







ISSN: 2617-6548

URL: www.ijirss.com



Wearable sensor-based analysis of heart rate variability as a biomarker for ischemic heart disease

 Nurdaulet Tasmurzayev^{1,2},  Bibars Amangeldy^{1,2*},  Imanbek Baglan¹,  Zhanel Baigarayeva^{1,3}, Akzhan Konysbekova⁴

¹Al-Farabi Kazakh National University, Almaty 050040, Kazakhstan.

²LLP «DigitAlem», Almaty 050042, Kazakhstan.

³LLP «Kazakhstan R&D Solutions», Almaty 050056, Kazakhstan.

⁴JSC «Research Institute of Cardiology and Internal Diseases», Almaty 050000, Kazakhstan.

Corresponding author: Bibars Amangeldy (Email: a.s.bibars@gmail.com)

Abstract

Cardiovascular diseases are the leading global cause of death, with ischemic heart disease (IHD) being the most prevalent and deadly subtype, emphasizing the need for accessible, non-invasive screening tools. Heart rate variability (HRV) is a recognized indicator of autonomic imbalance and a predictor of adverse cardiac events; however, standard diagnosis relies on 24-hour Holter electrocardiography, which limits its widespread application. This study introduces the Zhurek fingertip IoT device, a photoplethysmography (PPG)-based tool for measuring HRV parameters, and evaluates its ability to distinguish between healthy autonomic function and IHD-related dysregulation. A comparative analysis with a three-lead Holter ECG showed clinically acceptable mean deviations: -0.601 bpm for heart rate, $+33.1$ ms for SDNN, and -4.8 ms for RMSSD. HRV segment data were obtained from patients at the Cardiology Center in Almaty, Kazakhstan, with angiographically confirmed IHD. Eight features were measured, including SDNN, RMSSD, LF, HF, LF/HF ratio, Max HR, BMI, and age. Mann–Whitney U tests revealed significant differences between groups for SDNN, LF, HF, Max HR, BMI, and age ($p < 0.05$). Principal component analysis indicated that the first two components, accounting for 49.5% of the total variance, effectively separated the cohorts without labels. Feature importance analysis using CatBoost demonstrated that LF power had the greatest discriminative weight ($\sim 44\%$), followed by age ($\sim 19\%$) and HF power, with smaller contributions from maximum heart rate and RMSSD. These findings demonstrate that diagnostically relevant autonomic signatures are preserved in PPG recordings, reducing monitoring time compared to standard Holter studies while maintaining physiological fidelity. The pilot study confirms the suitability of the Zhurek wearable device and healthy-baseline machine learning pipelines for large-scale, ambulatory IHD risk stratification, supporting a shift from reactive intervention to proactive prevention.

Keywords: Cardiovascular diseases, Heart rate variability, IoT device, Ischemic heart disease, Machine learning, Photoplethysmography, Risk stratification.

DOI: 10.53894/ijirss.v8i5.9317

Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26103523).

History: Received: 2 July 2025 / Revised: 1 August 2025 / Accepted: 5 August 2025 / Published: 15 August 2025

Copyright: © 2025 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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

Cardiovascular diseases (CVDs) are the leading cause of death worldwide. According to estimates by the World Health Organization, in 2019, 17.9 million people died from CVDs, accounting for 32% of all global deaths. Of these deaths, 85% were due to heart attacks and strokes. Among the 17 million premature deaths (under the age of 70) from non-communicable diseases recorded in 2019, 38% were caused by CVDs [1]. IHD is one of the most common forms of CVD and a major cause of mortality [2]. According to 2022 data, diseases of the circulatory system are the most prevalent among the adult population of Kazakhstan, with 3,962.5 cases per 100,000 population. Of these, ischemic heart disease (IHD) accounts for 560.7 cases per 100,000 population. These figures confirm the high prevalence of IHD within the structure of cardiovascular pathology [3].

IHD continues to pose a significant burden on individuals and healthcare systems worldwide. The impact of this condition is considerable, contributing substantially to both mortality and morbidity [4]. Coronary artery disease (CAD), most commonly resulting from atherosclerosis, is the leading cause of ischemic heart disease (IHD), which manifests as myocardial ischemia. The primary mechanism underlying IHD is obstructive atherosclerosis of the coronary arteries, leading to impaired blood flow to the heart muscle [5]. Given the growing global demand on healthcare systems, there is an urgent need to develop early diagnostic approaches capable of detecting IHD before the occurrence of irreversible complications such as myocardial infarction or chronic heart failure (CHF) [6].

Heart rate variability (HRV) is defined as the fluctuation in the duration of cardiac cycles [7]. It is a non-invasive indicator obtained through heart rhythm monitoring that provides valuable insights into the overall health status of the body [8]. HRV reflects the dynamic capacity of the heart and the general physiological ability of an individual to adapt to varying environmental conditions through compensatory mechanisms [9]. It is directly influenced by the primitive components of the autonomic nervous system (ANS), particularly the parasympathetic branch, and also reflects the combined activity of both the sympathetic and parasympathetic divisions. Low HRV values have been associated with adverse cardiac events such as myocardial infarction, atherosclerosis progression, heart failure, ischemic heart disease (IHD), and sudden cardiac death [10]. HRV analysis is essential for assessing the functional state of the ANS [11].

Current clinical tests used to assess coronary health are often expensive, invasive, and insufficiently effective for the timely detection of progressing coronary ischemic conditions [12]. Although analytical angiography is considered one of the most accurate procedures for identifying heart abnormalities, it is associated with high costs, potential side effects, and requires significant technological expertise. Traditional diagnostic methods are time-consuming, prone to human error, and may lead to inaccurate diagnoses, making them costly and labor-intensive [13]. This highlights a critical gap in research: the need for reliable, non-invasive, and early detection techniques such as those potentially offered by HRV analysis.

2. Literature Review

CVDs remain the leading cause of morbidity and mortality worldwide, highlighting the critical importance of early diagnosis in high-risk individuals and the development of effective preventive and interventional strategies. Recent research has focused on creating multifactorial models that integrate physiological indicators, lifestyle factors, and clinical history to improve risk assessment reliability and enable a more personalized approach [14]. HRV, which measures the variation in time intervals between consecutive heartbeats (RR or NN intervals), is a non-invasive indicator widely used to assess cardiovascular health [15]. In the context of ischemic heart disease, reductions in the time-domain indices SDNN, RMSSD, and pNN50, along with changes in the LF/HF ratio, correlate with myocardial injury and a higher risk of adverse events, while an imbalance between LF and HF components reflects impaired autonomic regulation during ischemic episodes. This review systematizes these key HRV parameters and underscores their clinical relevance for monitoring and managing patients with IHD [6, 16]. In patients with IHD and arrhythmias, HRV metrics are significantly reduced compared to healthy individuals. Notably, time-domain parameters such as SDNN, SDANN, RMSSD, pNN50, and the triangular HRV index, along with nonlinear measures like α , α_1 , α_2 , SD1, SD2, Approximate Entropy (AppEn), and Sample Entropy (SampEn), show marked decreases in these patients [17]. These changes reflect impaired autonomic regulation of the heart and underline the utility of HRV analysis for evaluating cardiac function and disease progression in IHD [15]. HRV, defined as the fluctuation in ventricular response intervals in patients with atrial fibrillation (AF), is recognized as a non-random phenomenon. Its nonlinear characteristics, particularly multiscale entropy (MSE), are

considered to carry clinically relevant information. Numerous studies have demonstrated associations between HRV parameters and the risk of ischemic stroke in patients with AF. Multiscale entropy analysis of HRV has been proposed as a potential predictor of ischemic stroke in this patient population [18]. Higher values of sample entropy at specific time scales, derived from 24-hour Holter ECG recordings, were found to be associated with a greater likelihood of ischemic stroke in AF patients without prior history of stroke. HRV has also been utilized to evaluate cerebral hemisphere involvement in acute ischemic stroke (AIS). Comparative analysis of HRV in patients with right-hemispheric and left-hemispheric strokes revealed significantly higher sample entropy values in the left-hemispheric group. This finding indicates a lower HRV complexity in right-hemispheric strokes, which may reflect increased sympathetic nervous system activity. These differences remained statistically significant during the analysis of daytime ECG segments, supporting the potential of HRV measures in lateralizing cerebral ischemic lesions [19]. Beyond diagnostic stratification, HRV has been explored as a basis for developing indices to predict favorable or unfavorable short-term outcomes in the acute phase of ischemic stroke [20].

Parallel to HRV, electrocardiographic alternans (ECGA) has emerged as a promising non-invasive marker for assessing ischemic conditions and arrhythmic risk. ECGA includes analysis of T-wave alternans (TWA), QRS alternans (QRSA), and P-wave alternans (PWA), derived from standard ECG recordings. Data from the STAFF III database, which involved controlled balloon inflation to induce coronary artery occlusion, revealed significant increases in the magnitude of ECGA signals measured via correlation-based methods. These increases were time-dependent, with PWA detected during the first minute, QRSA in the second, and TWA in the third minute of occlusion [21]. ECGA is being investigated as a non-invasive electrophysiological indicator for risk stratification in patients with IHD [20]. While TWA has been studied in the context of IHD and heart failure, inconsistencies in experimental conditions and analytical methods have limited its interpretability [21]. Some studies suggest that combining TWA with HRV metrics may enhance diagnostic accuracy for detecting the progression of chronic heart failure. However, further research is required to establish its prognostic value specifically in IHD. Recent advancements emphasize the diagnostic potential of assessing all three alternans types TWA, QRSA, and PWA, simultaneously [22].

In addition to physiological metrics, growing attention has been directed toward genetic markers, particularly single-nucleotide polymorphisms (SNPs), which complement conventional risk factors. It has been demonstrated that SNP panels combined with clinical data such as the SCORE index, age, and coronary angiography results can achieve diagnostic accuracies of up to 93% [23]. Moreover, several studies have identified candidate genes involved in inflammation, lipid metabolism, and thrombosis that show potential for enhancing CVD risk prediction. Genetic markers provide prognostic value independent of traditional variables and are especially relevant in the context of type 2 diabetes, with which CVD shares common genetic pathways linked to metabolic predisposition [24]. Incorporating genetic data into clinical risk models thus facilitates the advancement of personalized medicine.

Traditionally, the assessment of cardiovascular function relies on ECG, which records the heart's electrical activity from the skin surface via electrodes [25]. However, over the past decade, the growing demand for continuous, user-friendly, and cost-effective solutions has sparked an active search for alternative monitoring methods [26]. Against this backdrop, photoplethysmography (PPG) has attracted particular attention due to its simple hardware implementation and ease of integration into consumer-grade devices. Moreover, PPG offers a significantly cheaper and more convenient alternative to ECG for continuous heart monitoring in both clinical and consumer settings [27].

Conventional ECG systems, despite their high accuracy, require clinical supervision, meticulous electrode placement, and regular calibration, which increases operating costs and inconveniences users [25]. Breakthroughs in microelectronics and sensor technologies have led to the sharp miniaturization of PPG sensors, allowing their integration into wristbands, smartwatches, mobile phones, and in-ear devices, thereby democratizing access to continuous cardiovascular monitoring [28]. Further integration of PPG with wireless data-transfer protocols and cloud-based analytics platforms provides a unique combination of affordability, portability, and user convenience that conventional ECG systems cannot match [29].

Methods of machine learning demonstrate high efficiency in detecting anomalies associated with IHD. Comparative analyses of algorithms such as support vector machines, artificial neural networks, and deep learning models have shown that, with proper preprocessing and feature selection, classification accuracy may exceed 90% [30]. Unsupervised clustering methods, particularly k-means, have been applied to detect outliers in heart disease datasets, which enhances the performance of subsequent supervised classifiers [31]. Deep learning models trained on ECG signals automatically extract discriminative features that differentiate normal from ischemic patterns with near-perfect accuracy [32]. Visual diagnostic data derived from non-contrast CT, echocardiograms, and CT angiograms are increasingly analyzed using deep learning models. These models form hierarchical representations of coronary artery anatomy and myocardial motion, identifying subtle variations in vessel caliber and wall motion indicative of ischemia [33, 34]. Architectures such as autoencoders and encoder-decoder networks compress and reconstruct multidimensional data, producing interpretable latent features [34]. Unsupervised anomaly detection further leverages clustering to segment data by similarity metrics, where deviating samples are labeled as anomalies [31]. This is especially valuable in medical datasets with sparse or imbalanced labels. Synthetic oversampling methods such as SMOTE are employed to balance class distributions, notably improving support vector machine performance [30]. Experimental studies confirm the reliability of deep learning models in ECG-based detection of IHD and myocardial infarction, reaching classification accuracy above 98% compared to healthy individuals [32]. These models detect fine deviations in ST-segment morphology and QRS complex duration, key indicators of ischemic events. Hybrid architectures combining convolutional and recurrent layers further enhance performance by capturing both spatial and temporal dependencies in cardiovascular datasets [35].

Demographic and anthropometric characteristics, particularly age and Body Mass Index (BMI), are also crucial components in risk models. Age is one of the most powerful independent predictors of CVD, as reflected in the Framingham Risk Score, while BMI as an indicator of obesity, is closely associated with metabolic and inflammatory processes contributing to cardiovascular pathology [36]. Another significant prognostic factor is the history of clinical interventions, including coronary angiography and revascularization procedures, which, when considered alongside other variables, can further refine risk stratification [37]. Behavioral factors such as smoking and alcohol consumption are equally important. Excessive alcohol intake and tobacco use are strongly associated with an elevated risk of IHD, whereas moderate alcohol consumption in non-smokers has occasionally shown a protective effect. However, this relationship is complex and may be confounded by additional variables such as physical activity and baseline health status, necessitating cautious interpretation [38, 39]. A comprehensive analysis of these elements suggests that incorporating data on harmful habits, genetic predisposition, demographics, clinical history, and HRV parameters provides the most holistic and personalized understanding of cardiovascular risk.

Diagnosis refers to the process of determining the presence or absence of a specific disease, such as IHD, at a given point in time [40]. The primary goal of diagnosis is to confirm disease status, enabling timely medical intervention and appropriate treatment [41]. Diagnosing IHD remains a complex clinical task [40]. Although invasive coronary angiography is considered the gold standard for confirming the presence of coronary artery disease, there is an ongoing search for non-invasive, cost-effective, rapid, and reliable diagnostic alternatives [42].

Risk prediction focuses on estimating the likelihood of future adverse events, such as the onset of IHD, myocardial infarction, or cardiovascular mortality [43-45]. This process often relies on the identification and analysis of risk factors. Long-term risk prediction for IHD is a major area of research, aiming to identify individuals at high risk before disease onset or clinical manifestation. The purpose is to enable preventive strategies and early interventions. Machine learning models integrate clinical variables, imaging data, and other biomarkers to enhance predictive accuracy [43]. The analysis of HRV temporal irreversibility during sleep has been proposed as an independent predictor of cardiovascular events. Similarly, combining machine learning with CCT data has shown promise in improving cardiac risk prediction [44]. Closely related to risk prediction is prognostication, which refers to forecasting the probable course or outcome of a disease after diagnosis. This includes predicting adverse cardiovascular events or mortality in patients with established disease [46]. Ischemia detected by stress cardiac magnetic resonance imaging (CMR) has been strongly associated with adverse cardiovascular outcomes, regardless of patient sex [47].

Machine learning algorithms have been widely utilized for the classification of cardiovascular conditions using HRV features derived from ECG signals [48]. Among these methods, the ensemble technique Random Forest (RF) is noted for its ability to handle the nonlinear patterns inherent in HRV data and achieved 95.1% accuracy in binary discrimination (normal vs diseased) and 84.8% accuracy when distinguishing among five specific heart-disease categories. The K-Nearest Neighbors (KNN) algorithm is also referenced in related studies focused on cardiovascular disease detection using ECG and biosignal data. For instance, KNN has been applied to classify the cardiac health status of drivers based on HRV parameters [49]. Decision Trees (DT) are among the machine learning techniques used in HRV-based heart disease detection, with one study reporting a classification accuracy of 92.86% when assessing the cardiovascular health status of drivers [50, 51].

Current research on cardiovascular disease detection using HRV and ECG-based machine learning methods faces several notable limitations. Small sample sizes and narrow population representations are common, reducing the statistical validity and strength of the findings [51]. Class imbalance within datasets, where the number of healthy versus diseased individuals or different disease subtypes is uneven, can impair model performance and introduce bias [48]. A major concern is the limited external validation of proposed models on independent datasets, which undermines their reliability and reproducibility [52]. Many algorithms are developed under controlled clinical conditions and lack applicability in real-world settings, where early diagnosis is crucial. The technical complexity and poor interpretability of extracted features also hinder the integration of such methods into clinical practice, posing challenges for trust and acceptance among healthcare professionals [48, 51].

Many studies lack adequate consideration of model uncertainty, which is a critical aspect when using machine learning algorithms. Neglecting this factor reduces the reliability of the results and limits their reproducibility [48]. Linear methods for HRV analysis have limited sensitivity and fail to detect complex patterns in physiological signals. Modified nonlinear methods demonstrate greater effectiveness in identifying atypical patterns, including U-shaped dependencies [18-20]. In a study focused on the diagnosis of IHD, the AUC value for exercise testing was significantly lower compared to results obtained through the analysis of volatile organic compounds (VOCs). This indicates a lower diagnostic relevance of exercise testing in this context [52]. Several studies focus only on short-term outcomes. A study aimed at predicting stroke outcomes based on HRV indicators did not consider long-term cardiovascular events.

While the link between Heart-Rate Variability (HRV) and cardiovascular disease is well established, a practical method for ischemic heart disease (IHD) screening with consumer-grade photoplethysmography (PPG) sensors is still lacking, and the most informative HRV biomarkers obtainable from such devices have yet to be defined. To address this gap, the present pilot introduces Zhurek, a fingertip PPG device designed in-house that performs on-board, real-time analysis of HRV metrics and securely streams the data to a cloud repository. Bench tests against a three-lead Holter ECG show clinically acceptable differences of -0.601 bpm for mean heart rate, $+33.1$ ms for SDNN, and -4.8 ms for RMSSD. Using Zhurek, HRV recordings were collected from 20 healthy volunteers and 20 angiographically confirmed IHD patients (18-70 years old) drawn from the Cardiology center in Almaty, Kazakhstan. Eight candidate features, SDNN, RMSSD, LF, HF, LF/HF, Max_HR, BMI, and age were explored. Mann-Whitney testing revealed significant group differences for

SDNN, LF, HF, Max_HR, BMI, and age with $p < 0.05$. Principal component analysis showed that the first two components explain 49.5% of total variance and already separate the cohorts without labels. CatBoost feature ranking identified LF power at about 44 percent importance, age at about 19 percent, and HF power as the most discriminative factors. The results of this pilot study demonstrate that measurements obtained with an affordable PPG device allow for the analysis of key HRV markers. While this method is not intended to replace the "gold standard" 24-hour Holter monitoring it shows the potential of low-cost wearable sensors for HRV analysis. This research lays the foundation for developing future scalable and cost-effective screening and monitoring systems for ischemic heart disease (IHD).

3. Materials and Methods

3.1. System Description

The proposed hybrid physiological monitoring system is engineered for continuous HRV analysis to enable the ambulatory assessment of autonomic nervous system status. The system integrates a wearable sensing device with on-device signal processing and remote data logging, as shown in Figure 1.

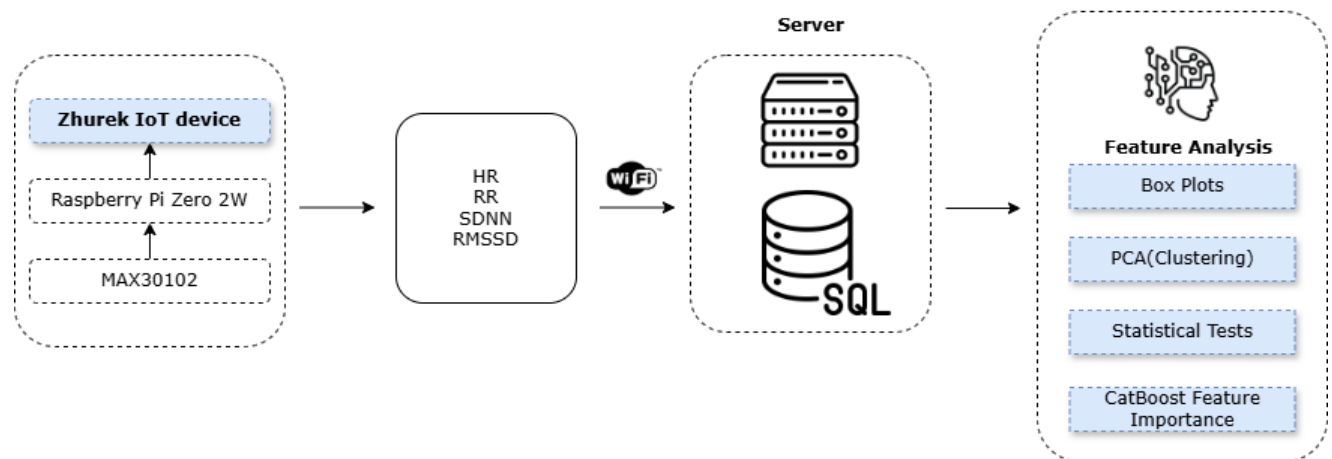


Figure 1.
Architecture and description of the system.

The overall system, as shown in Figure 1, is composed of three integrated subsystems: the sensing and processing module, the communication and storage layer, and the analytics and classification domain.

At the sensing module, the Zhurek IoT device is responsible for acquiring photoplethysmographic (PPG) signals and computing core HRV metrics in real time. The device captures fingertip-based PPG data, processes it locally to extract time-domain features, and prepares it for wireless transmission. The embedded software processes the raw signal in real time and extracts several well-established HRV metrics, including heart rate (HR), RR intervals, SDNN, and RMSSD. A detailed description of Zhurek's hardware and firmware architecture is provided in Section 3.2.

Processed HRV features are encapsulated in JavaScript Object Notation (JSON) format and transmitted via Wi-Fi using the MQTT protocol. The device publishes data to the topic zhurek/ppg/hrv, which is managed by Mosquitto 2.0 MQTT broker hosted on a centralized server. All communication is secured using TLS 1.3 with mutual certificate-based authentication to ensure data integrity and privacy.

At the storage layer, incoming MQTT messages are parsed and stored in a relational SQL database. Each record is time-stamped using a synchronized internal real-time clock, which is regularly updated via Network Time Protocol (NTP) to maintain temporal accuracy across devices. In parallel, the wearable device retains a local backup log in CSV format, providing redundancy in case of connectivity loss.

At the analytics domain, our focus was on exploring the relative importance of physiological and behavioral features in distinguishing between healthy individuals and patients with IHD, rather than predictive modeling. We employed a multifaceted approach, including statistical hypothesis testing with the Mann–Whitney U test to identify features with significant distributional differences, and Principal Component Analysis (PCA) as an unsupervised technique to explore latent structures and visual separation between groups. Specifically, CatBoost was utilized for feature importance analysis across the pilot dataset, which comprised 40 samples in total. Given the limited sample size, we deliberately excluded model performance metrics to avoid statistical overfitting and misinterpretation. Instead, the emphasis was on identifying which HRV and clinical variables, such as LF power, age, HF power, and Max_HR showed the strongest discriminative signals. This exploratory approach supports hypothesis generation and informs future large-scale studies aiming to develop robust, generalizable models for cardiovascular health monitoring.

In the analytics and classification domain, the extracted HRV features are utilized to predict the risk of autonomic dysfunction associated with IHD. The stored data is periodically processed using various machine learning algorithms, including gradient boosting methods (such as XGBoost and CatBoost), random forests (RF), interpretable generalized additive models (EBM), and hybrid architectures combining deep neural networks (DNN) with least-mean-square support vector machines (LMSVM). These models are trained on labeled datasets to classify patients by risk level and to detect

early patterns of dysfunction. This approach enables automated preliminary diagnostics and supports wellness assessment and risk stratification in remote monitoring scenarios.

This system enables continuous monitoring and structured data analysis by integrating embedded signal processing, secure wireless transmission, and modular analytics. The use of open-source tools and commercially available components supports reproducibility and facilitates deployment in remote monitoring scenarios.

3.2. Zhurek IoT Device

Zhurek, shown in Figure 2, is a custom-engineered, non-invasive wearable device designed to capture and process photoplethysmographic (PPG) signals in real time. Its compact form factor, self-contained electronics, and on-device analytics make it suitable for long-term ambulatory monitoring outside clinical environments.

The device integrates the MAX30102 optical sensor (DFRobot Gravity: SEN0344) with a Raspberry Pi Zero 2 W microcontroller (ARM Cortex-A53, 1 GHz, 512 MB RAM) running Raspberry Pi OS Lite (64-bit). Only the infrared channel is used for signal acquisition, sampled at 100 Hz over a hardware I²C bus (address 0x57). The sensor is enclosed in a 3D-printed PLA shell with an IR-shielded finger clip and soft elastomer padding to reduce motion artifacts and ambient interference.

All acquisition and processing code is implemented in Python 3.11, with `smbus2` used for I²C communication. The raw PPG signal undergoes baseline correction and smoothing (via moving average filtering). A derivative-based peak detection algorithm derived from HeartPy identifies cardiac cycles, with physiological validation applied to exclude outliers. RR intervals are calculated from peak timestamps, and time-domain HRV metrics, heart rate (HR), SDNN, and RMSSD, are computed in 30-s windows with a 5-s step. Frequency-domain metrics (LF, HF, LF/HF) and Max_HR are obtained offline, together with anthropometric features BMI and age, forming an eight-item feature vector.

Each result is serialized as a JSON object and published via MQTT. In addition to live data transmission, the device logs results locally in CSV format as a fallback mechanism. All data points are timestamped using a real-time clock (RTC) synchronized periodically via NTP.

Zhurek delivers optimal signal quality and high physiological accuracy under resting conditions. Resting monitoring reduces motion artefacts and yields stable autonomic patterns, ensuring reliable HRV computation. Validation results align with published evidence showing that resting acquisition provides the greatest accuracy and reproducibility for HRV, gas exchange, and metabolic rate measurements [53-55]. Remote HR and HRV recorded in this state correlate closely with ECG readings, and baseline metabolic rate, together with respiratory exchange ratio, remains stable and accurate during steady rest [55]. Physiological data collected under these conditions faithfully reflect autonomic nervous system activity and serve as a robust baseline for IHD risk surveillance.

The clinical utility and predictive value of numerous Heart Rate Variability (HRV) parameters are well-established in existing literature. Prior research [56] primarily using the gold-standard electrocardiograph (ECG), has confirmed that metrics such as SDNN, RMSSD, and the triangular index are significantly associated with patient outcomes. The primary challenge, however, lies in translating this diagnostic power from clinical-grade ECGs to convenient, non-invasive wearable devices.

To address this, the present validation study was conducted to assess the suitability of a modern wearable device for HRV analysis by comparing its readings against a reference clinical-grade three-lead Holter monitor. The results demonstrated a high degree of concordance between the two systems. Key time-domain parameters showed strong alignment, validating the tested device as a reliable surrogate for structured HRV analysis.

These findings are consistent with a growing body of research focused on validating PPG-based sensors. Authors of the study [56] also validated a wearable device, demonstrating that its HRV readings closely align with reference ECG data and can serve as a valid substitute for longer, standard measurements. Further reinforcing this, another study [57] found that the accuracy of certain PPG-derived HRV parameters could be adequate for patient monitoring, underscoring the importance of parameter-specific evaluation.

3.3. Data collection

To ensure the robust training and evaluation of machine learning models for IHD prediction, HRV data were collected from two independent cohorts: a clinical group of patients with confirmed cardiovascular diseases and a healthy control group. This dual-cohort design allows the models to learn patterns of autonomic dysfunction observed in real-world pathological states while distinguishing them from the normal physiological variability found in healthy individuals.

Monitoring of both cohorts was conducted using high-quality RR-interval acquisition devices to ensure the consistency and reliability of HRV measurements. Recordings from participants diagnosed with IHD and from healthy volunteers were collected under controlled laboratory conditions using a three-lead ECG device and our custom-built PPG-based Zhurek device. While ECG remains the "gold standard" for RR-interval detection, our study demonstrates that with proper signal processing and validation, PPG-based devices like Zhurek can provide HRV metrics with sufficient accuracy for machine learning-based risk stratification. Using both modalities within a unified framework helps to avoid data bias and reflects the practical realities of wearable cardiovascular monitoring.

3.3.1. Clinical Cohort

HRV data of adult inpatients with confirmed cardiovascular conditions were acquired at the Research Institute of Cardiology and Internal Diseases (Almaty, Kazakhstan). Diagnoses were established according to clinical protocols under

the supervision of the institute's cardiology department. Continuous recordings were collected using clinical-grade multi-lead Holter ECG monitors and the Zhurek device.

Participants represented both early and advanced stages of cardiovascular pathology. This broad distribution increases population heterogeneity and supports the development of generalizable machine-learning models. All recordings were stored as high-resolution numerical RR-interval files, and the resulting dataset already includes the key HRV variables heart rate (HR), RR intervals, SDNN, and RMSSD automatically computed and ready for downstream analysis.

3.3.2. Healthy Control Group

To establish a physiological baseline for HRV under normal autonomic conditions, data were collected from healthy volunteers. All participants reported no history of cardiovascular, neurological, or metabolic disorders. To minimize confounding factors, participants were instructed to abstain from alcohol, tobacco, caffeine, and intense physical activity for at least 24 hours prior to data collection, and to maintain regular sleep (7–8 hours) the night before. Participants were excluded if they had an acute illness, failed to meet the preparation criteria, or if signal recordings showed excessive artifacts.

Measurements were performed using a three-lead ECG Holter and the "Zhurek" IoT device, which integrates a MAX30102 PPG sensor with a Raspberry Pi Zero 2 W microcontroller. The infrared PPG channel was attached to the participant's fingertip using a shielded clip. Data were collected at 100 Hz via the hardware I²C interface. The device computed HRV features, including HR, RR intervals, SDNN, and RMSSD in real time using a 30-second rolling window. All measurements were stored in CSV format with high-resolution timestamps for later processing.

Demographic data for the healthy cohort and HRV feature distributions for this group are summarized in Table 1.

Table 1.
HRV Metrics – Healthy Group.

No.	Height	Weight	Age	BMI	Harmful habits	Gender	Genetic marker
1	168	63	19	22.3	No	Female	No
2	166	42	20	15.2	No	Female	No
3	150	82	21	36.4	No	Female	No
4	189	75	20	21	No	Male	No
5	192	102	20	27.7	Yes	Male	No
6	173	56	20	18.7	No	Male	No
7	179	78	19	24.3	No	Male	No
8	180	67	19	20.7	No	Male	No
9	175	69	20	21.5	No	Male	No
10	182	104	22	31.4	Yes	Male	No
11	162	65	18	24.8	No	Female	No
12	158	46	20	18.4	No	Female	No
13	180	74	20	22.8	No	Male	No
14	168	93	21	33	No	Female	No
15	170	50	17	17.3	No	Female	No
16	178	57	22	18	No	Male	No
17	184	102	19	30.1	Yes	Male	No
18	182	83	19	25.1	No	Male	No
19	168	53	20	18.8	No	Female	No
20	172	75	20	25.4	Yes	Female	No

3.4. Data Preprocessing

In this pilot study, data preprocessing involved the integration of two distinct cohorts: a healthy control group and a clinical group of patients diagnosed with IHD. The healthy group consisted of 20 adults aged between 18 and 22 years, while the IHD group comprised 20 patients selected from a larger dataset of exactly 300 individuals with confirmed IHD diagnosis, whose ages ranged from 18 to 92 years. For the purposes of this pilot study, only patients within the 18 to 71 age range were included. To ensure consistency in temporal resolution, one hour of continuous HRV data was extracted from each IHD patient's 24-hour ECG recording, aligning with the one-hour PPG-based recordings collected from the healthy cohort using the Zhurek device.

Several categorical variables were encoded numerically to facilitate analysis. Gender was binary-coded as Male (1) and Female (0). The "Bad Habits" variable was encoded as 1 for individuals who reported engaging in behaviors such as alcohol consumption, smoking, or regular intake of energy drinks, and 0 for those with no such behaviors. Genetic predisposition was captured through a "Genetic Marker" variable with the following encoding: 0 indicated no reported CVD history in family or a history limited to non-CVD conditions; 1 denoted a confirmed family history of cardiovascular conditions such as hypertension, IHD, myocardial infarction, or stroke; and 2 indicated a history of only non-cardiovascular illnesses among relatives. Surgical history was recorded in the "Operations" variable, with 0 representing no prior

operations, 1 indicating operations unrelated to cardiovascular pathology, and 2 signifying cardiovascular-related surgical interventions.

The pilot study's dataset included key physiological features such as SDNN, PNN50, RMSSD, LF, HF, and the LF/HF ratio all of which are established HRV metrics reflecting autonomic nervous system activity [58]. Additionally, Max_HR (maximum heart rate observed during recording) and BMI were included, with BMI calculated using the standard formula [59]:

$$BMI = \frac{Weight (kg)}{Height (m)^2} \quad (1)$$

All data were checked for consistency, and entries with missing, malformed, or outlier values were excluded. Non-numeric values were converted to numerical form where applicable, including the transformation of decimal commas to periods. This preprocessing ensured that the dataset was clean, consistent, and suitable for downstream statistical analysis and visualization.

3.5. Feature Analysis and Selection

Boxplots were initially applied to provide an intuitive and compact visualization of the distributional characteristics of physiological features across subject groups. Introduced by Tukey in 1977, the boxplot is a widely used exploratory tool for visualizing the five-number summary: the minimum, first quartile (Q1), median, third quartile (Q3), and maximum. Its ability to simultaneously present measures of central tendency and variability makes it particularly suitable for the rapid comparison of multiple distributions [60]. To further explore the underlying relationships among these variables, a correlation heatmap was subsequently employed.

A correlation heatmap is a visual representation of a correlation matrix, where each cell shows the strength and direction of a linear relationship between two variables. The correlation coefficient, denoted as r , ranges from -1 to $+1$: values near $+1$ indicate a strong positive correlation, values near -1 indicate a strong negative correlation, and values close to 0 suggest little or no linear association [61].

The correlation coefficient r between two variables x and y is calculated as:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

where x_i and y_i are individual sample points, \bar{x} and \bar{y} are the means of x and y , respectively, and n is the number of data points.

The correlation matrix R for n variables is given by:

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix} \quad (3)$$

where r_{ij} is the correlation coefficient between variable i and variable j .

A perfect self-correlation appears along the diagonal, where each variable is correlated with itself. This method allows for quick visual assessment of how variables interact or contrast within a dataset.

We applied PCA as an unsupervised technique to reduce the dimensionality of the feature space and explore latent structure within the dataset. PCA transforms correlated variables into a smaller set of orthogonal principal components that retain the majority of the dataset's variance, allowing for more compact and interpretable representations without relying on class labels. By projecting high-dimensional HRV and physiological features into a reduced space, PCA facilitates visual separation between groups and supports downstream analysis while minimizing noise and redundancy [62].

To assess whether the distributions of individual features differ significantly between healthy individuals and patients with IHD, the Mann–Whitney U test was employed. This non-parametric test is ideal for comparing two independent groups when the data may not follow a normal distribution, which is often the case in physiological and clinical datasets [63]. The core idea behind the test is to evaluate whether one group tends to have higher or lower values than the other. All observations from both groups are combined and ranked. The test statistic U is calculated by comparing the sum of ranks in each group. For a group of size n_1 with rank sum R_1 , the U statistic is computed as [64]:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad (4)$$

where n_1 and n_2 are the sample sizes of the two groups, and R_1 is the sum of the ranks for the first group. U_2 is calculated similarly. The smaller of U_1 and U_2 is used as the test statistic. The test was implemented using Python, specifically the `mannwhitneyu` function from the `scipy.stats` module. Healthy and IHD datasets were first loaded separately. For each feature, the Mann–Whitney U test was applied using a two-sided alternative hypothesis to determine whether its

distribution differed significantly between the two groups. The script reported the U statistic and associated p-value for each variable, and features with p-values below 0.05 were considered statistically significant [65].

Feature importance analysis using the CatBoost algorithm was conducted to explore which physiological and clinical variables contribute most to differentiating healthy individuals from patients with IHD. This was implemented as part of a pilot study, where the primary goal was not prediction but rather to identify features with the strongest discriminatory signals between the two groups. The pilot dataset consisted of 40 samples, comprising 20 healthy participants and 20 IHD patients. Given the limited sample size, we deliberately avoided model evaluation or performance metrics, as any results would be statistically unreliable and potentially misleading. Instead, the focus was on identifying promising features worth retaining and expanding upon in the larger cohort analysis.

4. Results

4.1. Participant Characteristics and HRV Data Collection

HRV data were collected from two distinct groups: individuals clinically diagnosed with IHD and healthy control participants with no known cardiovascular conditions. All subjects underwent standardized HRV recording sessions. Descriptive statistics of key HRV metrics for both the healthy control group and the IHD group are summarized in Table 2 and Table 3, respectively. The analyzed features include the time-domain measures SDNN, PNN50, and RMSSD, along with the frequency-domain measures LF, HF, and the LF/HF ratio. These values provide an overview of the autonomic nervous system activity in both populations and highlight potential differences in HRV patterns between healthy individuals and those with IHD [66].

Table 2.
Summary of HRV Metrics in Healthy Control Group.

No	SDNN	PNN50	RMSSD	LF	HF	LF/HF
1	72.4	24.45	46.4	0.08	0.06	1.39
2	53.93	15.59	37.76	0.06	0.04	1.54
3	39.33	1.71	18.63	0.05	0.03	1.54
4	47.82	5.9	26.55	0.06	0.04	1.53
5	68.21	36.37	58.38	0.06	0.09	0.71
6	65.65	20.89	43.3	0.07	0.07	1.01
7	95.05	40.89	64.28	0.09	0.09	1.02
.....						
.....						
.....						
19	93.95	33.4	58.35	0.1	0.09	1.18
20	33.96	7.7	23.57	0.03	0.03	0.99

Table 3.
Summary of HRV Metrics in IHD Group.

No.	SDNN	PNN50	RMSSD	LF	HF	LF/HF
1	89	36.34	56	0.37	0.35	1.05
2	99	1.26	18	0.30	0.16	1.84
3	80	2.08	22	0.19	0.16	1.20
4	61	1.02	15	0.18	0.12	1.55
5	124	16.67	41	0.44	0.26	1.67
6	69	7.47	31	0.29	0.27	1.08
7	88	2.79	28	0.25	0.19	1.30
.....						
.....						
.....						
19	72	15.63	39	0.23	0.22	1.04
20	59	3.17	23	0.26	0.26	1.00

As shown in Figure 2, the boxplots reveal distinct patterns across key physiological and HRV-related variables. The IHD group exhibits elevated values for SDNN, LF, and HF, indicating greater total variability and power in both LF and HF bands. In contrast, RMSSD, a marker of parasympathetic activity, shows different patterns [67] that were elevated in the healthy group, indicating enhanced autonomic adaptability, as higher RMSSD values are associated with greater flexibility of the autonomic nervous system compared to lower values [65]. The LF/HF ratio remains relatively stable across both groups, with no marked difference in median values. Max HR was observed to be higher in participants with

IHD. BMI values show mild intergroup variability, with a higher median in the IHD group, aligning with the established role of excess body weight in cardiovascular risk [68].

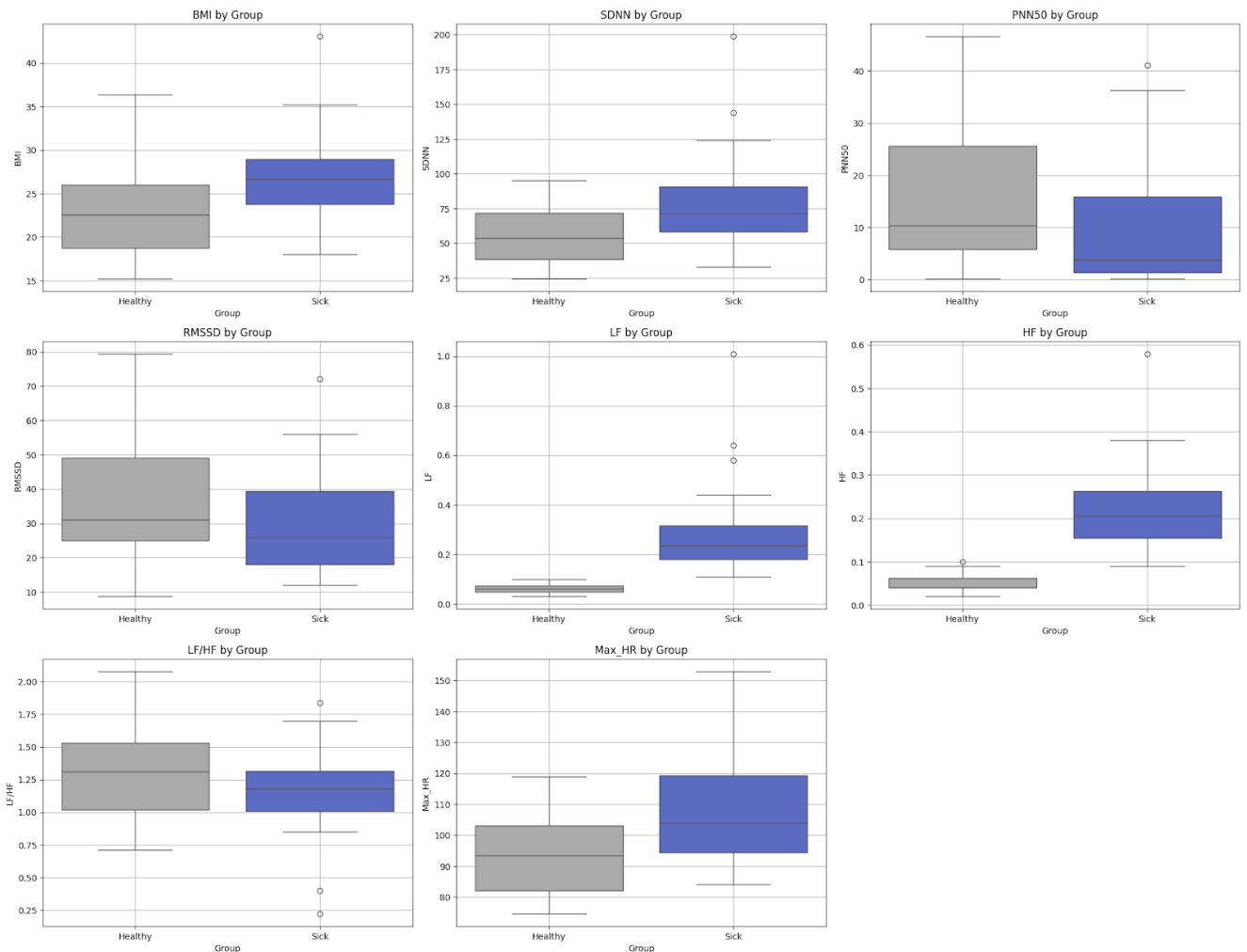


Figure 2.
Comparative distribution of physiological variables between healthy individuals and patients with IHD.

As illustrated in Figure 3, the distribution of categorical variables reveals key lifestyle and clinical differences between healthy individuals and patients with IHD. In the Genetic Marker category, the majority of healthy participants fall under class 0 (no reported family history of cardiovascular disease), whereas most IHD participants are classified as 1 (presence of cardiovascular disease in family history), supporting the role of heredity in CVD [69]. Regarding operations, healthy individuals predominantly report no surgical history (0), while IHD patients are more likely to have undergone cardiovascular-related (2) or non-cardiovascular (1) procedures, indicating greater clinical intervention in this group. In the bad habits category, unhealthy behaviors such as smoking, alcohol, or energy drink consumption (1) are more prevalent in the IHD group compared to the healthy cohort.

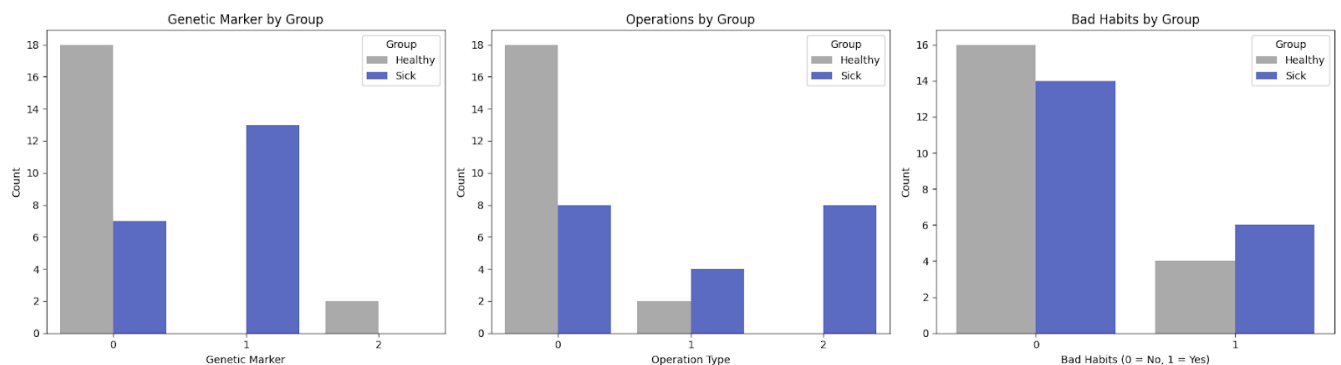


Figure 3.
Group-wise distribution of categorical variables related to genetic predisposition, surgical history, and behavioral risk factors.

The correlation heatmaps, as shown in Figure 4, illustrate the interrelationships between physiological, behavioral, and HRV features in healthy individuals and patients with IHD. The matrices reveal distinct patterns between the two cohorts, with healthy individuals exhibiting strong positive correlations among HRV features such as SDNN, RMSSD, and PNN50. In contrast, the IHD group exhibits a more disrupted correlation structure compared to the healthy group. Additionally, lifestyle factors such as unhealthy habits show positive correlations with impaired HRV in the IHD group, whereas they exhibit negative correlations with HRV features in the healthy group. As shown in the matrix, the diagonal line of red cells corresponds to perfect self-correlation, where each variable is perfectly correlated with itself.

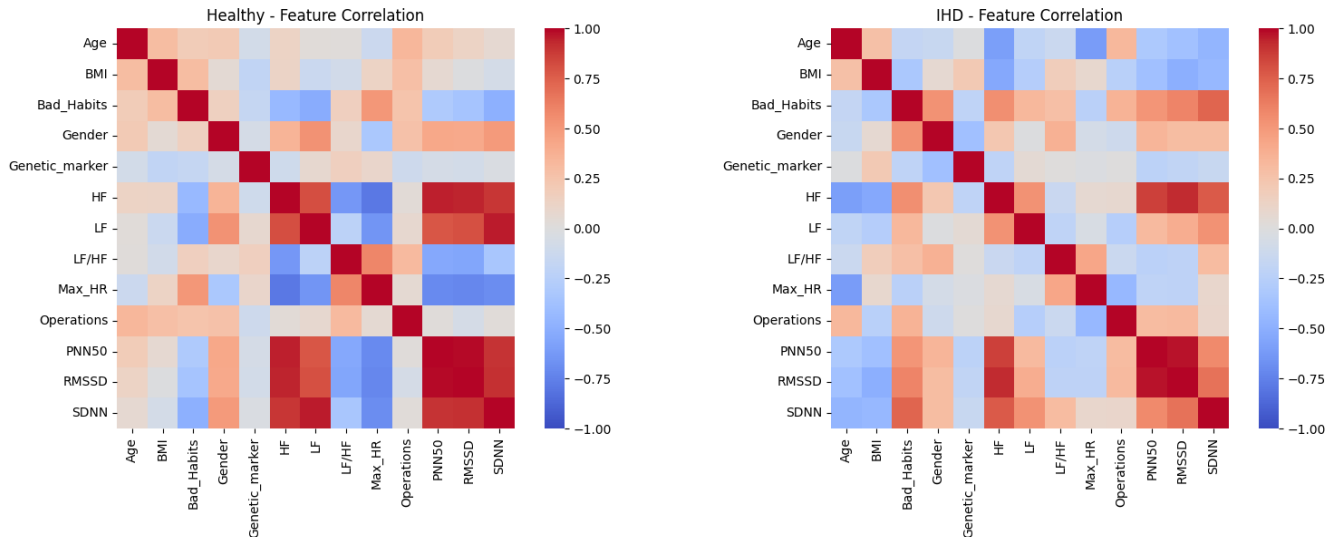


Figure 4. Correlation heatmaps of physiological and categorical features in healthy participants and patients with IHD.

Table 4 presents the results of the Mann–Whitney U test to evaluate whether the distributions of individual features differ significantly between healthy individuals and patients with IHD. Several features demonstrated statistically significant differences between the two groups, including BMI, SDNN, LF, HF, and Max_HR, all of which had p-values below 0.05. These findings highlight the discriminatory potential of autonomic nervous system activity and cardiovascular dynamics in distinguishing between healthy individuals and those with IHD. Other features, such as PNN50, RMSSD, and LF/HF ratio, did not reach statistical significance; however, they may still contribute valuable information in multivariate analyses or clinical interpretation.

Table 4. Statistical Comparison of Physiological Feature Distributions Between Healthy and IHD Groups Using the Mann–Whitney U Test.

Feature	p-value	Significance
BMI	0.042464	Yes
SDNN	0.025625	Yes
LF	6.49e-08	Yes
HF	9.47e-08	Yes
PNN50	0.072032	No
RMSSD	0.126377	No
LF/HF	0.432537	No
Max_HR	0.008343	Yes
Age	9.34e-07	Yes

To better understand the structure of the dataset and examine whether healthy individuals and IHD patients exhibit distinct physiological profiles, we applied PCA. For visualization purposes, we focused on the first two principal components, which together accounted for approximately 49.5% of the total variance, 27.3% from PC1 and 22.2% from PC2.

As illustrated in Figure 5, the projection onto the PC1–PC2 space reveals a visible separation between the two groups. Healthy participants (marked in grey) tend to cluster in a different region of the plot than those with IHD (shown in blue), suggesting that the original physiological and categorical features contain sufficient variation to differentiate the groups even in an unsupervised setting. This finding supports the relevance of the selected features and provides a strong rationale for their use in downstream supervised modeling and feature importance analysis.

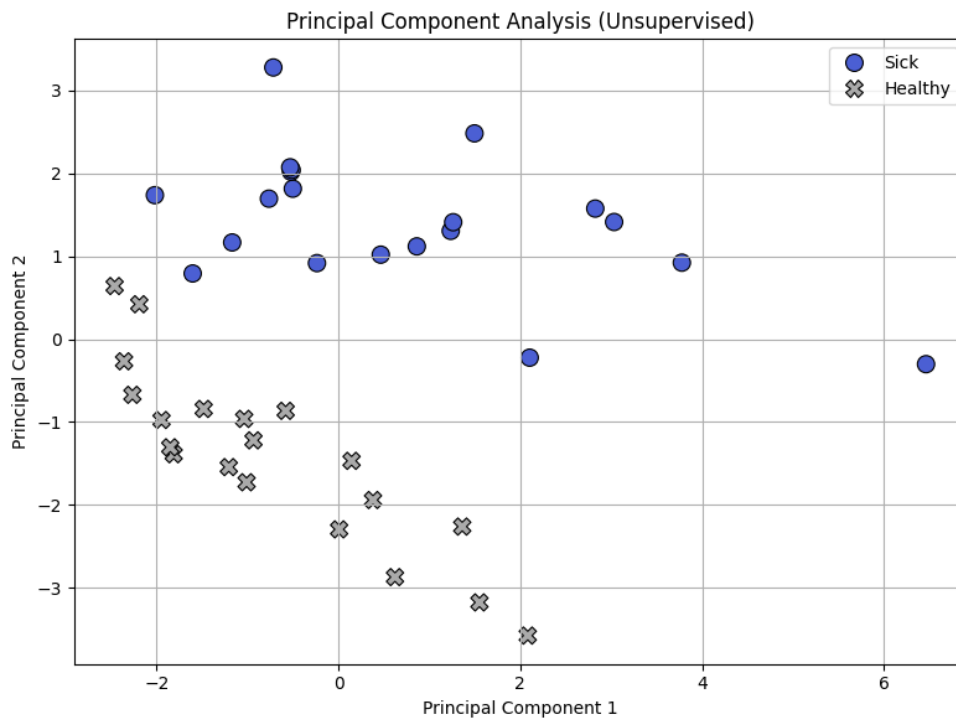


Figure 5.
Principal Component Analysis (PCA) of Physiological Features in Healthy and IHD Groups.

As part of our exploratory analysis, we examined the most influential features identified by the CatBoost algorithm to understand which physiological and clinical variables may contribute to group separation. While predictive performance was not the focus at this stage, this preliminary assessment highlights variables with the highest potential for distinguishing between healthy individuals and IHD patients.

As shown in Figure 6, the CatBoost classifier identified the five most important features for distinguishing between healthy individuals and patients with IHD. The LF component of HRV emerged as the top-ranked feature, accounting for approximately 44% of the total feature importance. Age followed with a relative importance of about 19%, while the high-frequency HF HRV component ranked third. Two additional features, Max_HR and tRMSSD, each contributed approximately 2–3% to the model's overall discriminative performance.

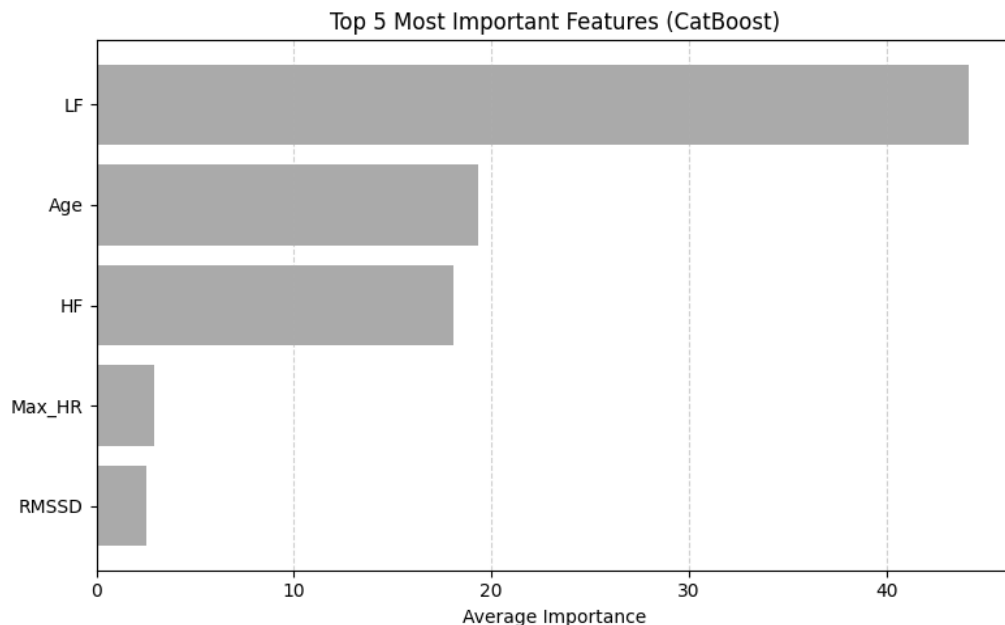


Figure 6.
Average Feature Importance Scores from CatBoost Across 5-Fold CV (Pilot Dataset).

To assess the model's reliance solely on physiological features, the age variable was removed, and the CatBoost classifier was applied to the pilot dataset. As no test set was used, the goal was not to evaluate predictive performance but to identify the most influential contributors to the model's internal decision process. As shown in Figure 7, the LF component of HRV emerged as the most important feature, accounting for approximately 54.4% of the total importance.

The HF component followed with 25.6%, indicating the model's sensitivity to autonomic nervous system indicators. Three additional features, Max_HR, RMSSD, and PNN50, each contributed between 2.8% and 3.1%, indicating a secondary but still relevant role in the model's discriminative capacity.

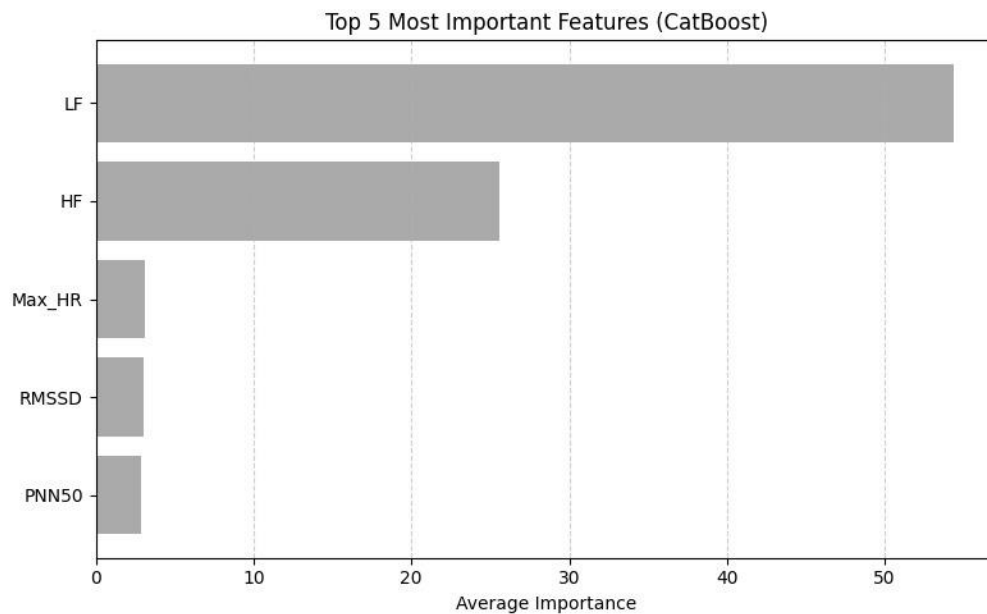


Figure 7.
Average Feature Importance Scores from CatBoost Across 5-Fold CV without age (Pilot Dataset).

5. Discussion

One of the primary objectives of this pilot study was to evaluate the performance and reliability of the Zhurek IoT device, a non-invasive tool engineered to capture and process PPG signals in real time under real-world, ambulatory conditions. HRV data from IHD patients and the healthy group were collected using the Zhurek device and standard three-lead Holter ECG monitors in a clinical setting. Several key indicators, such as SDNN, LF, and Max_HR showed statistically significant intergroup differences, supporting the diagnostic sensitivity of the "Zhurek" device. Moreover, principal component analysis (PCA) revealed a clear separation between healthy participants and IHD patients in the reduced-dimensional feature space, reinforcing the quality and informativeness of the PPG-derived signals. Importantly, the use of the Zhurek device did not introduce signal noise or measurement artifacts substantial enough to obscure group-level distinctions or lead to misclassification when compared with CVD patients. These findings confirm that our "Zhurek" device can be trusted to generate meaningful cardiovascular data and is well-suited for scaling up to larger population studies.

The prominence of LF as the most influential variable suggests that autonomic nervous system modulation, particularly sympathetic activity, plays a significant role [70] may be a key physiological marker differentiating individuals with IHD from healthy controls. Age ranking second aligns with well-established clinical evidence linking increasing age to elevated cardiovascular risk [71]. The HF power, typically associated with parasympathetic (vagal) activity, was also a major contributor [70]. Although Max_HR contributed relatively less to the model, it still ranked within the top five, suggesting that transient HRV and beat-to-beat interval variations carry some diagnostic weight, even if less strongly predictive on its own. RMSSD did not reach statistical significance in the Mann–Whitney U test; however, it was ranked among the top five most important features by the CatBoost model. This indicates that, while its standalone distribution may not differ substantially between groups, RMSSD plays a meaningful role when analyzed alongside other variables. Its contribution in the multivariate setting underscores the value of using machine learning to uncover nuanced patterns not captured by traditional statistical methods. To further explore the model's behavior without demographic influence, we repeated the feature importance analysis after removing the age variable. In this setting, LF remained the most dominant contributor, reinforcing its role as a key physiological marker. HF followed as the next most important feature, again highlighting the relevance of autonomic regulation.

Some limitations should be acknowledged in the current pilot study. The data for the two groups were collected using different protocols: PPG recordings for healthy individuals and ECG monitoring for patients, creating variability in signal quality that could bias the classifier. ECG captures the heart's electrical activity using electrodes placed on the skin surface [26]. Comparative studies consistently show that ECG outperforms other modalities in terms of signal quality and respiratory signal extraction, and even in scenarios involving spontaneous breathing, it demonstrates slightly better fidelity [72]. However, ECG is prone to interference and motion artifacts when skin contact is unstable [73]. In contrast, PPG is an optical technique that measures volumetric changes in blood flow within the microvascular bed using light absorption or reflection [26]. PPG sensors are typically placed on peripheral sites such as the fingertips, earlobes, or wrists, where they detect changes in light absorption corresponding to blood volume pulses [74]. Since PPG captures hemodynamic changes indirectly, factors such as poor peripheral perfusion or incorrect sensor placement can substantially degrade signal quality and reduce diagnostic specificity [26]. Despite these limitations, research suggests that PPG-based devices can perform

comparably to ECG in specific applications like atrial fibrillation detection, especially when advanced signal processing and algorithmic enhancements are applied [75].

Another notable limitation of this pilot study is the relatively narrow age distribution among participants, particularly within the healthy control group, which was largely composed of younger individuals. Despite this, age emerged as the second most important discriminative feature in the CatBoost model, highlighting its strong influence on HRV-related cardiovascular differences. This suggests that age may act as a confounding factor and, if not properly accounted for, could bias the interpretation of group differences. Therefore, in the larger-scale study, it will be essential to recruit a more age-diverse sample across the healthy groups.

6. Conclusion

The pilot study demonstrated that PPG recordings obtained with the "Zhurek" device accurately reproduce HRV metrics with clinically acceptable precision. In comparative assessments, the deviation from the three-lead Holter monitor was -0.601 bpm for mean heart rate, +33.1 ms for SDNN, and -4.8 ms for RMSSD. HRV signals were collected from 20 healthy volunteers and 20 patients with angiographically confirmed ischemic heart disease (IHD), and a total of 8 features were calculated. These features included SDNN, RMSSD, LF, HF, LF/HF, Max HR, BMI, and age.

The Mann–Whitney test revealed statistically significant differences for SDNN, LF, HF, Max_HR, BMI, and age with a p-value < 0.05. The first two principal components explained 49.5% of the total variance and separated the groups without labels, confirming the informativeness of the selected features. The CatBoost algorithm identified that LF made the largest contribution (about 44%), followed by age (about 19%) and HF, while Max_HR and RMSSD had a smaller impact. Thus, short sessions using the Zhurek device allow for the detection of autonomic markers previously available only through 24-hour monitoring and lay the foundation for scalable and affordable screening of ischemic heart disease.

Several factors limit the interpretation of these results. The control group had a narrow age range, yet age proved to be one of the main discriminative features, highlighting the risk of confounding effects. The sample size was limited to forty observations, so operational model metrics were not calculated to avoid overfitting and misleading conclusions.

Future research plans aim to expand the cohort to increase its size, and a multicenter longitudinal study is planned to evaluate the prognostic value of HRV indicators and to confirm the reproducibility of the identified features. The research plan also includes the integration of additional nonlinear HRV metrics and the expansion of automated analysis to ensure reliable risk stratification in ambulatory monitoring. Implementing these steps will enhance the clinical relevance of the Zhurek method and bring the transition from reactive treatment to proactive prevention of IHD closer to reality.

References

- [1] World Health Organization, *Cardiovascular diseases (CVDs)*. Geneva, Switzerland: World Health Organization, 2024.
- [2] M. A. Khan *et al.*, "Global epidemiology of ischemic heart disease: Results from the global burden of disease study," *Cureus*, vol. 12, no. 7, 2020.
- [3] k. Tengrinews, "The most common disease among Kazakhstanis has been named," Almaty, Kazakhstan, 2022. https://tengrinews.kz/kazakhstan_news/nazvana-samaya-rasprostranennaya-bolezn-sredi-kazahstantsev-503527/
- [4] P. Severino *et al.*, "Ischemic heart disease pathophysiology paradigms overview: From plaque activation to microvascular dysfunction," *International Journal of Molecular Sciences*, vol. 21, no. 21, p. 8118, 2020. <https://doi.org/10.3390/ijms21218118>
- [5] Ł. J. Janicki, W. Leoński, J. S. Janicki, M. Nowotarski, M. Dziuk, and R. Piotrowicz, "Comparative analysis of the diagnostic effectiveness of satro ECG in the diagnosis of ischemia diagnosed in myocardial perfusion scintigraphy performed using the spect method," *Diagnostics*, vol. 12, no. 2, p. 297, 2022. <https://doi.org/10.3390/diagnostics12020297>
- [6] Ş.-T. Duca *et al.*, "Enhancing comprehensive assessments in chronic heart failure caused by ischemic heart disease: The diagnostic utility of holter ECG parameters," *Medicina*, vol. 60, no. 8, p. 1315, 2024. <https://doi.org/10.3390/medicina60081315>
- [7] R. Banerjee, A. Ghose, and K. Muthana Mandana, "A hybrid CNN-LSTM architecture for detection of coronary artery disease from ECG," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020.
- [8] P. Ribeiro, J. Sá, D. Paiva, and P. M. Rodrigues, "Cardiovascular diseases diagnosis using an ECG multi-band non-linear machine learning framework analysis," *Bioengineering*, vol. 11, no. 1, p. 58, 2024. <https://doi.org/10.3390/bioengineering11010058>
- [9] S. P. Gaine *et al.*, "Multimodality imaging in the detection of ischemic heart disease in women," *Journal of Cardiovascular Development and Disease*, vol. 9, no. 10, p. 350, 2022. <https://doi.org/10.3390/jcdd9100350>
- [10] L. Wang, T. Bi, J. Hao, and T. H. Zhou, "Heart diseases recognition model based on HRV feature extraction over 12-lead ECG signals," *Sensors*, vol. 24, no. 16, p. 5296, 2024. <https://doi.org/10.3390/s24165296>
- [11] G. Doolub *et al.*, "Artificial intelligence as a diagnostic tool in non-invasive imaging in the assessment of coronary artery disease," *Medical Sciences*, vol. 11, no. 1, p. 20, 2023. <https://doi.org/10.3390/medsci11010020>
- [12] L. Verma and S. Srivastava, "A data mining model for coronary artery disease detection using noninvasive clinical parameters," *Indian Journal of Science and Technology*, vol. 9, no. 48, pp. 1-6, 2016.
- [13] M. Sayadi, V. Varadarajan, F. Sadoughi, S. Chopannejad, and M. Langarizadeh, "A machine learning model for detection of coronary artery disease using noninvasive clinical parameters," *Life*, vol. 12, no. 11, p. 1933, 2022. <https://doi.org/10.3390/life12111933>
- [14] A. M. Alaa, T. Bolton, E. Di Angelantonio, J. H. F. Rudd, and M. van der Schaar, "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants," *PLOS ONE*, vol. 14, no. 5, p. e0213653, 2019. <https://doi.org/10.1371/journal.pone.0213653>

- [15] R. Alizadehsani *et al.*, "Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 119-127, 2018. <https://doi.org/10.1016/j.cmpb.2018.05.009>
- [16] C. Brinza *et al.*, "Heart rate variability in acute myocardial infarction: Results of the heart-v-ami single-center cohort study," *Journal of Cardiovascular Development and Disease*, vol. 11, no. 8, p. 254, 2024. <https://doi.org/10.3390/jcdd11080254>
- [17] A. Hazra, S. K. Mandal, A. Gupta, A. Mukherjee, and A. Mukherjee, "Heart disease diagnosis and prediction using machine learning and data mining techniques: A review," *Advances in Computational Sciences and Technology*, vol. 10, no. 7, pp. 2137-2159, 2017.
- [18] M. Trigka and E. Dritsas, "Long-term coronary artery disease risk prediction with machine learning models," *Sensors*, vol. 23, no. 3, p. 1193, 2023. <https://doi.org/10.3390/s23031193>
- [19] B. A. Marzooq *et al.*, "Machinelearning model discriminate ischemic heart disease using breathome analysis," *Biomedicines*, vol. 12, no. 12, p. 2814, 2024. <https://doi.org/10.3390/biomedicines12122814>
- [20] G. Sibrecht, J. Piskorski, T. Krauze, and P. Guzik, "Heart rate asymmetry, its compensation, and heart rate variability in healthy adults during 48-h holter ECG recordings," *Journal of Clinical Medicine*, vol. 12, no. 3, p. 1219, 2023. <https://doi.org/10.3390/jcm12031219>
- [21] X. Wu, Q. Yang, J. Li, and F. Hou, "Investigation on the prediction of cardiovascular events based on multi-scale time irreversibility analysis," *Symmetry*, vol. 13, no. 12, p. 2424, 2021.
- [22] B. M. Curtis and J. H. O'Keefe, Jr., "Autonomic tone as a cardiovascular risk factor: The dangers of chronic fight or flight," *Mayo Clinic Proceedings*, vol. 77, no. 1, pp. 45-54, 2002. <https://doi.org/10.4065/77.1.45>
- [23] R. A. Rather and V. Dhawan, "Genetic markers: Potential candidates for cardiovascular disease," *International Journal of Cardiology*, vol. 220, pp. 914-923, 2016. <https://doi.org/10.1016/j.ijcard.2016.06.251>
- [24] S. De Rosa, B. Arcidiacono, E. Chiefari, A. Brunetti, C. Indolfi, and D. P. Foti, "Type 2 diabetes mellitus and cardiovascular disease: Genetic and epigenetic links," *Frontiers in Endocrinology*, Review vol. Volume 9 - 2018, 2018. <https://doi.org/10.3389/fendo.2018.00002>
- [25] J. L. Moraes, M. X. Rocha, G. G. Vasconcelos, J. E. Vasconcelos Filho, V. H. C. De Albuquerque, and A. R. Alexandria, "Advances in photoplethysmography signal analysis for biomedical applications," *Sensors*, vol. 18, no. 6, p. 1894, 2018. <https://doi.org/10.3390/s18061894>
- [26] M. Elgendi *et al.*, "The use of photoplethysmography for assessing hypertension," *npj Digital Medicine*, vol. 2, no. 1, p. 60, 2019/06/26 2019. <https://doi.org/10.1038/s41746-019-0136-7>
- [27] M. A. Almarshad, M. S. Islam, S. Al-Ahmadi, and A. S. BaHammam, "Diagnostic features and potential applications of ppg signal in healthcare: A systematic review," *Healthcare*, vol. 10, no. 3, p. 547, 2022. <https://doi.org/10.3390/healthcare10030547>
- [28] K. B. Kim and H. J. Baek, "Photoplethysmography in wearable devices: A comprehensive review of technological advances, current challenges, and future directions," *Electronics*, vol. 12, no. 13, p. 2923, 2023. <https://doi.org/10.3390/electronics12132923>
- [29] M. Shabaan *et al.*, "Survey: smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, p. 177, 2020. <https://doi.org/10.1186/s12911-020-01199-7>
- [30] I. C. Dipto, T. Islam, H. M. Rahman, and M. A. Rahman, "Comparison of different machine learning algorithms for the prediction of coronary artery disease," *Journal of Data Analysis and Information Processing*, vol. 8, no. 2, pp. 41-68, 2020.
- [31] R. C. Ripan *et al.*, "Adata-driven heart disease prediction model through k-means clustering-based anomaly detection," *SN Computer Science*, vol. 2, no. 2, p. 112, 2021. <https://doi.org/10.1007/s42979-021-00518-7>
- [32] V. Jahmunah, E. Y. K. Ng, T. R. San, and U. R. Acharya, "Automated detection of coronary artery disease, myocardial infarction and congestive heart failure using GaborCNN model with ECG signals," *Computers in Biology and Medicine*, vol. 134, p. 104457, 2021. <https://doi.org/10.1016/j.combiomed.2021.104457>
- [33] K. Kenya *et al.*, "A deep learning approach for assessment of regional wall motion abnormality from echocardiographic images," *JACC: Cardiovascular Imaging*, vol. 13, no. 2_Part_1, pp. 374-381, 2020. <https://doi.org/10.1016/j.jcmg.2019.02.024>
- [34] M. Zreik *et al.*, "Deep learning analysis of coronary arteries in cardiac CT angiography for detection of patients requiring invasive coronary angiography," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1545-1557, 2020. <https://doi.org/10.1109/TMI.2019.2953054>
- [35] A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, and S. N. Qasem, "Machinelearning-based predictive models for detection of cardiovascular diseases," *Diagnostics*, vol. 14, no. 2, p. 144, 2024. <https://doi.org/10.3390/diagnostics14020144>
- [36] A. H. Khandoker, Y. Al Zaabi, and H. F. Jelinek, "What can tone and entropy tell us about risk of cardiovascular diseases?," in *2019 Computing in Cardiology (CinC)*, 2019: IEEE, 2019.
- [37] S. E. Chiuve *et al.*, "Lifestyle-based prediction model for the prevention of CVD," *Journal of the American Heart Association*, vol. 3, no. 6, p. e000954, 2014. <https://doi.org/10.1161/JAHA.114.000954>
- [38] M. Roerecke and J. Rehm, "Alcohol consumption, drinking patterns, and ischemic heart disease: A narrative review of meta-analyses and a systematic review and meta-analysis of the impact of heavy drinking occasions on risk for moderate drinkers," *BMC Medicine*, vol. 12, no. 1, p. 182, 2014. <https://doi.org/10.1186/s12916-014-0182-6>
- [39] R. Ng, R. Sutradhar, Z. Yao, W. P. Wodchis, and L. C. Rosella, "Smoking, drinking, diet and physical activity—modifiable lifestyle risk factors and their associations with age to first chronic disease," *International Journal of Epidemiology*, vol. 49, no. 1, pp. 113-130, 2019. <https://doi.org/10.1093/ije/dyz078>
- [40] F. Buccelletti *et al.*, "Heart rate variability and myocardial infarction: Systematic literature review and metanalysis," *European Review for Medical & Pharmacological Sciences*, vol. 13, no. 4, 2009.
- [41] M. Perulli *et al.*, "Short- vs long-term assessment of heart rate variability: Clinical significance in Dravet Syndrome," *Epilepsy & Behavior*, vol. 146, 2023. <https://doi.org/10.1016/j.yebeh.2023.109357>
- [42] A. Voss *et al.*, "Short-term versus long-term heart rate variability in ischemic cardiomyopathy risk stratification," (in English), *Frontiers in Physiology*, vol. Volume 4 - 2013, 2013. <https://doi.org/10.3389/fphys.2013.00364>

- [43] R. K. Agrawal, R. R. Sewani, D. Delen, and B. Benjamin, "A machine learning approach for classifying healthy and infarcted patients using heart rate variabilities derived vector magnitude," *Healthcare Analytics*, vol. 2, p. 100121, 2022. <https://doi.org/10.1016/j.health.2022.100121>
- [44] G. Georgieva-Tsaneva and E. Gospodinova, "Heart rate variability analysis of healthy individuals and patients with ischemia and arrhythmia," *Diagnostics*, vol. 13, no. 15, p. 2549, 2023. <https://doi.org/10.3390/diagnostics13152549>
- [45] G. Georgieva-Tsaneva, K. Cheshmedzhiev, Y.-A. Tsanev, M. Dechev, and E. Popovska, "Health care monitoring using an internet of things-based cardio system," *IoT*, vol. 6, no. 1, p. 10, 2025. <https://doi.org/10.3390/iot6010010>
- [46] W. Liu *et al.*, "Relationship between heart rate variability traits and stroke: A Mendelian randomization study," *Journal of Stroke and Cerebrovascular Diseases*, vol. 34, no. 4, 2025. <https://doi.org/10.1016/j.jstrokecerebrovasdis.2025.108251>
- [47] J. Aftyka *et al.*, "The hemisphere of the brain in which a stroke has occurred visible in the heart rate variability," *Life*, vol. 12, no. 10, p. 1659, 2022. <https://doi.org/10.3390/life12101659>
- [48] J. Aftyka, J. Staszewski, A. Dębiec, A. Pogoda-Wesołowska, and J. Żebrowski, "Can HRV predict prolonged hospitalization and favorable or unfavorable short-term outcome in patients with acute ischemic stroke?," *Life*, vol. 13, no. 4, p. 856, 2023. <https://doi.org/10.3390/life13040856B>
- [49] G. Chairina, K. Yoshino, K. Kiyono, and E. Watanabe, "Ischemic stroke risk assessment by multiscale entropy analysis of heart rate variability in patients with persistent atrial fibrillation," *Entropy*, vol. 23, no. 7, p. 918, 2021. <https://doi.org/10.3390/e23070918>
- [50] I. Marcantoni, E. Iammarino, A. Dell'Orletta, and L. Burattini, "Prognostic role of electrocardiographic alternans in ischemic heart disease," *Journal of Clinical Medicine*, vol. 14, no. 8, p. 2620, 2025. <https://doi.org/10.3390/jcm14082620>
- [51] D. K. Kaufmann, G. Raczak, M. Szwoch, E. Wabich, M. Świątczak, and L. Daniłowicz-Szymanowicz, "Baroreflex sensitivity but not microvolt T-wave alternans can predict major adverse cardiac events in ischemic heart failure," *Cardiology Journal*, vol. 29, no. 6, pp. 1004-1012, 2022.
- [52] M.-J. Wu *et al.*, "Exploring relationships of heart rate variability, neurological function, and clinical factors with mortality and behavioral functional outcome in patients with ischemic stroke," *Diagnostics*, vol. 14, no. 12, p. 1304, 2024. <https://doi.org/10.3390/diagnostics14121304>
- [53] R. M. Birn *et al.*, "The influence of physiological noise correction on test-retest reliability of resting-state functional connectivity," *Brain Connectivity*, vol. 4, no. 7, pp. 511-522, 2014. <https://doi.org/10.1089/brain.2014.0284>
- [54] G. Sanchez-Delgado *et al.*, "Reliability of resting metabolic rate measurements in young adults: Impact of methods for data analysis," *Clinical Nutrition*, vol. 37, no. 5, pp. 1618-1624, 2018. <https://doi.org/10.1016/j.clnu.2017.07.026>
- [55] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014.
- [56] A. Taoum, A. Bisiaux, F. Tilquin, Y. Le Guillou, and G. Carrauld, "Validity of ultra-short-term hrv analysis using ppg—a preliminary study," *Sensors*, vol. 22, no. 20, p. 7995, 2022. <https://doi.org/10.3390/s22207995>
- [57] C. Hoog Antink, Y. Mai, M. Peltokangas, S. Leonhardt, N. Oksala, and A. Vehkaoja, "Accuracy of heart rate variability estimated with reflective wrist-PPG in elderly vascular patients," *Scientific Reports*, vol. 11, no. 1, p. 8123, 2021/04/14 2021. <https://doi.org/10.1038/s41598-021-87489-0>
- [58] N. Gullett, Z. Zajkowska, A. Walsh, R. Harper, and V. Mondelli, "Heart rate variability (HRV) as a way to understand associations between the autonomic nervous system (ANS) and affective states: A critical review of the literature," *International Journal of Psychophysiology*, vol. 192, pp. 35-42, 2023. <https://doi.org/10.1016/j.ijpsycho.2023.08.001>
- [59] A. Zierle-Ghosh and A. Jan, *Physiology, body mass index. Instat pearls*. Treasure Island, FL, USA.: Stat Pearls Publishing, 2023.
- [60] M. Abt, T. Leuders, K. Loibl, and F. Reinhold, "Developing initial notions of variability when learning about box plots," *Mathematical Thinking and Learning*, pp. 1-24. <https://doi.org/10.1080/10986065.2024.2421412>
- [61] N. K. Singh and M. Nagahara, "LightGBM-, SHAP-, and correlation-matrix-heatmap-based approaches for analyzing household energy data: Towards electricity self-sufficient houses," *Energies*, vol. 17, no. 17, p. 4518, 2024. <https://doi.org/10.3390/en17174518>
- [62] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An overview of principal component analysis," *Journal of signal and information processing*, vol. 4, no. 3, pp. 173-175, 2013. <https://doi.org/10.4236/jsip.2013.43b031>
- [63] C. Kühnast and M. Neuhäuser, "A note on the use of the non-parametric wilcoxon-mann-whitney test in the analysis of medical studies," *GMS German Medical Science*, vol. 6, p. Doc02, 2008.
- [64] W. R. Emerson, "Mann-Whitney U test and t-test," *Journal of Visual Impairment & Blindness*, vol. 117, no. 1, pp. 99-100, 2023. <https://doi.org/10.1177/0145482x221150592>
- [65] J. Shreffler and M. R. Huecker, "Hypothesis testing, P values, confidence intervals, and significance," in *Stat Pearls*. Treasure Island, FL: Stat Pearls Publishing, 2023.
- [66] B. Vandenberk *et al.*, "Repeating noninvasive risk stratification improves prediction of outcome in ICD patients," *Annals of Noninvasive Electrophysiology*, vol. 25, no. 6, p. e12794, 2020. <https://doi.org/10.1111/anec.12794>
- [67] J. Zeng *et al.*, "High vagally mediated resting-state heart rate variability is associated with superior working memory function," *Frontiers in Neuroscience*, vol. Volume 17 - 2023, 2023. <https://doi.org/10.3389/fnins.2023.1119405>
- [68] M. Volpe *et al.*, "How cardiologists can manage excess body weight and related cardiovascular risk. An expert opinion," *International Journal of Cardiology*, vol. 381, pp. 101-104, 2023. <https://doi.org/10.1016/j.ijcard.2023.03.054>
- [69] S. Kathiresan and D. Srivastava, "Genetics of Human Cardiovascular Disease," *Cell*, vol. 148, no. 6, pp. 1242-1257, 2012. <https://doi.org/10.1016/j.cell.2012.03.001>
- [70] J. R. Ryder *et al.*, "Impaired cardiac autonomic nervous system function is associated with pediatric hypertension independent of adiposity," *Pediatric Research*, vol. 79, no. 1, pp. 49-54, 2016. <https://doi.org/10.1038/pr.2015.188>
- [71] R. Dhingra and R. S. Vasan, "Age as a risk factor," *Medical Clinics of North America*, vol. 96, no. 1, pp. 87-91, 2012. <https://doi.org/10.1016/j.mcna.2011.11.003>
- [72] C. Orphanidou, "Derivation of respiration rate from ambulatory ECG and PPG using ensemble empirical mode decomposition: Comparison and fusion," *Computers in Biology and Medicine*, vol. 81, pp. 45-54, 2017. <https://doi.org/10.1016/j.combiomed.2016.12.005>

- [73] P. H. Charlton *et al.*, "Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp. 2-20, 2018. <https://doi.org/10.1109/RBME.2017.2763681>
- [74] T. Leppänen *et al.*, "Pulse oximetry: The working principle, signal formation, and applications," in *Advances in the Diagnosis and Treatment of Sleep Apnea : Filling the Gap Between Physicians and Engineers*, T. Penzel and R. Hornero Eds. Cham: Springer International Publishing, 2022, pp. 205-218.
- [75] H. Gruwez *et al.*, "Accuracy of physicians interpreting photoplethysmography and electrocardiography tracings to detect atrial fibrillation: INTERPRET-AF," *Frontiers in Cardiovascular Medicine*, vol. 8, p. 734737, 2021. <https://doi.org/10.3389/fcvm.2021.734737>