# Using Mel-Frequency Cepstrum and Amplitude-Time Heart Variability as XGBoost Handcrafted Features for Heart Disease Detection

SS Krivenko[1], AA Pulavskyi[1], LS Kryvenko[2], OV Krylova[2], SA Krivenko[3]

[1]Scientific Company KOLIBRI LLC, Kharkiv, Ukraine
[2] Kharkiv National Medical University, Kharkiv, Ukraine
[3] Kharkiv National University of radioelectronics, Kharkiv, Ukraine

## Abstract

*We have developed the XGBoost model to identify 27 heart pathologies within the challenge Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/ Computing in Cardiology Challenge 2021. The technical part included several stages. At the first stage, the ECG was cut off to 10 seconds. At the second stage, resampling to frequencies 125 and 500 Hz was carried out and filtering in the 0.5-45 Hz bands. At the third stage, the features of HRV and symbolic dynamics were extracted from the signal with a sampling rate of 125 Hz. The melspectrograms were calculated based on a signal with a sampling frequency of 500 Hz. Then the features calculated for each lead were concatenated to obtain the final vector of features. We were faced with the task of constructing 27 independent binary classifiers, each of which defines a certain pathology. The fourth important step was to build balanced datasets for the algorithm. For the robustness of the models, the control groups for each contained almost all pathologies presented in the databases, except target disease.*

*Our team Sunset scored 0.22, 0.21, 0.22, 0.21, 0.20 for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead models, respectively, ranking 32 out of 39 teams for the first four lead combinations and 31 out of 39 teams for the last.*

## 1. Introduction

The automation of medical diagnostics is a very urgent task, especially for the tasks of mass screening. Proper diagnosis and treatment can reduce mortality, disability and hospitalization. The PhysioNet / Computing in Cardiology Challenge 2021 aims to develop open automated methods for the classification of various cardiac pathologies determined from a reduced set of ECGs [1, 2]. In real life, the process of determining the pathology of the heart requires the availability of a qualified cardiologist and carrying out a series of diagnostic procedures. This can be time consuming and expensive. A simple, cheap and automated solution can be helpful for screening diagnostics at the first, most massive, stage without involving a doctor. The early screening for diabetes mellitus is an example of such a solution [3]. One of the goals was to create a simple solution using the standard Python libraries for processing ECG / Heart rate variability (HRV) and obtaining a mel-frequency spectrogram (melspectrogram). It was decided to use three different groups of parameters as input to the model namely data of HRV, symbolic dynamics and melspectrogram. Each of these groups has proven itself well in its subject area.

## 2. Method

### 2.1. Data preprocessing

The competition used ECGs from 6 databases [4-8]. The main feature of these databases is that they contain short ECG segments, most of which are 10-30 seconds long. For our models, we used ECG segments with a baseline duration of 10 seconds. If an ECG segment was shorter than 10 seconds, it was increased to 10 seconds by adding a copy of the segment itself. Further use of such augmented segments may be incorrect for calculating HRV parameters. But due to the small number of such segments, we decided not to remove such segments from the general dataset. If an ECG segment was longer than 10 seconds, we used the first 10 seconds of the segment. Our models used two sampling rates: 125 Hz and 500 Hz. That was done to speed up some resource-intensive operations when generating features. Accordingly, each ECG segment was resampled to those two frequencies. The next stage of pre-processing was filtration. We used a 2nd order Butterworth bandpass filter. The lower frequency of the filter was 0.5 Hz and the upper frequency was 45 Hz. This frequency band is sufficient to preserve the characteristic sharp fragments of the ECG morphology inherent in some pathologies. At the same time, such a filter effectively removes power line noise and baseline drift. After that, the dynamic range of the signal was calculated. If the dynamic range was less than 0.5 mV, then the signal was normalized to the dynamic range. Such processing was necessary for

the correct further operation of a number of algorithms that identify a signal with a small dynamic range as noise. The next step was to determine the R-peaks and their amplitudes using a QRS detector.

## 2.2 Handcrafted features

To build an effective model, we used specific domain knowledge. Three groups of features were generated: HRV (57 features), symbolic dynamics (86 features) and melspectrograms (10 spectrograms, 64 elements each). The features of HRV and symbolic dynamics took into account the characteristics of the autonomic nervous system (ANS), but the melspectrograms were based on the ECG morphology. Besides, the age and gender of the patient were used as features. These features were generated for each lead used (i.e. 785 of features were generated for each lead). Then the vectors of features were concatenated to form the final vector of features. For the features of HRV and symbolic dynamics segments with a sampling rate of 125 Hz were used; the same segments were used for melspectrograms but with a sampling rate of 500 Hz. Let us describe these features in detail.

### 2.2.1 HRV features

Table 1 provides a list of the features used. All the features were divided into three classes. They are the following: time (CVNN, CVSD, MAD, MHR, MRRI, NNI parameters, ΔNNI parameters, MadNN, MeanNN, MedianNN, MCVNN, pNN20, pNN50, RMSSD, SDANN, SDNN, SDNN_index, SDSD, TINN), frequency (ULF, VLF, LF, LFn, LF Peak, LF Relative HF, HFnu, HF Peak, HF Relative, LF/HF), nonlinear (SD1, SD2, S, SD1/SD2, SampEn, DFA, CSI Modified, CSI, CVI). The features listed above are indicated by abbreviations as the use of full titles is impossible due to the limited volume of the publication. Makowski D. et al gives detailed description for each one [9]. An important fact is that most HRV parameters must be calculated using short-term ECG recordings (usually 5 minutes). Thus, the application of computational algorithms to ultrashort segments goes beyond their standard use. Nevertheless, these values have a certain semantic significance. Considering that, some of the parameters could not be calculated, and their values were replaced by 0.

### 2.2.2 Symbolic dynamics

HRV is usually calculated as the time difference of successive heartbeats. However, there is also amplitude variability of the ECG maxima, including the R-peak. To find differentiating properties of HRV, we used 5 methods of symbolic analysis based on temporal HRV and 1 method based on amplitude-time HRV. The frequency of

occurrence of different characters from finite alphabets or words of constant length, composed of several consecutive characters, results from the transformation of these methods. These frequencies are considered biomarkers of different pathological conditions. At the same time, unless otherwise stated, normalized frequencies were used to level the number of elements of the original IBI. A brief description of the methods is presented below.

### 2.2.2.1 Amplitude-time HRV

This method is the only method in the work that takes into account, in addition to the time, the amplitude variability of the R-peak [3]. This method encodes the increments of amplitudes, time intervals, as well as the tangent of the angle, which takes into account the conditional lines of the segments of the amplitude and time changes as legs in the calculation. In this case, the increase in amplitude in adjacent heart beats is defined as $dR = R_i - R_{i-1}$. The time increment is defined as $dT = IBI_i - IBI_{i-1}$. The tangent of an angle is defined as $tan(\alpha) = IBI_i/R_i$. Accordingly, the change in the tangent of an angle can be written as $dA = tan_i - tan_{i-1}$. Considering the above definitions, we can imagine the definition of a symbol of a new sequence from a finite alphabet A = {0, 1, 2, 3, 4, 5}:

$$TS_i = \begin{cases} 0: dT_i < dT_{i-1}, dR_i < dR_{i-1}, dA_i < dA_{i-1} \\ 1: dT_i \geq dT_{i-1}, dR_i \geq dR_{i-1}, dA_i < dA_{i-1} \\ 2: dT_i < dT_{i-1}, dR_i \geq dR_{i-1}, dA_i < dA_{i-1} \\ 3: dT_i \geq dT_{i-1}, dR_i < dR_{i-1}, dA_i \geq dA_{i-1} \\ 4: dT_i < dT_{i-1}, dR_i < dR_{i-1}, dA_i \geq dA_{i-1} \\ 5: dT_i \geq dT_{i-1}, dR_i \geq dR_{i-1}, dA_i \geq dA_{i-1} \end{cases} \quad (1)$$

The output properties are word frequencies. We used the values W = [1], T = [1, 2].

### 2.2.2.2 Ordinal patterns

This method refers to the analysis of time HRV and describes the relationship between short intervals of the time series [3]. This algorithm transforms the original time series into a new basis. This approach is robust to noise and computationally simple; however, it ignores the absolute values of the elements of the time series. For each W elements of the original row, the permutation index P of the ordered elements is calculated. Next, a new time series is formed from the permutation indices, which can be considered as a sequence of characters / words from a finite alphabet of size W!. In this case, it is also possible to introduce the concept of a time shift T, which determines the distance between the selected elements of the original time series. So, for T = 3, when forming the permutation indices, every 3 element of the initial time series (IBI) is used. Next, the frequencies of each of the permutation indices are calculated. We used a frequency analysis of

permutation indices for the sequences obtained for the values W = [3] and T = [1].

### 2.2.2.3 Delta coding

This method refers to temporary HRV [10]. Instead of absolute values, the difference between two adjacent IBI elements is used. This means an increase or decrease in heart rate. The element of the resulting sequence belongs to the finite alphabet [0, 1] and can be expressed by the formula:

$$TS_i = \begin{cases} 0, if\, IBI_i - IBI_{i-1} \geq 0 \\ 1, if\, IBI_i - IBI_{i-1} < 0 \end{cases}, \qquad (2)$$

where $TS_i$ is an element of the new binary time series, $IBI_i$, $IBI_{i-1}$ - current and previous elements of IBI. Further, from this binary series it is possible to form a time series of symbols, where each symbol is the union of several consecutive elements of the TS series. In this study, we used the following character lengths W = [4, 5, 6, 7]. It is possible to introduce a time offset T for the sequence for the original IBI row. In this work, we used T = [1, 2, 3]. The output properties are word frequencies. A special case of binary coding is the use of a specific time value τ instead of zero in comparison:

$$TS_i = \begin{cases} 0, if\, IBI_i - IBI_{i-1} \geq \tau \\ 1, if\, IBI_i - IBI_{i-1} < \tau \end{cases}, \qquad (3)$$

We used a frequency analysis of permutation indices for the sequences obtained for the values W = [3], T = [1] and τ= [5 ms].

### 2.2.2.4 Max-min method

This method belongs to the group of temporary methods [10]. To obtain a character series, the entire dynamic range of IBI is divided into γ equal quanta. In this case, one level of quantization is equal to $l = (max(IBI) - min(IBI))/\gamma$. Thus, the final alphabet consists of γ elements A = {0, 1, ..., γ-1}, and the element of the resulting symbolic sequence can be represented:

$$TS_i = \begin{cases} 0: min(IBI) \leq IBI_i < 1 \cdot l \\ 1: 1 \cdot l \leq IBI_i < 2 \cdot l \\ \qquad . \\ \qquad . \\ \gamma - 1: (\gamma - 1) \cdot l \leq IBI_i \leq max(IBI) \end{cases}, \qquad (3)$$

Usually, the value is γ = 6. The output properties are word frequencies. We used the values W = [1], T = [1, 2].

### 2.2.3 Mel-frequency spectrogram

A Mel-frequency spectrogram is a spectrogram where the frequencies are converted to the mel scale [11]. The mel scale is a perceptual scale of pitches judged by listeners to be equal in distance from one another. In 1937, a unit of pitch was proposed such that equal distances in pitch sounded equally distant to the listener. This is called the mel scale. We performed a mathematical operation on frequencies to convert them to the mel scale. Despite this approach is for an audio processing we used it for ECG processing. We generated 10 melspectrogram with 64 elements each. Length of the FFT window was 2048, and the number of samples between successive frames was 512, exponent for the magnitude melspectrogram was 2.

## 2.3 Imbalance problem

Due to the use of the XGBoost algorithm as the main classifier, it is necessary to separately consider the problem of class imbalance when building models. We were dealing with a multilabel multiclass classification task. Because this configuration is not provided by XGBoost natively, so we presented the problem as 27 independent binary classification problems. All ECGs were divided into two classes: class 0 (negative, control group) - this is the ECG class of patients who not suffering from THE target disease. Class 1 (positive, target group) is the ECG class of patients with target disease. To ensure the stability of the algorithm, in each control group, all pathologies were evenly represented, with the exception of the pathology of the target group. Where the same number of all pathologies could not be provided, the number of pathologies decreased. This approach guaranteed the search for differences precisely in the characteristics of the target group.

## 2.4 XGBoost model

The main machine learning algorithm in our work was XGBoost. This algorithm provides good characteristics of finding hidden dependencies with excellent reproducibility of results on new data. The hyperparameter settings are shown in Table 1. The rest of the hyperparameters were accepted by default.

| Parameter name | Value |
|---|---|
| n_estimators | 500 |
| max_depth | 7 |
| learning_rate | 0.005 |
| objective | binary:logistic |
| scale_pos_weight | 1 |
| eval_metric | aucpr |
| early_stopping_rounds | 100 |

Table 1. XGBoost model hyperparameter set

A stratified 5-folds cross-validator was used to check

correctness of the model. This decision is redundant due to the fact that the training data (target and control groups) were prepared (balanced) before cross-validation. Nevertheless, this approach is standard for solving problems of this kind. All metrics were calculated on a holdout set, randomly representing 20% of the total data.

## 3. Result and Discussion

In common medical practice, sensitivity and specificity values above 0.7 are acceptable for screening diagnostics. According to the results obtained, our method can be used for screening diagnostics purposes to detect 20 out of 27 diseases. The disadvantages of the method include the use of one type of spectra to take into account the morphology of the ECG. In the future, it will be possible to expand the range of methods that eliminate this problem and consider the features of the ECG morphology in more detail. Challenge scores are presented in Table 2. It is obvious that in spite of the adopted regularization, the model is overfitted. Further analysis revealed main reason for this behaviour. The reason is the large number of zero values in the sparse feature matrix. This is due to the fact that most of the methods used are designed for ECG segments 5 minutes long. Accordingly, for ultrashort segments, many properties were not defined and were assigned a value of zero.

| Leads | Training | Validation | Test | Ranking |
|---|---|---|---|---|
| 12 | 0.54±0.03 | 0.40 | 0.22 | 32 |
| 6 | 0.51±0.04 | 0.39 | 0.21 | 32 |
| 4 | 0.53±0.03 | 0.40 | 0.22 | 32 |
| 3 | 0.51±0.04 | 0.39 | 0.21 | 32 |
| 2 | 0.50±0.04 | 0.39 | 0.20 | 31 |

Table 2. Challenge scores for our final selected entry (team *Sunset*) using 5-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one-time scoring on the hidden test set as well as the ranking on the hidden test set.

## 4. Conclusions

A particularity of our XGBoost model is a set of features used as input to the model. These features are divided into 3 classes: HRV features, symbolic dynamics and melspectrograms. The name of our team is Sunset. Our algorithm achieved the following results on the final test data of the PhysioNet/CinC Challenge 2021: ranked 32nd (out of officially qualified 39 teams) for 12-lead model - 0.22, 6-lead model - 0.21, 4-lead model - 0.22, and 3-lead model - 0.21, and ranked 31st for 2-lead model - 0.20. This result allows us to say with confidence that the reduced-set ECG works no worse or slightly worse than the full-set ECG. This opens up prospects for the reliable use of such approaches in telemedicine.

## References

[1] Perez Alday EA, Gu A, Shah AJ, Robichaux C, Wong AI, Liu C, Liu F, Rad AB, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD*, Reyna MA*. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiol Meas. 2020 Nov 11. doi: 10.1088/1361-6579/abc960.

[2] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021; 48:1–4.

[3] Pulavskyi AA, Krivenko SS, Linskiy IV, Posokhov MF, Krivenko SA and Kryvenko LS, Determination of the Risk of Developing Diabetes Mellitus Based on the Patterns of Symbolic Dynamics of the Amplitude-Time Heart Rate Variability, 2021 10th Mediterranean Conference on Embedded Computing (MECO). 2021, 1-6, doi: 10.1109/MECO52532.2021.9460271.

[4] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, et al. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. Journal of Medical Imaging and Health Informatics 2018;8(7):1368—1373.

[5] Tihonenko V, Khaustov A, Ivanov S, Rivin A, Yakushenko E. St Petersburg INCART 12-lead Arrhythmia Database. PhysioBank PhysioToolkit and PhysioNet 2008; Doi:10.13026/C2V88N.

[6] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der EKG Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik 1995;40(S1):317–318.

[7] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, LunzeFI, Samek W, et al. PTB-XL, a Large Publicly Available Electrocardiography Dataset. Scientific Data 2020;7(1):1–15.

[8] Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead Electrocardiogram Database for Arrhythmia Research Covering More Than 10,000 Patients. Scientific Data 2020;7(48):1–8.

[9] Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA. NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing. Behav Res Methods. 2021 Aug;53(4):1689-1696. doi: 10.3758/s13428-020-01516-y.

[10] Cysarz D, Porta A, Montano N. et al. Quantifying Heart Rate Dynamics Using Different Approaches of Symbolic Dynamics. Eur. Phys. J. Spec. Top. 222. 2013; 487–500. https://doi.org/10.1140/epjst/e2013-01854-7

[11] McFee B, Raffel C, Liang D, PW Ellis D, McVicar M, Battenberg E, and Nieto O. Librosa: Audio and Music Signal Analysis in Python, in Proceedings of the 14th Python in Science Conference. 2015, 1-7.

Address for correspondence:
Sergey S Krivenko (krivenkos@ieee.org)
61038, Ukraine, Kharkiv, 50 Khalturina str, 2 app.