International Journal of Innovative Research and Scientific Studies, 8(4) 2025, pages: 2365-2376

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Environmental factors and their influence on heart rate variability for cardiovascular disease risk classification: A machine learning approach

Zhanel Baigarayeva[1*], Assiya Boltaboyeva[1], Mergul Kozhamberdiyeva[1], Gulmira Dikhanbayeva[1], Gulshat Amirkhanova[1]

[1]*Faculty of Information Technologies, Al-Farabi Kazakh National University, Almaty 050040, Kazakhstan.*

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

Abstract

Air pollution is a critical global health issue, affecting nearly the entire human population and contributing to millions of premature deaths annually. One pathway through which pollution impacts human health is via the autonomic nervous system, measurable through HRV, a sensitive marker of cardiovascular regulation. Prior studies have demonstrated that fine PM2.5, CO_2, and VOCs are associated with HRV reduction, but most rely on large cohort data or lack fine temporal resolution. Here, we present a real-time, multimodal sensing platform that combines wearable physiological monitoring with environmental air quality sensors to assess short-term HRV responses under varying ambient conditions. Using our in-house Zhurek IoT device, we synchronized physiological data with environmental parameters across three contrasting urban and natural settings. Machine learning models, especially XGBoost, accurately classified levels of HRV change (low, moderate, high) based on air quality metrics with up to 86.71% accuracy. This confirms that pollutant levels can predict subtle changes in autonomic function even in healthy young adults. These findings extend prior knowledge by demonstrating that short-term fluctuations in air quality can measurably affect HRV, even in the absence of chronic illness. In contrast to earlier research focusing on long-term exposure or clinical populations, our study highlights the vulnerability of healthy individuals to environmental stressors and shows how machine learning enhances the detection of such effects. The results underscore the utility of combining wearable technology and artificial intelligence in environmental health monitoring. They provide a foundation for personalized risk assessment and targeted public health interventions, especially in urban areas with fluctuating pollution levels.

Keywords: Air pollution, Air quality, Heart rate variability, Internet of Things, Machine learning, Wearable sensors.

DOI: 10.53894/ijirss.v8i4.8391
Funding: This research is supported by the Committee of Science of the Republic of Kazakhstan (Grant number: AP23488586).
History: Received: 26 May 2025 / **Revised:** 30 June 2025 / **Accepted:** 2 July 2025 / **Published:** 8 July 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

In 2019, approximately 99% of the world's population resided in areas where air pollution levels surpassed the recommendations of the World Health Organization (WHO), directly affecting public health. Both outdoor and indoor air pollution contribute to approximately 6.7 million premature deaths worldwide each year, with about 4.2 million of these deaths solely attributable to ambient air pollution [1]. HRV serves as a sensitive marker of autonomic control of the cardiovascular system, quantitatively reflecting the balance between the sympathetic and parasympathetic nervous systems. A decrease in HRV parameters is associated with an increased risk of arrhythmias, myocardial infarction, and overall cardiovascular mortality. Measurements of HRV in time and frequency domains show that exposure to environmental pollutants reduces these parameters, indicating autonomic dysfunction [2].

Air pollution is a complex mixture of particulate matter of various fractions, including coarse, fine (PM2.5), and ultrafine particles (UFP), as well as gaseous pollutants such as ozone (O_3) and carbon monoxide (CO). Special attention is given to PM2.5 and UFP, which are capable of penetrating deep into the respiratory tract and systemic circulation, causing inflammatory and oxidative reactions that impair cardiovascular function [3]. Inhalation of fine particles induces oxidative stress in the lungs and local inflammation, which may lead to systemic inflammatory responses. This cascade affects autonomic regulation of the heart by disrupting the balance of sympathetic and parasympathetic activity. Exposure to PM2.5 is linked to increased regulation of reactive oxygen species in the myocardium and impaired ion channel function, contributing to reduced HRV [4, 5]. Experimental studies confirm that acute PM2.5 exposure causes a significant decrease in HRV parameters in both time and frequency domains, indicating a key role of oxidative stress in these processes [6].

Meta-analyses demonstrate a statistically significant decrease in HRV parameters with each increase in PM2.5 concentration, establishing a clear dose-dependent relationship between air pollution and autonomic dysfunction [7]. Epidemiological data are supported by individual studies using wearable monitoring devices that assess the direct impact of pollutants on HRV [8, 9]. Experimental animal models provide opportunities for controlled studies of exposure mechanisms. Studies on rodents show that short-term exposure to particulate matter by inhalation or instillation leads to decreased HRV, accompanied by increased systemic inflammation and oxidative stress, thereby enhancing the biological impact of pollutants on autonomic cardiac regulation [6].

The effect of air pollution on HRV varies depending on demographic and clinical factors. Vulnerable groups, including elderly individuals, patients with cardiovascular diseases, metabolic syndrome, and ischemic heart disease, exhibit more pronounced HRV reductions at comparable pollutant exposure levels. Epidemiological studies in elderly cohorts demonstrate significant HRV decreases even with slight increases in PM2.5 concentrations, indicating increased susceptibility due to age-related changes in autonomic regulation [7, 10]. Obesity acts as an important modifier of pollution effects, with individuals having excess weight exhibiting a more pronounced decline in HRV upon exposure to particulate matter [9]. The influence of air pollution on HRV significantly depends on environmental conditions and daily activities. Particular attention is paid to commuting conditions in urban transport, where elevated concentrations of particles and CO cause immediate HRV changes with variable effects depending on the type of transport and pollutant profile [11]. Exposure to indoor pollution sources, such as cooking and incense burning, is associated with significant reductions in HRV, emphasizing the importance of indoor air quality for cardiovascular risk [8, 9]. Psychosocial stress is an additional factor interacting with air pollution, exacerbating HRV reduction. Studies show that the combination of psychological stress and ozone exposure leads to a more pronounced HRV decrease than each factor alone [12]. This indicates that environmental and behavioral interventions aimed at reducing air pollution and stress may have a synergistic effect on improving cardiovascular regulation.

The use of air purification devices, such as GENANO filters, demonstrates significant improvements in HRV parameters in elderly individuals, providing direct evidence that improving indoor air quality can bring measurable benefits to the cardiovascular system [13]. Longitudinal studies with extended observation periods will be important to assess whether improvements in air quality lead to sustained increases in HRV and reductions in cardiovascular morbidity and mortality [14, 15]. Physiological responses to air pollution occur rapidly: even short-term exposure can cause measurable HRV changes within minutes or hours [10, 14]. Studies conducted under real exposure conditions in public transport cabins emphasize the importance of contextual factors, as the combination of source and exposure environment plays a decisive role in the cardiovascular effects of air pollution. Elevated concentrations of particles and carbon monoxide in enclosed cabin spaces are associated with significant HRV changes [11]. Heart rate variability is an indicator reflecting changes in the intervals between successive heartbeats, widely used to assess the state of the autonomic nervous system and cardiovascular regulation. High HRV values indicate balanced activity of the sympathetic and parasympathetic systems, whereas decreased HRV indicates dysfunction and is linked to an increased risk of chronic diseases and mortality [16].

The main pollutants affecting HRV include PM, O_3, CO, NOx, polycyclic aromatic hydrocarbons (PAHs), 1,2-naphthoquinone (1,2-NQ), CO_2, and volatile organic compounds (VOCs) [17, 18]. Most studies report a negative association between concentrations of PM, fine (PM2.5), and ultrafine particles (UFP), and HRV parameters [19, 20]. Exposure to these particles is linked to decreased overall HRV power and frequency domain parameters (LF, HF), although compensatory HRV increases occur in some cases with certain fractions (PM1–2.5), possibly reflecting different mechanisms of action [20, 21]. Ozone exerts a suppressive effect on HRV, reducing time-domain and frequency-domain measures, as confirmed by meta-analyses, indicating an increased risk of cardiovascular events, especially in vulnerable groups [22]. Exposure to carbon monoxide at concentrations above 2.7 ppm also leads to HRV reduction, although effects at lower levels are less clear [23, 24]. Polycyclic aromatic hydrocarbons are associated with increased heart rate and HRV changes, as studied using personal exposimeters in industrial settings [25]. 1,2-Naphthoquinone, a secondary aerosol component, causes HRV disturbances in animals by activating the sympathetic nervous system and increasing the risk of arrhythmia [26]. Nitrogen oxides affect autonomic regulation, as confirmed by portable sensor measurements in daily life conditions [27, 28]. Indoor carbon dioxide concentrations may also influence HRV under poor ventilation conditions [16, 29]. Volatile organic compounds have been studied in office and residential environments, showing possible effects on autonomic cardiac control [28]. Air pollutants cause both decreases and increases in HRV depending on the mechanism of action. HRV reduction is often related to sympathetic activation in response to stress, inflammation, or oxidative stress, while activation of airway receptors can enhance vagal activity, manifesting as HRV increases [30].

In this study, a comprehensive field experiment was conducted to investigate how outdoor and indoor environmental factors affect short-term HRV in healthy young adults across urban, natural, and controlled settings. The primary goal was to determine the extent to which fluctuations in air quality parameters such as PM2.5, PM10, and CO_2 can influence autonomic nervous system activity, as reflected through HRV measurements. Due to the complexity of capturing synchronized environmental and physiological data in real-life conditions, a custom IoT-based solution, named Zhurek, was developed. This device allowed for real-time acquisition of HRV indicators and was complemented by an external environmental monitoring system. The synchronized data streams were analyzed using machine learning models, including Decision Tree, Random Forest, DNN, and XGBoost classifiers, to evaluate their ability to predict HRV changes based on pollutant levels. Among these, XGBoost demonstrated the highest accuracy (86.71%) in classifying HRV change into low, moderate, and high categories. This integrated approach enabled the identification of the most critical environmental features influencing HRV and highlighted the immediate physiological impacts of air pollution, even in healthy individuals. The findings provide new evidence that air quality fluctuations often occurring within minutes can affect cardiovascular regulation, emphasizing the need for contextual monitoring beyond traditional long-term studies. By combining wearable health technology, environmental sensors, and data-driven modeling, this research offers valuable insights for public health strategies, smart city planning, and personalized exposure risk assessment in urban environments.

2. Methodology

2.1. System Architecture

The system architecture is designed to enable synchronized acquisition, transmission, and analysis of physiological and environmental data. It consists of three functional layers: the sensing layer, the communication layer, and the processing and analytics layer. The complete structure is illustrated in Figure 1.

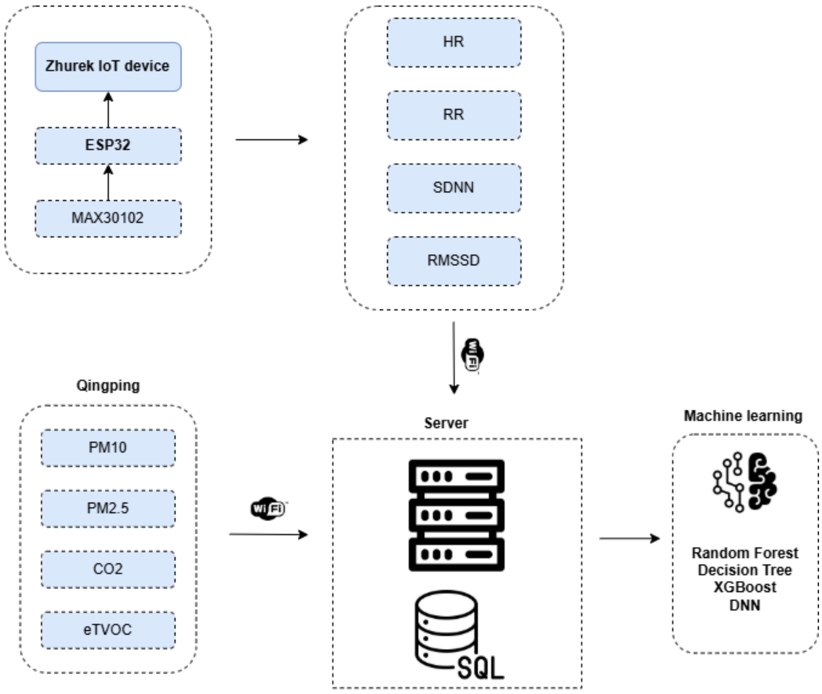

Figure 1.
The architecture of the system.

The sensing layer comprises two parallel subsystems: physiological signal acquisition and environmental monitoring. For physiological measurements, the Zhurek wearable device was developed in-house using an ESP32-WROOM-32 microcontroller and a DFRobot MAX30102 photoplethysmographic sensor, connected via an I²C bus. The MAX30102 captures dual-wavelength optical signals from the fingertip at a sampling rate of 100 Hz. These signals are filtered using a moving average and high-pass digital filter to remove baseline drift and noise. Peak detection is applied to extract inter-beat intervals (IBIs), from which the device computes heart rate (HR), the standard deviation of NN intervals (SDNN), and the root mean square of successive differences (RMSSD). All HRV features are computed directly on the ESP32 using an embedded C++ implementation optimized for real-time execution under memory constraints.

Electrocardiographic (ECG) signals were concurrently recorded using the Polar H10 chest strap to serve as a reference standard. The data were transmitted via Bluetooth Low Energy (BLE) to a mobile device running the HRV Logger application, which computed benchmark HRV metrics. These metrics were used to validate the accuracy of the metrics derived from the Zhurek device.

Environmental data are collected using the Qingping CGS1 smart air quality monitor, a commercially available, fully integrated sensing device. It measures PM2.5, PM10, carbon dioxide (CO_2), total volatile organic compounds (TVOC), temperature, and humidity, and computes an onboard air quality index (AQI). The device transmits environmental readings in real time via Wi-Fi using the MQTT protocol.

Physiological and environmental data are both published to a Mosquitto MQTT broker hosted on a secure laboratory server. Data packets are serialized in JSON format and include UTC timestamps, assigned by each device's real-time clock, which is synchronized using the Network Time Protocol (NTP) before each measurement session. This configuration ensures sub-second temporal alignment between physiological and environmental streams.

The processing layer consists of a central server running a structured query language (SQL) database for persistent data storage and indexing. Data ingestion scripts subscribe to MQTT topics, parse incoming telemetry, and write structured entries to the database. After collection, the synchronized physiological and environmental data are used to train and evaluate a series of supervised machine learning models, including deep neural networks (DNN), extreme gradient boosting (XGBoost), random forests, and TabNet.

This modular architecture supports reliable real-time data acquisition and tightly synchronized multimodal measurements, enabling accurate assessment of short-term autonomic responses to environmental air quality fluctuations. By computing HRV metrics locally and aligning them with high-resolution environmental data streams, the system facilitates robust modeling of stress-related physiological effects under varying ambient conditions.

2.2. Experiment Description

The objective was to investigate the effect of air quality on heart rate variability (HRV) in healthy young adults. Measurements were conducted across three contrasting environments: Al-Farabi Avenue (a densely trafficked urban location), the Botanical Garden (a clean, low-pollution green area), and a laboratory with controlled environmental conditions. These locations were selected to represent varying levels of air pollution and to assess corresponding physiological responses under both real-world and controlled settings.

Ten healthy volunteers aged between 19 and 22 participated in the experiment. All participants received detailed instructions and provided written informed consent. To minimize potential confounding influences on HRV, participants were required to abstain from consuming caffeine, alcohol, and tobacco for at least 12 hours prior to each measurement session.

Each subject completed three measurement sessions, one in each location, with each session lasting 10 minutes. Physiological signals were continuously recorded using the Zhurek IoT device, a wearable system developed for real-time monitoring of cardiac function. The device measured heart rate (HR), RR intervals, and standard HRV metrics, including SDNN (standard deviation of RR intervals) and RMSSD (root mean square of successive differences between RR intervals). All data were transmitted to a central gateway and securely stored in a remote database for further analysis.

Air quality parameters were measured in parallel every 5 minutes, including concentrations of fine particulate matter (PM2.5), carbon dioxide (CO_2), volatile organic compounds (VOCs), ambient temperature, and relative humidity. Identical measurement intervals and standardized sensor placement protocols were applied across all environments to ensure consistency. The collected data enabled the analysis of potential associations between environmental conditions and HRV indicators, providing insights into how air quality may influence the autonomic regulation of cardiovascular function in young, healthy individuals.

Table 1.
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

The data acquisition protocol adopted in this study utilized a dual-channel measurement configuration, enabling simultaneous recording of physiological signals and environmental parameters to assess short-term autonomic nervous system responses to changes in air quality. Physiological data were obtained using a custom-built portable Internet of Things (IoT) device referred to as Zhurek (Figure 2), which incorporates an ESP32 microcontroller interfaced with a MAX30102 photoplethysmographic (PPG) sensor via the I²C bus. The components are housed within a compact enclosure fabricated using 3D printing, providing mechanical protection and facilitating deployment in field or laboratory settings. The MAX30102 sensor continuously acquires PPG waveforms, which are processed locally on the ESP32 using embedded firmware written in C++. The onboard signal processing pipeline includes digital bandpass filtering, motion artifact suppression, and peak detection, enabling reliable extraction of RR intervals and computation of time-domain heart rate variability (HRV) metrics, such as heart rate (HR), the standard deviation of NN intervals (SDNN), and the root mean square of successive differences (RMSSD). All computed metrics are timestamped in UTC and transmitted wirelessly to a remote research server via MQTT over a 2.4 GHz Wi-Fi connection. In parallel, ambient environmental parameters were continuously recorded using the Qingping CGS1 (Xiaomi Air Detector, 9-in-1), a multi-sensor instrument capable of measuring concentrations of PM2.5, PM10, CO_2, formaldehyde (CH_2O), total volatile organic compounds (TVOC), as well as temperature, relative humidity, and barometric pressure. The device also calculates the Air Quality Index (AQI) based on onboard algorithms. Telemetry from the CGS1 was transmitted in JSON format to the same MQTT broker, with routing parameters pre-configured via the manufacturer's cloud interface (developer.qingping.co). A dedicated Python script, subscribed to the topic qingping/#, received and parsed all environmental data and stored the results in a PostgreSQL relational database indexed by timestamp, device ID, and spatial context. To ensure temporal alignment between data sources, both the Zhurek and CGS1 devices were synchronized using the Network Time Protocol (NTP) before each session. The experimental protocol involved ten healthy adult participants (n = 10), each of whom completed three 10-minute recording sessions under contrasting environmental conditions: a heavily trafficked urban roadway (Al-Farabi Avenue), a humid green space with dense vegetation (Botanical Garden), and a controlled indoor laboratory environment. All participants were instructed to abstain from caffeine, alcohol, nicotine, and strenuous activity for at least 12 hours prior to each session in order to minimize confounding influences. Data quality assurance procedures included artifact rejection, validation of physiological plausibility, and removal of incomplete or noisy sessions. Temporal data fusion based on synchronized timestamps enabled accurate integration of physiological and environmental parameters, providing a coherent dataset suitable for statistical analysis and predictive modeling.

Figure 2.
Zhurek IoT device.

The lower section of the 3D-printed enclosure houses the ESP32 microcontroller responsible for local signal processing and data transmission. The upper section contains the MAX30102 PPG sensor connected via I²C. The design allows for modular deployment in both laboratory and field conditions.

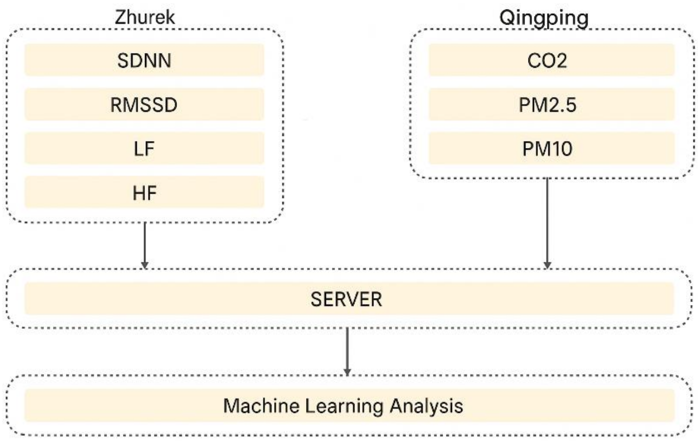

Figure 3.
Data Flow for Physiological and Environmental Monitoring.

The presented Figure 3 architecture of an integrated research platform is designed to investigate the impact of air quality parameters on heart rate variability (HRV). The Zhurek device, based on an ESP32 microcontroller and a MAX30102 optical sensor, performs continuous acquisition of physiological signals and extraction of key HRV metrics, including SDNN, RMSSD, LF, and HF. In parallel, the Qingping module monitors environmental indicators such as carbon dioxide (CO_2), particulate matter (PM2.5 and PM10), providing real-time measurements. All data are transmitted to a central server, where they are temporally synchronized and aggregated. The processed data are then analyzed using machine learning algorithms aimed at identifying statistically significant relationships between environmental pollution and alterations in autonomic cardiac regulation. This architecture enables quantitative analysis of the interplay between environmental exposures and physiological responses, which is essential for advancing research in environmental epidemiology and cardiovascular risk assessment.

2.4. Data Preprocessing

Heart rate fluctuations were assessed by calculating a moving average of the pulse using a window of three measurements for each subject individually. Following this, the absolute difference between consecutive moving average pulse values was computed to capture the dynamics of heart rate variability.

Thresholds for classifying the extent of pulse changes were then established based on the 33rd and 66th percentiles of the absolute change distribution across the full dataset. The moving average pulse changes were categorized into three levels: low change (below the 33rd percentile), moderate change (between the 33rd and 66th percentiles), and high change (above the 66th percentile). This approach enabled the identification of variability levels and facilitated the study of their association with environmental factors.

Gaussian noise enhances the generalization ability of the model by introducing natural variations during data augmentation and ensures its robustness by accounting for noise. This method was chosen due to its proven effectiveness in improving the performance of machine learning systems, especially with small datasets in clinical and environmental applications. During data augmentation, Gaussian noise creates diverse variants of the original data by adding random noise with zero mean and a known variance. This process increases data variability and helps the model become resilient to small perturbations while accounting for noise. As a result, the model better adapts to real-world variations and noise, improving training quality.

Synthetic data generation with Gaussian noise is performed according to the following formula 1:

$$x=\mu+\sigma\cdot z \tag{1}$$

where x is the data value after adding noise, μ is the mean (in this case, 0), σ is the standard deviation derived from the real data distribution, and z is a random variable drawn from the standard normal distribution: $Z_N(0,1)$. This approach allows controlled random variations to be introduced while preserving the natural variability of the original data.

A thorough data inspection was conducted to ensure data integrity. Additionally, standardization was applied to balance variables with different scales. This process transforms each variable's values to a scale with a mean of 0 and a standard deviation of 1. Standardization improves training efficiency by eliminating disproportion among parameters, thereby enhancing prediction accuracy.

The applied preprocessing methods and synthetic data generation using Gaussian noise significantly improved the quality of the main dataset, enhancing the overall performance and predictive capability of the models.

2.5. Machine Learning Classification

For the classification of heart rate variability based on environmental parameters, Random Forest, Decision Tree, XGBoost, and DNN models were employed. These algorithms were selected for the three-class classification task due to their ability to effectively capture complex dependencies, handle nonlinear relationships, and work with structured data.

Random Forest and Decision Tree utilize ensemble methods and tree-based structures to improve model accuracy and interpretability. XGBoost employs gradient boosting, which provides high performance and robustness against overfitting. DNN enables the discovery of hidden patterns and complex dynamic dependencies directly from the data without the need for explicitly defined rules.

Data preprocessing included normalization of features using StandardScaler from the Scikit-learn library, ensuring a uniform scale of input variables. The dataset contained only environmental parameters (CO_2, PM2.5, PM10), and the target variable represented classes of heart rate variability changes: low, moderate, and high.

Classification models were developed using Scikit-learn for Random Forest, Decision Tree, and XGBoost, as well as PyTorch for DNN. Hyperparameter tuning was performed using GridSearch, allowing for parameter optimization and improved classification performance.

Model evaluation was conducted using classification accuracy, precision, recall, and F1-score metrics to comprehensively assess model performance. Stratified k-fold cross-validation was employed to ensure robust and reliable validation results, which is particularly important for addressing potential class imbalance.

The developed models demonstrate promising potential for real-time monitoring of physiological responses influenced by environmental factors, which could be applied in health risk assessments and personalized environmental interventions.

3. Results

3.1. Environmental Measurements Across Experimental Settings

Environmental conditions were recorded in real-time at all three experimental locations. Measurements included temperature, humidity, and levels of air pollutants. The mean concentrations of CO_2, PM2.5, and PM10 measured at Al-Farabi Avenue, the Botanical Garden, and the Laboratory were documented, as shown in Table 2. The botanical garden showed the highest levels of PM2.5 (57.5 μg/m³) and PM10 (46.5 μg/m³), while the laboratory recorded the lowest concentrations of both particulate matter. In contrast, CO_2 levels were substantially higher in the laboratory (1478.975 ppm) compared to the other two locations. Al-Farabi Avenue exhibited intermediate values for all pollutants. These measurements indicate distinct air quality profiles across the three experimental environments. Humidity levels varied notably across the three study environments, as indicated in Table 3. The Botanical Garden recorded the highest mean relative humidity at 83.54 ± 0.97%, followed by Al-Farabi Avenue with 48.33 ± 5.08%. The laboratory exhibited the lowest average humidity, measured at 32.61 ± 4.11%. These differences in atmospheric moisture highlight distinct environmental conditions present at each location.

Table 2.
Environmental pollutant levels were recorded at the three sites.

Location	**Mean CO2, ppm**	**Mean PM2.5, μg/m³**	**Mean PM10, 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

Table 3.
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

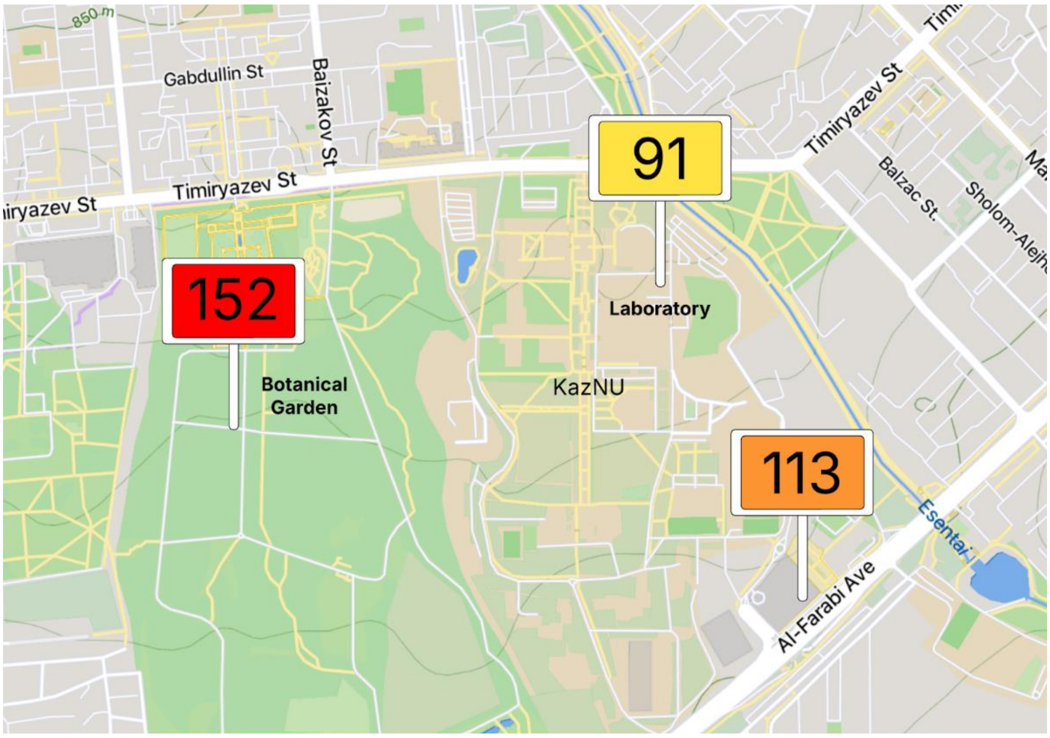

Figure 4.
AQI measurements were recorded at each of the three sites during the experiment, based on data obtained from environmental sensors.

The map in Figure 4 displays the Air Quality Index (AQI) levels at the three experimental locations, which range from 0 to 500 [31]. Lower AQI values correspond to good air quality with little to no health risk, while higher values indicate air quality that ranges from unhealthy to hazardous. The highest AQI was recorded at the Botanical Garden (152), indicating unhealthy air quality. Al-Farabi Avenue showed a moderate AQI value of 113, while the Laboratory had the lowest AQI reading of 91, suggesting better air conditions relative to the other two locations. These values were derived from real-time sensor data collected during the study.

3.2. Data Augmentation

To overcome the limitations of a small sample size, Gaussian noise-based data augmentation was applied to the original dataset, increasing the number of data points from 30 to 1500. Figure 5. The text discusses the distribution of original and augmented data for three key environmental features: CO_2, PM_{10}, and $PM_{2.5}$. The generated samples closely resemble the original data in terms of density, shape and central tendency, indicating that the Gaussian noise was appropriately calibrated. For CO_2, the augmented distribution replicates the primary peak of the original data with slightly increased sharpness, suggesting a higher sample density near the mean. The distributions of PM_{10} and $PM_{2.5}$ also align well with their original counterparts, preserving the overall form and range while introducing minor variability that enhances model robustness. These findings confirm that Gaussian noise injection effectively expanded the dataset while maintaining key distributional properties.

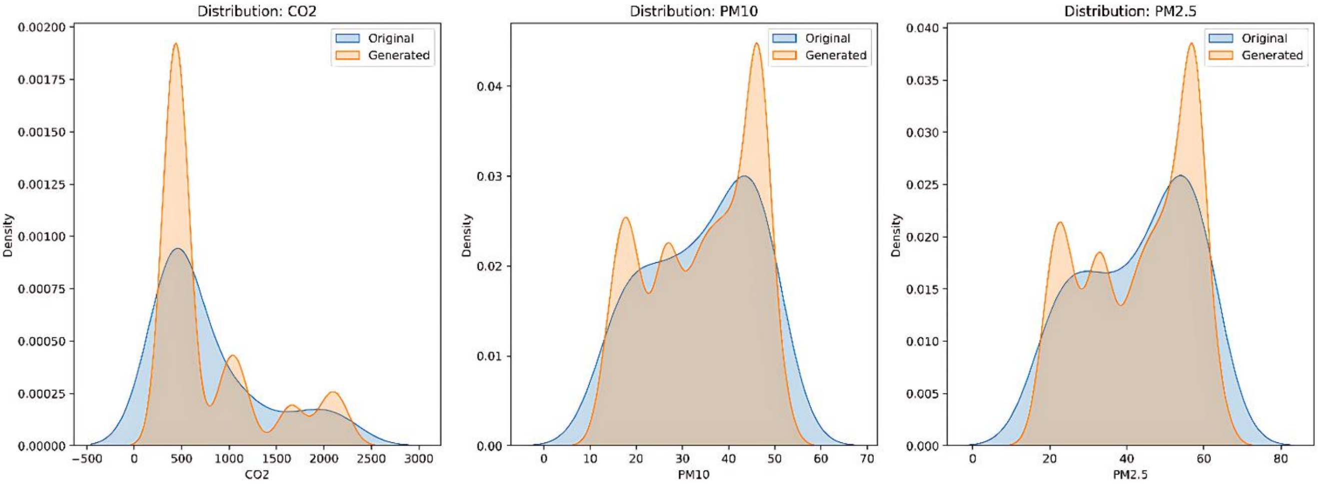

Figure 5.
Kernel density estimation (KDE) comparing original and augmented data for three environmental features.

3.3. Results of Machine Learning Classification

To assess how well air pollution indicators can predict changes in heart rate variability (HRV), several machine learning models were trained using CO_2, PM_{10}, and $PM_{2.5}$ levels as input features. The target was the degree of change in HRV, categorized into three classes: low (bottom 33rd percentile), moderate (34th to 66th percentile), and high (top 67th percentile) HRV change. The performance of four classifiers was evaluated, including DNN, RF, XGBoost, and DT. Table 4 presents the results across four key evaluation metrics: accuracy, precision, recall, and F1-score.

Among the tested models, XGBoost achieved the best overall performance. It recorded an accuracy of 86.71 percent, a precision of 86.86%, a recall of 86.96%, and an F1-score of 86.74%. These values indicate a strong ability to capture the underlying associations between pollutant levels and physiological responses. Random Forest demonstrated competitive performance, with an accuracy of 85.13% and closely matched values in precision and recall. The DT and DNN models yielded slightly lower performance, although they still maintained overall accuracy above 84%.

As shown in Figure 6, XGBoost classified most samples correctly across all three classes. For the low change category (class 0), the model correctly predicted 122 out of 134 instances, with 12 samples misclassified as moderate and none as high. In the case of moderate change (class 1), 123 out of 157 samples were correctly classified. The majority of misclassifications in this category occurred with the low class, where 30 samples were incorrectly labeled, and only 4 were misclassified as high. For the high change category (class 2), 140 out of 153 instances were accurately classified, while 13 were misclassified as moderate and none as low.

Table 4.
Evaluation Metrics for Machine Learning Models.

Evaluation metrics	**DNN**	**RF**	**XGBoost**	**DT**
Accuracy	84.23%	85.13%	86.71%	84.68%
Precision	86.12%	86.29%	86.86%	85.28%
Recall	84.73%	85.80%	86.96%	85.20%
F1-Score	84.55%	85.06%	86.74%	84.56%

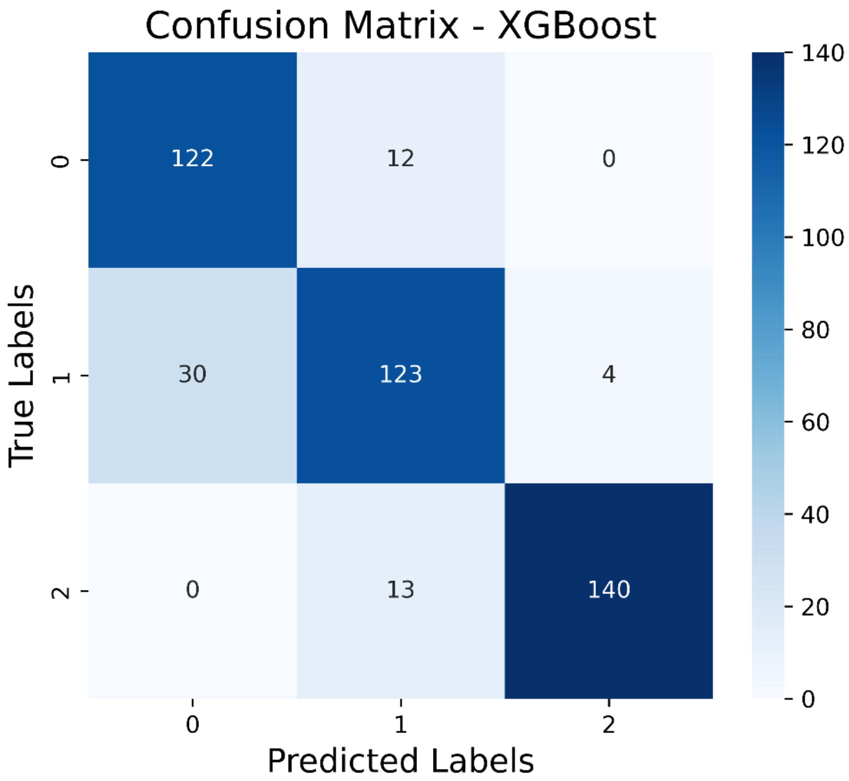

Figure 6.
XGBoost Confusion Matrix Based on Air Quality Parameters.

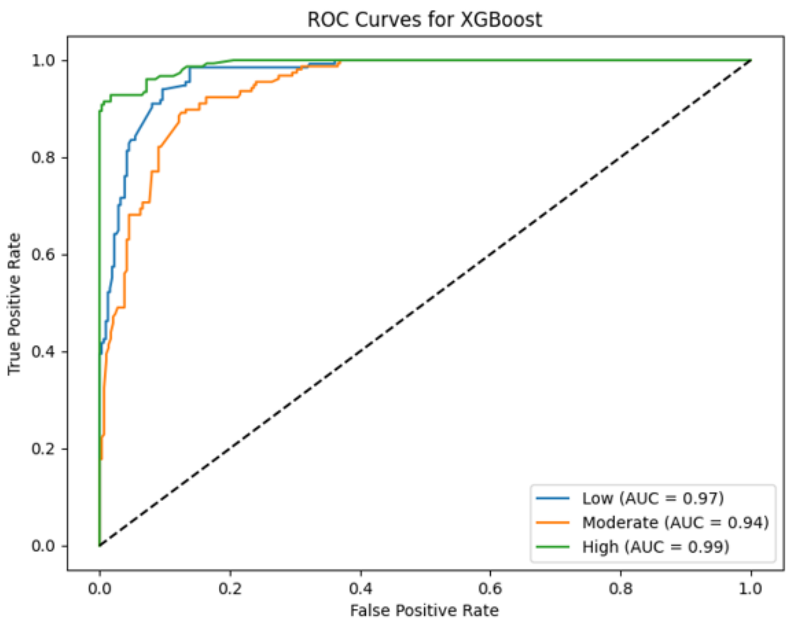

Figure 7.
Receiver Operating Characteristic (ROC) Curve for Evaluating Model Accuracy.

The ROC curves for the XGBoost model demonstrate strong discriminative performance across all three HR change categories, as indicated in Figure 7. The Area Under the Curve (AUC) values are high for each class, indicating a high level of classification reliability.

The high change class achieved the best result with an AUC of 0.99, suggesting that the model can almost perfectly distinguish high HRV instances from the others. The low change class followed closely with an AUC of 0.97, while the moderate change class yielded a slightly lower AUC of 0.94. Although the moderate class shows marginally reduced separability, its AUC still reflects a high level of predictive accuracy. These curves confirm the model's robustness and its ability to maintain a favorable balance between true positive and false positive rates across all target categories.

4. Discussion

The increased AQI recorded at the botanical garden (Figure 4) is mainly due to the elevated humidity levels noted in Table 3. High humidity, often a result of dense plant life causing transpiration and limited air circulation, leads to greater moisture retention in the area [32, 33]. This added moisture causes fine particulate matter ($PM_{2.5}$ and PM_{10}) to absorb water, increasing their weight and concentration. As a result, the botanical garden exhibited higher levels of these particles, which negatively impacted air quality. These specific environmental factors created an important setting to examine physiological effects, such as variations in HRV, in response to air pollution.

Previous research Park et al. [34] and Li et al. [13] has consistently shown that air pollutants such as PM2.5 and ozone negatively impact HRV. HRV is a key indicator of autonomic nervous system function and cardiovascular health. Decreases in HRV have been observed following exposure to air pollution. These associations tend to be stronger in individuals with pre-existing conditions like hypertension, diabetes, or ischemic heart disease. Our study found that even in healthy individuals, environmental factors can influence HRV. This highlights that pollution can influence autonomic function regardless of baseline health status. These findings are consistent with earlier research showing pollution-related autonomic changes. The alignment strengthens the credibility of our results. Machine learning helped uncover these patterns with greater precision and clarity. In our study, HRV was measured over a 5-minute window to capture short-term autonomic responses. Short-term HRV analysis is recognized as a convenient and effective method for estimating autonomic status and detecting dynamic changes in cardiac function within minutes. In contrast, long-term HRV is more suitable for assessing trends over hours and predicting prognosis [35]. Since our goal was to investigate the immediate effects of air pollution on autonomic function, a short-term approach was most appropriate. Selecting the correct time window is crucial for spectral HRV analysis, and our choice aligns with best practices for studying rapid physiological responses.

The research was subject to some limitations. Wind speed and vegetation type in the botanical garden were not measured or included in the analysis, which may have influenced the recorded environmental conditions.

5. Conclusion

This study successfully demonstrated the feasibility and effectiveness of using an integrated IoT-based platform to monitor the short-term effects of ambient air quality on HRV in healthy young individuals. By conducting field experiments in three different environments, a polluted urban street, a relatively clean green zone, and a controlled indoor laboratory, we were able to assess real-time changes in HRV in response to fluctuations in pollutant concentrations, including PM2.5, PM10, and CO_2. Experimental data showed that the highest concentrations of PM2.5 - 57.5 µg/m³ and PM10 - 46.5 µg/m³ were recorded in the Botanical Garden, likely due to high humidity and limited air circulation, while CO_2 levels peaked in the indoor laboratory environment (1478.98 ppm). These environmental conditions were found to significantly affect HRV, which was assessed using SDNN and RMSSD metrics calculated by the custom-developed wearable device Zhurek.

To predict HRV changes based on environmental data, several machine learning models were applied. Among them, XGBoost outperformed the others, achieving an accuracy of 86.71%, a precision of 86.86%, a recall of 86.96%, and an F1-score of 86.74%. The model was particularly effective in classifying high HRV change states, with an AUC of 0.99, demonstrating its robustness in identifying strong physiological responses to air pollution. The Random Forest model also showed strong performance with an accuracy of 85.13%, while Decision Tree and DNN models achieved slightly lower, yet still competitive, results. The study highlights that even short-term exposure to elevated air pollutants can lead to measurable changes in autonomic cardiac regulation, as evidenced by significant shifts in HRV. This study lays the foundation for the future development of intelligent health monitoring systems aimed at the early detection of cardiovascular diseases. By continuously assessing physiological responses to environmental conditions, such systems will be able to identify subtle signs of autonomic nervous system imbalance that may precede the development of cardiovascular disorders. Integrating such solutions into everyday urban environments will enhance the accuracy of personalized risk assessment and contribute to early-stage disease prevention, ultimately improving the cardiovascular health of the population.

References

[1] World Health Organization, "Ambient (outdoor) air quality and health," 2024. Retrieved: https://www.who.int/ru/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health, 2024.

[2] Y. W. Jeong, H. M. Choi, Y. Park, Y. Lee, J. Y. Jung, and D. R. Kang, "Association between exposure to particulate matter and heart rate variability in vulnerable and susceptible individuals," *npj Digital Medicine,* vol. 8, no. 1, p. 52, 2025. https://doi.org/10.1038/s41746-024-01373-5

[3] S. Breitner *et al.*, "Ambient and controlled exposures to particulate air pollution and acute changes in heart rate variability and repolarization," *Scientific Reports,* vol. 9, no. 1, p. 1946, 2019. https://doi.org/10.1038/s41598-019-38531-9

[4] A. Fiordelisi, P. Piscitelli, B. Trimarco, E. Coscioni, G. Iaccarino, and D. Sorriento, "The mechanisms of air pollution and particulate matter in cardiovascular diseases," *Heart Failure Reviews,* vol. 22, no. 3, pp. 337-347, 2017. https://doi.org/10.1007/s10741-017-9606-7

[5] J. Hu *et al.*, "Fine particulate matter constituents and heart rate variability: A panel study in Shanghai, China," *Science of The Total Environment,* vol. 747, p. 141199, 2020. https://doi.org/10.1016/j.scitotenv.2020.141199

[6] F. Huang, P. Wang, X. Pan, Y. Wang, and S. Ren, "Effects of short-term exposure to particulate matters on heart rate variability: A systematic review and meta-analysis based on controlled animal studies," *Environmental Pollution,* vol. 256, p. 113306, 2020. https://doi.org/10.1016/j.envpol.2019.113306

[7] F. Wang *et al.*, "The relationship between exposure to PM2.5 and heart rate variability in older adults: A systematic review and meta-analysis," *Chemosphere,* vol. 261, p. 127635, 2020. https://doi.org/10.1016/j.chemosphere.2020.127635

[8] D.-H. Lee *et al.*, "Personal exposure to fine particulate air pollutants impacts blood pressure and heart rate variability," *Scientific Reports,* vol. 10, no. 1, p. 16538, 2020. https://doi.org/10.1038/s41598-020-73205-x

[9] M.-C. M. Tsou *et al.*, "A community-based study on associations between PM2.5 and PM1 exposure and heart rate variability using wearable low-cost sensing devices," *Environmental Pollution,* vol. 277, p. 116761, 2021. https://doi.org/10.1016/j.envpol.2021.116761

[10] Z. Niu *et al.*, "Acute effect of ambient fine particulate matter on heart rate variability: An updated systematic review and meta-analysis of panel studies," *Environmental Health and Preventive Medicine,* vol. 25, no. 1, p. 77, 2020. https://doi.org/10.1186/s12199-020-00912-2

[11] C.-S. Tang *et al.*, "Impacts of in-cabin exposure to size-fractionated particulate matters and carbon monoxide on changes in heart rate variability for healthy public transit commuters," *Atmosphere,* vol. 10, no. 7, p. 409, 2019. https://doi.org/10.3390/atmos10070409

[12] E. Tirosh and I. Schnell, "The relationship between ambient carbon monoxide and heart rate variability—a systematic world review—2015," *Environmental Science and Pollution Research,* vol. 23, no. 21, pp. 21157-21164, 2016. https://doi.org/10.1007/s11356-016-7533-0

[13] L. Li *et al.*, "Effect of short-term exposure to particulate air pollution on heart rate variability in normal-weight and obese adults," *Environmental Health,* vol. 20, no. 1, p. 29, 2021. https://doi.org/10.1186/s12940-021-00707-0

[14] M. Riediker, Y. Franc, M. Bochud, R. Meier, and V. Rousson, "Exposure to fine particulate matter leads to rapid heart rate variability changes," *Frontiers in Environmental Science,* vol. 6, 2018. https://doi.org/10.3389/fenvs.2018.00002

[15] X. Jia *et al.*, "Short-term effects of particulate matter in metro cabin on heart rate variability in young healthy adults: Impacts of particle size and source," *Environmental Research,* vol. 167, pp. 292-298, 2018. https://doi.org/10.1016/j.envres.2018.07.017

[16] T. Nie, G. Zhang, Y. Sun, W. Wang, T. Wang, and H. Duan, "Effects of indoor air quality on human physiological impact: A review," *Buildings,* vol. 15, no. 8, p. 1296, 2025. https://doi.org/10.3390/buildings15081296

[17] Z. Zong, M. Zhang, K. Xu, Y. Zhang, and C. Hu, "Association between short-term exposure to ozone and heart rate variability: A systematic review and meta-analysis," *International Journal of Environmental Research and Public Health,* vol. 19, no. 18, p. 11186, 2022. https://doi.org/10.3390/ijerph191811186

[18] M. Meina *et al.*, "Heart rate variability and accelerometry as classification tools for monitoring perceived stress levels—a pilot study on firefighters," *Sensors,* vol. 20, no. 10, p. 2834, 2020. https://doi.org/10.3390/s20102834

[19] C.-S. Tang *et al.*, "Effects of personal exposures to micro- and nano-particulate matter, black carbon, particle-bound polycyclic aromatic hydrocarbons, and carbon monoxide on heart rate variability in a panel of healthy older subjects," *International Journal of Environmental Research and Public Health,* vol. 16, no. 23, p. 4672, 2019. https://doi.org/10.3390/ijerph16234672

[20] C.-C. Chan, L.-Y. Lin, C.-H. Lai, K.-J. Chuang, M.-T. Wu, and C.-H. Pan, "Association of particulate matter from cooking oil fumes with heart rate variability and oxidative stress," *Antioxidants,* vol. 10, no. 8, p. 1323, 2021. https://doi.org/10.3390/antiox10081323

[21] G. Wu, H. Liu, S. Wu, G. Liu, and C. Liang, "Can heart rate variability (HRV) be used as a biomarker of thermal comfort for mine workers?," *International Journal of Environmental Research and Public Health,* vol. 18, no. 14, p. 7615, 2021. https://doi.org/10.3390/ijerph18147615

[22] J. D'Angelo *et al.*, "Using heart rate variability methods for health-related outcomes in outdoor contexts: A scoping review of empirical studies," *International Journal of Environmental Research and Public Health,* vol. 20, no. 2, p. 1330, 2023. https://doi.org/10.3390/ijerph20021330

[23] A. Ilie, J. Vasilescu, C. Talianu, C. Iojă, and A. Nemuc, "Spatiotemporal variability of urban air pollution in Bucharest City," *Atmosphere,* vol. 14, no. 12, p. 1759, 2023. https://doi.org/10.3390/atmos14121759

[24] D.-W. Wu *et al.*, "Anxiety is a mediator between heart rate variability and quality of life in chronic obstructive pulmonary disease," *Journal of Personalized Medicine,* vol. 12, no. 6, p. 960, 2022. https://doi.org/10.3390/jpm12060960

[25] M. Laurino *et al.*, "Sport in town: The smart healthy ENV project, a pilot study of physical activity with Multiparametric monitoring," *International Journal of Environmental Research and Public Health,* vol. 18, no. 5, p. 2432, 2021. https://doi.org/10.3390/ijerph18052432

[26] A. G. Soares *et al.*, "Disruption of atrial rhythmicity by the air pollutant 1,2-Naphthoquinone: Role of beta-adrenergic and sensory receptors," *Biomolecules,* vol. 14, no. 1, p. 57, 2024. https://doi.org/10.3390/biom14010057

[27] L. I. Lecca *et al.*, "Response of the cardiac autonomic control to exposure to nanoparticles and noise: A cross-sectional study of airport ground staff," *International Journal of Environmental Research and Public Health,* vol. 18, no. 5, p. 2507, 2021. https://doi.org/10.3390/ijerph18052507

[28] X. Lin *et al.*, "Wearable sensor-based monitoring of environmental exposures and the associated health effects: A review," *Biosensors,* vol. 12, no. 12, p. 1131, 2022. https://doi.org/10.3390/bios12121131

[29] J. Oh, H. Lee, and H. Park, "Effects on heart rate variability of stress level responses to the properties of indoor environmental colors: A preliminary study," *International Journal of Environmental Research and Public Health,* vol. 18, no. 17, p. 9136, 2021. https://doi.org/10.3390/ijerph18179136

[30] S. González-Rojas *et al.*, "Air pollution and endurance exercise: A systematic review of the potential effects on cardiopulmonary health," *Life,* vol. 15, no. 4, p. 595, 2025. https://doi.org/10.3390/life15040595

[31] 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

[32] Z. Zhang, Y. Lv, and H. Pan, "Cooling and humidifying effect of plant communities in subtropical urban parks," *Urban Forestry & Urban Greening,* vol. 12, no. 3, pp. 323-329, 2013. https://doi.org/10.1016/j.ufug.2013.03.010

[33] A. J. Wright and R. M. Francia, "Plant traits, microclimate temperature and humidity: A research agenda for advancing nature-based solutions to a warming and drying climate," *Journal of Ecology,* vol. 112, no. 11, pp. 2462-2470, 2024. https://doi.org/10.1111/1365-2745.14313

[34] S. K. Park, M. S. O'Neill, P. S. Vokonas, D. Sparrow, and J. Schwartz, "Effects of air pollution on heart rate variability: The VA normative aging study," *Environmental Health Perspectives*, vol. 113, no. 3, pp. 304-309, 2004. https://doi.org/10.1289/ehp.7447

[35] K. Li, H. Rüdiger, and T. Ziemssen, "Spectral analysis of heart rate variability: Time window matters," *Frontiers in Neurology*, vol. 10, 2019. https://doi.org/10.3389/fneur.2019.00545