictive power.

Table 5.
Accuracy, precision, recall and F1-score values obtained for ML classifiers in the prediction of stress levels based on air parameters.

Evaluation metrics	DNN	XGBoost	RF	TabNet
Accuracy	90.60%	91.92%	89.9%	83.89%
F1-Score	0.8767	0.9042	0.8944	0.7949
Precision	0.8725	0.9182	0.8816	0.8825
Recall	0.8828	0.8928	0.9110	0.7673

Table 5 presents a comparison of the performance of four machine learning models: DNN, XGBoost, Random Forest, and TabNet in a classification task. Four standard evaluation metrics were used to assess classification quality: Accuracy, F1-Score, Precision, and Recall. The highest Accuracy (91.92%) and F1-score (0.9042) were achieved by the XGBoost model, demonstrating its overall effectiveness and balanced performance between precision and recall. The results indicate that ensemble methods (XGBoost, Random Forest) outperform both the deep neural network architecture (DNN) and the specialized model for tabular data (TabNet) in the context of this task.

3.3.3. Feature Importance Analysis Using SHAP

To assess the contribution of different air parameters to the evaluation of stress levels, SHapley Additive exPlanations (SHAP) were used to interpret the XGBoost model. The SHAP summary plot, Figure 13, provides insights into the relative importance of each environmental factor in determining stress levels.

Among the features analyzed, $PM_{2.5}$ emerged as the most influential factor, indicating that fine particulate matter had the strongest impact on stress assessment. This was followed by PM_{10}, suggesting that larger airborne particles also played a significant role in affecting physiological responses. CO_2 ranked third in importance, showing a moderate influence on stress levels.

These findings highlight the substantial role of air quality in physiological stress responses, with fine particulate matter ($PM_{2.5}$ and PM_{10}) being more critical than CO_2 in assessing variations in stress levels. This suggests that exposure to higher

levels of air pollutants may contribute more strongly to stress-related physiological changes than carbon dioxide concentration.

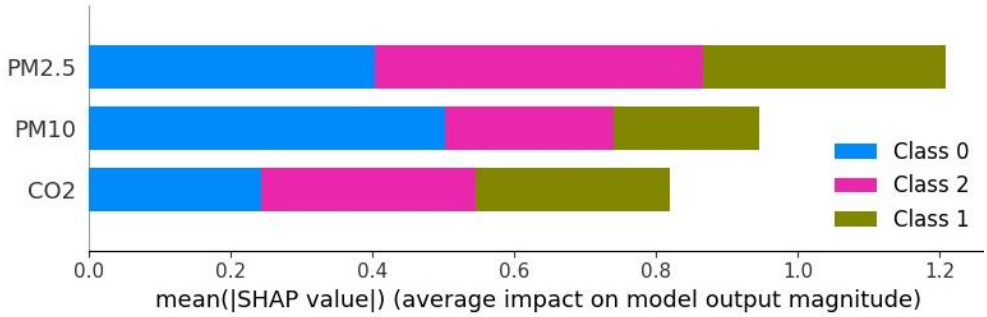


Figure 13.
Feature Importance Analysis Using SHAP.

Emerging research consistently supports a strong link between air pollution and physiological stress responses. Fine particulate matter such as $PM_{2.5}$ and PM_{10} has been shown to induce oxidative stress, inflammation, and redox signaling disruptions, all contributing to biological stress. These pollutants can penetrate deeply into the body and affect neurological and systemic functions through mechanisms including neuroinflammation and epigenetic changes [56-58]. This aligns with our finding that particulate matter significantly impacts autonomic nervous system activity and stress levels, highlighting the importance of air quality monitoring in stress assessment and public health efforts.

There were some limitations in the current research. Wind speed and vegetation type in the botanical garden were not measured or included in the analysis, which might have influenced the environmental data. Before the experiments, only short-term lifestyle factors such as 24-hour dietary restrictions were considered, leaving out other important aspects of participants' routines.

4. Conclusion

In this study, air quality parameters and heart rate variability (HRV) parameters were collected in real-time, while RR intervals were simultaneously recorded from 10 volunteers using PPG in three different locations. The obtained results confirmed the significant impact of environmental factors on physiological responses, particularly stress levels, successfully demonstrating this relationship. A detailed HRV analysis revealed variations in autonomic nervous system activity in response to different environmental conditions, further reinforcing the link between air quality and physiological stress markers. These findings contribute to the growing body of research on environmental stressors and their effects on human health. Notably, the XGBoost classification algorithm achieved the highest assessment accuracy of 91.92%, confirming its effectiveness in stress classification based on air quality parameters. SHAP feature analysis identified $PM_{2.5}$ as the most significant environmental factor influencing HRV and stress levels. Integrating HRV and air quality data in an IoT system could enable city dwellers to receive real-time alerts during high pollution periods, helping them avoid outdoor activities or modify routines. Wearable-based feedback systems could even suggest breathing exercises or stress-relief actions, making urban living healthier and more sustainable.

Future research will address current limitations by measuring wind speed and identifying vegetation types in the study environment. It will also consider long-term lifestyle factors such as sleep quality, eating habits, and other behavioral patterns. Additionally, plans include increasing the volume of real data collected and exploring further data augmentation strategies to test and enhance the robustness of the current approach.

As part of this work, the developed machine learning models will be integrated into an intelligent and adaptive environment that dynamically responds to users' stress levels and air quality conditions. This integration will improve real-time stress monitoring and mitigation strategies, potentially enhancing overall well-being. Further research in cardiovascular diseases (CVD) will investigate stress as a key contributing factor to predicting CVD risks. The research will include participants across different age groups, with continuous 24-hour HRV and air quality monitoring conducted for each individual to capture comprehensive physiological responses to environmental stressors. Analyzing the interplay between chronic stress and cardiovascular risk factors will provide deeper insights into preventive measures and personalized medical interventions, ultimately contributing to advancements in stress management and cardiovascular health research.

References

[1] A. Kurasz, G. Y. Lip, S. Dobrzycki, and Ł. Kuźma, "A breath of trouble: Unraveling the impact of air pollution on atrial fibrillation," *Journal of Clinical Medicine*, vol. 13, no. 23, p. 7400, 2024. https://doi.org/10.3390/jcm13237400

[2] Z. Leni, L. Künzi, and M. Geiser, "Air pollution causing oxidative stress," *Current Opinion in Toxicology*, vol. 20-21, pp. 1-8, 2020. https://doi.org/10.1016/j.cotox.2020.02.006

[3] S.-Y. Cho and H.-T. Roh, "Combined effects of particulate matter exposure and exercise training on neuroplasticity-related growth factors and blood–brain barrier integrity," *Atmosphere*, vol. 16, no. 2, p. 220, 2025. https://doi.org/10.3390/atmos16020220

[4] P. Thangavel, D. Park, and Y.-C. Lee, "Recent insights into particulate matter (PM2. 5)-mediated toxicity in humans: An overview," *International Journal of Environmental Research and Public Health*, vol. 19, no. 12, p. 7511, 2022. https://doi.org/10.3390/ijerph19127511
[5] World Health Organization, *Air quality, energy and health*. Geneva: World Health Organization, 2025.
[6] J. Li, L. Huang, B. Han, T. J. van der Kuijp, Y. Xia, and K. Chen, "Exposure and perception of PM2. 5 pollution on the mental stress of pregnant women," *Environment International*, vol. 156, p. 106686, 2021. https://doi.org/10.1016/j.envint.2021.106686
[7] K. Leontjevaite, A. Donnelly, and T. E. MacIntyre, "Air pollution effects on mental health relationships: Scoping review on historically used methodologies to analyze adult populations," *Air*, vol. 2, no. 3, pp. 258-291, 2024. https://doi.org/10.3390/air2030016
[8] A. Lischke, R. Jacksteit, A. Mau-Moeller, R. Pahnke, A. O. Hamm, and M. Weippert, "Heart rate variability is associated with psychosocial stress in distinct social domains," *Journal of Psychosomatic Research*, vol. 106, pp. 56-61, 2018. https://doi.org/10.1016/j.jpsychores.2018.01.005
[9] T. Senoner and W. Dichtl, "Oxidative stress in cardiovascular diseases: Still a therapeutic target?," *Nutrients*, vol. 11, no. 9, p. 2090, 2019. https://doi.org/10.3390/nu11092090
[10] M. G. Scioli *et al.*, "Oxidative stress and new pathogenetic mechanisms in endothelial dysfunction: Potential diagnostic biomarkers and therapeutic targets," *Journal of Clinical Medicine*, vol. 9, no. 6, p. 1995, 2020. https://doi.org/10.3390/jcm9061995
[11] M. Czlapka-Matyasik, L. Wadolowska, P. Gut, and A. Gramza-Michałowska, "Changes in oxidative stress, inflammatory markers, and lipid profile after a 6-week high-antioxidant-capacity dietary intervention in CVD patients," *Nutrients*, vol. 17, no. 5, p. 806, 2025. https://doi.org/10.3390/nu17050806
[12] C. Penna and P. Pagliaro, "Endothelial dysfunction: Redox imbalance, NLRP3 inflammasome, and inflammatory responses in cardiovascular diseases," *Antioxidants*, vol. 14, no. 3, p. 256, 2025. https://doi.org/10.3390/antiox14030256
[13] M. Bolpagni, S. Pardini, M. Dianti, and S. Gabrielli, "Personalized stress detection using biosignals from wearables: A scoping review," *Sensors*, vol. 24, no. 10, p. 3221, 2024. https://doi.org/10.3390/s24103221
[14] I. B. Messaoud and O. Thamsuwan, "Heart rate variability-based stress detection and fall risk monitoring during daily activities: A machine learning approach," *Computers*, vol. 14, no. 2, p. 45, 2025. https://doi.org/10.3390/computers14020045
[15] A. G. Polak, B. Klich, S. Saganowski, M. A. Prucnal, and P. Kazienko, "Processing photoplethysmograms recorded by smartwatches to improve the quality of derived pulse rate variability," *Sensors*, vol. 22, no. 18, p. 7047, 2022. https://doi.org/10.3390/s22187047
[16] A. Agorastos, A. C. Mansueto, T. Hager, E. Pappi, A. Gardikioti, and O. Stiedl, "Heart rate variability as a translational dynamic biomarker of altered autonomic function in health and psychiatric disease," *Biomedicines*, vol. 11, no. 6, p. 1591, 2023. https://doi.org/10.3390/biomedicines11061591
[17] I. Martínez-González-Moro, I. Albertus Camara, and M.-J. Paredes Ruiz, "Influences of intense physical effort on the activity of the autonomous nervous system and stress, as measured with photoplethysmography," *International Journal of Environmental Research and Public Health*, vol. 19, no. 23, p. 16066, 2022. https://doi.org/10.3390/ijerph192316066
[18] G. Volpes *et al.*, "Feasibility of ultra-short-term analysis of heart rate and systolic arterial pressure variability at rest and during stress via time-domain and entropy-based measures," *Sensors*, vol. 22, no. 23, p. 9149, 2022. https://doi.org/10.3390/s22239149
[19] P. Lebamovski and E. Gospodinova, "Investigating stress during a virtual reality game through fractal and multifractal analysis of heart rate variability," *Applied System Innovation*, vol. 8, no. 1, p. 16, 2025. https://doi.org/10.3390/asi8010016
[20] A. Tonacci, L. Billeci, E. Burrai, F. Sansone, and R. Conte, "Comparative evaluation of the autonomic response to cognitive and sensory stimulations through wearable sensors," *Sensors*, vol. 19, no. 21, p. 4661, 2019. https://doi.org/10.3390/s19214661
[21] R. Mateo-Reyes, I. A. Cruz-Albarran, and L. A. Morales-Hernandez, "Intelligent stress detection using ECG signals: Power spectrum imaging with continuous wavelet transform and CNN," *Journal of Experimental and Theoretical Analyses*, vol. 3, no. 1, p. 6, 2025. https://doi.org/10.3390/jeta3010006
[22] K. M. Dalmeida and G. L. Masala, "HRV features as viable physiological markers for stress detection using wearable devices," *Sensors*, vol. 21, no. 8, p. 2873, 2021. https://doi.org/10.3390/s21082873
[23] T. Chalmers *et al.*, "Stress watch: The use of heart rate and heart rate variability to detect stress: A pilot study using smart watch wearables," *Sensors*, vol. 22, no. 1, p. 151, 2021. https://doi.org/10.3390/s22010151
[24] K. Y. Lim, M. A. Nguyen Duc, M. T. Nguyen Thien, R. Yuvaraj, and J. S. Fogarty, "Investigating the effects of microclimate on physiological stress and brain function with data science and wearables," *Sustainability*, vol. 14, no. 17, p. 10769, 2022. https://doi.org/10.3390/su141710769
[25] M. A. Grandner *et al.*, "Development and initial validation of the assessment of sleep environment (ASE): Describing and quantifying the impact of subjective environmental factors on sleep," *International Journal of Environmental Research and Public Health*, vol. 19, no. 20, p. 13599, 2022. https://doi.org/10.3390/ijerph192013599
[26] S. G. Aras, J. Runyon, J. B. Kazman, J. F. Thayer, E. M. Sternberg, and P. A. Deuster, "Is greener better? Quantifying the impact of a nature walk on stress reduction using HRV and Saliva cortisol biomarkers," *International Journal of Environmental Research and Public Health*, vol. 21, no. 11, p. 1491, 2024. https://doi.org/10.3390/ijerph21111491
[27] A. Almadhor *et al.*, "Wrist-based electrodermal activity monitoring for stress detection using federated learning," *Sensors*, vol. 23, no. 8, p. 3984, 2023. https://doi.org/10.3390/s23083984
[28] K. E. Speer, S. Semple, and A. J. McKune, "Acute physiological responses following a bout of vigorous exercise in military soldiers and first responders with PTSD: An exploratory pilot study," *Behavioral Sciences*, vol. 10, no. 2, p. 59, 2020. https://doi.org/10.3390/bs10020059
[29] S. Qu and R. Ma, "Exploring multi-sensory approaches for psychological well-being in urban green spaces: Evidence from Edinburgh's diverse urban environments," *Land*, vol. 13, no. 9, p. 1536, 2024. https://doi.org/10.3390/land13091536
[30] S. Campanella, A. Altaleb, A. Belli, P. Pierleoni, and L. Palma, "A method for stress detection using empatica E4 bracelet and machine-learning techniques," *Sensors*, vol. 23, no. 7, p. 3565, 2023. https://doi.org/10.3390/s23073565
[31] A. Georgas, A. Panagiotakopoulou, G. Bitsikas, K. Vlantoni, A. Ferraro, and E. Hristoforou, "Stress monitoring in pandemic screening: Insights from GSR sensor and machine learning analysis," *Biosensors*, vol. 15, no. 1, p. 14, 2025. https://doi.org/10.3390/bios15010014

[32] M. K. Moser, M. Ehrhart, and B. Resch, "An explainable deep learning approach for stress detection in wearable sensor measurements," *Sensors*, vol. 24, no. 16, p. 5085, 2024. https://doi.org/10.3390/s24165085
[33] A. A. Al-Atawi *et al.*, "Stress monitoring using machine learning, IoT and wearable sensors," *Sensors*, vol. 23, no. 21, p. 8875, 2023. https://doi.org/10.3390/s23218875
[34] M. B. Bin Heyat *et al.*, "Wearable flexible electronics based cardiac electrode for researcher mental stress detection system using machine learning models on single lead electrocardiogram signal," *Biosensors*, vol. 12, no. 6, p. 427, 2022. https://doi.org/10.3390/bios12060427
[35] M. Abd Al-Alim, R. Mubarak, N. M. Salem, and I. Sadek, "A machine-learning approach for stress detection using wearable sensors in free-living environments," *Computers in Biology and Medicine*, vol. 179, p. 108918, 2024. https://doi.org/10.1016/j.compbiomed.2024.108918
[36] A. Nechyporenko *et al.*, "Galvanic skin response and photoplethysmography for stress recognition using machine learning and wearable sensors," *Applied Sciences*, vol. 14, no. 24, p. 11997, 2024. https://doi.org/10.3390/app142411997
[37] T. Johnson, "DigitalExposome: A dataset for wellbeing classification using environmental air quality and human physiological data," *Data in Brief*, vol. 59, p. 111442, 2025. https://doi.org/10.1016/j.dib.2025.111442
[38] Q. Chen, Y. Chi, Q. Zhu, N. Ma, L. Min, and S. Ji, "Effects of perfluorooctane sulfonic acid exposure on intestinal microbial community, lipid metabolism, and liver lesions in mice," *International Journal of Molecular Sciences*, vol. 26, no. 6, p. 2648, 2025. https://doi.org/10.3390/ijms26062648
[39] J. Zhang *et al.*, "Revealing the potential of biochar for heavy metal polluted seagrass remediation from microbial perspective," *Ecotoxicology and Environmental Safety*, vol. 292, p. 117991, 2025. https://doi.org/10.1016/j.ecoenv.2025.117991
[40] N. Tasmurzayev, B. Amangeldy, Z. Baigarayeva, M. Mansurova, B. Resnik, and G. Amirkhanova, "Improvement of HVAC system using the intelligent control system," presented at the 2022 IEEE 7th International Energy Conference (ENERGYCON), 2022.
[41] J. Brame *et al.*, "Health effects of a 12-week web-based lifestyle intervention for physically inactive and overweight or obese adults: Study protocol of two randomized controlled clinical trials," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, p. 1393, 2022. https://doi.org/10.3390/ijerph19031393
[42] L. Haarmann, E. Kalbe, G. Anapa, D. Kurt, and Ü. S. Seven, "Subjective health literacy and personality in older adults: Conscientiousness, neuroticism, and openness as key predictors—a cross-sectional study," *International Journal of Environmental Research and Public Health*, vol. 22, no. 3, p. 392, 2025. https://doi.org/10.3390/ijerph22030392
[43] G. R. Marconi *et al.*, "Assessing nutritional knowledge and physical health among football players: A pilot study from three sports clubs in Western Romania," *Sports*, vol. 13, no. 1, p. 16, 2025. https://doi.org/10.3390/sports13010016
[44] J. Simões *et al.*, "Non-Intrusive monitoring of vital signs in the lower limbs using optical sensors," *Sensors*, vol. 25, no. 2, p. 305, 2025. https://doi.org/10.3390/s25020305
[45] A. Taoum, A. Bisiaux, F. Tilquin, Y. Le Guillou, and G. Carrault, "Validity of ultra-short-term hrv analysis using ppg—a preliminary study," *Sensors*, vol. 22, no. 20, p. 7995, 2022. https://doi.org/10.3390/s22207995
[46] R. Tiwari, R. Kumar, S. Malik, T. Raj, and P. Kumar, "Analysis of heart rate variability and implication of different factors on heart rate variability," *Current Cardiology Reviews*, vol. 17, no. 5, pp. 74-83, 2021. https://doi.org/10.2174/1573403X16999201231203854
[47] T. Peters and C. Zhen, "Evaluating indoor air quality monitoring devices for healthy homes," *Buildings*, vol. 14, no. 1, p. 102, 2023. https://doi.org/10.3390/buildings14010102
[48] O. Hahad, J. Lelieveld, F. Birklein, K. Lieb, A. Daiber, and T. Münzel, "Ambient air pollution increases the risk of cerebrovascular and neuropsychiatric disorders through induction of inflammation and oxidative stress," *International Journal of Molecular Sciences*, vol. 21, no. 12, p. 4306, 2020. https://doi.org/10.3390/ijms21124306
[49] J. Beinecke and D. Heider, "Gaussian noise up-sampling is better suited than SMOTE and ADASYN for clinical decision making," *BioData Mining*, vol. 14, no. 1, p. 49, 2021. https://doi.org/10.1186/s13040-021-00283-6
[50] A. E. Bilali, A. Taleb, M. A. Bahlaoui, and Y. Brouziyne, "An integrated approach based on Gaussian noises-based data augmentation method and AdaBoost model to predict faecal coliforms in rivers with small dataset," *Journal of Hydrology*, vol. 599, p. 126510, 2021. https://doi.org/10.1016/j.jhydrol.2021.126510
[51] M. Li, H. Sun, Y. Huang, and H. Chen, "Shapley value: From cooperative game to explainable artificial intelligence," *Autonomous Intelligent Systems*, vol. 4, no. 1, p. 2, 2024. https://doi.org/10.1007/s43684-023-00060-8
[52] S. Ha and H. Jeong, "Deep learning reveals hidden interactions in complex systems," *arXiv preprint arXiv:2001.02539*, 2020. https://doi.org/10.48550/arXiv.2001.02539
[53] Y. Wang, Z. Pan, J. Zheng, L. Qian, and M. Li, "A hybrid ensemble method for pulsar candidate classification," *Astrophysics and Space Science*, vol. 364, no. 8, p. 139, 2019. https://doi.org/10.1007/s10509-019-3602-4
[54] S. Ö. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 8, pp. 6679-6687, 2021. https://doi.org/10.1609/aaai.v35i8.16826
[55] W. Xu, Y. Tian, Y. Liu, B. Zhao, Y. Liu, and X. Zhang, "Understanding the spatial-temporal patterns and influential factors on air quality index: The case of north China," *International Journal of Environmental Research and Public Health*, vol. 16, no. 16, p. 2820, 2019. https://doi.org/10.3390/ijerph16162820
[56] M. P. Sierra-Vargas, J. M. Montero-Vargas, Y. Debray-Garcia, J. C. Vizuet-de-Rueda, A. Loaeza-Roman, and L. M. Teran, "Oxidative stress and air pollution: Its impact on chronic respiratory diseases," *International Journal of Molecular Sciences*, vol. 24, no. 1, p. 853, 2023. https://doi.org/10.3390/ijms24010853
[57] M. Lane, E. Oyster, Y. Luo, and H. Wang, "The effects of air pollution on neurological diseases: A narrative review on causes and mechanisms," *Toxics*, vol. 13, no. 3, p. 207, 2025. https://doi.org/10.3390/toxics13030207
[58] S. Kalenik, A. Zaczek, and A. Rodacka, "Air pollution-induced neurotoxicity: The relationship between air pollution, epigenetic changes, and neurological disorders," *International Journal of Molecular Sciences*, vol. 26, no. 7, p. 3402, 2025. https://doi.org/10.3390/ijms26073402