Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia, 2026, Iss. 104, pp. 63–73

UDC 616.1-073:004.8

Selection of Amplitude-Temporal and Spectral Parameters of PPG Signals for Diagnostics Based on Machine Learning Methods

Mosiichuk V. S., Sharpan O. B., Guseva O. V.

National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

E-mail: *mosv@ros.kpi.ua*

The problem of increasing the efficiency of non-invasive express diagnostics of the functional state of the organism based on the analysis of photoplethysmographic (PPG) signals using machine learning methods is considered. Based on the analysis of the morphological features of the pulse wave, which contain important information about hemodynamic processes, the feasibility of switching from the traditional analysis of the amplitude-time characteristics of the signal to a complex approach is justified, which includes additional extraction of spectral-phase parameters within individual cardiocycles. A method of signal preprocessing is proposed, which includes band-pass filtering, normalization and segmentation, as well as an approach to "soft" filtering, which allows preserving diagnostically significant features of the signal shape. The system of amplitude-time parameters (duration of anacrota, intervals between characteristic points, area under the curve, etc.) and spectral-phase characteristics obtained using Fourier series expansion is considered in detail. Particular attention is paid to the procedure for selecting informative features for machine learning models. A multi-step approach is proposed, which includes eliminating the dependence on heart rate, multicollinearity analysis, assessing the importance of features using the permutation importance method, and interpreting the results using the PCA and SHAP methods. It was found that the greatest contribution to the accuracy of the models is provided by spectral-phase parameters, in particular group delay and phase shifts of harmonics, which are directly related to the physiological properties of the vascular system. As a result, a set of features was formed that combines the temporal and spectral characteristics of the signal and provides improved classification and prediction quality. The results obtained confirm the prospects of using complex analysis of PPG signals in combination with machine learning methods for medical diagnostics and human condition monitoring.

Keywords: non-invasive diagnostics; physiological signal analysis; biosignal processing; feature extraction; feature selection; frequency-domain analysis; phase spectrum; photoplethysmography (PPG); pulse wave analysis; machine learning

DOI: 10.64915/RADAP.2026.104.63-73

Introduction

The development of medical and biological express diagnostic methods is largely associated with the transition to non-invasive, rapid, and scalable approaches for assessing the functional state of the human body. In this context, the analysis of pulse biosignals, which contain hidden information about hemodynamic and biochemical processes, is of particular importance. Unlike traditional approaches based on invasive measurements or laborious anthropometric data collection, pulse biosignals can be obtained rapidly and non-invasively using available sensor devices.

Pulse waves, recorded by any method, comprehensively reflect changes in the volume and velocity of blood movement through the vessels and contain information on the period of blood ejection from the cardiac ventricles. The shape and parameters of the pulse wave reflect the state of the blood vessels and the heart, as well as the dynamics of intravascular and intracardiac pressure. This makes it possible to assess the integrity and functional activity of the regulatory centers of the cardiovascular system. These biosignals are also used indirectly to assess the functional integrity of the central and autonomic nervous systems, as well as other human organs and systems [1].

In this context, an in-depth study of the PPG waveform is promising, since the morphological characteristics of the pulse wave, such as amplitude, time, and derivative parameters, potentially correlate with a wide range of physiological indicators. In particular, published studies demonstrate statistical relationships between PPG parameters and blood pressure (BP) [2–5], glucose level [6], lipid profile (cholesterol) [7], and other parameters. However, these results are currently

mostly limited to the analysis of integral characteristics, primarily amplitude-time (contour) parameters of the signal, without a comprehensive in-depth study of it as a source of informative features based on detailed morphological analysis of the PPG pulse waveform [8].

Some approaches, for example [9], use regular biosignal plots with a fixed number of samples for CNN-based biosignal analysis without attempting to isolate cardiac cycles.

Modern machine learning methods create fundamentally new opportunities for processing such signals. Their effectiveness directly depends on the quality and representativeness of the input features. In this case, the key task is the systematic extraction of numerical descriptors of the PPG signal shape (feature extraction), which can be used as input variables for regression and classification models. These include, in particular, the parameters of the systolic and diastolic phases, vascular stiffness indices, and time intervals between characteristic points of the signal, which are also obtained from the characteristic points of its first and second derivatives.

Modern machine learning methods create fundamentally new opportunities for processing such signals. Their effectiveness directly depends on the quality and representativeness of the input features. In this case, the key task is the systematic extraction of numerical descriptors of the PPG signal shape (feature extraction), which can be used as input variables for regression and classification models. These include, in particular, the parameters of the systolic and diastolic phases, vascular stiffness indices, and time intervals between characteristic points of the signal, which are also obtained from the characteristic points of its first and second derivatives.

However, the potential of using PPG waveform features as an additional data source for building generalized rapid diagnostic models has not yet been sufficiently explored. In particular, despite significant progress in the application of machine learning algorithms, most existing approaches to assessing biochemical parameters (such as glucose or cholesterol) rely on indirect or partially invasive data rather than primary biosignals.

Thus, a relevant scientific task is to further develop approaches to analyzing the shape of PPG signals, extracting informative numerical features, and using these features in machine learning models for non-invasive prediction of physiological and biochemical indicators. In this context, spectral amplitude and phase parameters should be considered complementary to, rather than replacements for, conventional amplitude-time descriptors. Particular attention should be paid to the formation and use of specialized datasets that combine "raw" PPG signals with reference medical measurements, thereby enabling objective model training and comparison.

1 Features and diagnostic capabilities of PPG signal analysis

Traditionally, in medical and biological practice, the most widely used diagnostic features are amplitude-time (contour) features of pulse signals. Based on the analysis of pulse wave parameters, primarily rheographic parameters, dozens of indicators have been proposed to characterize the state of the cardiovascular system. This direction continues to develop. The study in [2] proposed extracting five different features of the PPG signal form from preprocessed segments within a single cardiac cycle. These features include the time and nature of the pulse rise (duration of the anacrota) and its width at the levels of 25%, 50%, and 75%. The signal shape indicator in the form of the "area under the curve" of the PPG cardiocycle reflects changes in vascular tone [10]. It has been established that the rise time of the pulse wave amplitude is associated with changes in blood pressure. This parameter is reported to be more pronounced in younger than in older individuals [11]. In particular, in [3], this parameter is used to estimate arterial pressure without a cuff. The period of the PPG signal is associated with systemic vascular resistance [12]. The studies in [6, 7] examined the possibility of non-invasively assessing glucose levels in human blood (Raman theory). In addition, in the last two decades, new diagnostic results and capabilities have been obtained based on the analysis of the amplitude and phase spectral characteristics of pulse signals [13, 14].

A number of studies have reported reliable correlations between systolic and diastolic blood pressure values and pulse wave velocity values based on combined analysis of central and peripheral blood flow signals. Usually, ECG is used synchronously with the PPG signal to determine pulse wave velocity. The time difference between the signals is associated with the tone (stiffness) of the arterial vessels [2].

The studies in [5, 6] indicate that slope transit time (STT; the period of rapid growth of the pulse wave during systole, from the minimum value to the peak) is almost the only PPG parameter that correlates with blood pressure.

It should be noted that some algorithms for extracting detailed amplitude-time parameters for morphological analysis of PPG using the first and second derivatives are possible in the case of "hard" filtering [15]. However, such filtering significantly distorts the signal shape, which can lead to the loss of diagnostically valuable features and practically makes it impossible to reliably extract characteristic points from the second derivative. To prevent this, we propose a procedure that involves applying "soft" filtering, cardiocycle separation, and a subsequent amplitude-phase method of parameter extraction, followed by the use of machine learning methods for analysis, interpretation, and identification of dependencies with metadata.

Based on these observations [15], the present approach focuses on preserving the pulse waveform during preprocessing so that both amplitude-time and spectral-phase descriptors can be extracted consistently.

In this study, the main attention is focused on the local isolation of morphological patterns of pulse signals using the segmentation method and the extraction of spectral amplitude and phase parameters within individual cardiocycles in addition to amplitude-time parameters. The proposed parameters are not treated as universal indicators; rather, they are considered as an additional group of descriptors that may improve the completeness of PPG waveform representation.

The contribution of the study is as follows:

- The feasibility of using the spectral amplitude and phase method for extracting diagnostically significant normalized parameters of PPG signal segments as a complement to the amplitude-time (contour) parameters is substantiated.
- A feature selection analysis of the set of amplitude-time and spectral-phase parameters obtained from real signals is performed, and their effect on the formation of models for determining the functional state using machine learning methods is assessed.

2 Methodology for detecting and extracting a set of parameters unique to PPG signals

2.1 Signal preprocessing

Before parameter extraction, we used three preprocessing techniques: bandpass filtering, normalization, and region segmentation.

A 0.5–30 Hz bandpass filter was used to remove unwanted frequency components and noise from the recorded raw signal. This filter is intended to reduce the influence of external and internal interference, including power line interference (50/60 Hz) and baseline drift (low-frequency noise), rather than to correct motion artifacts. Segments with pronounced motion artifacts should be rejected or classified as unsuitable before morphological and spectral analysis. The Butterworth filter was chosen because it provides minimal signal distortion while preserving important signal properties due to its maximally flat frequency response across the filter's bandwidth.

2.2 Justification of the choice of amplitude-time parameters of pulse signals

In practice, a significant number of indicators are used in the analysis of PPG signals, which allow, to a greater or lesser extent, to assess the nature of hemodynamics and the state of the vessels of the studied area.

A characteristic feature of PPG signals is their high susceptibility to motion artifacts, which can reduce their morphological quality and, therefore, affect the reliability of the extracted information. Thus, accounting for motion artifacts is especially relevant when signals are recorded during everyday activities, for example, using smartwatches. In this case, it is advisable to create a separate classifier for the PPG pulse wave signal quality index; in particular, three classes are recommended: unsuitable for use due to interference and artifacts, suitable for assessing heart rate (basic quality classifier), and suitable for morphological analysis (high-quality biosignal classifier) [16,17]. The development of such a quality pre-classifier is a separate task and is outside the scope of the present study.

While heart rate (HR) and some morphological features related to systolic peak detection can be assessed from a basic-quality pulse wave, more complex morphological features require detection of both systolic and diastolic peaks [15,18,19], for which only a high-quality pulse wave is suitable.

The pulse wave corresponding to one period of heart contraction consists of an ascending part, a maximum, and a descending part, on which the incisura and dicrotic tooth are located. Additional teeth may also be present, and their occurrence is explained by bifurcations in the vascular system. Each PPG period can be divided into two phases: a fast anacrotic phase, which corresponds to the systolic contraction of the heart, and a catacrotic phase, which depends on both the diastolic phase of the heart and the influence of blood pumping in the vascular bed of peripheral arteries [20].

Fig. 1 shows an example of one period of a real recorded pulse wave signal with the result of preprocessing by spline interpolation based on a third-degree polynomial and the selection of its individual amplitude and time parameters. Traditionally, during the analysis and identification of the functional state by contour analysis methods, the time and amplitude parameters of characteristic points within each period of the PPG signal are used:

- systolic base: the beginning of the systolic phase and the minimum of the pulse wave;
- maximum of the steepness of the anacrotic;
- systolic peak: the most noticeable maximum and its amplitude;
- time marker of the incisura, the dicrotic notch that is caused by the closure of the aortic valve and is most noticeable in healthy young people [1];
- diastolic peak: the second noticeable maximum of the pulse wave.

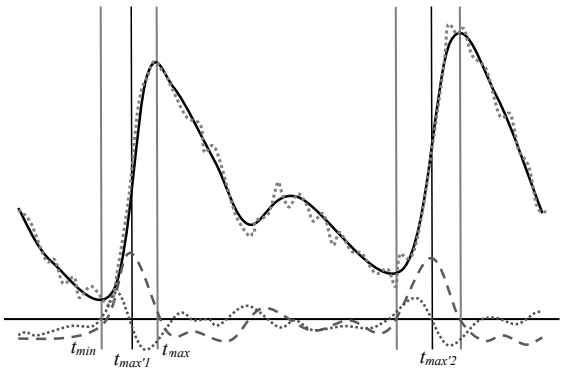

Fig. 1. Determining characteristic points of the PPG signal

Some of the listed features are physiologically grounded, i.e., they characterize the hemodynamics of the cardiovascular system. The study in [4] demonstrates the effectiveness of using such biosignal characteristics in prognostic models in comparison with conventional indicators identified as characteristic points of amplitude-time analysis of biosignals.

Characteristic points of amplitude-time analysis, such as maxima, troughs, and notches, are usually detected more accurately by the second derivative using differentiation. This is especially important for determining the position and visualization of the dicrotic notch (incisura), which in some cases may not be clearly expressed in the primary PPG signal but is important for cuffless blood pressure determination.

When analyzing a PPG signal, it is advisable to calculate the following parameters in the time domain: α_1, α_2, β, T, HR, the times of maxima, minima, and characteristic points of the pulse wave quasi-periods, as well as the delay between the ECG parameters and the PPG quasi-periods to calculate the pulse wave propagation velocity. Here, α_1 is the time from the valley to the point of maximum acceleration of the anacrota, α_2 is the time from the point of maximum acceleration of the anacrota to the peak, β is the duration of the catacrota, and T is the period of the pulse signal. The minimum condition for using signals in further analysis is the recognition of periods, in particular the maxima of the periodic signal. The parameters α_1, α_2, and β are defined as:

$$\alpha_1 = \frac{t_{max'1} - t_{min}}{T},$$

$$\alpha_2 = \frac{t_{max} - t_{max'1}}{T},$$

$$\beta = \frac{t_{max'2} - t_{max}}{T}.$$

The points $t_{max'1}$ and $t_{max'2}$ are the coordinates of the neighboring largest maxima of the pulse curve steepness, the difference between which determines the duration of the quasi-period T; t_{max} and t_{min} are the coordinates of the maximum and minimum of the pulse curve.

For the set of typical parameters in the study [4], six time parameters were selected, namely: 1) heart rate; 2) systolic peak; 3) catacrotic, or peak-to-valley, interval; 4) valley-to-valley interval; 5) peak-to-peak interval; and 6) pulse wave width at the incisura level.

The area under the PPG signal within the cardiac cycle is a rather informative feature (pulse area/AUC). The AUC parameter of the PPG signal is related to the blood volume per cycle (analogous to stroke volume) and tissue perfusion.

For each cardiac interval:

$$AUC = \int_{t_{max'1}}^{t_{max'2}} PPG(t)dt$$

In discrete form:

$$AUC \approx \sum t_i \Delta t,$$

where t_i – sampling times, Δt – sampling period.

The most appropriate integration method is the trapezoidal rule, as it provides a linear approximation that is stable in the presence of noise in "real" signals. $\int f(t)dt \approx \frac{\Delta t}{2} \sum (f_i + f_{i+1})$.

Another integration option is Simpson's method, which uses the Cotes formula for smooth functions with a quadratic approximation. This method provides higher accuracy but is sensitive to the presence of noise:

$$\int_a^b f(t)dt \approx \frac{\Delta t}{3}(f(t_0) + 4f(t_1) + 2f(t_2) + \ldots).$$

Here $\Delta t = (b - a)/n$, $t_0 = a$ is the beginning of the segment, $t_n = b$ is the end of the segment, and $t_0, \ldots, t_n$ denote the cardiac interval of the pulse wave divided into n equal subintervals.

2.3 Justification of the choice of amplitude and phase spectral parameters of pulse signals

In [13,14], an approach for assessing the functional state of the human cardiovascular system is proposed based on determining the spectral amplitude and phase parameters of pulse signals. These parameters can be determined in two ways.

The first is based on spectral analysis of a pulse-wave segment of a certain duration selected using window processing. An example of an amplitude spectrum obtained in this way is shown in Fig. 2. Diagnostic parameters of the amplitude spectrum for this analysis variant are the coordinates of the pulse-signal harmonics on the frequency axis (i.e., heart rate), the ratio of the harmonic amplitudes, and the "blurring" of the spectrum around the harmonics, which characterizes heart rate variability.

The second option is based on periodic analysis of the pulse wave signal. In this case, the amplitude and phase spectrum parameters are calculated for each signal pulse. The period T of each pulse is determined by measuring the pulse duration between the points

$t'_{max}1$ and $t'_{max}2$ (the time points of the maxima of the first derivative of the anacrotic parts of two consecutive pulses), and the spectral composition is determined by the nature of the pulse curve between these points. Determining each quasi-period from the maxima of the first derivative also reduces the influence of instability in the start and end points of cardiac cycles, although the quality of segmentation remains an important prerequisite for reliable phase estimation. By expanding this function into a Fourier series, the amplitude c_n (1) and phase φ_n spectra are determined (see Fig. 2 for the amplitude spectrum):

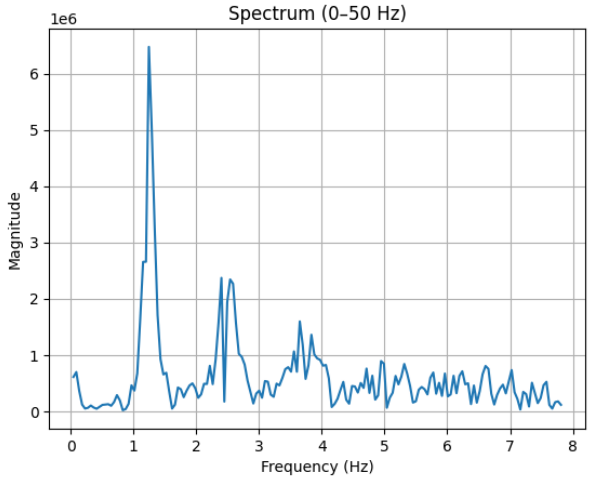

Fig. 2. Amplitude spectrum of the windowed portion of the pulse wave over an integer number of cardiac cycles

It should be noted that a discrete Fourier transform is chosen to form the phase spectral characteristic within one quasi-period. The fast Fourier transform option is not suitable, since the number of samples from which the transform is calculated is not necessarily equal to 2^N. Given the requirement for the most accurate formation of the spectral characteristic, this limitation cannot be neglected. For diagnostic analysis, the first four or five harmonics are usually sufficient; therefore, the relevant upper-frequency range is approximately 10–15 Hz, while higher harmonics have a substantially smaller influence on the considered parameters.

Spectral components are obtained using the known Fourier transform dependencies:

$$\begin{aligned} F(t) &= \sum_{n=0}^{N} c_n \cos\left(\frac{2\pi nt}{N} + \varphi_n\right), \\ c_n &= \sqrt{a_n^2 + b_n^2}, \\ \varphi_n &= -\arctan\left(\frac{b_n}{a_n}\right), \end{aligned} \tag{1}$$

where

$$\begin{aligned} a_n &= \frac{1}{T}\int_0^T F(t)\cos\frac{2\pi nt}{T}dt, \\ b_n &= \frac{1}{T}\int_0^T F(t)\sin\frac{2\pi nt}{T}dt. \\ a_0 &= \frac{1}{2T}\int_0^T F(t)\,dt. \end{aligned} \tag{2}$$

As an example, Fig. 3 shows a typical graph of the phase spectrum of a pulse signal. The graph is obtained by superimposing the phase spectra of 25 periods of the pulse wave. The phases of the harmonics are plotted as positive values for the convenience of further calculation of the phase and group delay times.

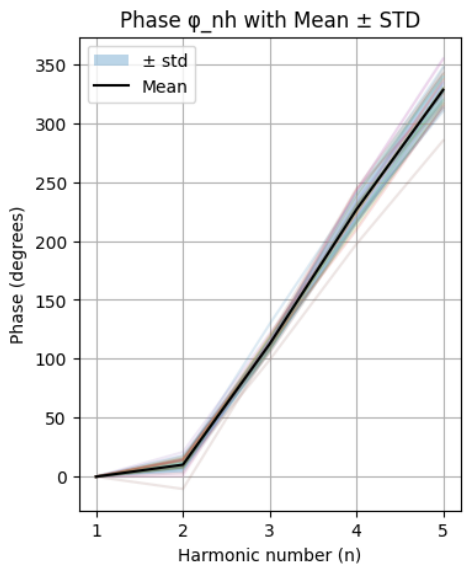

Fig. 3. Variability of the phase spectrum parameters fi

The parameters in Fig. 3 characterize the relative shift of the initial phases (φ_n) of the first five harmonics ($n = 1, 2, \ldots, 5$) relative to the initial phase of the first harmonic, which is taken as zero in the graph. Here, "blurring" is also observed, which is due to the variability of the initial phases of the harmonics under the influence of heart rate variability. Diagnostic indicators of the phase spectrum are the levels and ratios of the initial phases of the harmonics, their variability, and the slope of the spectrum.

In [13, 14], new time parameters of the pulse-signal phase spectrum, τ_{gr} and τ_{ph}, were proposed for determining the dynamic functional state of a person. These parameters are equivalent to the group and phase propagation times of the signal, respectively. They characterize the relative propagation speed of the pulse wave and are calculated by the formulas:

$$\tau_{gr_{n,n-1}} = \frac{T(\varphi_n - \varphi_{n-1})}{2\pi}, \tag{3}$$

$$\tau_{ph_n} = \frac{T(f_{i_n} - f_{i_{n-1}})}{2\pi(n-1)}, \tag{4}$$

where φ_n is the phase of the n-th harmonic of the spectrum, and $f_{i_n} = \varphi_n - n\varphi_1$.

The study in [14] notes that it is also advisable to select parameters that characterize the normalized phase spectrum indicators (in degrees) for the first five harmonics of the signal, in particular:

$$\begin{gathered} f_{i_n} = \varphi_n - n\varphi_1, \\ \varphi_n - \varphi_{n-1}. \end{gathered} \tag{5}$$

In general, the parameters $c_1 \dots c_n$ (1) characterize the relative dependence of the amplitudes of the pulse harmonics (reduced to the amplitude of the first harmonic, the relative value of which is taken to be 100). The parameters $\tau_{gr2} \dots \tau_{gr5}$ characterize the dependence of the time parameter of the phase spectrum, with units of seconds, equivalent to the group delay time of the signal in the spectral interval between the corresponding harmonics of the pulse signal. The parameters $\tau_{ph2}, \dots, \tau_{ph5}$ characterize the dependence of the time parameter of the phase spectrum, with units of seconds, equivalent to the phase delay time of the signal in the spectral interval between the higher and first harmonics of the pulse signal. These spectral parameters reflect both vascular structural properties and dynamic hemodynamic changes; therefore, their interpretation should be made together with amplitude-time parameters and metadata rather than in isolation.

Examples of assessing the dynamics and diagnostic sensitivity of spectral parameters of pulse signals under various test influences are given in [13, 14] and other works by the authors.

Studies assessing the diagnostic sensitivity of the specified spectral parameters of PPG signals were performed using various test influences: orthostatic and cold tests, tea and alcohol consumption, smoking, breath holding, sports training, and recovery after abdominal surgery. These studies showed high sensitivity of the specified parameters to such influences. However, certain features were observed. Thus, in the presence of test influences that are not related to mechanical muscle tension or psychological stress, the amplitude spectrum is more suitable for visual observation, whereas under influences involving muscle tension or psychological stress, it is more suitable to use the phase spectrum and its time parameters. It was also found that when assessing the functional state of a person under postoperative rehabilitation conditions (in the first hours of the day after surgery), the dynamics of changes in the functional state are monitored more clearly using spectral parameters than by blood pressure measurements, which are traditionally used in practice.

The ability to postprocess original biosignals to extract new amplitude-temporal parameters or refine existing ones is also important.

3 Selection of biosignal parameters to form a dataset for machine learning

From the above list of analytically formalized diagnostically valuable indicators, it is advisable for machine-learning-based express diagnostics to choose multiparameter neural network models and to consider the potential use of deep neural networks. When convolutional neural networks are used to represent morphological features, it is also possible to automate the detection and extraction of characteristic morphological diagnostic features of the pulse wave in deeper network layers. Thus, CNN-based processing of raw or weakly segmented PPG fragments should be regarded as a promising complementary direction rather than as an approach excluded by the present feature-based methodology.

To determine the functional state of a person based on biosignals, it is important to establish parameters and their normalized numerical indicators that can be classified as normal for the population, as well as those that deviate from normal values and may indicate the presence or development of disease. In other words, certain simple rules or threshold functions can be used to carry out express diagnostics.

Determining features (biosignal parameters) that are informative for a specific task is important for reliable assessment of the functional state of a person.

To identify truly informative features from the PPG shape based on the considered parameters, correlation and statistical analyses of the extracted biosignal parameters must be conducted. The key issue is to determine which of the considered parameters actually carry physiological information and which are only derivatives of heart rate (HR) or noise.

The informativeness of a parameter is acceptable if:

- there is no trivial dependence on heart rate (HR), as in the example with AUC;
- there is small and stable dispersion between cycles within one functional state;
- there is sensitivity to changes in the functional state and variability under load or pathology;
- there is a possibility of explaining the physiological meaning of the parameter in terms of hemodynamics.

To select the most significant features for blood pressure monitoring based on the PPG waveform, a multicollinearity test was applied in [2]. The presence

of multicollinearity among predictor variables affects the generalizability of the algorithm by increasing the estimated mean square error.

The information content of the parameters was assessed using the following algorithm for all amplitude-time and spectral-phase parameters considered above, including (c_n, τ_{gr}, τ_{ph}, $\Delta\varphi$, α/β, AUC, etc.); in total, 21 parameters were assessed.

The selection of informative features included normalization, HR filtering, correlation filtering, permutation importance, and Recursive Feature Elimination (final selection).

3.1 Multicollinearity test

Heart rate (HR) is a highly informative physiological feature that should be retained in the feature set. At the same time, features that are strongly correlated with HR should be removed to reduce redundancy and mitigate multicollinearity:

$$\|corr(f_i, HR)\| < \gamma.$$

Removing multicollinearity:

$$|corr(f_i, f_j)| < \delta,$$

where γ and δ are threshold values.

3.2 Feature importance analysis

To assess the contribution of the extracted photoplethysmographic (PPG) features, a random forest importance analysis was performed [21–23]. This algorithm can tolerate moderate noise in extracted features, but it does not replace signal-quality screening for PPG segments recorded by wearable devices, such as smartwatches. The importance of each feature was quantified as the increase in prediction error after a random permutation of its values.

$$I_i = E[L(f) - L(f_{\pi(i)})],$$

where L is the loss function, $f_{\pi(i)}$ is the feature with permuted values.

The results showed that spectral and phase features provide the most significant contribution to model performance in the considered feature set. In particular, the group delay τ_{gr} and the phase differences between harmonics $\Delta\varphi_n$ showed the highest importance scores. These features reflect temporal and structural properties of the pulse wave and may be associated with vascular stiffness and wave propagation dynamics.

In contrast, amplitude features such as the raw area under the curve (AUC) showed less importance due to their strong dependence on heart rate (HR). After normalization for cardiac cycle length, the contribution of AUC increased but remained secondary compared to phase-related descriptors.

3.3 SHAP interpretation

To further interpret the model predictions, SHAP (SHapley Additive exPlanations) values were calculated. This approach allows the model output to be decomposed into the additive contributions of individual features.

Feature contribution:

$$\phi_i = \sum_{S \subseteq Fi} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)],$$

where F is the set of all features, S is a subset of features, and $f(S)$ is the model prediction if only S is used.

The results of SHAP analysis are shown in Figs. 4 and 5, which show the dominant role of spectral features of PPG signals over amplitude-time features in the analyzed model. The summary plot showed that higher values of τ_{gr} usually lead to an increase in the predicted output signal, indicating an association with the underlying physiological parameter within this dataset. Similarly, the phase difference $\Delta\varphi_n$ showed a nonlinear effect, indicating complex interactions between harmonic components. These results should be interpreted as model-specific feature-importance estimates and require confirmation on independent datasets and sensors.

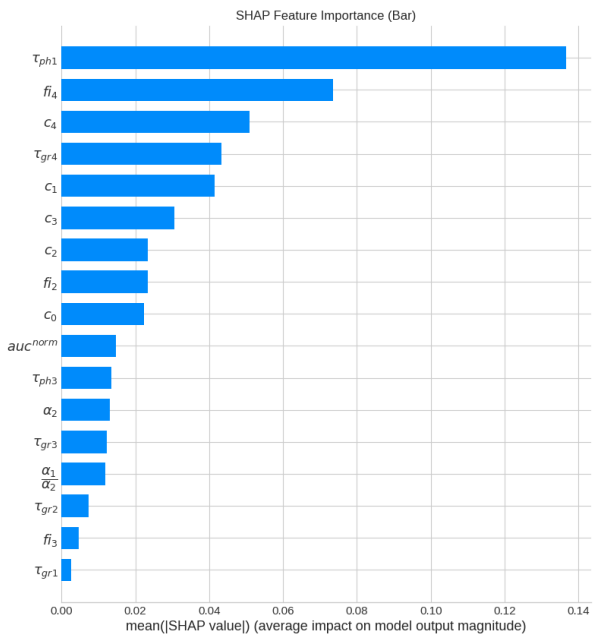

Fig. 4. Potential contribution of PPG signal parameters in machine learning models

Time-domain features, including the ratios α_1/α_2 and $(\alpha_1 + \alpha_2)/\beta$, also showed moderate importance. These parameters characterize the morphology of the PPG waveform and reflect the balance between systolic and diastolic phases.

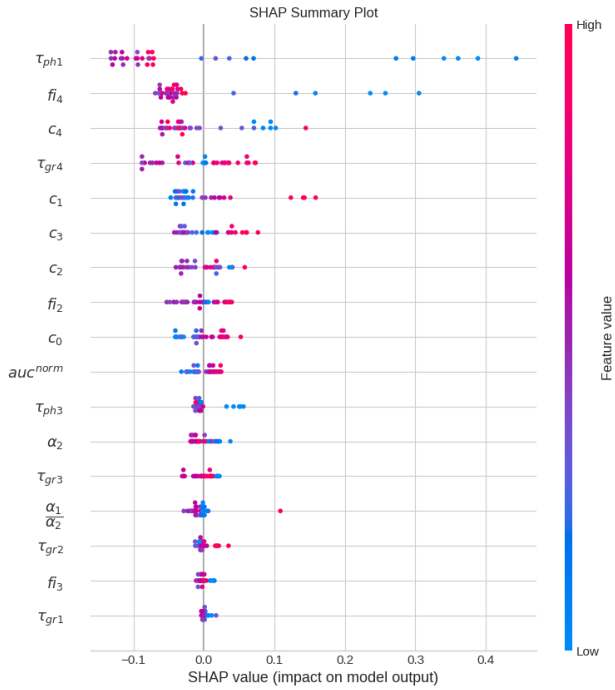

Fig. 5. Visualization of the potential contribution of PPG signal parameters in a machine learning model using the SHAP method

3.4 Stability and redundancy of features

Additional filtering steps were applied to improve the robustness of the selected features. Features with high correlation with heart rate ($|\mathrm{corr}(f_i, HR)| > 0.7$) were excluded to reduce trivial dependencies. In addition, highly correlated features ($|\mathrm{corr}(f_i, f_j)| > 0.9$) were removed to reduce redundancy.

The analysis and removal of trivial dependencies resulted in the remaining features showing improved stability across cardiac cycles and reduced multicollinearity. However, because several phase-derived descriptors are mathematically related through the same Fourier representation and cardiac period, multicollinearity cannot be considered fully eliminated by threshold filtering alone. The coefficient of variation (CV) for the selected features was significantly lower compared to the original set, indicating consistent behavior within each record.

3.5 PCA analysis

The PCA biplot is shown in Fig. 6. In this figure, the length of the arrow determines the importance of the feature: long indicates strong influence, and short indicates weak influence. The angle between the arrows, $\cos(\theta)$, indicates correlation. Small angles indicate a significant correlation between the parameters, 90° indicates independence, and 180° indicates anticorrelation.

For PPG signals, τ_{gr} and $\Delta\varphi$ are expected to group together (wave structure), α/β is expected to form a separate direction (shape), and AUC often appears separately (HR influence). Such grouping helps visualize the structure of the feature space but does not by itself prove physiological universality.

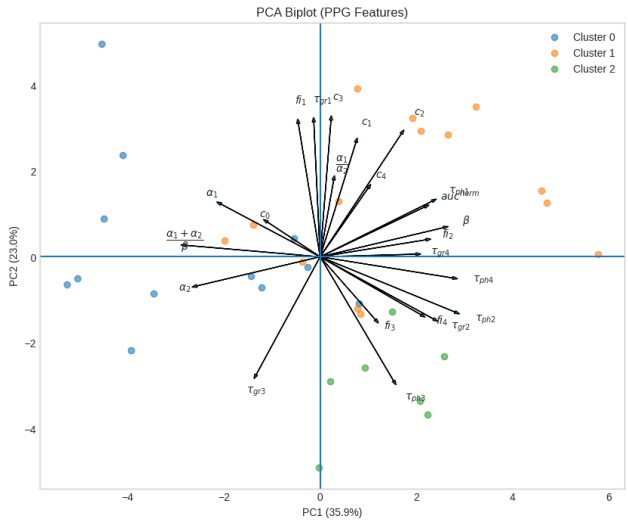

Fig. 6. PCA analysis of PPG signal parameters

From Fig. 6, it follows that phase-related features cluster along the first principal component, while parameters related to morphology mainly affect the second component. In the biplot, the rightward direction indicates influence on PC_1, and the upward direction indicates influence on PC_2.

3.6 Final set of features

After selection, an optimal feature subspace is formed:

$$f^* = \{f_i | I_i^{max}, corr(f_i, HR)^{min}\}.$$

Based on a combined analysis of permutation importance, SHAP values, and statistical filtering, the following features were identified as the most informative:

- Group and phase delay τ_{gr} and τ_{ph};
- Phase difference $\Delta\varphi_n$;
- Normalized amplitudes of harmonics c_n;
- Ratios α_1/α_2 and $(\alpha_1 + \alpha_2)/\beta$;
- Normalized area under the curve AUC^{norm}.

These features represent the PPG signal in terms of both amplitude-time and spectral characteristics and can be effectively used for further analysis and classification under appropriate signal-quality control.

4 Discussion

Among the considered amplitude-time parameters, there are no parameters related to the dicrotic notch. From a practical point of view, its detection is difficult, since it is not clearly present in all signals. In addition, some PPG signals contain several similar dicrotic notches, which can reduce the accuracy of automatic parameter determination using amplitude-time methods. Instead, the spectral representation uses information from the entire pulse waveform, including cases in which the dicrotic wave is absent or poorly expressed. Therefore, the presence or absence of the dicrotic notch can still affect the spectral-phase parameters and their evolution over cardiac periods, without requiring the notch to be identified as a separate time-domain point.

An inverse relationship was observed between AUC and heart rate, which is consistent with the dependence of the integration interval on the duration of the cardiac cycle. This emphasizes the need for normalization procedures when using features based on integral measures.

Overall, the results show that phase- and delay-based features are informative within the analyzed feature set and may provide physiologically meaningful information that complements amplitude-based descriptors. The proposed feature selection pipeline identifies a compact subset of parameters suitable for modeling and interpreting PPG signals, but the stability of this subset should be verified on independent data.

The considered methods of extracting and representing diagnostic features demonstrate the potential of using biological signals for rapid diagnostics. However, although individual parameters extracted from biosignals and their graphical representations may be diagnostically significant, their interpretation is complex and requires the experience of expert physicians. This is especially relevant for spectral-phase parameters, whose physiological interpretation is less direct than that of many contour parameters. Processing and analysis of biosignal parameters can be performed manually based on the previous experience of physicians. Despite the above limitations, this approach to diagnostics and analysis of biosignal parameters should not be neglected. Moreover, the development of machine learning methods and neural networks has significant potential for the automatic search for hidden patterns, for which it is necessary to train a neural network based on expert physicians' assessments and to create appropriate datasets.

Several limitations should be emphasized. First, spectral-phase analysis requires high-quality PPG segments; therefore, a pre-classifier for signal quality and rejection of motion-contaminated fragments is necessary for practical use with wearable devices. Second, the dominance of spectral features in SHAP analysis may partly depend on the dataset, sensor characteristics, and the transformation into the Fourier domain; therefore, external validation is required before generalizing these findings. Third, the relationship between spectral parameters and specific biochemical or pathophysiological changes should be verified in targeted studies rather than inferred from feature importance alone. Accordingly, curated expert-labeled datasets are required for reliable training and validation of neural network models.

Conclusions

The feasibility of using a comprehensive approach to the analysis of PPG signals, which combines amplitude-time and spectral amplitude and phase parameters, is substantiated. It is shown that traditional amplitude-time characteristics do not provide a complete description of the pulse wave morphology and may have limited informativeness in machine learning tasks when used alone.

A signal preprocessing method based on "soft" filtering and segmentation by cardiac cycles is proposed, which helps preserve diagnostically significant signal features and supports reliable parameter extraction from high-quality PPG segments.

A generalized set of features is formed, which includes amplitude-time and spectral amplitude and phase parameters obtained within individual cardiac cycles. The information analysis showed that, within the considered feature set, the most significant features include the parameters of the phase spectrum and group and phase delay, which reflect the dynamics of pulse wave propagation and may be associated with the physiological properties of the vascular system.

It was found that the amplitude-time (contour) characteristics of pulse signals, in particular the area under the curve, require normalization to increase the reliability of diagnostics due to their dependence on the heart rate, while spectral-phase features appeared more stable and informative within the considered analysis.

As a result of feature selection, a compact subspace of parameters was formed, which supports increased accuracy and stability of machine learning models. The results obtained confirm the potential effectiveness of the proposed approach for non-invasive express diagnostics tasks, provided that signal quality is controlled and the models are validated on representative datasets.

Further research should be directed at expanding the experimental database, improving methods for automatic feature extraction and signal-quality classification, comparing feature-based analysis with CNN-based processing, and validating the physiological interpretation of spectral-phase parameters across sensors and populations.

References

[1] Kyriacou P. and Allen J. (2021) *Photoplethysmography: Technology, Signal Analysis and Applications*, 1st ed. Elsevier.

[2] Khalid S. G., Zhang J., Chen F. and Zheng D. (2018) Blood pressure estimation using photoplethysmography only: Comparison between different machine learning approaches, *Journal of Healthcare Engineering*, vol. 2018, 1548647, doi: 10.1155/2018/1548647

[3] Visvanathan A., Sinha A. and Pal A. (2013) Estimation of blood pressure levels from reflective photoplethysmograph using smart phones, *13th IEEE International Conference on BioInformatics and BioEngineering*, pp. 1-5. doi: 10.1109/BIBE.2013.6701568

[4] Ali N. F. and Atef M. (2022) Lstm multi-stage transfer learning for blood pressure estimation using photoplethysmography, *Electronics*, vol. 11, 22, doi: 10.3390/electronics11223749

[5] El-Hajj C. and Kyriacou P. (2020) A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure, *Biomedical Signal Processing and Control*, vol. 58, p. 101 870. doi: 10.1016/j.bspc.2020.101870

[6] Priyadarshini R. G., Kalimuthu M., Nikesh S. and Bhuvaneshwari M. (2021) Review of ppg signal using machine learning algorithms for blood pressure and glucoseestimation, *IOP Conference Series: Materials Science and Engineering*, vol. 1084, pp. 012-031. doi: 10.1088/1757-899X/1084/1/012031

[7] Arg¨uello-Prada E. J., Ojeda A. V. V., and Ojeda M. Y. V. (2025) Non-invasive prediction of cholesterol levels from photoplethysmogram (ppg)-based features using machinelearning techniques: A proof-of-concept study, *Cogent Engineering*, vol. 12, no. 1, p. 2 467 153. doi: 10.1080/23311916.2025.2467153

[8] Elgendi M. (2020) *PPG Signal Analysis*. 1st ed. CRC Press. [Online]. Available: https://www.perlego.com/book/1974664/ppg-signal-analysis-an-introduction-using-matlab-pdf

[9] Seenivasan D. and Sakthivel K. (2025) Patch-based ecg segmentation for arrhythmia classification using hybrid deep learning, *Information Technology and Control*, vol. 54, no. 4, pp. 1189-1205. doi: 10.5755/j01.itc.54.4.41549

[10] Seitsonen E. et al. (2005) EEG spectral entropy, heart rate, photoplethysmography and motor responses to skin incision during sevoflurane anaesthesia, *Acta anaesthesiologica Scandinavica*, vol. 49, pp. 284-292. doi: 10.1111/j.1399-6576.2005.00654.x

[11] Alty S. R., Angarita-Jaimes N.,Millasseau S. C. and Chowienczyk P. J. (2007) Predicting arterial stiffness from the digital volume pulse waveform, *IEEE Transactions on Biomedical Engineering*, vol. 54, pp. 2268-2275. doi: 10.1109/TBME.2007.897805

[12] Awad A. A. et al. (2007) The relationship between the photoplethysmographic waveform and systemic vascular resistance, *Journal of Clinical Monitoring and Computing*, vol. 21, pp. 365-372. doi: 10.1007/s10877-007-9097-5

[13] Sharpan O. B. (2003) Diagnostic sensitivity of the phase spectrum of the pulse signal, *Elektronika i svyaz'*, No. 18, pp. 91-95.

[14] Sharpan O. B., Husieva O. V., Mahliovana N. I. and Rybin O. I (2007) The program-measuring photoplethismographic system of pulse spectrometry. *Bulletin of Vinnytsia Polytechnic Institute*, no. 5, pp. 34-40, 2007.

[15] Rinkevi˘cius M. et al. (2019) Photoplethysmogram signal morphology-based stress assessment, *2019 Computing in Cardiology (CinC)*, pp. 1-4. doi: 10.22489/CinC.2019.126

[16] Elgendi M. (2016) Optimal signal quality index for photoplethysmogram signals, *Bioengineering*, vol. 3, no. 4. doi: 10.3390/bioengineering3040021

[17] Moscato S., Lo Giudice S., Massaro G., and Chiari L. (2022) Wrist photoplethysmography signal quality assessment for reliable heart rate estimate and morphological analysis, *Sensors*, vol. 22, 15, doi: 10.3390/s22155831

[18] Yousef Q., Reaz M.B.I. and Ali M. (2012) The analysis of ppg morphology: Investigating the effects of aging on arterial compliance, *Measurement Science Review*, vol. 12, pp. 266-271. doi: 10.2478/v10048-012-0036-3

[19] Hickey M. and Phillips J. (2015) Investigation of peripheral photoplethysmographic morphology changes induced during a hand-elevation study, *Journal of Clinical Monitoring and Computing*, vol. 30. doi: 10.1007/s10877-015-9761-0

[20] Reguig F. B. (2016) Photoplethysmogram signal analysis for detecting vital physiological parameters: An evaluating study, *2016 International Symposium on Signal, Image, Video and Communications (ISIVC)*, pp. 167-173. doi: 10.1109/ISIVC.2016.7893981

[21] Gonzalez-Landaeta R., Ramirez B. and Mejia J. (2022) Estimation of systolic blood pressure by random forest using heart sounds and a ballistocardiogram, *Scientific Reports*, vol. 12, no. 1, p. 17196, doi: 10.1038/s41598-022-22205-0

[22] Benchekroun M. et al. (2022) Comparison of stress detection through ecg and ppg signals using a random forest-based algorithm, *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* vol. 2022, pp. 3150-3153. doi: 10.1109/EMBC48229.2022.9870984

[23] Lee S.-W., Woo D.-K., Son Y.-K. and Mah P. S. (2019) Wearable bio-signal(ppg)-basedpersonal authentication method using random forest and period setting consideringthe feature of ppg signals, *J. Comput.*, vol. 14, pp. 283-294. doi: 10.17706/jcp.14.4.283-294.

Вибір амплітудно-часових і спектральних параметрів PPG сигналів для діагностики на основі методів машинного навчання

Мосійчук В. С., Шарпан О. Б., Гусева О. В.

Розглянуто задачу підвищення ефективності неінвазивної експрес-діагностики функціонального стану організму на основі аналізу фотоплетизмографічних (PPG) сигналів із використанням методів машинного навчання.

На основі аналізу морфологічних особливостей пульсової хвилі, які містять важливу інформацію про гемодинамічні процеси, обґрунтовано доцільність переходу від традиційного аналізу амплітудно-часових характеристик сигналу до комплексного підходу, що включає додаткове вилучення спектрально-фазових параметрів у межах окремих кардіоциклів.

Запропоновано методику попередньої обробки сигналів, що включає смугову фільтрацію, нормалізацію та сегментацію, а також підхід до фільтрації в смузі частот до 0,5 - 30 Гц та сплайнової ітерполяції сусідніх вибіро дискретизації в зареєстрованих сигналів, який дозволяє зберегти діагностично значущі особливості форми сигналу. Детально розглянуто систему амплітудно-часових параметрів (тривалість анакроти, інтервали між характерними точками, площа під кривою тощо) та спектрально-фазових характеристик, отриманих за допомогою розкладу у ряд Фур'є.

Особливу увагу приділено процедурі відбору інформативних ознак для моделей машинного навчання. Запропоновано багатокроковий підхід, що включає усунення залежності від частоти серцевих скорочень, аналіз мультиколінеарності, оцінку важливості ознак методом permutation importance та інтерпретацію результатів за допомогою методів PCA та SHAP. Встановлено, що найбільший внесок у точність моделей забезпечують спектрально-фазові параметри, зокрема групова затримка та фазові зсуви гармонік, які безпосередньо пов'язані з фізіологічними властивостями судинної системи.

У результаті сформовано набір ознак, що поєднує часові та спектральні характеристики сигналу та забезпечує підвищення якості класифікації і прогнозування. Отримані результати підтверджують перспективність використання комплексного аналізу PPG-сигналів у поєднанні з методами машинного навчання для задач медичної діагностики та моніторингу стану людини.

Ключові слова: неінвазивна діагностика; аналіз фізіологічних сигналів; обробка біосигналів; виділення ознак; вибір ознак; аналіз у частотній області; фазовий спектр; фотоплетизмографія (ФПГ); аналіз пульсової хвилі; машинне навчання