Automatic recognition of congestive heart failure signs in heart rate variability data

A.A. Pulavskyi, S.S. Krivenko
Scientific Company KOLIBRI LLC
Kharkiv, Ukraine
info@kolibri.one

S.A. Krivenko
Kharkiv National University of radioelectronics
Kharkiv, Ukraine
l1stanislaw2015@gmail.com

I.V. Linskiy, M.F. Posokhov
Institute of Neurology, Psychiatry and Narcology of the National Academy of Medical Sciences of Ukraine
Kharkiv, Ukraine
inpn@ukr.net

L.S. Kryvenko
Kharkiv National Medical University
Kharkiv, Ukraine
ls.kryvenko@knmu.edu.ua

2022 11th Mediterranean Conference on Embedded Computing (MECO) | 978-1-6654-6828-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/MECO55406.2022.9797210

Abstract—**Automatic screening of the population for congestive heart failure (CHF) is a matter of pressing concern due to the severity of the health consequences resulting in disability and death of people. On the one hand, portable devices working with ECG signals become convenient tools for the lay user due to the simplicity. On the other hand, analyzing the specific behavior of the R-peaks sequence (analysis of heart rate variability) in cardiac pathologies allows identifying the patterns inherent in particular heart dysfunction. Such patterns are effectively differentiated using symbolic dynamics methods and the subsequent application of machine learning methods. In this study, a highly specific model was obtained (sensitivity 0.71, specificity 0.96), suitable for automatic screening of CHF. Its operability and performance characteristics have been verified through testing in several publicly available databases.**

Keywords-ECG; HRV; congestive heart failure; noninvasive screening; symbolic dynamics.

I. INTRODUCTION

Cardiovascular diseases are still the leading cause of death and disability according to WHO [1]. Early diagnosis of health threats during population screening can help minimize severe health consequences and save the budgets of national healthcare systems. Currently used a regular outpatient questionnaire of people from risk groups cannot yet claim a quick mass solution [2–4]. This is the niche of telemedicine, particularly personal tools for health conditions assessment [5, 6]. Usually, such equipment does not require professional skills from the user to conduct a test while the interpretation of data is provided by medical personnel, often remotely [7]. At the same time, the use of telemedicine devices by laypersons poses new challenges for their developers [8–10]. So, for example, a poor-quality ECG signal becomes a significant obstacle to calculating acceptable final results. Therefore, automatic pretreatment and preprocessing of physiological signals (ECG, PPG and EEG, etc.) recorded by a non-professional user is a primary task of remote monitoring in mass screening of the population. When working with an ECG signal, the analysis of the R-peak sequence comes to the fore (IBI - interbeat interval sequence). First of all, because the R-peak is the most specific and clearly-defined ECG point, that ensures maximum efficiency and robustness of signal processing, especially when distorted by unremovable noises of various nature (electrode motion, muscle artefacts, etc.). Accordingly, the division of patterns into physiologically normal and pathological becomes more accurate since it is based on heart rate variability (HRV) originated from IBI analysis. It is almost impossible to achieve the same separation efficacy when working with raw ECG [11]. The purpose of this study was to create a highly specific diagnostic model capable of automatically identifying patients with congestive heart failure (without the involvement of a professional). Such a model included all stages of preprocessing the initial ECG signal and algorithms for correcting ectopic beats. At the same time, the input data have been taken from the existing databases, without preliminary selection of "suitable" (in the sense of quality) ECGs.

II. HANDCRAFTED FEATURES

The right choice of discriminating features determines the quality of created classifiers. These features have been created for our system based on IBI. The processing of time series using the methods of information theory, in particular, symbolic dynamics, has proven to be promising [12]. A time series was represented as a sequence of symbols from a finite alphabet. This sequence could then be statistically processed. The most common technique is to combine consecutive symbols into words. In this case, for combining into words, both consecutive symbols and those taken with a certain lag (skipping one or two symbols, etc.) can be used. For analysis, words with a length of 2-4 symbols are important. However, in some situations, longer combinations are also used. Our study used the occurrence (occurrence frequency) of various words in the IBI, transformed by a specific method. The occurrence of particular words was a biomarker of certain pathological

978-1-6654-6828-2/22/$31.00 ©2022 IEEE

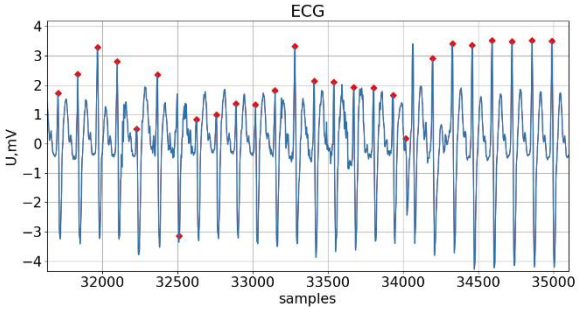
conditions, making it possible to differentiate various patterns of HRV. Normalization of the occurrence was used to neutralize the factor of different IBI lengths. We have used two IBI transformation methods: ordinal patterns and binary delta coding [13]. For further analysis, the following structure has been used to decode the designations to the normalized occurrence of symbols: sym_ (method) _ (word_length) _ (time_lag) __ (ordinal number of a character in the alphabet). For example, sym_dcbin_5_1__29 means the normalized occurrence of the 29th symbol obtained by the binary delta coding method with a word length of 5 (such an alphabet contained 2^5 = 32 symbols) and time lag equal to 1 (i.e., consecutive symbols were used to compose words).

The fundamental element in the architecture of our system was a sequence of RR intervals (IBI) obtained from an ECG segment with a duration of 5 minutes. In the case when the duration of the ECG segment was less than 5 minutes, the entire available ECG has been used (an extension of the signal up to 5 minutes was not applied). When the ECG time exceeded 5 minutes, the entire ECG signal has been divided into non-overlapping segments of 5 minutes each. If IBI and information about the sampling rate of the initial ECG were available in the database (DB), then the IBI corresponding to 5-minute ECG has been considered a key element. If only IBI were available in the database, then an IBI with 300 items has been taken as a basis.

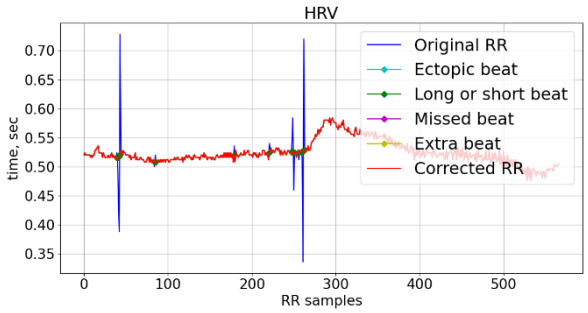
The Mann–Whitney U-test has been used to assess the statistical significance of the distribution characteristics of the obtained features [14]. The study results were statistically significant at the $p < 0.05$, and the median and interquartile range (IQR) were used as distribution parameters.

III. DATABASES AND PREPROCESSING

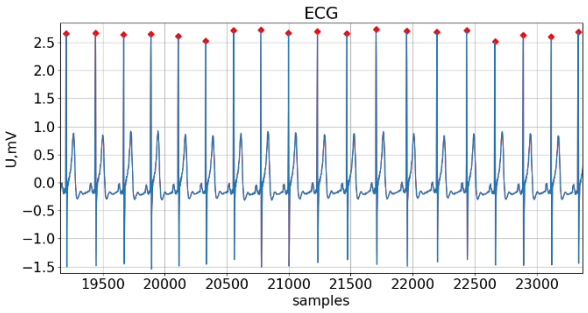
The preprocessing of the ECG segment included using notch filters at 50 and 60 Hz and a bandpass filter of 0.5–45 Hz. This approach helped remove interference caused by the powerline net, baseline wander and high-frequency noise. A standard QRS detector [15] has been used to obtain R-peaks (raw IBI). IBI preprocessing used, among others, a modern adaptive algorithm (Fig. 1), which corrected ectopic beats, long or short beats, missed beats, extra beats [16].



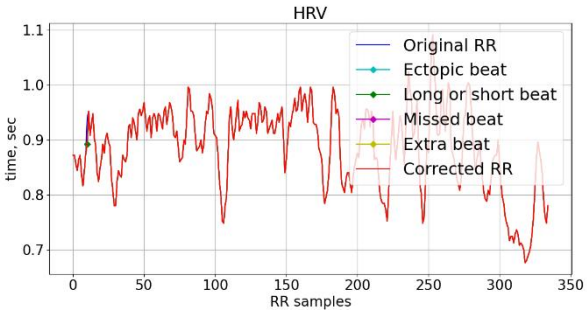
a



b



c



d

Figure 1. Automatically preprocessed ECG and IBI segments: a, b - a patient with CHF, c, d - a healthy subject (important: the red curve covers the blue one on graphs b and d, i.e, Corrected RR and Original RR are equal, that's why the curves overlapped).

It should be noted that all stages of pretreatment, including splitting into segments, has been carried out automatically and have not been subjected to subsequent manual adjustments. Therefore, training and test datasets included some quantity of "bad" ECG segments, i.e., data usually excluded from training or validation but encountered in practice. Thus, the process of datasets formation can be compared to real data collection. The building of reliable models directly depends on the quality and versatility of the data. When constructing models for use in medical diagnostics, it is also vital to ensure future performance in various conditions (geographic, ethnic, gender). In other words, the similarity (ideally equality) of the distributions of the training and validation data shall be ensured. The easiest way to achieve this was to use several databases obtained in different conditions and geographical

areas by independent teams of researchers. To meet the given requirements, we used 5 different databases taken from the resources of PhysioNet and Kaggle for training and validation[17]. These databases contained either ECG signals or RR intervals. Considering that the construction of models that extract HRV features was in main focus (i.e., only the information contained in the RR intervals was used), such a choice of databases is justified.

A. Kaggle ECG Database

This DB has been taken from the publicly available source (https://www.kaggle.com/c/ecg-diagnosis). It contains information on 2515 patients, presented in the form of 600 heart beats (R-peak and its amplitude) for each of them. Along with the data of healthy volunteers, patients suffering from diabetes mellitus, peptic ulcer disease, somatoform autonomic dysfunction, ischemic heart disease, and nodular goitre were included in the data the database. This DB has been used for the training and test datasets.

B. BIDMC Congestive Heart Failure Database

This DB has been taken from the publicly available source PhysioNet (https://physionet.org/content/chfdb/1.0.0/). It includes long-term ECG of severe congestive heart failure, NYHA class 3–4: 15 records, incl.11 males, aged 22 to 71, and 4 females, aged 54 to 63. This DB has been used for the training and test datasets.

C. Congestive Heart Failure RR Interval Database (CHFRR)

This DB has been taken from the publicly available source PhysioNet (https://physionet.org/content/chf2db/1.0.0/). It includes annotations of R-peaks for long-term ECGs of 29 patients of 34-79 years old (8 males and 2 females, gender of others unknown) suffering from congestive heart failure of NYHA class 1–3. This DB has been used for the dataset aside for validation.

D. MIT-BIH Normal Sinus Rhythm Database (MBN)

This DB has been taken from the publicly available source PhysioNet (https://physionet.org/content/nsrdb/1.0.0/). It includes 18 long-term ECGs belonging to patients of 20-50 years old (5 males and 13 females) without visible/evident arrhythmias. This DB has been used for the dataset aside for validation.

E. Combined measurement of ECG, Breathing and Seismocardiograms (CEBS)

This DB has been taken from the publicly available source PhysioNet (https://physionet.org/content/cebsdb/1.0.0/). It includes 20 long-term ECGs belonging to presumably healthy patients. This DB has been used for the dataset aside for validation.

F. Fantasia Database

This DB has been taken from the publicly available source PhysioNet (https://physionet.org/content/fantasia/1.0.0/). It includes 40 long-term ECGs belonging to presumably healthy subjects. Among them, 20 patients were young - 21-34 years old (10 males and 10 females), and 20 patients were elderly - 68-85 years old (10 males and 10 females). This DB has been used for the dataset aside for validation.

G. PTB Diagnostic ECG Database

This DB has been taken from the publicly available source PhysioNet (https://www.physionet.org/content/ptbdb/1.0.0/). It includes 549 short-term ECGs of 290 presumably healthy subjects and patients suffering from myocardial infarction, cardiomyopathy, bundle branch block, dysrhythmia, myocardial hypertrophy, valvular heart disease etc. The age of subjects was in the range of 17-87 years. This DB has been used for the dataset aside for validation.

IV. MODEL AND DISCUSSION OF RESULTS

LightGBM has been used as the base method of model construction. It has proven itself well for the treatment of extensive tabular data. In addition, LightGBM allowed determining the degree of significance (influence) of various features in the model. The main dataset has been generated from the Kaggle and BIDMC datasets. The ECG segments duration was 5 minutes. Subsequently, the ECG records of all patients in the Kaggle and BIDMC sets were divided into non-overlapping 5-minute segments. Each segment obtained from ECGs of patients suffering from CHF has been labelled 1 (class 1, positive). The label 0 (class 0, negative) has been assigned to segments that did not satisfy the above condition. The final dataset included 1353 ECG segments of the control group and 1187 ECG segments of the target group. Further, the primary dataset has been randomly divided into training (75%) and test (25%) subsets. When dividing, stratification by target label has been used. It is important to note that the ECG segments belonging to one patient were included in only one set: training or test. It has been done on purpose to avoid the ECG segments of the same patient getting into both sets, i.e., randomization has been performed by patients rather than by the ECG segments. In such a situation, forming a target group raised no questions. However, the control group formation required additional adjustment. Due to the presence of several pathologies in the databases and the quantitative predominance of one pathology over others, there was a risk of overfitting the model. That's why all pathologies available in the databases were evenly represented in the control group. This allowed the model to highlight the features of the target group rather than the differences between the target and control groups (e.g., between healthy subjects and patients with CHF). The performance assessment has been done by sensitivity (Sen) and specificity (Spec) of the model. For that purpose, the resulting model has been tested on a test set and several databases, namely MBN, CHFRR, CEBS, Fantasia, PTB (Table 1). ECGs from the databases were divided into 5-minute segments (except ECGs from the PTB, where the maximum duration of the ECG segments did not exceed two minutes). As follows from the Table 1, the model has demonstrated high specificity for almost all databases, more than 0.95. A feature of the PTB database, whose specificity was 0.78, was the short duration of ECG segments (less than two minutes). Therefore, it affected the quality and completeness of the calculated HRV features. However, many pathologies presented in this database became an argument for its use as a test one. The sensitivity level in all

sets was higher than 0.7, which is acceptable for screening [18].

TABLE I. ASSESSMENT OF LGB-MODEL PERFORMANCE

	Databases						
	Test set		***MBN+CHFRR***		***CEBS***	***Fant asia***	***PTB***
Metric	Sen	Spec	Sen	Spec	Spec	Spec	Spec
Model	0.71	0.99	0.86	0.96	1.00	0.97	0.78

Fig. 2 demonstrates the degree of influence of various features on the model calculated using Shapley values [19]. It is evident that features obtained based on coding the difference of neighboring IBI intervals prevail - the three most influential features have been obtained using the *dcbin* method. This approach allowed using the above features as clear explicable markers in identifying a given disease. The distributions of significant HRV features for the target and control groups are shown in Fig. 3. In some cases, there is a visual difference in the distributions of the target and control groups. Table 2 contains statistically significant parameters of the distributions, which confirms the excellent separating ability of the found nonlinear HRV features.

TABLE II. CHARACTERISTICS OF DISTRIBUTIONS OF THE MOST SIGNIFICANT FEATURES

Feature	**Class 0 (no CHF)**		**Class 1 (CHF)**	
	Median	***IQR***	***Median***	***IQR***
sym_dcbin_5_1__29	0.01	0.02	0.06	0.03
sym_dcbin_6_1__59	0.003	0.01	0.04	0.03
sym_dcbin_7_1__118	0.0	0.0	0.02	0.01
sym_dcbin_6_1__29	0.003	0.01	0.04	0.02
sym_op_4_2__0	0.05	0.06	0.1	0.1
sym_dcbin_5_1__23	0.02	0.02	0.06	0.03
sym_op_3_1__5	0.2	0.07	0.1	0.09

At the current stage, the weakness of the final model lies in the low sensitivity value and inability to use the classical HRV features from the time and frequency domains as input parameters, which will become a further step in the development and improvement of the model.

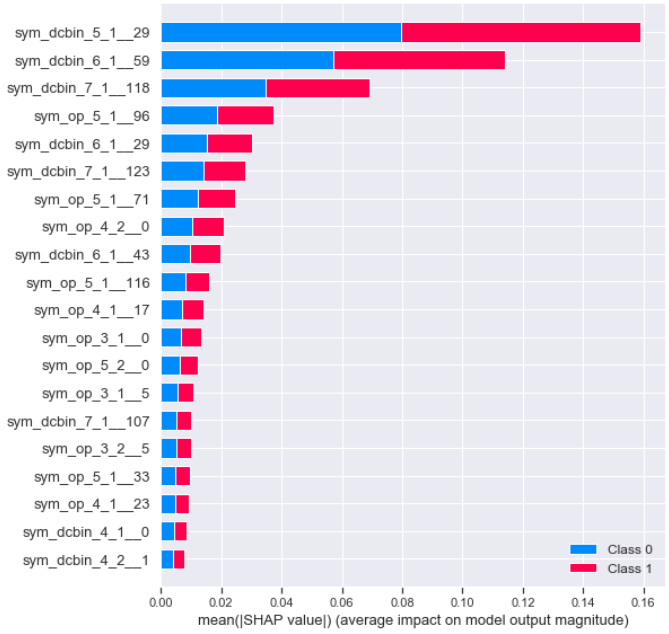
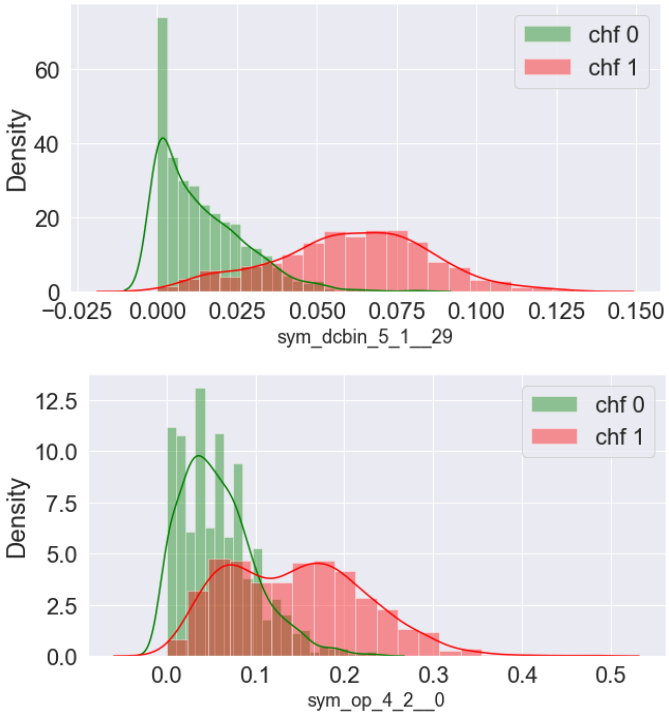


Figure 2. Influence of HRV features on the LGB-model.



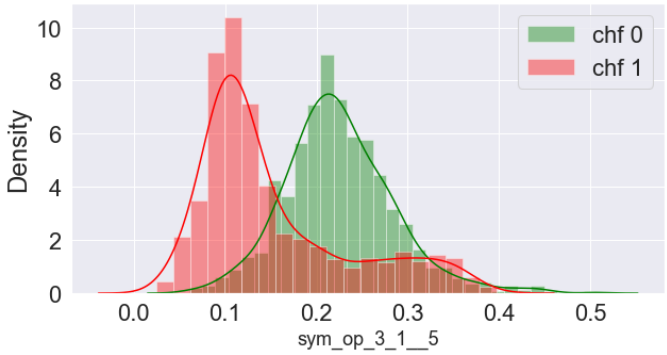


Figure 3. Distributions of some significant HRV features.

V. CONCLUSIONS

A highly specific LightGBM model of automatic screening for congestive heart failure was obtained during the study. The model's sensitivity was not less than 0.71, and the specificity was not less than 0.95 on test databases. Automatic screening implies the determination of signs of the disease without manual error correction of RR intervals, including the correction of ectopic beats, long or short beats, missed beats, extra beats. Thus, the proposed approach can be used for personal telemedicine devices in the early detection of health threats.

ACKNOWLEDGMENT

This study is the part of scientific research work "Optimization of early diagnosis, prevention and treatment of oral tissue diseases with smoking addiction", KPKVK 2301020, 0120U102057, Kharkiv National Medical University, and is partly funded by Ministry of Health of Ukraine. The authors are grateful to Mrs. Oksana Kolyesnik and Dr. Olha Krylova for their valuable contribution to the experiment conduction. Their professional assistance, expert advice, and support were very helpful in the process of preparing article.

REFERENCES

[1] 'Quality health services: a planning guide'. https://www.who.int/publications-detail-redirect/9789240011632 (accessed Jan. 12, 2021).

[2] R. Nazaryan, L. Kryvenko, Y. Zakut, O. Karnaukh, and V. Gargin, 'Application of estimated oral health indices in adolescents with tobacco addiction', *Pol Merkur Lekarski*, vol. 48, no. 287, pp. 327–330, Oct. 2020.

[3] V. Gargin, R. Radutny, G. Titova, D. Bibik, A. Kirichenko, and O. Bazhenov, 'Application of the computer vision system for evaluation of pathomorphological images', in *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, Ukraine, Apr. 2020, pp. 469–473. doi: 10.1109/ELNANO50318.2020.9088898.

[4] R. S. Nazaryan, L. S. Kryvenko, and V. V. Gargin, 'The role of nitric oxide synthase in the modulation of the immune response in atopic disease', *New Armenian Medical Journal*, vol. 11, no. 2, pp. 52–57, 2017.

[5] V. Gargin *et al.*, 'Relationship between bone density of paranasal sinuses and adrenal steroids pattern in women during menopausal transition', *Anthropological Review*, vol. 83, no. 4, pp. 407–418, Dec. 2020, doi: 10.2478/anre-2020-0031.

[6] S. S. Krivenko, A. A. Pulavskyi, and S. A. Krivenko, 'Determination of low hemoglobin level in human using the analysis of symbolic dynamics of the heart rate variability', in *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Kiev, May 2017, pp. 271–274. doi: 10.1109/UKRCON.2017.8100490.

[7] 'KOLIBRI'. https://kolibri.healthentire.com/ (accessed Jan. 12, 2021).

[8] V. Bezruk, S. Krivenko, and L. Kryvenko, 'Salivary lipid peroxidation and periodontal status detection in ukrainian atopic children with convolutional neural networks', in *2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*, Kharkov, Oct. 2017, pp. 122–124. doi: 10.1109/INFOCOMMST.2017.8246364.

[9] V. Bezruk, S. Krivenko, and L. Kryvenko, 'The Pareto chart of caries intensity evaluation for children with allergic diseases', in *2015 Second International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T)*, Kharkiv, Ukraine, Oct. 2015, pp. 110–111. doi: 10.1109/INFOCOMMST.2015.7357285.

[10] S. Krivenko and A. Pulavskyi, 'Accuracy improvement of noninvasive determination of glucose concentration in human blood', in *2013 12th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, Feb. 2013, pp. 181–183.

[11] M. Porumb, E. Iadanza, S. Massaro, and L. Pecchia, 'A convolutional neural network approach to detect congestive heart failure', *Biomedical Signal Processing and Control*, vol. 55, p. 101597, Jan. 2020, doi: 10.1016/j.bspc.2019.101597.

[12] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, and N. Wessel, 'Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics', *Computers in Biology and Medicine*, vol. 42, no. 3, pp. 319–327, Mar. 2012, doi: 10.1016/j.compbiomed.2011.03.017.

[13] D. Cysarz, A. Porta, N. Montano, P. V. Leeuwen, J. Kurths, and N. Wessel, 'Quantifying heart rate dynamics using different approaches of symbolic dynamics', *Eur. Phys. J. Spec. Top.*, vol. 222, no. 2, pp. 487–500, Jun. 2013, doi: 10.1140/epjst/e2013-01854-7.

[14] S. Krivenko, V. Lukin, O. Krylova, L. Kryvenko, and K. Egiazarian, 'A Fast Method of Visually Lossless Compression of Dental Images', *Applied Sciences*, vol. 11, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/app11010135.

[15] J. Pan and W. J. Tompkins, 'A Real-Time QRS Detection Algorithm', *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: 10.1109/TBME.1985.325532.

[16] J. A. Lipponen and M. P. Tarvainen, 'A robust algorithm for heart rate variability time series artefact correction using novel beat classification', *Journal of Medical Engineering & Technology*, vol. 43, no. 3, pp. 173–181, Apr. 2019, doi: 10.1080/03091902.2019.1640306.

[17] A. L. Goldberger *et al.*, 'PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals', *Circulation*, vol. 101, no. 23, Jun. 2000, doi: 10.1161/01.CIR.101.23.e215.

[18] M. A. Bujang, 'Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis', *JCDR*, 2016, doi: 10.7860/JCDR/2016/18129.8744.

[19] E. Kalai and D. Samet, 'On weighted Shapley values', *Int J Game Theory*, vol. 16, no. 3, pp. 205–222, Sep. 1987, doi: 10.1007/BF01756292.