IRSTI 28.23.15 https://doi.org/10.26577/jpcsit202542

Nurdaulet Tasmurzayev[1*], Dinara Turmakhanbet[1], Adilet Kakharov[1], Mukhamejan Aitkazin[1], Aliya Baidauletova[1], Mergul Kozhamberdiyeva[1]

Al-Farabi Kazakh National University, Almaty, Kazakhstan
*e-mail: tasmurzayev.n@gmail.com

CVD PREDICTION FROM HRV DERIVED FROM WEARABLE PPG

Abstract. Cardiovascular disease is the leading global cause of death; ischemic heart disease (IHD) is its most common and lethal form, motivating scalable, non-invasive screening. We tested whether a single 60-minute photoplethysmography (PPG) recording from the Zhurek fingertip wearable can distinguish healthy autonomic control from IHD-related dysregulation. Agreement with a three-lead Holter reference was clinically acceptable (HR −0.601 bpm; SDNN +33.1 ms; RMSSD −4.8 ms). Forty hour-long sessions were analyzed (20 healthy, 18–22 years; 20 angiography-confirmed IHD) using eight HRV/demographic features. Mann–Whitney tests showed significant differences for SDNN, LF, HF, Max_HR, BMI, and age ($p < 0.05$), and a two-component PCA (49.5% variance) separated cohorts without labels. SHAP for a CatBoost model highlighted LF and age as strongest positive contributors and HF as protective. Thus, one-hour PPG preserves diagnostically useful autonomic signatures, enabling ~24× shorter monitoring than Holter and supporting scalable ambulatory IHD risk stratification.

Keywords: CVD, IHD, HRV, Machine learning, PPG, wearable sensor.

1. Introduction

Cardiovascular diseases (CVDs) remain the top cause of mortality globally. WHO estimates indicate that in 2019, 17.9 million people died from CVDs–32% of all deaths–with heart attacks and strokes accounting for 85% of these losses. Among the 17 million premature deaths (under 70) from non-communicable diseases that year, 38% were attributable to CVDs [1]. Ischemic heart disease (IHD) is among the most prevalent CVD entities and a principal driver of mortality [2]. In Kazakhstan, 2022 statistics show circulatory diseases as the most widespread among adults (3,962.5 per 100,000), of which IHD contributes 560.7 per 100,000, underscoring its substantial share within cardiovascular morbidity [3].

IHD imposes a heavy clinical and economic burden, substantially elevating both mortality and morbidity worldwide [4]. Coronary artery disease (CAD)–predominantly a consequence of atherosclerosis–is the leading cause of IHD and culminates in myocardial ischemia. The core pathophysiologic mechanism is obstructive atherosclerosis of the coronary vessels, which compromises myocardial perfusion [5]. In view of rising pressure on health systems, there is a pressing need for early, non-invasive diagnostic strategies that can flag IHD before irreversible outcomes such as myocardial infarction or chronic heart failure (CHF) occur [6].

Heart rate variability (HRV)–the beat-to-beat fluctuation in cardiac cycle duration [7]–is a non-invasive rhythm-based marker that yields clinically useful information about overall physiological status [8]. HRV indexes the heart's adaptive capacity and an individual's ability to respond to environmental challenges via compensatory mechanisms [9]. It is shaped by autonomic inputs–particularly parasympathetic tone–while reflecting the joint activity of sympathetic and parasympathetic branches. Depressed HRV has been linked to adverse endpoints including myocardial infarction, progression of atherosclerosis, heart failure, IHD, and sudden cardiac death [10]. Accordingly, HRV analysis is central to evaluating autonomic nervous system (ANS) function [11]. Conventional coronary assessment tools are frequently costly, invasive, and suboptimal for timely detection of evolving ischemia [12]. Although diagnostic angiography is among the most definitive techniques for identifying cardiac abnormalities, it carries high expense, potential complications, and requires specialized expertise; traditional

workflows can be time-intensive, error-prone, and resource-heavy, risking misclassification and higher costs [13]. This motivates a shift toward reliable, non-invasive, early detection methods–HRV-based approaches being a prime candidate.

CVDs continue to dominate global morbidity and mortality statistics, reinforcing the importance of early identification in high-risk groups and the development of effective preventive and therapeutic interventions. Recent efforts emphasize multifactorial risk models that fuse physiological metrics, lifestyle variables, and medical history to improve predictive performance and enable personalization [14]. HRV–the variability in RR (NN) intervals–is a widely used non-invasive indicator of cardiovascular status [15]. In IHD, reductions in time-domain indices (SDNN, RMSSD, pNN50) and alterations in the LF/HF ratio derived from frequency-domain analysis (FFT of RR intervals) associate with myocardial injury and higher adverse-event risk; an LF/HF imbalance signals disrupted autonomic control during ischemic episodes. This review consolidates key HRV features and highlights their clinical utility in monitoring and managing IHD [6], [16].

Patients with IHD and arrhythmias generally exhibit lower HRV than healthy controls. Time-domain measures such as SDNN, SDANN, RMSSD, pNN50, and the triangular index, together with nonlinear descriptors (α, α1, α2, SD1, SD2, Approximate Entropy, Sample Entropy), are markedly diminished in these populations [17]. These patterns reflect impaired autonomic regulation and support the role of HRV analytics in tracking cardiac function and disease trajectory in IHD [15].

In atrial fibrillation (AF), HRV–defined as fluctuations in ventricular response intervals–is not random; its nonlinear structure, especially multiscale entropy (MSE), carries clinical meaning. Numerous studies link HRV parameters to ischemic stroke risk in AF, and MSE of HRV has been proposed as a predictor in this group [18]. Notably, higher sample-entropy values at specific time scales from 24-hour Holter data correlate with increased stroke likelihood in AF patients without prior stroke. HRV has also been applied to assess hemispheric involvement in acute ischemic stroke (AIS): sample entropy was significantly higher in left-hemispheric than right-hemispheric strokes, implying reduced HRV complexity (and possibly heightened sympathetic drive) on the right; these differences persisted in daytime segments, suggesting value for lesion lateralization [19]. Beyond diagnosis, HRV-based indices have been explored to forecast short-term outcomes in the acute phase of ischemic stroke [20].

Alongside HRV, electrocardiographic alternans (ECGA) provides a promising non-invasive electrophysiological marker of ischemia and arrhythmic risk. ECGA encompasses T-wave (TWA), QRS (QRSA), and P-wave alternans (PWA) derived from standard ECG. Evidence from the STAFF III study – using controlled balloon occlusion – showed time-ordered increases in alternans magnitude measured by correlation methods: PWA within the first minute, QRSA by the second, and TWA by the third minute of coronary occlusion [21]. ECGA is under active evaluation for IHD risk stratification [20]. Although TWA has been examined in IHD and heart failure, heterogeneity in protocols and analytics complicates interpretation [21]. Some reports suggest that combining TWA with HRV may enhance detection of chronic heart failure progression; however, its prognostic role in IHD requires further validation. Recent work argues for concurrent assessment of TWA, QRSA, and PWA to maximize diagnostic yield [22].

Beyond physiology, genetic markers–particularly single-nucleotide polymorphisms (SNPs) – increasingly complement traditional risk factors. Panels that integrate SNPs with clinical variables (e.g., SCORE, age, angiography) have achieved diagnostic accuracies up to 93% [23]. Candidate genes implicated in inflammation, lipid metabolism, and thrombosis further improve CVD risk prediction, offering value independent of standard predictors and showing special relevance in type 2 diabetes, where shared metabolic pathways link to cardiovascular risk [24]. Incorporating genetics into clinical models advances individualized prevention and care. ECG has long been the mainstay for cardiac assessment, capturing electrical activity via surface electrodes [25]. Yet the last decade's push for continuous, user-friendly, and affordable monitoring has accelerated exploration of alternatives [26]. Photoplethysmography (PPG) stands out for simple hardware and seamless integration into consumer devices, providing a substantially cheaper and more convenient path to continuous monitoring in both clinical and everyday contexts [27].

While highly accurate, conventional ECG systems demand clinical oversight, careful electrode placement, and periodic calibration–factors that raise costs and reduce convenience [25]. Advances in microelectronics have miniaturized PPG sensors for wearables (wrist, watch, phone, in-ear), broadening access to continuous cardiovascular tracking

[28]. Coupled with wireless data transfer and cloud-based analytics, PPG offers a distinctive blend of affordability, portability, and usability that traditional ECG cannot easily match [29].

Machine-learning approaches are highly effective for detecting IHD-related anomalies. Head-to-head evaluations of support vector machines, artificial neural networks, and deep models report accuracies above 90% when robust preprocessing and feature selection are applied [30]. Unsupervised routines–most notably k-means–are used to flag outliers in cardiac datasets, which in turn improves the performance of downstream supervised classifiers [31]. On ECG signals, deep networks learn discriminative representations that separate normal from ischemic patterns with near-perfect performance [32]. In imaging, deep learning applied to non-contrast CT, echocardiography, and CT angiography builds hierarchical encodings of coronary anatomy and myocardial motion, capturing subtle lumen-caliber and wall-motion abnormalities consistent with ischemia [33], [34]. Representation-learning schemes such as autoencoders and encoder–decoder frameworks further compress high-dimensional data into interpretable latent features [34]. For label-sparse or imbalanced cohorts, unsupervised anomaly detection segments by similarity and marks deviants as anomalies [31], while synthetic oversampling (SMOTE) rebalances classes and often boosts SVM performance [30]. ECG-based studies frequently exceed 98% accuracy in distinguishing IHD or myocardial infarction from healthy controls by exploiting minute ST-segment deviations and QRS-duration changes–canonical ischemic markers [32]. Hybrid architectures that combine convolutional and recurrent layers enhance results by jointly modeling spatial morphology and temporal dynamics in cardiovascular datasets [35].

Although the association between heart-rate variability (HRV) and cardiovascular disease is well established, a practical workflow for screening ischemic heart disease (IHD) with consumer-grade photoplethysmography (PPG) remains undefined, as do the most informative HRV biomarkers obtainable from such sensors. To address this, we present a pilot using Zhurek–an in-house fingertip PPG device that records 60-minute signals, computes HRV features on board, and transmits encrypted data to a cloud store. Bench comparison with a three-lead Holter ECG showed clinically acceptable mean biases: −0.601 bpm for heart rate, +33.1 ms for SDNN, and −4.8 ms for RMSSD. With Zhurek, one-hour recordings were obtained from 20 healthy volunteers and 20 angiographically confirmed IHD patients sampled from a 300-case registry. Eight candidate variables were evaluated (SDNN, RMSSD, LF, HF, LF/HF, Max_HR, BMI, age). Mann–Whitney tests indicated significant group differences for SDNN, LF, HF, Max_HR, BMI, and age ($p<0.05$). Principal component analysis showed that the first two components accounted for 49.5% of variance and already separated the cohorts in an unsupervised projection. CatBoost feature importance ranked LF power highest (~44%), followed by age (~19%), with HF also strongly discriminative. Collectively, these results show that short, point-of-care PPG acquisitions from an affordable wearable can recover key autonomic signatures previously accessible mainly via 24-hour Holter monitoring, establishing a concrete basis for scalable, low-cost IHD screening grounded in clearly defined HRV biomarkers.

2. Materials and Methods

The hybrid physiological monitoring platform is built for continuous heart-rate variability (HRV) assessment to support ambulatory evaluation of autonomic nervous system function. It couples a wearable sensor that performs on-device processing with secure remote data logging, as shown in Figure 1. The architecture brings together three tightly linked layers: the sensing and on-device processing layer, the communication and storage layer, and the analytics and classification layer.

In the sensing tier, the Zhurek IoT device acquires fingertip photoplethysmography (PPG) and computes core HRV indices in real time. The embedded firmware transforms the raw waveform into time-domain features and prepares them for transmission. In particular, it derives heart rate (HR), pulse period (PP), SDNN, and RMSSD on device; Section 3.2 provides a detailed description of the hardware and firmware stack.

Computed HRV features are serialized as JSON and sent over Wi-Fi via MQTT. The device publishes to the topic zhurek/ppg/hrv, served by a Mosquitto 2.0 broker on a central server. All links are protected with TLS 1.3 and mutual certificate-based authentication to preserve integrity and confidentiality.

Incoming MQTT payloads are parsed and persisted in a relational SQL database. Each entry carries an accurate timestamp from an on-board real-time clock synchronized by Network Time Protocol (NTP) to maintain cross-device temporal consistency. For resilience during network outages, the wearable simultaneously keeps a local CSV log.

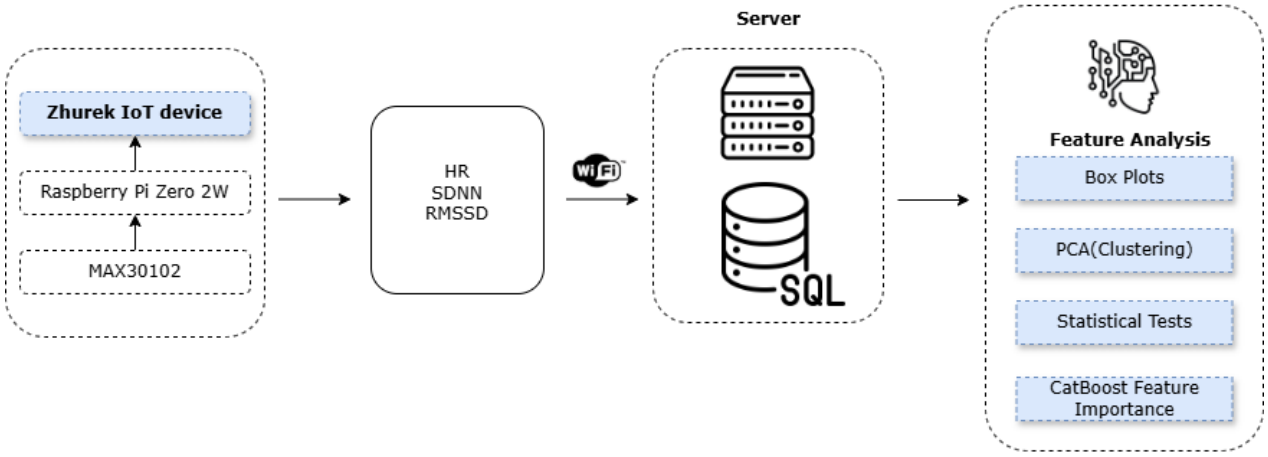
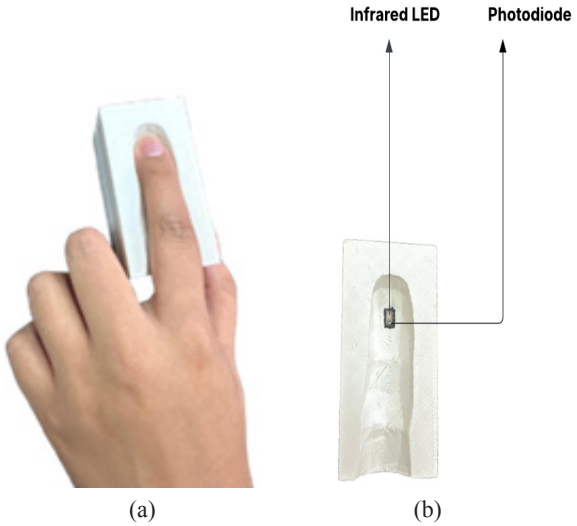


Figure 1 – System Architecture of the Zhurek HRV Pipeline.

In this pilot, the analytics layer prioritized understanding which physiological and clinical variables differentiate healthy controls from patients with ischemic heart disease (IHD), rather than optimizing predictive metrics. Mann–Whitney U tests were applied to detect distributional shifts, and principal component analysis (PCA) was used to explore latent structure and label-free separation between cohorts. Feature importance was estimated with CatBoost on the 40-sample dataset, highlighting variables such as LF power, age, HF power, and Max_HR as the strongest discriminators. Given the limited sample size, model accuracy metrics were intentionally omitted to avoid overfitting and misinterpretation; the emphasis was on hypothesis generation for larger studies.

In the classification track of the analytics layer, stored HRV features can be processed periodically with machine-learning models including gradient boosting methods (XGBoost, CatBoost), random forests (RF), interpretable generalized additive models (EBM), and hybrid designs that combine deep neural networks (DNN) with least-mean-square support vector machines (LMSVM). Trained on labeled data, these models assign risk levels and flag early signs of autonomic dysfunction, enabling automated preliminary triage and risk stratification in remote-monitoring workflows.

By unifying embedded signal processing, encrypted wireless transport, and modular analytics, the system supports round-the-clock monitoring with structured downstream analysis. Reliance on open-source software and off-the-shelf components enhances reproducibility and simplifies deployment in distributed settings.



(a) (b)

Figure 2 – Zhurek Fingertip PPG Sensor: IR LED–Photodiode Layout.

Zhurek (see Figure 2) is a custom, non-invasive wearable for real-time PPG capture and processing. The device integrates a MAX30102 optical sensor (DFRobot Gravity: SEN0344) with a Raspberry Pi Zero 2 W (ARM Cortex-A53, 1 GHz, 512 MB RAM) running Raspberry Pi OS Lite (64-bit). Acquisition uses only the infrared channel at 100 Hz over hardware I²C (address 0x57). The sensor resides in a 3D-printed PLA enclosure with an IR-shielded finger clip and soft elastomer pads to limit motion artefacts and ambient light.

All logic is written in Python 3.11. I²C transactions use smbus2. The raw PPG stream undergoes baseline correction and moving-average smoothing. Cardiac cycles are detected by a derivative-based peak finder adapted from HeartPy, followed by physiological plausibility checks to remove outliers. RR intervals are derived from peak times; HR, SDNN, and RMSSD are computed in 30-s windows with a 5-s hop. Frequency-domain indices (LF, HF, LF/HF) and Max_HR are computed offline, then combined with BMI and age to form an eight-feature vector.

Each result is packaged as a JSON object and published via MQTT; a concurrent CSV log on the device acts as a fail-safe. Timestamps are generated by an RTC that is periodically synchronized using NTP.

The device delivers its best signal quality and physiological fidelity at rest. Resting acquisitions reduce motion artefacts and yield stable autonomic patterns, supporting reliable HRV computation–consistent with evidence that resting protocols maximize accuracy and reproducibility for HRV, gas-exchange, and metabolic-rate measurements [36], [37], [38]. Under these conditions, remote HR and HRV derived from PPG closely track ECG-based readings [38], providing a robust baseline for IHD risk surveillance.

To determine whether wearable ECGs are suitable for resting-state HRV, we carried out a 24-hour comparison between a clinical three-lead Holter and the Polar H10 chest strap. The two systems showed close concordance on key time-domain metrics: mean heart rate differed from the Holter by 0.601 bpm (1.77%), SDNN by 33.088 ms (6.77%), and RMSSD by 4.778 ms (14.57%). The confidence intervals were narrow–e.g., ±1.239 bpm for heart rate–supporting the stability and consistency of both devices during rest.

To build and validate machine-learning models for IHD prediction, HRV data were gathered from two distinct sources: a clinical cohort with confirmed cardiac disease and a healthy control cohort. This split design lets models capture autonomic patterns characteristic of pathology while learning to separate them from normal variability in healthy subjects.

Both groups were recorded with high-fidelity RR-interval sensors to ensure consistent HRV measurement. The clinical set used long-duration, multi-lead Holter ECGs from participants diagnosed with IHD or related disorders. Healthy volunteers were monitored in controlled laboratory sessions with either a single-lead chest-strap ECG (Polar H10) or the custom PPG-based Zhurek device. Although ECG is the reference for RR detection, our results show that, with appropriate preprocessing and validation, PPG from Zhurek attains HRV accuracy adequate for ML-driven risk stratification. Using both modalities within a unified pipeline mitigates dataset bias and mirrors real-world wearable cardiovascular monitoring.

HRV records from 20 adult inpatients with verified cardiovascular disease were obtained at the Research Institute of Cardiology and Internal Diseases (Almaty, Kazakhstan). Diagnoses followed institutional clinical protocols under cardiology department oversight. Each participant underwent continuous 24-hour monitoring with diagnostic-grade, multi-lead Holter ECG systems that provide high-resolution RR-interval outputs appropriate for rigorous HRV assessment. The cohort included patients across a spectrum of disease severity, from early to advanced stages, increasing heterogeneity and supporting the development of models with better external validity. Data were stored as numerical RR-interval series rather than raw ECG, and core HRV variables–heart rate (HR), RR intervals, SDNN, and RMSSD–were computed automatically and supplied for downstream analysis.

To characterize baseline autonomic function, HRV data were collected from 20 healthy volunteers who reported no cardiovascular, neurological, or metabolic conditions. To reduce confounding, participants refrained from alcohol, tobacco, caffeine, and vigorous exercise for at least 24 hours before recording and maintained regular sleep (7–8 hours) the preceding night. Individuals with acute illness, nonadherence to preparation, or excessive signal artefacts were excluded. Measurements used the Zhurek IoT device, combining a MAX30102 PPG sensor with a Raspberry Pi Zero 2 W. The MAX30102 is a low-power optical module with integrated red/IR LEDs, photodiode, and low-noise AFE with ambient-light suppression; in this setup only the infrared

channel was sampled at 100 Hz over hardware I²C. PPG was recorded at the fingertip using a shielded spring-loaded clip with elastomer padding to mitigate motion artefact. Sessions lasted 60 minutes under resting conditions, with a Polar H10 chest-strap ECG worn concurrently for validation. Zhurek computed HR, inter-beat intervals (IBIs), SDNN, and RMSSD in real time using a 30-second rolling window and saved all outputs to CSV with high-resolution time-stamps for subsequent processing.

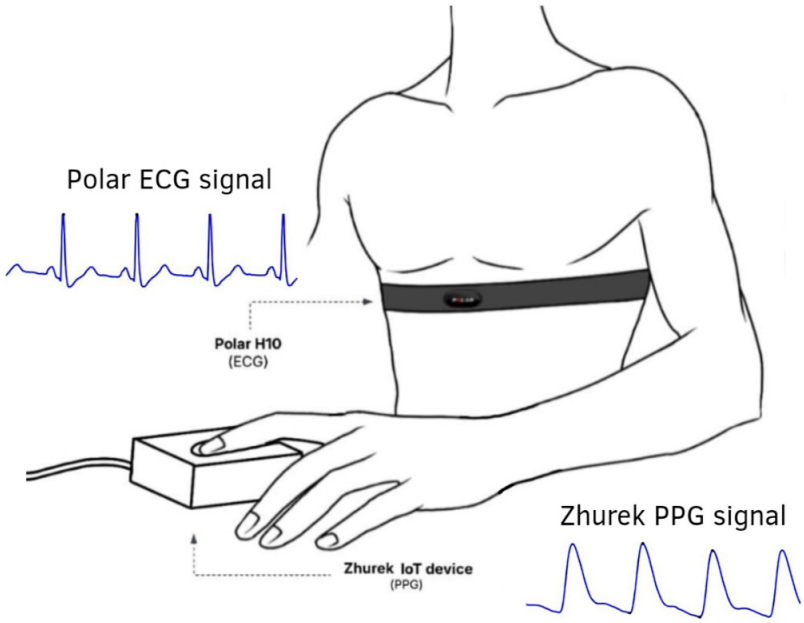


Figure 3 – Simultaneous Acquisition: Polar H10 ECG and Zhurek Fingertip PPG.

In this pilot, preprocessing began by merging two distinct cohorts: a healthy control set and a clinical set with ischemic heart disease (IHD). The control group included 20 adults aged 18–22. The IHD pool was drawn from a registry of exactly 300 confirmed cases spanning 18–92 years; for the present analysis, only patients aged 18–71 were retained. To align temporal resolution across cohorts, a continuous 60-minute segment was extracted from each patient's 24-hour Holter ECG, matching the one-hour PPG recordings collected from healthy participants with the Zhurek device.

Several categorical attributes were numerically encoded to streamline analysis. Sex was coded as Male = 1, Female = 0. "Bad Habits" was set to 1 for respondents reporting alcohol use, smoking, or routine consumption of energy drinks, and 0 otherwise. Familial predisposition was represented by a "Genetic Marker" variable: 0 indicated no known CVD in relatives or a family history limited to non-CVD conditions; 1 denoted a verified family history of cardiovascular disorders (e.g., hypertension, IHD, myocardial infarction, stroke); 2 indicated only non-cardiovascular illnesses among relatives. Surgical history ("Operations") was encoded as 0 for no prior procedures, 1 for non-cardiac surgeries, and 2 for cardiovascular-related interventions.

The feature set comprised established HRV measures–SDNN, pNN50, RMSSD, LF, HF, and the LF/HF ratio–reflecting autonomic nervous system dynamics [39]. Two additional variables were included: Max_HR (the maximum heart rate observed within the recording) and BMI, computed by the standard Equation (1) [40]:

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

Quality control steps removed entries with missing fields, malformed values, or extreme outliers. Non-numeric strings were normalized to numeric form where applicable (including converting decimal commas to periods). These procedures yielded a clean, consistent dataset suitable for downstream statistical analysis and visualization.

3. Results

HRV was measured in two cohorts: patients with clinically confirmed IHD and healthy controls without known cardiovascular disease. All participants completed standardized recording sessions. Summary statistics for each group are reported in Table 1 (controls) and Table 2 (IHD). The analysis covers time-domain indices (SDNN, pNN50, RMSSD) and frequency-domain indices (LF, HF, LF/HF), offering a snapshot of autonomic nervous system dynamics [39] and revealing potential contrasts in HRV patterns between healthy and IHD populations.

Table 1 – Descriptive HRV statistics – healthy control group.

№	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 2 – HRV summary in the IHD cohort.

№	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

Figure 4 boxplots indicate clear groupwise differences in physiological and HRV features. SDNN, LF, and HF are higher in the IHD cohort, reflecting greater overall variability and increased spectral power in both bands. By contrast, RMSSD–an index of parasympathetic tone [41]–is higher in healthy participants, consistent with greater autonomic flexibility associated with elevated RMSSD values [42]. The LF/HF ratio shows similar medians across groups. Max_HR is higher among IHD subjects. BMI varies modestly between groups, with a higher median in the IHD cohort, in line with the recognized contribution of excess body weight to cardiovascular risk [43].

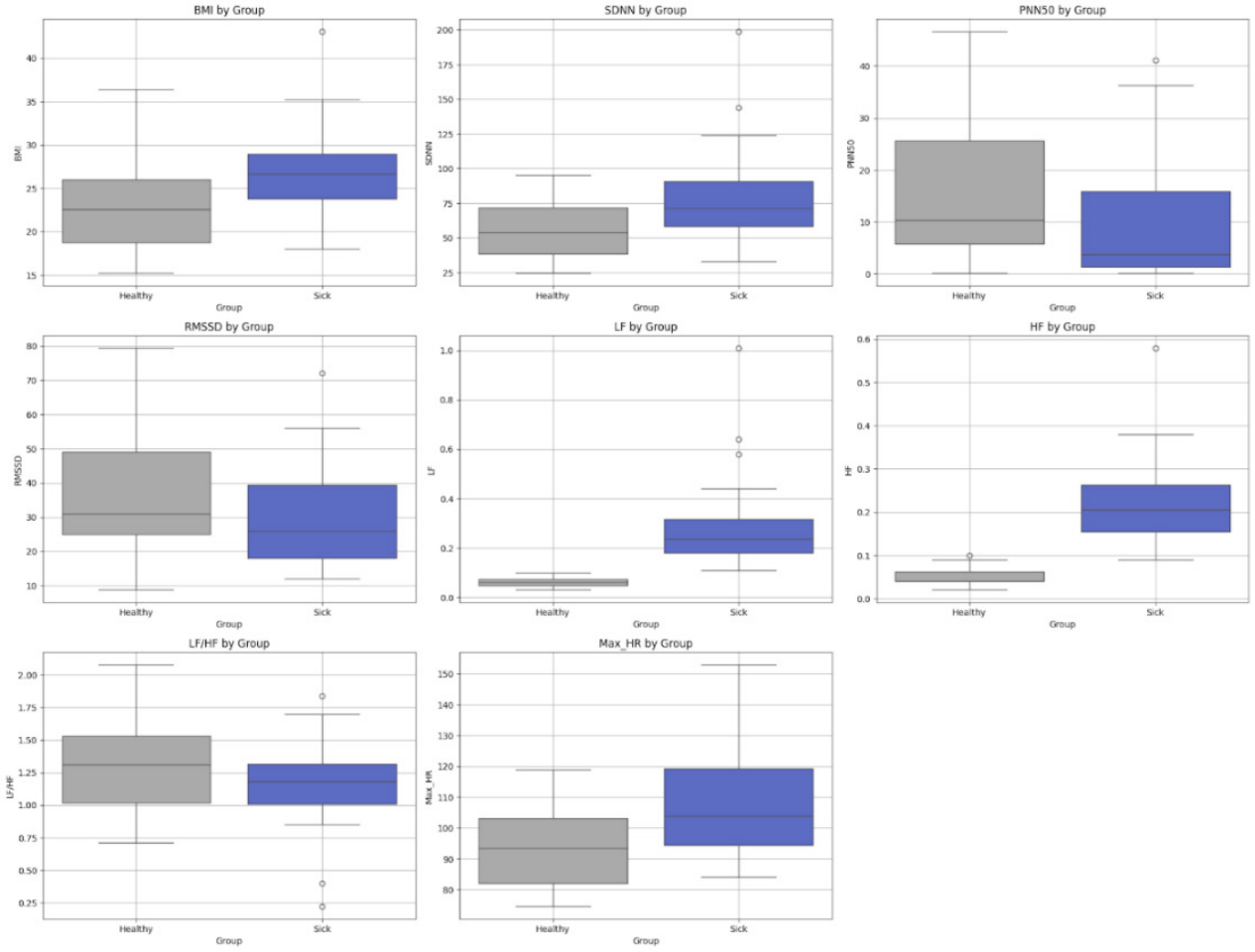


Figure 4 – Groupwise Boxplots of Physiological and HRV Metrics (Healthy vs. IHD).

Continuous physiological and HRV metrics–BMI, SDNN, RMSSD, pNN50, LF, HF, LF/HF, and Max_HR–are summarized for healthy controls and IHD patients. As depicted in Figure 5, the Genetic Marker variable is predominantly 0 among healthy participants, whereas most IHD cases are labeled 1, emphasizing hereditary risk in cardiovascular disease [44]. For Operations, controls are almost entirely 0, while the IHD cohort shows markedly higher rates of both non-cardiac (1) and cardiac (2) procedures, reflecting greater clinical intervention. The Bad Habits indicator (smoking, alcohol, energy drinks) is also more frequent in the IHD group than in controls, underscoring modifiable lifestyle contributions. Collectively, these distributions show that genetic predisposition, surgical history, and behavioral risk factors jointly separate the cohorts and provide salient predictors for IHD outcome modeling.

The correlation heatmaps, as shown in Figure 6, 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. In the healthy group, the heatmap reveals a structured and physiologically intuitive set of relationships. Strong positive correlations are evident among the time-domain HRV metrics, specifically SDNN, RMSSD, and PNN50. In contrast, the IHD group exhibits a more disrupted correlation structure compared to the healthy group, which suggests a fundamental loss of autonomic coherence. Additionally, lifestyle factors such as un-healthy 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. These heatmaps offer more than just a summary of feature relationships; they visualize the integrity of the autonomic nervous system. The pattern in the healthy group reflects physiological harmony and adaptability, while the disrupted pattern in the IHD group visually rep-resents the autonomic dysregulation and loss of resilience that is a hallmark of cardiovascular disease.

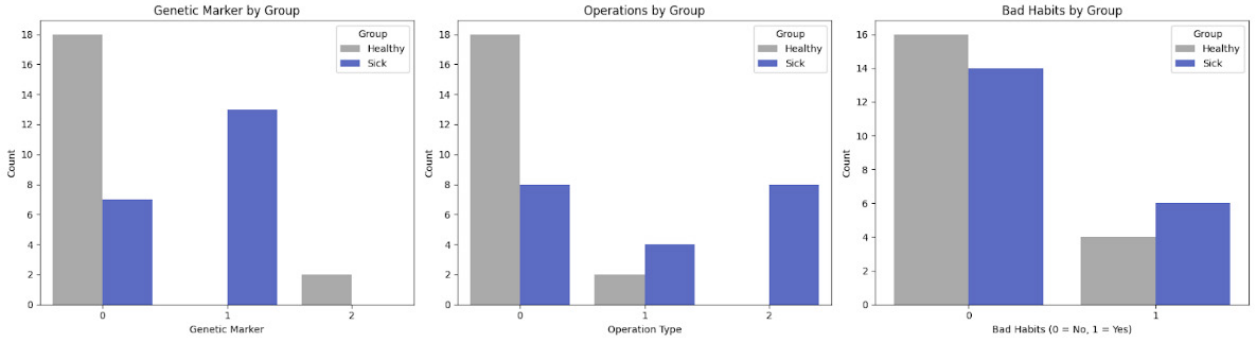


Figure 5 – Categorical Risk Factors by Group (Genetic Marker, Operations, Bad Habits).

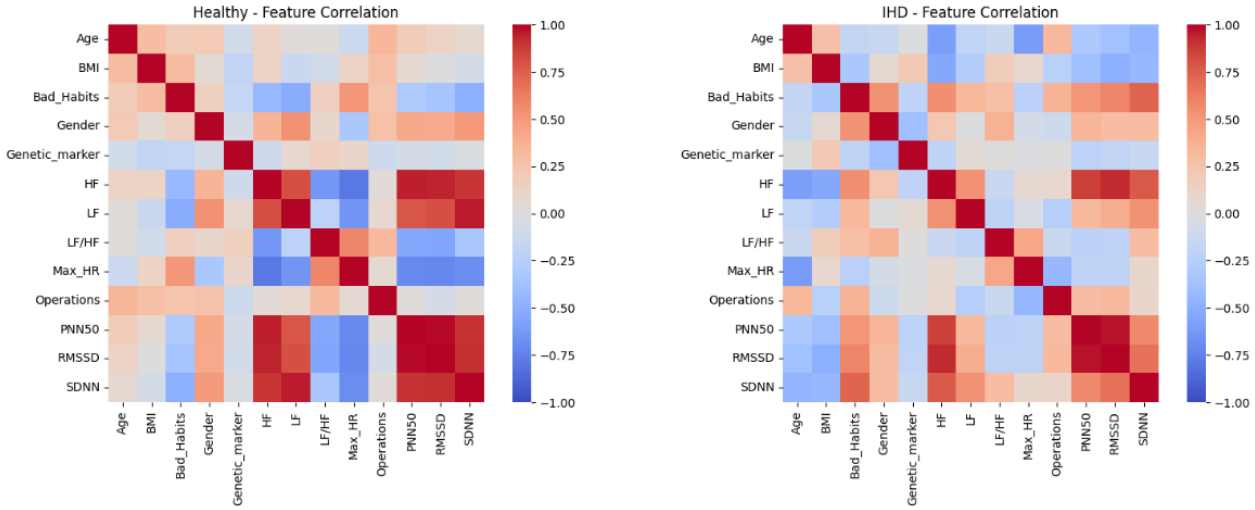


Figure 6 – Feature Correlation Heatmaps (Healthy vs. IHD)

Table 3 summarizes distributional comparisons between healthy controls and IHD patients. Significant group differences ($p < 0.05$) were observed for BMI, SDNN, LF, HF, Max_HR, and Age, indicating that autonomic activity and cardiovascular dynamics provide strong discriminatory signal. By contrast, PNN50, RMSSD, and the LF/HF ratio were not individually significant, though they may still add value in multivariate models or clinical interpretation.

Table 3 – Mann–Whitney u comparison of physiological feature distributions (Healthy vs. IHD).

Feature	P-VALUE	Significance
BMI	0.042464	Yes
SDNN	0.025625	Yes
LF	6.49e-08	Yes

Continuation of the table

Feature	p-value	Significance
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 probe cohort-level physiology, we applied principal component analysis. The first two components accounted for ~49.5% of the variance (PC1: 27.3%, PC2: 22.2%). As shown in Figure 7, the PC1–PC2 projection exhibits a clear separation between groups–healthy participants cluster apart from IHD cases–suggesting that the selected physiological and categorical features contain sufficient signal for unsupervised differentiation and motivating their use in downstream supervised modeling and feature-importance analysis.

As part of the exploratory workflow, we inspected CatBoost feature importances to see which physiological and clinical variables most strongly drive group separation. Model accuracy was not the objective at this stage; the goal was to flag the variables most promising for differentiating healthy participants from patients with IHD.

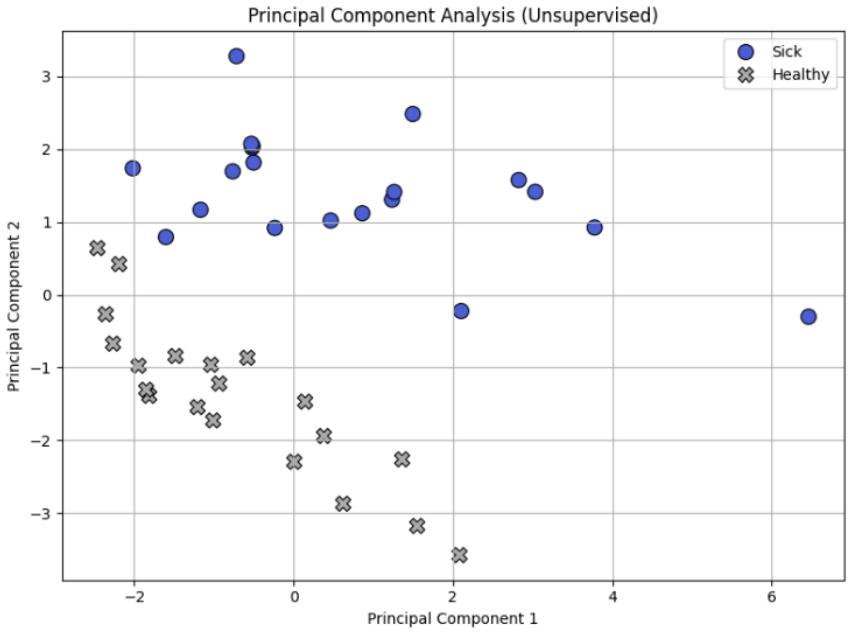


Figure 7 – Unsupervised PCA Projection of HRV Features (Healthy vs. IHD)

As shown in Figure 8, SHAP analysis of the CatBoost model ranks LF (low-frequency HRV power) as the top contributor (highest mean SHAP value), with HF (high-frequency power) next–underscoring the central role of autonomic dynamics. Additional, smaller but non-trivial contributions come from Operations, Genetic_marker, and PNN50, suggesting that surgical history, hereditary risk, and beat-to-beat variability also aid in distinguishing the two cohorts.

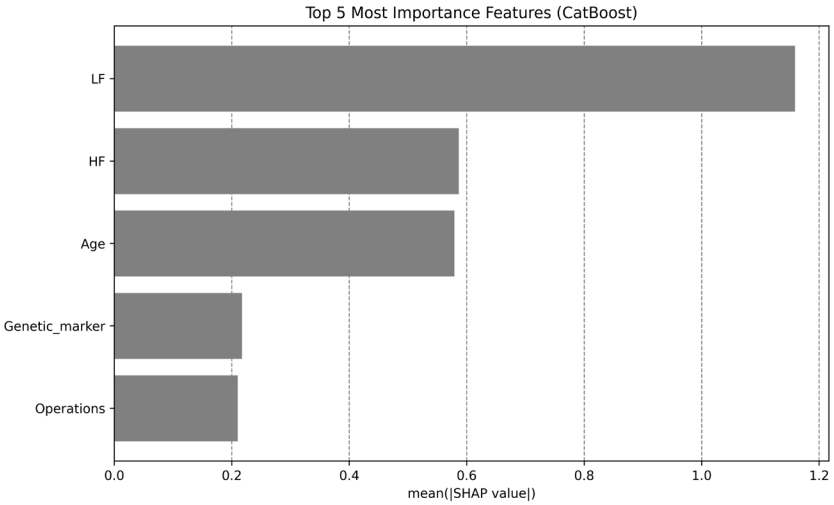


Figure 8 – Unsupervised PCA Projection of HRV Features (Healthy vs. IHD)

To gauge how much the model depends on physiology alone, we re-ran CatBoost on the pilot dataset after removing the age variable. Because no hold-out set was used, the aim was not to score accuracy but to probe which inputs drive the model's internal decisions. As shown in Figure 9, SHAP analysis identifies five leading contributors to IHD prediction. The low-frequency (LF) HRV component shows the largest mean absolute SHAP value, indicating the strongest influence, followed by the high-frequency (HF) component and Age, pointing to major roles for autonomic dynamics and age-related effects. Genetic_marker and Operations contribute less overall but still affect certain cases. Force-plot views further illustrate that higher LF/HF ratios or greater age tend to push individual estimates toward the "Ischaemia" class.

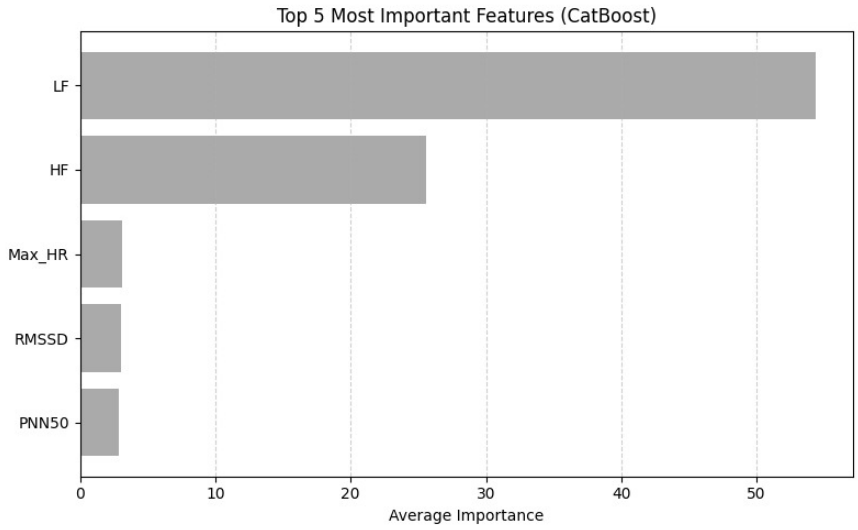


Figure 9 – CatBoost Feature Importance (physiology-only model).

4. Conclusions

The principal contribution of this pilot is an end-to-end system and protocol for scalable, ambulatory screening of ischemic heart disease (IHD). We developed and validated Zhurek, a custom, non-invasive fingertip PPG sensor that enables rapid, point-of-care acquisition. A key design element is showing that a single one-hour recording is sufficient to capture diagnostically meaningful autonomic signatures. Hour-long sessions were obtained from 20 healthy volunteers aged 18–22 and 20 angiographically confirmed IHD patients, and eight features were derived: SDNN, RMSSD, LF, HF, LF/HF, Max_HR, BMI, and age. Against a three-lead Holter ECG, the system demonstrated clinically acceptable mean absolute errors–0.601 bpm for heart rate, 33.1 ms for SDNN, and 4.8 ms for RMSSD.

Nonparametric testing (Mann–Whitney) indicated significant between-group differences for SDNN, LF, HF, Max_HR, BMI, and age ($p < 0.05$). The first two principal components explained 49.5% of total variance and already separated cohorts without labels, supporting the informativeness of the chosen variables. Methodologically, the study used machine learning not to overfit predictions on a small sample but to conduct exploratory ranking of short-duration biomarkers. CatBoost importance scores placed LF at ~44% of total contribution, age at ~19%, followed by HF, with smaller effects for Max_HR and RMSSD. Together with the PCA separation, this pinpoints which markers carry the greatest diagnostic weight and shows that brief Zhurek acquisitions can recover autonomic signals traditionally accessed via 24-hour Holter, laying the groundwork for affordable, large-scale screening.

Incorporating SHAP improved the transparency of the Zhurek–CatBoost pipeline and aligns with emerging explainability standards; at deployment scale, individualized SHAP profiles could inform genuinely personalized prevention.

Several caveats qualify these findings. Data modalities differed by cohort–healthy participants were recorded with optical PPG, whereas IHD patients were recorded with ECG–introducing potential signal-quality biases. The control group's narrow age range contrasts with age being a strong discriminator, raising the risk of confounding. The sample size ($n = 40$) is limited; therefore, formal performance metrics were intentionally omitted to avoid overinterpretation.

Future work will enlarge and age-diversify the cohorts–especially the healthy group–and standardize acquisition by collecting parallel PPG and single-lead ECG from all participants. A multicenter longitudinal study is planned to test the prognostic utility of short-term markers and to confirm reproducibility. The roadmap also includes adding nonlinear HRV measures and expanding automated analysis to enable reliable ambulatory risk stratification. These steps should enhance the clinical relevance of the Zhurek approach and accelerate the shift from reactive care to proactive IHD prevention.

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

Conflicts of Interest

The authors declare no conflict of interest.

References

1. ‘Cardiovascular diseases (CVDs)’. Accessed: Jul. 21, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
2. M. A. Khan *et al.*, ‘Global Epidemiology of Ischemic Heart Disease: Results from the Global Burden of Disease Study’, *Cureus*, Jul. 2020, doi: 10.7759/cureus.9349.
3. tengrinews.kz, ‘Названа самая распространенная болезнь среди казахстанцев’, Главные новости Казахстана – Tengrinews.kz. Accessed: Jul. 21, 2025. [Online]. Available: 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’, *IJMS*, vol. 21, no. 21, p. 8118, Oct. 2020, doi: 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, Jan. 2022, doi: 10.3390/diagnostics12020297.
6. Ștefania-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, Aug. 2024, doi: 10.3390/medicina60081315.
7. N. Singh *et al.*, ‘Heart Rate Variability: An Old Metric with New Meaning in the Era of using mHealth Technologies for Health and Exercise Training Guidance. Part One: Physiology and Methods’, *Arrhythmia & Electrophysiology Review*, vol. 7, no. 3, p. 193, 2018, doi: 10.15420/aer.2018.27.2.
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, Jan. 2024, doi: 10.3390/bioengineering11010058.
9. S. P. Gaine *et al.*, ‘Multimodality Imaging in the Detection of Ischemic Heart Disease in Women’, *JCDD*, vol. 9, no. 10, p. 350, Oct. 2022, doi: 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, Aug. 2024, doi: 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, Feb. 2023, doi: 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, Dec. 2016, doi: 10.17485/ijst/2016/v9i48/105707.
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, Nov. 2022, doi: 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, May 2019, doi: 10.1371/journal.pone.0213653.
15. H. ChuDuc, K. NguyenPhan, and D. NguyenViet, ‘A Review of Heart Rate Variability and its Applications’, *APCBEE Procedia*, vol. 7, pp. 80–85, 2013, doi: 10.1016/j.apcbee.2013.08.016.
16. C. Brinza *et al.*, ‘Heart Rate Variability in Acute Myocardial Infarction: Results of the HeaRt-V-AMI Single-Center Cohort Study’, *JCDD*, vol. 11, no. 8, p. 254, Aug. 2024, doi: 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’.
18. M. Trigka and E. Dritsas, ‘Long-Term Coronary Artery Disease Risk Prediction with Machine Learning Models’, *Sensors*, vol. 23, no. 3, p. 1193, Jan. 2023, doi: 10.3390/s23031193.
19. B. A. Marzoog *et al.*, ‘Machine Learning Model Discriminate Ischemic Heart Disease Using Breathome Analysis’, *Biomedicines*, vol. 12, no. 12, p. 2814, Dec. 2024, doi: 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’, *JCM*, vol. 12, no. 3, p. 1219, Feb. 2023, doi: 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, Dec. 2021, doi: 10.3390/sym13122424.
22. B. M. Curtis and J. H. O’Keefe, ‘Autonomic Tone as a Cardiovascular Risk Factor: The Dangers of Chronic Fight or Flight’, *Mayo Clinic Proceedings*, vol. 77, no. 1, pp. 45–54, Jan. 2002, doi: 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, Oct. 2016, doi: 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’, *Front. Endocrinol.*, vol. 9, Jan. 2018, doi: 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 Photopletysmography Signal Analysis for Biomedical Applications’, *Sensors*, vol. 18, no. 6, p. 1894, Jun. 2018, doi: 10.3390/s18061894.
26. M. Elgendi *et al.*, ‘The use of photoplethysmography for assessing hypertension’, *npj Digit. Med.*, vol. 2, no. 1, Jun. 2019, doi: 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, Mar. 2022, doi: 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, Jul. 2023, doi: 10.3390/electronics12132923.

29. M. Shabaan *et al.*, 'Survey: smartphone-based assessment of cardiovascular diseases using ECG and PPG analysis', *BMC Med Inform Decis Mak*, vol. 20, no. 1, Dec. 2020, doi: 10.1186/s12911-020-01199-7.

30. I. C. Dipto, T. Islam, H. M. M. Rahman, and M. A. Rahman, 'Comparison of Different Machine Learning Algorithms for the Prediction of Coronary Artery Disease', *JDAIP*, vol. 08, no. 02, pp. 41–68, 2020, doi: 10.4236/jdaip.2020.82003.

31. R. C. Ripan *et al.*, 'A Data-Driven Heart Disease Prediction Model Through K-Means Clustering-Based Anomaly Detection', *SN COMPUT. SCI.*, vol. 2, no. 2, Apr. 2021, doi: 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, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104457.

33. K. Kusunose *et al.*, 'A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images', *JACC: Cardiovascular Imaging*, vol. 13, no. 2, pp. 374–381, Feb. 2020, doi: 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 Trans. Med. Imaging*, vol. 39, no. 5, pp. 1545–1557, May 2020, doi: 10.1109/tmi.2019.2953054.

35. A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, and S. N. Qasem, 'Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases', *Diagnostics*, vol. 14, no. 2, p. 144, Jan. 2024, doi: 10.3390/diagnostics14020144.

36. 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, Sep. 2014, doi: 10.1089/brain.2014.0284.

37. 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, Oct. 2018, doi: 10.1016/j.clnu.2017.07.026.

38. 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*, Chicago, IL: IEEE, Aug. 2014, pp. 2957–2960. doi: 10.1109/embc.2014.6944243.

39. 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, Oct. 2023, doi: 10.1016/j.ijpsycho.2023.08.001.

40. A. Zierle-Ghosh and A. Jan, 'Physiology, Body Mass Index', in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. Accessed: Jul. 21, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK535456/

41. J. Zeng *et al.*, 'High vagally mediated resting-state heart rate variability is associated with superior working memory function', *Front. Neurosci.*, vol. 17, Feb. 2023, doi: 10.3389/fnins.2023.1119405.

42. J. Hart, 'OPTIMAL LEVEL OF HEART RATE VARIABILITY FOR SPINAL ADJUSTMENT: A CASE REPORT', *JCC*, vol. 2, no. 1, pp. 103–108, Sep. 2019.

43. 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, Jun. 2023, doi: 10.1016/j.ijcard.2023.03.054.

44. S. Kathiresan and D. Srivastava, 'Genetics of Human Cardiovascular Disease', *Cell*, vol. 148, no. 6, pp. 1242–1257, Mar. 2012, doi: 10.1016/j.cell.2012.03.001.

Information About Authors:

Nurdaulet Tasmurzayev, PhD. Dr. Tasmurzayev is a Research Engineer at DigitAlem LLP (Almaty, Kazakhstan). He received his PhD (Candidate of Technical Sciences) in Intelligent Control Systems (Big Data and Machine Learning) from al-Farabi Kazakh National University in 2025, his M.Sc. in Intelligent Control Systems in 2022, and his B.Sc. in Automation and Control (Information Systems Department) in 2020. He has authored and co-authored 10 peer-reviewed journal and conference papers. Research interests: artificial intelligence and machine learning; intelligent control systems; Internet of Things (including industrial IoT); smart cities; intelligent building systems and automation; big-data analytics for control and monitoring. ORCID iD: 0000-0003-3039-6715.

Dinara Turmakhanbet is a 2025 B.Sc. graduate in Intelligent Control Systems from 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: 0009-0004-8388-4979.

Adilet Kakharov is a fourth-year B.Sc. student in Intelligent Control Systems at al-Farabi Kazakh National University (Almaty, Kazakhstan). His academic interests include intelligent control, the applied machine learning and AI. In this study, he assisted with technical support, data processing, and visualization. ORCID iD: 0009-0005-3612-5678.

Mukhamejan Aitkazin is a fourth-year B.Sc. student in Intelligent Control Systems at al-Farabi Kazakh National 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: 0009-0004-0181-7351.

Aliya Baidauletova is a Candidate of Medical Sciences and a practicing somnologist and neurologist at al-Farabi Kazakh National University (Almaty, Kazakhstan, baidaulet123@gmail.com). She has extensive clinical and research experience in sleep medicine, neurology, and neurophysiological disorders. Her professional interests include sleep disorders, circadian rhythm regulation, neurological aspects of sleep pathology, and the application of modern diagnostic and monitoring technologies in clinical practice. ORCID iD: 0009-0000-5510-3590.

Mergul Kozhamberdiyeva is a Candidate of Pedagogical Sciences and a practicing pedagogist at al-Farabi Kazakh National University (Almaty, Kazakhstan, kozhamberdiyeva.m@outlook.com). She has extensive clinical and research experience in sleep medicine, neurology, and neurophysiological disorders. Her professional interests include sleep disorders, circadian rhythm regulation, neurological aspects of sleep pathology, and the application of modern diagnostic and monitoring technologies in clinical practice. ORCID iD: 0009-0001-0429-7919.

Submission received: 13 August, 2025.
Revised: 20 November, 2025.
Accepted: 20 November, 2025.