IRSTI 28.23.01 https://doi.org/10.26577/jpcsit202546

Zhanel Baigarayeva[1*], Assiya Boltaboyeva[1], Gulmira Dikhanbayeva[1], Marlen Maulenbekov[1,2], Aiman Bekturganova[1], Gulshat Amirkhanova[1]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan
[2]Satbayev University, Almaty, Kazakhstan
*e-mail: zhanel.baigarayeva@gmail.com

INTEGRATED ENVIRONMENTAL AND PHYSIOLOGICAL MONITORING FOR CARDIOVASCULAR RISK DETECTION USING IOT AND MACHINE LEARNING

Abstract. This study investigates the impact of air pollution on heart rate variability (HRV), a key physiological marker reflecting the state of the autonomic nervous and cardiovascular systems. Despite growing interest, the complex relationship between environmental exposure and HRV, especially in the context of early cardiovascular disease (CVD) detection, remains insufficiently explored. An integrated real-time monitoring system was developed using Internet of Things (IoT) devices and machine learning (ML) methods to collect and analyze data from 10 healthy participants (aged 18–22) in three different environments: a controlled laboratory, an urban roadside (Al-Farabi Avenue), and a natural setting (botanical garden). Physiological signals (RMSSD, SDNN, LF, HF) were obtained using Polar H10 ECG sensors and Zhurek PPG devices, while environmental data (PM2.5, PM10, CO_2) were recorded via Tynys and Qingping sensors. Three supervised ML models–deep neural networks (DNN), random forest (RF), and XGBoost–were used to classify HRV levels based on environmental parameters. Among them, XGBoost achieved the best performance with 91.92% accuracy, 91.82% precision, and a 90.42% F1-score. The results revealed a consistent negative correlation between higher levels of PM2.5 and PM10 and reduced HRV metrics, particularly SDNN and RMSSD, indicating potential autonomic dysfunction and increased cardiovascular risk. Although CO_2 levels showed weaker associations, their influence was still noted. These findings emphasize the importance of considering environmental factors in health monitoring and demonstrate the potential of IoT and ML technologies in enabling early detection of cardiovascular stress and supporting personalized healthcare strategies.

Keywords: heart rate variability (HRV), air pollution, particulate matter (PM2.5, PM10), carbon dioxide (CO_2), autonomic nervous system, cardiovascular risk, machine learning, Internet of Things (IoT), real-time monitoring, environmental exposure.

1. Introduction

Heart rate variability (HRV) is a key physiological indicator reflecting the activity of the autonomic nervous system (ANS) and overall cardiovascular health. HRV analysis, including metrics such as the root mean square of successive differences (RMSSD), standard deviation of all NN intervals (SDNN), and the high-frequency to low-frequency ratio (HF/LF), is widely used to assess the balance between sympathetic and parasympathetic activity. This balance is directly related to the overall physical readiness of the body and its ability to respond to various stressors. Changes in HRV are associated with a wide range of health conditions, including cardiovascular disorders, metabolic syndrome, and psychological issues.

Changes in environmental conditions, such as air quality deterioration and temperature fluctuations, can significantly affect human health, especially the function of the autonomic nervous system. It has been established that exposure to adverse environmental factors, including CO_2, PM2.5, PM10, and extreme temperature conditions, can negatively affect HRV parameters. A decrease in these parameters is often observed in people living in areas with high air pollution levels, which increases the risk of cardiovascular diseases. The HF/LF ratio, reflecting the balance between parasympathetic and sympathetic nervous system activity, can also serve as an indicator of stress, the intensity of which is heightened under adverse environmental conditions.

Despite growing interest in studying the relationship between environmental exposure and HRV, this interaction remains insufficiently explored. Previous studies often focused on isolated health conditions, without considering the broader ecological context that contributes to changes in HRV. A significant limitation in current research is the lack of integration of environmental data, such as air pollution, into machine learning models. This hinders the comprehensiveness and accuracy of predictive models, especially those designed to assess the impact of environmental factors on health outcomes.

The individual effects of pollutants and the utility of HRV are well documented, yet there is a critical gap in research regarding the insufficient integration of environmental data into machine learning models and a limited understanding of the combined effects of various pollutants. This points to the need for a shift from isolated studies to a holistic systems approach. The focus of this study on integrated monitoring and machine learning represents a significant step forward in achieving a more comprehensive understanding of ecological health. It goes beyond identifying isolated correlations and aims to create predictive models capable of accounting for complex, multifactorial environmental influences on physiological responses, thereby paving the way for truly personalized and preventive medicine in the context of ecological stressors. The relationship between environmental factors and HRV parameters provides valuable data for early diagnosis and prevention of various diseases, which is especially important for developing personalized treatment strategies aimed at improving environmental conditions for individual patients.

The objectives of this research include the development and validation of an integrated system for real-time monitoring of physiological parameters (HRV) and environmental conditions (air quality), analysis of the impact of specific environmental factors (PM2.5, PM10, CO_2) on HRV metrics in different conditions, and the application and evaluation of machine learning models to identify correlations between environmental exposure and physiological responses, with a focus on early detection of cardiovascular dysfunction. This research also aims to demonstrate the potential for integrating environmental data into health monitoring systems for personalized health recommendations and public health strategies.

2. Literature Review

Modern healthcare is undergoing a transformation, increasingly focusing on proactive, predictive, and personalized medicine through the integration of digital technologies. One of the most promising directions in this paradigm shift is the use of digital twin technology. Digital twins are virtual models of patients created based on individual physiological data, allowing for the simulation of disease progression, prediction of clinical outcomes, and optimization of therapeutic strategies. This approach enhances diagnostic accuracy and helps develop personalized medical interventions, making treatment more effective and tailored to individual patient needs [1], [2].

Telemedicine also plays an important role in modern healthcare, providing remote monitoring and consultations for patients, which is especially relevant in situations where access to traditional healthcare services is limited. During the COVID-19 pandemic, when in-person visits to doctors became impossible for most people, telemedicine became a vital tool for maintaining health, especially for managing chronic diseases [3], [2]. Its ability to bridge access gaps in remote and underserved areas highlights the importance of this technology in improving accessibility and healthcare efficiency [4],[5].

Wearable devices, such as smartwatches and fitness trackers, have become an essential part of digital healthcare. These devices not only track basic parameters such as heart rate and physical activity but also collect real-time data, allowing doctors to intervene promptly and adjust treatment plans. These devices are indispensable for health monitoring and can detect early signs of disease, especially in the cardiovascular system. One of the most informative parameters for monitoring is heart rate variability (HRV), which allows assessing the autonomic nervous system and cardiovascular function of the patient [5],[6]. Studies have shown that wearable devices can detect early signs of cardiovascular risk and optimize therapeutic strategies, contributing to personalized healthcare interventions [5].

Artificial intelligence (AI) is increasingly being integrated into clinical practice, enhancing diagnostic capabilities, medical image analysis, and personalized treatment planning. Machine learning algorithms help more accurately assess the condition of patients, taking into account their individual characteristics and overall health. However, the use of AI is also associated with several challenges, such as data reliability, ethical issues, and the need for model interpretability. These issues require special attention to ensure patient safety and clinical effectiveness of such technologies [7], [6].

Furthermore, the pandemic emphasized the importance of monitoring mental health, particularly in the context of social isolation and stress. Solutions such as virtual therapy and remote mental health monitoring have provided significant support to patients with anxiety disorders, depression, and other psychological issues [1]. In this context, the integration of Internet of Things (IoT) technologies in healthcare systems played an important role by providing continuous monitoring of vital signs and real-time feedback, which helps prevent disease exacerbations and reduce patient risks [8]. Research indicates that IoT technologies enhance patient engagement in health management, leading to better outcomes in both physical and mental health domains [9].

However, despite advancements in digital technologies, there are still many issues that hinder a full understanding and effective use of health data. One of the major limitations is the insufficient integration of environmental data, such as air pollution, into machine learning models. This hinders the creation of more accurate and comprehensive predictive models that could account for the impact of environmental factors on human health. Most current studies focus on specific aspects of pollution or limited patient groups, reducing the general applicability of the results and complicating the broader use of such models [10]. The need to incorporate diverse environmental parameters such as CO_2, particulate matter, and temperature into predictive healthcare models is becoming increasingly evident as environmental factors play a significant role in shaping public health outcomes [11], [12].

It is also worth noting that the combined effects of various pollutants on human health, particularly on the cardiovascular system, are not well studied. Specifically, the impact of factors such as carbon dioxide (CO_2) and temperature on heart rate variability requires further research. More in-depth studies are needed that will consider multiple environmental factors to understand their impact on health and create more accurate and effective models for predicting and managing disease risks [11], [13]. This will help to develop comprehensive health monitoring systems that can track the effects of environmental pollutants in real-time and adjust health recommendations accordingly [12].

Thus, current research shows a fragmented understanding of the relationships between digital technologies, environmental factors, and human health. Despite significant successes in health monitoring and the development of predictive models, there is a need for a more thorough and integrated approach to combining these data in order to provide more accurate and personalized methods of disease treatment and prevention [14]. The integration of environmental data with wearable technologies and machine learning has the potential to revolutionize healthcare by offering highly personalized and predictive solutions for a range of diseases [9].

3. Methodology

3.1. System Architecture

The developed system integrates the monitoring of physiological parameters, environmental conditions, and data analysis based on machine learning techniques. The data flow within the system involves collecting health indicators from various devices, transmitting them to a central server, storing them in an SQL database, and subsequently processing the data using machine learning algorithms to derive informative insights and forecasts.

The system architecture is designed to seamlessly integrate physiological and environmental monitoring. Physiological data, including heart rate variability (HRV) metrics such as SDNN, RMSSD, LF, and HF, are captured using various IoT devices, such as the Polar H10 ECG sensor, the Zhurek device equipped with the MAX30102 sensor for photoplethysmographic (PPG) signals, and the Samsung Watch 6. The physiological data are transmitted via Bluetooth to a central server, where they are processed and analyzed in real-time. The system's architecture ensures that environmental data, such as CO_2, PM2.5, and PM10, are also collected from dedicated IoT devices (Tynys and Qingping Air Quality Monitor CGS1) and transmitted through MQTT and Wi-Fi protocols.

3.2. Physiological Parameter Monitoring Subsystem

The Polar H10 ECG sensor is used to record ECG signals and measure key HRV metrics, which are considered the "gold standard" for HRV assessment due to its high accuracy. Additionally, the Zhurek IoT device, equipped with the MAX30102 sensor, detects changes in blood volume, crucial for cardiovascular health monitoring.

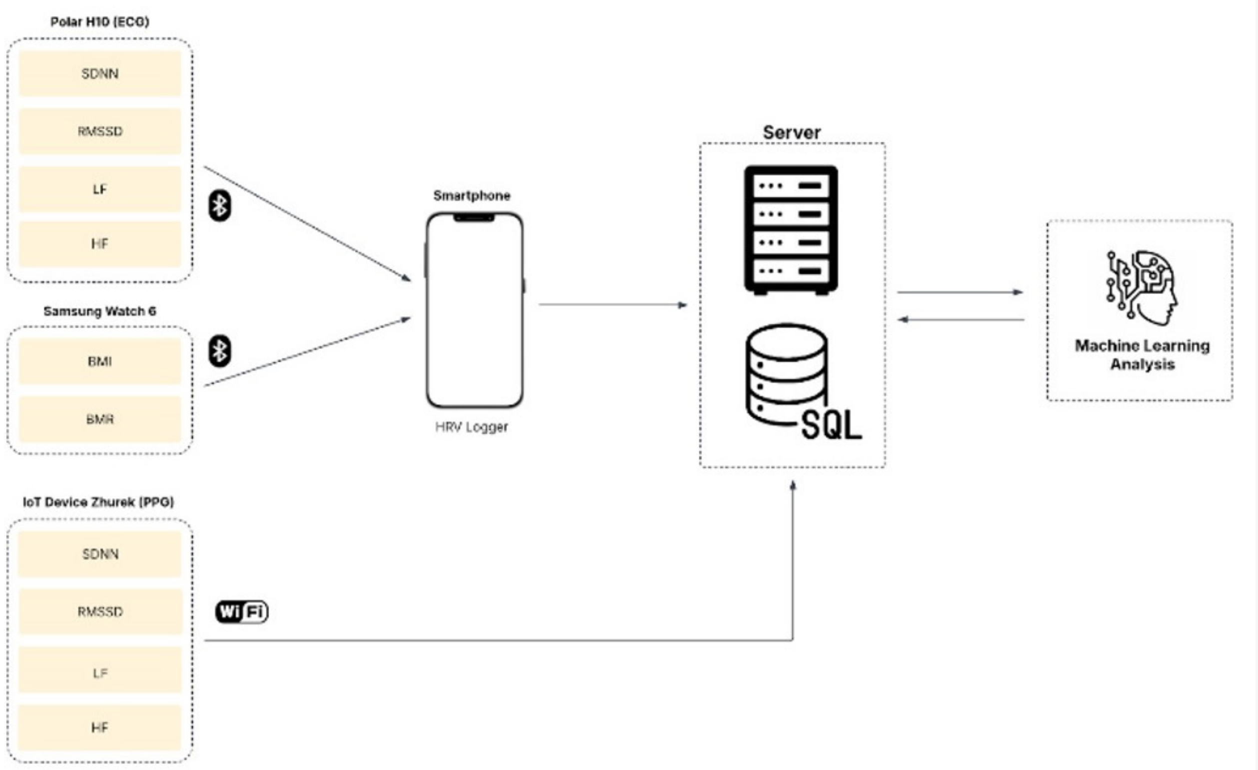

Figure 1 – Architecture of the System.

Figure 2 – MAX30102 GY PPG sensor – IoT Device Zhurek.

The Samsung Watch 6 provides additional health data such as body mass index (BMI), basal metabolic rate (BMR), and blood pressure (BP), transmitting this information via Bluetooth for comprehensive health tracking.

Data validation is a key aspect in ensuring the reliability and accuracy of collected data. The Polar H10, as a medical-grade ECG device, undergoes clinical validation to verify its accuracy in measuring heart rate and HRV and its compliance with data privacy and security standards. Similarly, the Samsung Watch 6 undergoes validation for consumer health monitoring, meeting industry standards and regulatory requirements to ensure the accuracy of health metrics such as BMI and BMR.

The system adopts a multimodal approach, using both ECG (Polar H10) and PPG (Zhurek with MAX30102) sensors for HRV and commercial smartwatches (Samsung Watch 6) for additional health metrics. This approach, along with explicit validation steps for each device, emphasizes the system's commitment to data quality and reliability. This robust data collection strategy enhances the credibility of research outcomes, validating new IoT devices against established medical standards, which increases trust in wearable technologies for clinical applications and facilitates their integration into preventive medicine and remote patient monitoring.

3.3. Environmental Monitoring Subsystem

The IoT device Tynys is used to monitor air quality parameters such as PM2.5, PM10, and CO_2 in real-time. The data is transmitted via MQTT over Wi-Fi to a central server for further processing. Additionally, the Qingping Air Quality Monitor CGS1 was used to gather environmental data in specific external settings, such as a botanical garden and Al-Farabi Avenue, providing insights into air quality under different conditions.

3.4. Data Transmission, Storage, and Processing

Data from all devices, including the Polar H10, Samsung Watch 6, Zhurek, Tynys, and Qingping, are transmitted to a central server. A Raspberry Pi gateway aggregates data from the Tynys sensors and synchronizes the timestamps with physiological measurements to ensure accurate correlation analysis. All data streams–both physiological and environmental–are stored in a structured SQL database on the server, providing scalable access and enabling further processing through machine learning pipelines [1].

Figure 3 – The IoT device Tynys.

The emphasis on real-time monitoring, timestamp synchronization, and continuous monitoring is critical. The system is not merely designed for data collection but for contextualized data collection. The system is designed to understand when certain physiological changes occur in relation to specific environmental exposures. This real-time contextualization is fundamental for developing truly predictive cardiovascular disease (CVD) models. Instead of simply identifying correlations, the system enables the creation of early warning systems that can alert individuals or healthcare professionals about potential health risks as environmental conditions change, facilitating proactive interventions rather than reactive treatment. This contributes to a dynamic and adaptive health management model.

3.5. Experimental Protocol

The research was conducted in three distinct environmental settings to analyze the influence of environmental factors on heart rate variability (HRV) [1]. The first setting was a controlled indoor laboratory environment, where air purifiers and humidifiers were used to minimize external influences and maintain stable parameters. The second environment was an urban roadside location along Al-Farabi Avenue, where participants were exposed to moderate stressors, including traffic noise, air pollution, and dense pedestrian activity. The third environment was a botanical garden, characterized by minimal acoustic disturbances and abundant vegetation, which created a calm and restorative atmosphere.

In each of these environments, HRV measurements were taken over five-minute intervals while environmental data, including air quality indicators, were continuously monitored. Participants were instructed to remain seated and breathe naturally throughout the recordings to avoid signal artifacts that might result from controlled breathing or physical movement.

The intentional selection of these contrasting settings–controlled, high-stress urban, and low-stress natural–represents a key strength of the experimental design. This quasi-experimental approach allows for a clearer observation of how HRV responds to different environmental exposures and supports more robust conclusions regarding causal relationships between specific environmental stressors and physiological responses. The methodological consistency enhances the reliability of the findings and lays the groundwork for future research aimed at quantifying physiological stress loads across various urban and natural settings, potentially guiding urban planning and environmental policy to better protect public health.

3.6. Participants

The study included a cohort of 10 participants. Strict inclusion criteria were established to ensure the reliability of HRV measurements and minimize the impact of confounding variables. The study cohort was limited to individuals aged 18 to 22 years, with no history of cardiovascular diseases, and not taking medications that could affect HRV. Participants were required to abstain from alcohol and caffeine for 24 hours before data collection. Exclusion criteria included insufficient sleep (<6 hours) the night prior to the assessment, exposure to significant psychological or physiological stress on the day of the evaluation, or the presence of technical artifacts identified during the preliminary data analysis.

The stringent inclusion and exclusion criteria (age, health status, medication, substance use, sleep, stress) were designed to minimize confounding variables that could independently affect HRV. By controlling these internal factors, the study aims to more effectively isolate the influence of environmental factors on HRV. This careful participant selection enhances the internal validity of the study, making the observed correlations between environmental parameters and HRV more attributable to environmental exposure rather than individual physiological variations or lifestyle choices. This is crucial for generating reliable evidence that can inform public health recommendations and personalized interventions.

3.5. Data Collection

Physiological data (HRV metrics: SDNN, RMSSD, LF, HF) were recorded using the IoT Zhurek device (MAX30102 sensor) and supplemented with data from the Samsung Watch 6 (BMI, BMR, BP). Environmental parameters (CO_2, PM2.5, PM10) were collected in real-time using IoT devices Tynys and Qingping. Data collection was conducted in controlled sessions with fixed five-minute intervals to ensure adequate temporal resolution. The Zhurek IoT device collected PPG signals at a specified frequency, converting raw pulse waveforms into RR intervals and calculating HRV indices in near real-time. The embedded ESP32 controller locally buffered these data to prevent loss during temporary network outages, ensuring data integrity before transmission via Wi-Fi. The Raspberry Pi gateway aggregated data from the Tynys sensors, applying timestamp synchronization with physiological measurements for precise correlation analysis. All data streams were stored in a structured SQL database on the server.

4. Results and Discussion

4.1. Data Preprocessing

The data preprocessing involved normalizing the values of CO_2, PM2.5, and PM10, as well as encoding the levels of HRV parameters using Label Encoder: ‘0’ for low, ‘1’ for medium, and ‘2’ for high stress. The dataset contained three features (environmental attributes) and the target variable (HRV levels). To improve class distribution and increase variability in the training dataset, synthetic data were generated and combined with the original samples.

4.2. Environmental Conditions and Air Quality Measurements

Environmental parameters were tracked in three different locations to assess their impact on heart rate variability (HRV). These locations included a natural open environment, an urban environment with high traffic, and a controlled indoor space. The environmental factors measured included temperature, humidity, particulate matter (PM2.5, PM10), and carbon dioxide (CO_2) levels.

Table 1 – Average Air Quality Parameters in Different Environments.

Location	PM2.5 (μg/m³)	PM10 (μg/m³)	CO□ (ppm)	Temperature (°C)	Humidity (% RH)
Al-Farabi Avenue	15.6	28.3	450	5	44
Botanical Garden	21.2	35.1	400	1	83
Laboratory	10.1	18.5	1200	2	49

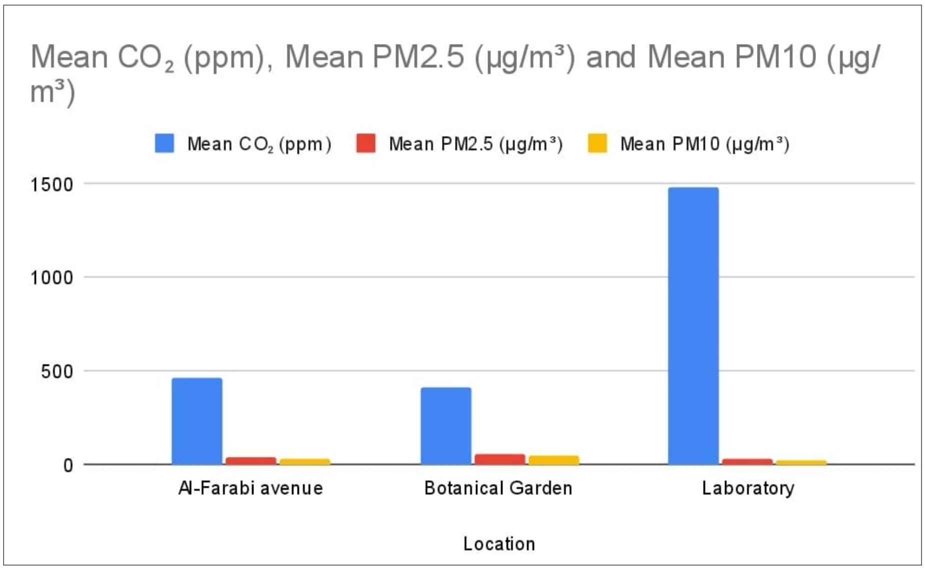

Figure 5 – Mean Air Quality Parameters Across Different Environments.

As seen in Table 1, the concentrations of PM2.5 and PM10 were noticeably highest in the botanical garden, while significantly elevated CO_2 levels were observed in the closed laboratory due to limited ventilation. The higher levels of PM in the botanical garden may be related to pollutants being trapped by vegetation and higher humidity, which helps retain particles.

The observation that the highest concentrations of PM2.5 and PM10 were in the botanical garden is counterintuitive when assuming that natural environments are inherently "cleaner." The explanation, involving pollutants being trapped by vegetation and the higher humidity helping to retain particles, points to the complex dynamics of the local environment. This means that a "natural" environment does not always equate to "low pollution" for all types of pollutants. This result underscores the need for detailed environmental monitoring and public health recommendations. It shows that even seemingly favorable conditions may present specific pollution risks, and understanding local atmospheric conditions and ecological interactions is critical for accurate health risk assessment and targeted interventions.

4.3. The Impact of Air Quality on HRV Metrics

HRV data were collected to assess the autonomic nervous system (ANS) response to different environmental conditions. Parameters such as SDNN and RMSSD were analyzed.

Participants exposed to higher concentrations of PM2.5 showed a significant decrease in SDNN and RMSSD values, indicating a shift towards increased sympathetic dominance and decreased parasympathetic activity. Despite the high CO_2 levels, laboratory conditions were associated with higher HRV values, suggesting a more stable autonomic response.

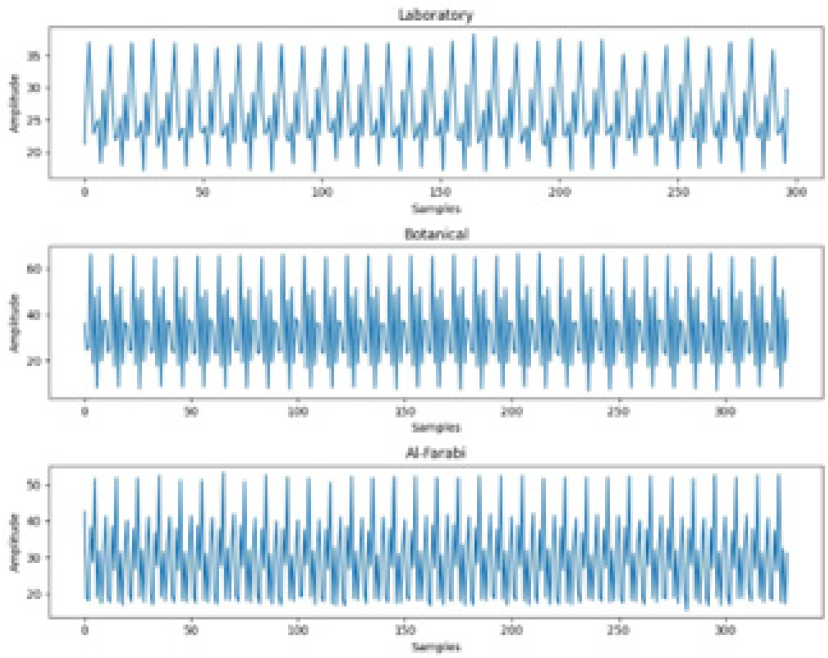

Figure 6 – HRV Metrics Across Experimental Locations.

Table 2 – Individual HRV Parameters by Experimental Locations.

Location	Conditions (°C / % RH)	Mean HR (bpm)	Mean RR Interval (ms)	SDNN (ms)	RMSSD (ms)
Al-Farabi Avenue	5 / 44	92.687	666.333	42.724	29.241
Botanical Garden	1 / 83	93.819	650.662	48.758	32.979
Laboratory	2 / 49	94.318	650.662	42.936	26.720

The observation that the laboratory, despite the highest CO_2 levels, maintained higher or more stable HRV values compared to the other locations is an important nuance. This suggests that, while CO_2 may impact HRV, other environmental stressors common in urban or even natural settings (such as noise, other pollutants, and general urban stress) might have a more dominant negative effect on HRV. The controlled nature of the laboratory, even with elevated CO_2 levels, might mitigate other stressors. This means that health risk assessments related to environmental factors should not solely rely on individual pollutant concentrations. Instead, a holistic approach, considering the entire environmental context (such as noise, temperature, other co-pollutants, and psychological stressors), is essential for accurate physiological response prediction and effective intervention development. This highlights the complexity of environmental health and the need for multifactorial analysis.

4.3. Correlation Analysis

4.3.1 Correlation Between SDNN, RMSSD, HF, LF

Figure 5 demonstrates the analysis of the relationship between various environmental parameters, including CO_2, PM10, and PM2.5, and heart rate variability (HRV) metrics such as SDNN and RMSSD, as well as frequency components including high-frequency (HF) and low-frequency (LF) components. All graphs represent scatter plots with a regression line for each pair of variables, accompanied by a correlation calculation.

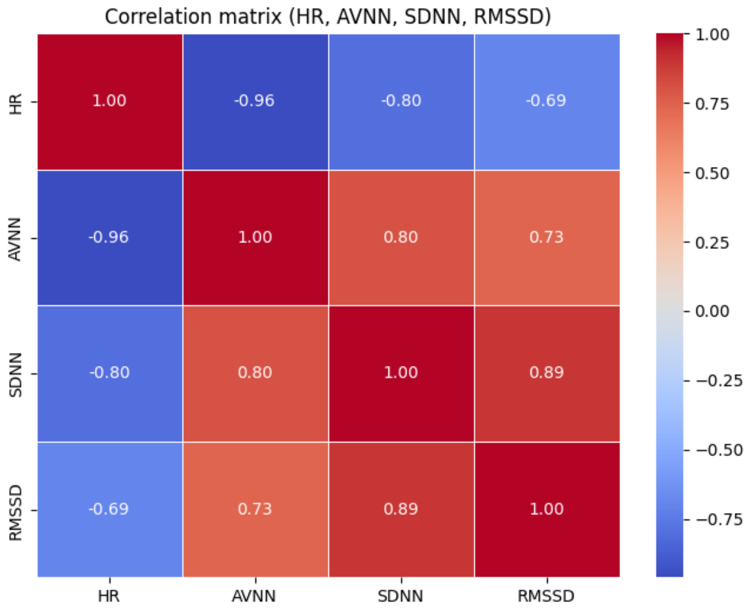

Figure 7 – Correlation matrix illustrating the relationships among HR parameters (HR, AVNN, SDNN, RMSSD).

Figure 6 illustrates the relationships between HRV metrics (SDNN, RMSSD, LF, and HF). The strong positive correlation between SDNN and RMSSD (0.89) confirms their similar role in evaluating RR interval variability. The correlation between SDNN and LF (0.84) and HF (0.74) indicates the influence of both sympathetic and parasympathetic regulation. RMSSD shows a strong correlation with HF (0.78), confirming its connection with parasympathetic activity. Finally, the positive correlation between LF and HF (0.74) may indicate coordination between the sympathetic and parasympathetic nervous systems. These results highlight the complex interactions between HRV parameters and may be useful for deeper analysis of autonomic regulation.

4.3.2 Correlation Between Environmental Parameters (CO_2, PM2.5, PM 10)

The correlation matrix in Figure 7 illustrates the relationships between environmental parameters, including CO_2, PM2.5, and PM10. The negative correlation between CO_2 and PM2.5 (-0.74), as well as between CO_2 and PM10 (-0.75), suggests that higher CO_2 levels may be associated with lower particulate concentrations in the air. Furthermore, the negative correlation between PM2.5 and PM10 (-0.74) indicates that an increase in one type of particulate matter may be linked to a decrease in the other, potentially reflecting differences in their sources, dispersion patterns, or atmospheric interactions.

Strong negative correlations observed between CO_2 and particulate matter (PM2.5, PM10), as well as between PM2.5 and PM10 themselves, are counterintuitive if one assumes that all pollutants come from similar sources or disperse in the same way. This suggests that these pollutants may have different sources (e.g., CO_2 from combustion, PM from road traffic/industrial emissions or even natural sources) or have different atmospheric dispersion models. This finding emphasizes the complexity of air quality management. It means that strategies aimed at reducing one type of pollutant may not automatically reduce others and, in some cases, may even be inversely related. A comprehensive approach to air quality monitoring and management must consider the unique sources, chemical interactions, and atmospheric behavior of various pollutants.

4.3.3 Correlation Between HRV Metrics and Environmental Parameters

The set of plots in Figure 8 concerns the RMSSD metric. A weak negative correlation is also observed between CO2 and RMSSD ($r = -0.127$), confirming the lack of a significant impact of CO2 on this metric. For PM10 and PM2.5, the correlations are positive but also weak ($r = 0.280$ and $r = 0.287$, respectively), indicating a minor link between the levels of these pollutants and heart rate variability. Figure 9, on the frequency components, shows a significantly stronger positive correlation between HF and SDNN (r = 0.739), as well as between LF and SDNN (r = 0.838). These results indicate the significant impact of low-frequency and high-frequency components on heart rate variability.

Figure 10 complements this analysis, including environmental factors and showing that increasing concentrations of particulate matter (PM2.5 and PM10) are associated with a decrease in HRV parameters such as SDNN and RMSSD. This result suggests the potential negative impact of air pollution on the autonomic nervous system. Furthermore, Figure 10 reveals a negative correlation between CO_2 levels and PM2.5 (-0.74), as well as PM10 (-0.75), which may point to differences in the sources or mechanisms of dispersion of these pollutants. The obtained results underline the importance of considering environmental factors when analyzing HRV, as air pollution can adversely affect the cardiovascular system and autonomic regulation. Statistical analysis also showed a moderate negative correlation between PM2.5 and HRV parameters (SDNN and RMSSD), indicating that higher exposure to fine particles may lead to reduced HRV, reflecting increased physiological stress. PM10 also showed a weak negative correlation with HRV, confirming that airborne particles can influence autonomic regulation. In contrast, CO_2 levels showed a weaker correlation with HRV, meaning that their impact is less pronounced compared to particulate matter.

Despite some minor internal discrepancies in one specific plot (Figure 8) regarding interpretation, the overwhelming evidence from both sources [1] consistently points to a negative correlation between particulate matter and HRV. This convergence of results from various analyses confirms that air pollution is a significant stressor for the autonomic nervous system. This compelling evidence of the negative impact of PM on HRV is crucial for public health. It provides a solid scientific basis for policies aimed at reducing air pollution, as such measures may directly contribute to improving cardiovascular health and reducing the risk of autonomic dysfunction in the general population.

4.5 Classification with Machine Learning

The objective of this analysis was to assess the relationship between environmental conditions and heart rate variability by applying machine learning classification models. The dataset included three environmental attributes: carbon dioxide concentration, PM2.5, and PM10, which were used as input features. The target variable represented HRV levels based on SDNN and RMSSD measurements. Each HRV instance was assigned to one of three categories indicating the degree of physiological regulation related to cardiovascular function. These categories were encoded with integer values: ‘0’ for low, ‘1’ for medium, and ‘2’ for high variability.

Classification was performed using three supervised learning algorithms: deep neural network (DNN), XGBoost, and random forest (RF). Each of these models is capable of detecting complex interactions between features and identifying patterns that are not captured by linear approaches. Prior to training, the environmental input data were normalized to ensure equal contribution from all variables. To improve class distribution and increase variability in the training dataset, synthetic data were generated and combined with the original samples. The classification models were developed using the Python programming language. PyTorch was used for implementing the deep neural network model. Scikit-learn was used for the random forest model, and the XGBoost library was used for building the gradient boosting model. The performance of each classifier was evaluated based on four standard classification metrics: accuracy, precision, recall, and F1-score.

As shown in Table 3, the XGBoost model demonstrated the highest overall performance, achieving an accuracy of 91.92%, precision of 91.82%, and F1-score of 90.42%, indicating its high ability to effectively classify HRV levels based on air quality parameters, outperforming the deep neural network (DNN) and random forest (RF) models. The random forest model achieved an accuracy of 88.30%, precision of 86.70%, and F1-score of 86.02%, while the DNN model achieved an accuracy of 89.47%, precision of 88.12%, and F1-score of 87.55%. These results highlight that machine learning models, especially XGBoost, can effectively capture complex relationships between environmental factors and physiological responses, offering a more accurate and precise approach to identifying cardiovascular disease risk compared to traditional methods.

For further validation of classification results, a confusion matrix was constructed using the predictions from the XGBoost model.

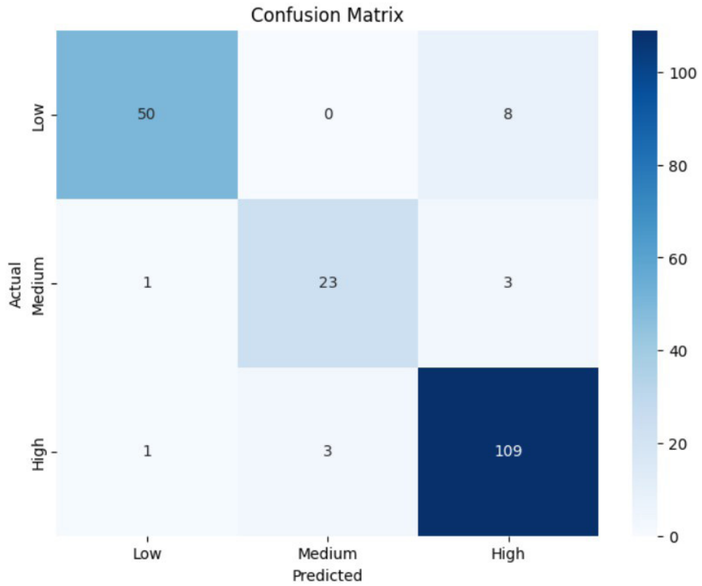

Figure 8 – Classification outcome matrix based on environmental features using XGBoost.

Table 1 – Classification Metrics of Applied Machine Learning Models.

Models	Accuracy	Precision	Recall	F1-Score
DNN	0.8947	0.8812	0.8700	0.8755
XGBoost	0.9192	0.9182	0.8928	0.9042
RF	0.8830	0.8670	0.8540	0.8602

Figure 11 illustrates this matrix, which compares the predicted HRV categories (Low, Medium, High) with the actual labels. The model demonstrated high accuracy, correctly classifying 50 instances of low HRV, 23 instances of medium HRV, and 109 instances of high HRV. The classification errors were minimal, mostly occurring between the "Low" and "High" categories, suggesting minor overlaps in the physiological signals associated with these groups. These results confirm the model's ability to effectively distinguish HRV variations based on environmental exposure, emphasizing the potential of integrating air pollution indicators with physiological monitoring systems for early cardiovascular risk assessment.

The superior performance of XGBoost compared to DNN and RF is a significant finding. XGBoost, as a gradient boosting algorithm, is known for its ability to handle complex, nonlinear relationships and interactions between features, which is characteristic of environmental and physiological data. This indicates that for this type of predictive task, ensemble methods may be more effective than deep learning or simpler tree-based models. This result has direct implications for the development of deployable health monitoring systems. Identifying the most accurate machine learning model (XGBoost) means that the proposed system can provide more reliable and precise cardiovascular risk predictions based on environmental exposure. This is crucial for transforming research into practical tools for early detection and personalized interventions in clinical settings and public health.

5. Conclusions

This study clearly highlights the significant impact of environmental factors, particularly air pollution (PM2.5, PM10, CO_2), on heart rate variability (HRV). Elevated concentrations of particulate matter were consistently associated with reductions in HRV–especially in the SDNN and RMSSD metrics–suggesting impaired autonomic nervous system regulation and an increased risk of cardiovascular dysfunction. The developed integrated IoT-based monitoring system effectively captured both physiological and environmental data in real time, offering a comprehensive view of how external conditions influence autonomic function. Among the machine learning models applied, XGBoost demonstrated the highest classification performance, achieving an accuracy of 91.92%, precision of 91.82%, and an F1-score of 90.42%, indicating strong potential for HRV prediction based on environmental exposure.

These findings underscore the importance of incorporating environmental data into health monitoring systems to enhance early disease prediction and preventive care. The study confirms the value of advanced machine learning algorithms in identifying subtle physiological changes triggered by environmental stressors. The proposed system facilitates continuous real-time assessment of cardiovascular health through wearable technology, supporting personalized and timely health interventions. Furthermore, this integrated approach provides unprecedented insights into early signs of autonomic imbalance and reinforces the need for improved air quality management, particularly in urban areas.

Despite its contributions, the study has several limitations. The participant cohort consisted solely of healthy individuals aged 18–22, which may restrict the generalizability of the findings to broader age groups or populations with pre-existing health conditions. The research was limited to short-term monitoring, making it difficult to assess the chronic effects of air pollution or seasonal variability in HRV. Additionally, the analysis did not account for other potentially influential environmental factors, such as ambient noise levels or humidity.

Future research should aim to address these limitations by including a more diverse participant pool representing various age groups and health statuses. Expanding the scope of monitored environmental parameters–such as noise pollution and humidity–will enable a more holistic understanding of factors affecting autonomic nervous system regulation. Long-term monitoring across different seasons will also be critical for evaluating cumulative exposure effects. The system itself will continue to be improved for better portability and ease of use, enabling deployment in both clinical and home environments.

By proposing clear next steps, this research lays the foundation for developing reliable, clinically validated predictive models capable of integration into everyday healthcare practices. The future direction outlined here reflects a forward-thinking approach that moves beyond identifying correlations to building actionable, personalized health monitoring systems. Such progress has the potential to transform current models of care, shifting from reactive to truly preventive and environmentally informed medicine.

Funding

This work was funded by Committee of Science of the Republic of Kazakhstan AP23488586 “Development of an intelligent system for monitoring and prevention of cardiovascular diseases using deep learning and IoMT (Internet of Medical Things)” (2024-2026).

Author Contributions

Conceptualization, Zh.B., A.B. and M.M.; Methodology, Zh.B., A.B. and A.B.; Software, D.G. and M.M.; Validation, Zh.B. and A.B.; Formal Analysis, Zh.B. and G.A.; Investigation, Zh.B., A.B.; Resources, G.A., G.D. and A.B.; Data Curation, G.D. and M.M.; Writing – Original Draft Preparation, Zh.B., A.B., A.B. and M.M.; Writing – Review & Editing, Zh.B., G.D. and G.A.; Visualization, A.B.; Supervision, Zh.B.; Project Administration, G.A.; Funding Acquisition, G.A.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. C. A. Sîrbu *et al.*, ‘Air Pollution and Its Devastating Effects on the Central Nervous System’, *Healthcare*, vol. 10, no. 7, p. 1170, Jun. 2022, doi: 10.3390/healthcare10071170.
2. S. K. Park *et al.*, ‘Air Pollution and Heart Rate Variability: Effect Modification by Chronic Lead Exposure’, *Epidemiology*, vol. 19, no. 1, pp. 111–120, Jan. 2008, doi: 10.1097/EDE.0b013e31815c408a.
3. P. Stachteas, C. Stachteas, E. K. Symvoulakis, and E. Smyrnakis, ‘The Role of Telemedicine in the Management of Patients with Chronic Diseases in Primary Care During the COVID-19 Pandemic’, *Maedica*, vol. 17, no. 4, Dec. 2022, doi: 10.26574/maedica.2022.17.4.931.
4. C. Ejiyi *et al.*, ‘The internet of medical things in healthcare management: a review’, *JDH*, pp. 30–62, Jun. 2023, doi: 10.55976/jdh.22023116330-62.

5. S. Huhn *et al.*, 'The Impact of Wearable Technologies in Health Research: Scoping Review', *JMIR Mhealth Uhealth*, vol. 10, no. 1, p. e34384, Jan. 2022, doi: 10.2196/34384.

6. S. Boyd, 'Wearable Technology for Health Monitoring and Diagnostics', *IJCE*, vol. 5, no. 5, pp. 33–44, Jul. 2024, doi: 10.47941/ijce.2041.

7. G. Battineni, G. G. Sagaro, N. Chinatalapudi, and F. Amenta, 'Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis', *JPM*, vol. 10, no. 2, p. 21, Mar. 2020, doi: 10.3390/jpm10020021.

8. J. Moon and B.-K. Ju, 'Wearable Sensors for Healthcare of Industrial Workers: A Scoping Review', *Electronics*, vol. 13, no. 19, p. 3849, Sep. 2024, doi: 10.3390/electronics13193849.

9. I. Surenther, K. P. Sridhar, and M. Kingston Roberts, 'Maximizing energy efficiency in wireless sensor networks for data transmission: A Deep Learning-Based Grouping Model approach', *Alexandria Engineering Journal*, vol. 83, pp. 53–65, Nov. 2023, doi: 10.1016/j.aej.2023.10.016.

10. S. Hammoud, R. Karam, R. Mourad, I. Saad, and M. Kurdi, 'Stress and Heart Rate Variability during University Final Examination among Lebanese Students', *Behavioral Sciences*, vol. 9, no. 1, p. 3, Dec. 2018, doi: 10.3390/bs9010003.

11. A. I. Pérez-Alcalde *et al.*, 'The Effects of a Single Vagus Nerve's Neurodynamics on Heart Rate Variability in Chronic Stress: A Randomized Controlled Trial', *Sensors*, vol. 24, no. 21, p. 6874, Oct. 2024, doi: 10.3390/s24216874.

12. M. M. Bennett, C. W. Tomas, and J. M. Fitzgerald, 'Relationship between heart rate variability and differential patterns of cortisol response to acute stressors in mid-life adults: A data-driven investigation', *Stress and Health*, vol. 40, no. 3, p. e3327, 2024, doi: 10.1002/smi.3327.

13. M. Benchekroun, P. E. Velmovitsky, D. Istrate, V. Zalc, P. P. Morita, and D. Lenne, 'Cross Dataset Analysis for Generalizability of HRV-Based Stress Detection Models', *Sensors*, vol. 23, no. 4, p. 1807, Feb. 2023, doi: 10.3390/s23041807.

14. M. Bahameish, T. Stockman, and J. Requena Carrión, 'Strategies for Reliable Stress Recognition: A Machine Learning Approach Using Heart Rate Variability Features', *Sensors*, vol. 24, no. 10, p. 3210, May 2024, doi: 10.3390/s24103210.

Information About Authors:

Zhanel Baigarayeva is 3-rd year PhD student, received her B.Sc. degree in Automation and Control in 2020 and her M.Sc. degree in Intelligent Control Systems in 2022. She is currently pursuing a PhD in Automation and Internet of Things (IoT). Her research interests include artificial intelligence, intelligent automation systems, IoT technologies, and applied machine learning. In this study, she contributed to the conceptualization, methodology, software development, validation, investigation, data curation, and original draft writing. She is actively engaged in research related to the development of smart systems for real-time monitoring and decision-making. ORCID iD: 0000-0003-1919-3570.

Assiya Boltaboyeva is currently pursuing her PhD in Intelligent Control Systems at al-Farabi Kazakh National University (Almaty, Kazakhstan, boltaboyeva_assiya3@kaznu.edu.kz). She received her B.Sc. degree in Automation and Control in 2021 and her M.Sc. degree in Intelligent Control Systems in 2023. Her research interests include artificial intelligence, intelligent control systems, industrial IoT technologies, and the application of machine learning in process automation. In this study, she contributed to the conceptualization, methodology, data preparation and analysis, validation, and original draft writing. She actively participates in research projects focused on developing smart solutions for control and automated monitoring. ORCID iD: 0000-0002-7279-9910.

Gulmira Dikhanbayeva, PhD. Dr. Gulmira Dikhanbayeva is an Associate Professor at the Department of Neurology, Psychiatry, and Narcology at the International Kazakh-Turkish University named after Khoja Akhmet Yassawi (Turkistan, Kazakhstan). She received her M.D. degree in Pediatrics from Aktobe State Medical Institute in 1993 and completed her specialization in Pediatric Neurology in 1995. She earned her PhD in Medical Sciences in 2000 and was awarded the academic title of Associate Professor in 2007. Her research interests include pediatric and adult neurology, rehabilitation, epilepsy, and neuropsychiatric disorders. In this study, she contributed to validation, supervision, and review. Dr. Dikhanbayeva has authored over 120 scientific publications and has supervised numerous Master's and PhD students. She frequently presents at scientific conferences and delivers professional training sessions. ORCID iD: 0009-0002-1374-3236.

Marlen Maulenbekov is a fourth-year B.Sc. student in Intelligent Control Systems at Satbayev University (Almaty, Kazakhstan). His academic interests include intelligent control, the Internet of Things (IoT), and applied machine learning. In this study, he assisted with technical support, data processing, and visualization. ORCID iD: 0009-0002-0703-7634.

Aiman Bekturganova is a 3rd-year undergraduate student at al-Farabi Kazakh National University (Almaty, Kazakhstan). Her academic interests include intelligent control systems, the Internet of Things (IoT), and applied machine learning. In this study, she contributed to data collection, literature review, and initial analysis. ORCID iD: 0009-0002-8289-4505.

Gulshat Amirkhanova, PhD. Dr. Gulshat Amirkhanova is a Senior Lecturer at al-Farabi Kazakh National University (Almaty, Kazakhstan). She holds a PhD in her field and is actively engaged in academic and research activities at the university. Her research interests include intelligent systems, automation technologies, industrial IoT, and the integration of machine learning methods into control and monitoring processes. Dr. Amirkhanova contributes to both teaching and applied research, focusing on the development of innovative approaches for smart environments and automated solutions. She has co-authored several publications and is involved in guiding student research. ORCID iD: 0000-0003-3933-5476.

Submission received: 29 May, 2025.
Revised: 30 September, 2025.
Accepted: 30 September, 2025.