Distribution Analysis of Long-Term Heart Rate Variability Versus Blood Glucose

I. Vishinov*, M. Gusev**, L. Poposka***, M. Vavlukis***, I. Ahmeti****
* Innovation DOOEL, Skopje, North Macedonia
** Sts. Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, Skopje, North Macedonia
*** Sts. Cyril and Methodius University in Skopje, Clinic of Cardiology, Skopje, North Macedonia
**** Sts. Cyril and Methodius University in Skopje, Clinic of Endocrinology, Skopje, North Macedonia
E-mail: ilija.vishinov@innovation.com.mk, marjan.gushev@finki.ukim.mk,
lidijapoposka@gmail.com, marija.vavlukis@gmail.com, iahmeti@yahoo.com

***Abstract*—This research explores the class-distributions of long-term heart rate variability (HRV) parameters compared to the distribution of glycated haemoglobin (HbA1c) which depicts the long-term blood glucose regulation ability. The goal is to find the optimal HRV parameter that correlates to class distribution based on HbA1c and to check which measurement duration achieves better performance The class-distribution separability will provide an answer if future highly accurate, precise, and sensitive machine learning classification can be constructed and if so, to aid their interaction with the input data. We found that removing a dataset sample in which at least one feature value is considered an outlier made leads to a much better results. The strongest point-biserial correlations for the class-distribution separation, were found for 24-hour SDRMSSD-3 (r = -0.43), 20-hour SDRMSSD-3 (r = -0.34) and 24-hourARMSSD-3 (r = -0.33) satisfying the significance p-value threshold (p $\leq$ 0.01). All correlations were negative, showcasing that lower HRV is associated with bad blood glucose regulation. We observed that the longer the measurement period is the better correlation is achieved. The best performance of class distribution based on univariate threshold is achieved for SDRMSSD-3 with ACC = 86.52% and weighted F1 score of 86.71%, making it stand out as the single most valuable HRV parameter when it comes to distinguishing good from bad blood glucose regulation.**

***Index Terms*—heart rate variability, blood glucose, HbA1c, classification**

I. Introduction

This research analysis is part of the Glyco [1] research project, with the end goal of predicting blood glucose levels from an electrocardiogram (ECG). The project currently counts 155 patients with heart problems and arrhythmia. The connection between heart rate variability (HRV) and blood glucose [2], [3] has researchers striving to develop an accurate and non-invasive detector of blood glucose levels [4].

In this research, we present a detailed analysis of the class-distributions of the HRV parameters for the good and bad glucose regulation ability classes and present the motivation for a future study with its objective bieng the correlation with, and prediction from the HRV parameters of the glycated haemoglobin (HbA1c).

The HRV parameters can be in one of two categories. Time-domain (TD) HRV parameters include SDNN, ASDNN, SDANN, NN50, pNN50, RMSSD, ARMSSD, SDRMSSD which can have Prima, Secunda, or Tertia variations. Non-linear HRV parameters include: SD1, SD2, SD1/SD2.

The inter-quartile range method was applied to clean the dataset from outliers. Histograms, one-dimensional and two-dimensional Gaussian kernel density estimators are used to describe the univariate and bivariate distributions in the dataset. Point-biserial correlation is used for quantifying class distribution separability in terms of correlations and their significance. Classification performance measures include accuracy, macro, micro and weighted values of F1 score, and corresponding confusion matrix metrics of sensitivity, specificity, positive and negative predictive rate.

The rest of this paper is structured as follows. Other research papers with similar work are presented in Section II. Section III describes the implemented methodology to remove outliers in the data, find correlations, and evaluate the results. The results from the applied methods are elaborated in Section IV. The results and other specifics about the patient selection process, limitations of the research, etc. are discussed in Section V, along with visual observations through the displayed figures. The conclusions from this study and plans for extending it in future work are presented in Section VI.

II. Related Work

HbA1c has long been considered a diabetes marker[5] in the management of blood glucose in diagnosed patients with diabetes and as a diagnostic marker as well. It is therefore relevant to explore the difference in HRV for diabetic and control groups in the research literature.

Coopmans et al. [6] have concluded that both prediabetes and type 2 diabetes were independently associated with lower HRV. Imthiaz [7] states that even in offsprings of diabetic patients who did not have diabetes, autonomic modulations if any, can be picked up by HRV analysis. Further on, the author reports significantly lower RMSSD and pNN50 values for the study group in comparison to the control group and a non-significant difference between the two groups for HRV expressed in SDNN. A similar conclusion that HRV decreases with the development of diabetes mellitus is presented by Schroeder et al. [8].

With the HRV parameters contained in our study, Kudat et al. [9] report a significantly lower SDNN, SDANN, pNN50, and RMSSD for diabetic patients in comparison to a control group.

III. Methodology

A. Statistical distribution methods

For analyzing the univariate distributions of the HRV parameters, histograms with adaptive binning and a Gaussian kernel density estimation to smooth and approximate the distribution are used.

To analyze the bivariate distributions between the HRV parameters and HbA1c, two-dimensional Gaussian kernel density estimate plots to smooth and approximate the bivariate distribution were utilized. To deal with the bounded data, since HRV ≥ 0 and in some cases very small values for the long-term HRV parameters are almost impossible, the density estimations were truncated at the data limits.

B. Outlier-removal methods

Since the univariate distribution of the HRV parameters is non-gaussian, to handle the extreme values in the data for the HRV parameters, the inter-quartile range (IQR) outlier-removal technique was applied. The three quartile-bounding points are defined such that the median is equivalent to the second quartile-bounding point (Q2). The median of the data below Q2 is the first quartile-bounding point (Q1) and the median of the data above Q2 is the third quartile-bounding point (Q3). The inter-quartile range is defined as Q3-Q1 . As outliers are considered the points that lie outside the interval [Q 1-1.5 IQR, Q3 +1.5 IQR].

Two variations of the IQR outlier-removal method were used. In the first, the outliers in one HRV parameter did not influence the other feature values in the sample, i.e. it was replaced with a null value for that feature alone. In the second variation, if an outlier was found for one HRV parameter, the sample was removed from the dataset with the assumption that an outlier in one feature might indicate that there are irregularities in the measurement and it as a whole might be an outlier. The two variations are named IQR-Feature and IQR-Sample correspondingly.

C. Statistical correlation methods

The correlation methods used in this paper is the *Point-biserial correlation* to quantify the separability of two classes based on a correlation method. The point-biserial correlation is identical with the Pearson correlation, which measures the linear dependency between two continuous variables, and in this case, one of the variables is a dichotomous one instead of a continuous variable.

Correlation matrices are shown only for the point-biserial correlation analysis. The maximum correlation found with this method was 0.43, so the color mapping was made relative to this maximum value. The significant correlations ($p \leq 0.05$) have the biggest cell area, and only those cells are annotated with the corresponding correlation coefficients. A visual guide to the interpretation of the correlation matrices is shown in Figure 1.

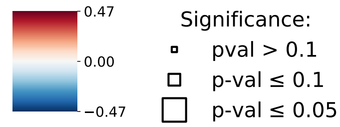

Fig. 1. Correlation matrices scales: color scale (left) and significance scale (right)

D. Confusion matrix metrics for classification performance

The number of true positives, true negatives, false positives, and false negatives are labeled with TP, TN, FP, and FN respectively. The confusion matrix classification performance metrics are calculated as:

- Accuracy (ACC) = (TP + TN) / (TP + TN + FP + FN)
- Precision (PPV) = TP / (TP + FP)
- Recall (SEN) = TP / (TP + FN)
- Specificity (SPC) = TN / (TN + FP)
- Negative predictive value (NPV) = TN / (TN + FN)
- F1 = 2 (PREC $\times$ REC) / (PREC + REC)
- F0 = 2 (SPEC $\times$ NPV) / (SPEC + NPV)
- F-macro (F1M) = (F1 + F0) / 2
- F-weighted (F1w) = F1 $\times$ 1_samples + F0 $\times$ 0_samples

Note that, F0 is the F1 measure when the positive and negative classes are reversed. F-macro is the macro average of these values to represent the overall classification ability. F-micro is equivalent to accuracy for multi-class performance dependent on class distribution, while F1 macro is average of performances to detect each class. Te eliminate the class distribution anomaly we use weighted F score (F1w).

E. Clinical research and dataset

The ECG signal strip, produced by a single lead wearable sensor [4], was processed such that ectopic beats, artifacts, lost signals, noise, and other problematic signal types were removed or handled appropriately with preprocessing, to reduce their negative influence on the HRV parameters.

Nine of the HRV parameters are defined in one of our previous research papers [10] and can be viewed there. The definitions of the two new HRV parameters that were defined recently goes as follows, where RR denotes an interval between two heartbeats, and NN is an interval between two normal (N) beats:

- **ARMSSD** - an average of the root mean square of successive NN interval differences for all analyzed segments within the defined time period; The value is expressed in milliseconds.
- **SDRMSSD** - standard deviation of the root mean square of successive NN interval differences for all analyzed segments within the defined time period; The value is expressed in milliseconds.

Additional HRV parameters are calculated with the following identification:

TABLE I
NUMBER OF SAMPLES AND PATIENTS IN THE ORIGINAL DATASET

Dataset	Samples	B	G	Patients	B	G
DS-16h	7345	1583	5762	123	37	86
DS-20h	5488	1106	4382	121	36	85
DS-24h	4362	813	3549	120	36	84

TABLE II
NUMBER OF SAMPLES AND PATIENTS AFTER IQR REMOVAL

Dataset	Samples	B	G	Patients	B	G
DS-IQR-16h	4366	1094	3272	86	27	59
DS-IQR-20h	3205	741	2464	80	25	55
DS-IQR-24h	2515	613	1902	77	24	53

- **Prima (1)** calculations apply a 5-minute segment.
- **Secunda (2)** calculations apply a 30-seconds segment.
- **Tertia (3)** calculations apply a continuous uninterrupted ECG segment within a predefined duration.

The interval lengths included in the research for which the long-term HRV parameters are calculated are the following:

- **16h** - 16-hour intervals
- **20h** - 20-hour intervals
- **24h** - 24-hour intervals

The classes derived from a simple threshold imposed upon HbA1c are the following:

- **Good blood glucose regulation (G)**: HbA1c(%) ≤ 7
- **Bad blood glucose regulation (B)**: HbA1c(%) > 7

This defines the original dataset (DS) with number of patients and samples shown in Table I.

Histograms with adaptive bin sizes as well as the correlation methods used in the study, except for the Spearman rank correlation, are extremely sensitive to outliers in the data. The bivariate kernel density estimators require more outliers to be significantly influenced, but are still prone to providing unclear results when outliers are present in the data. The outlier removal resulted in two more datasets: DS_IQR-Feature and DS_IQR-Sample.

Applying the IQR-Feature method to the dataset does not change the number of samples and patients. However, applying the IQR-Sample method changes the number of samples and patients shown in Table II.

IV. RESULTS

One of the main points of interest of the study is the classification problem of whether a patient regulates the glucose well or badly. The point-biserial correlation is the most relevant for this purpose. This analysis also helps in distinguishing which outlier removal method improves the class-distribution separability or if it only does harm when aiming for better separability. Thus, the point-biserial correlation coefficients and their associated p-values were calculated.

The point-biserial correlation matrices for DS, DS_IQR-Feature, and DS_IQR-Sample datasets are presented in Figure **??**, Figure **??** and Figure **??** respectively.

The p-values of most of the correlations are less than 0.05 because of the large number of samples for which the correlations are calculated. Overall the correlations are significant and negative. The fact that the whole correlation matrix depicts negative correlations is a promising indication of the connection between HRV and the glucose regulation class of a patient. There is an exception for the SD1/SD2 HRV parameter, where we observe a change in the sign of the correlation when the IQR-Sample method is applied. This is due to the lack of any correlation and the class-distribution for that feature being almost inseparable. Thus, the correlation direction is stochastic.

For DS_IQR-Feature, in some cases, the point-biserial correlations we observe a decline, and in others, there is an improvement in the strength of the correlation. However, for DS_IQR-Sample there is an obvious improvement in almost all correlations compared to the original dataset.

Since the bad blood glucose regulation class (B) is encoded with 1, and the good blood glucose regulation class (G) with 0, negative point-biserial correlation coefficients indicate that bad blood glucose regulation corresponds to lower HRV.

V. DISCUSSION

Our findings of data dependence analysis for negative correlation of HRV and ability to control diabetes align with the results provided by other researchers as analyzed in the related work.

In this research we explored two other important matters to find the optimal HRV parameter for efficient class-distribution and to find which measurement duration is better.

A. Optimal HRV parameter

In order to observe how the classification is improved with each of the outlier removal methods, the best-performing HRV parameter in the threshold classification, SDRMSSD-3 achieves the best performance in the class distribution data analysis, and is further analyzed visually through univariate and bivariate distribution plots with HbA1c.

Additionally, box-and-whiskers plots are presented for each of the classes. Note that the outliers shown by the box-and-whiskers plots are not really outliers, especially for the DS_IQR-Sample dataset, since the real outliers were removed beforehand when the dataset was cleaned. They only represent the "outliers" in the class-distribution. The role of the box-and-whiskers plots is to summarize and simplify the univariate class-distribution.

The class-distributions for the 24h SDRMMSSD-3 of the original dataset with no outliers removed is presented in the top part of Figure 3. In contrast, the class-distributions for the dataset with applied IQR-Sample removal is presented in the bottom part of Figure 3, so the changes of the outlier removal method can be observed visually. It is important to note that the horizontal axis limits change when going from the dataset with outliers present in the data to the dataset with removed outliers so that the class separation can still be observed when much of the data is discarded and the maximum points are

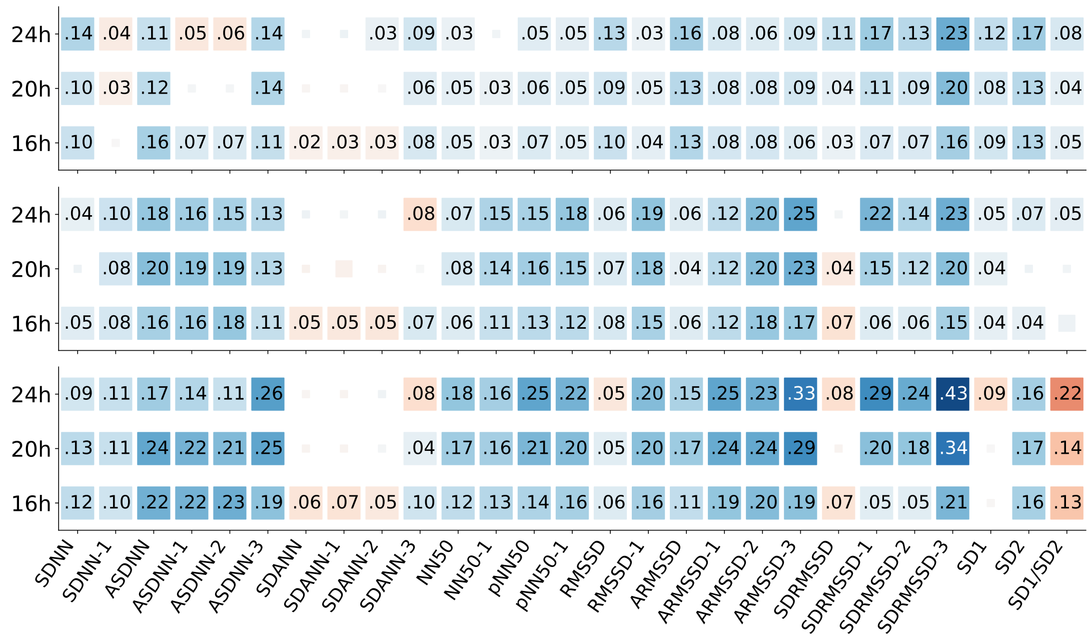

Fig. 2. Point-biserial correlations: the original dataset (top), after applying IQR feature removal (middle) and IQR sample removal (bottom).

shifted towards lower values. The change in horizontal axis limits will be present in all comparison plots that follow.

B. *Dependence on HRV measurement duration*

The dependence on the time duration for calculation of the HRV is analyzed in Fig. 4. As expected, we observe that the longer the measurement duration is, the better class-distribution.

To illustrate the performance dependence on the measurement duration and outlier removal we have provided a simple threshold cut off class distribution method. The results applied for a threshold of 9.4 for the SDRMSSD-3 are presented in Table III.

We observe a high accuracy and F1 scores over 85% for the most optimal HRV parameter affecting the class distribution. Note that an increase of up 10% is observed for accuracy between 16h and 20h measurements and additional 5% for 24h measurements. Since the HbA1c is a diabetes marker expressed usually as at least a two month average of the glucose levels, longer HRV may better express the correlation to HbA1c. This motivates us to continue measuring with even longer periods, such as 48 or 72h and reach even better correlation to HbA1c.

It is important to note that in this research, not all of the patients with ECG recordings were included in the study. Those who did not have HbA1c measurements were excluded, as well as some patients whose instantaneous blood glucose measurements clashed with the HbA1c(%) values and indicated that other external factors may have influenced the instantaneous blood glucose measurements. Also, the number of patient samples to the number of all samples ratios are not ideal and require further data collection and statistical analysis in extension to this research.

VI. Conclusion

The focus of our statistical analysis of the HRV datasets, coupled with HbA1c levels and a glucose regulation ability class based on the HbA1c levels, was the exploration of the class-distributions for each of the classes and how separable they are so that future highly accurate machine learning algorithms can be constructed.

Since the distribution of the HRV parameters is skewed positively and is rarely gaussian-like, two variations of the IQR outlier removal method were explored. IQR-Feature only removed the outliers for each feature separately, and IQR-Sample removed the complete sample when an outlier was found in it, assuming that an outlier in one HRV parameter might indicate that the ECG measurement as a whole might be an outlier. The latter showed a significant improvement in the class-distribution separability, which was demonstrated with the point-biserial correlations.

The point-biserial correlation analysis depicted the negative association between HRV and the regulation class. In other words, the bad glucose regulation class (B) was characterized with slightly lower values for the HRV parameters. The strongest correlations found were for the 24-hour SDRMSSD-3 (r = -0.43), 24-hour ARMSSD-3 (r = -0.33), and 20-hour SDRMSSD-3 (r = -0.34).

The univariate threshold classification resulted with relatively high accuracy and F1 scores over 85%. The best improvements in the point-biserial correlations were approximately mapped into the threshold classification performances, with the DS_IQR-Sample-24h SDRMSSD-3 (accuracy ACC =

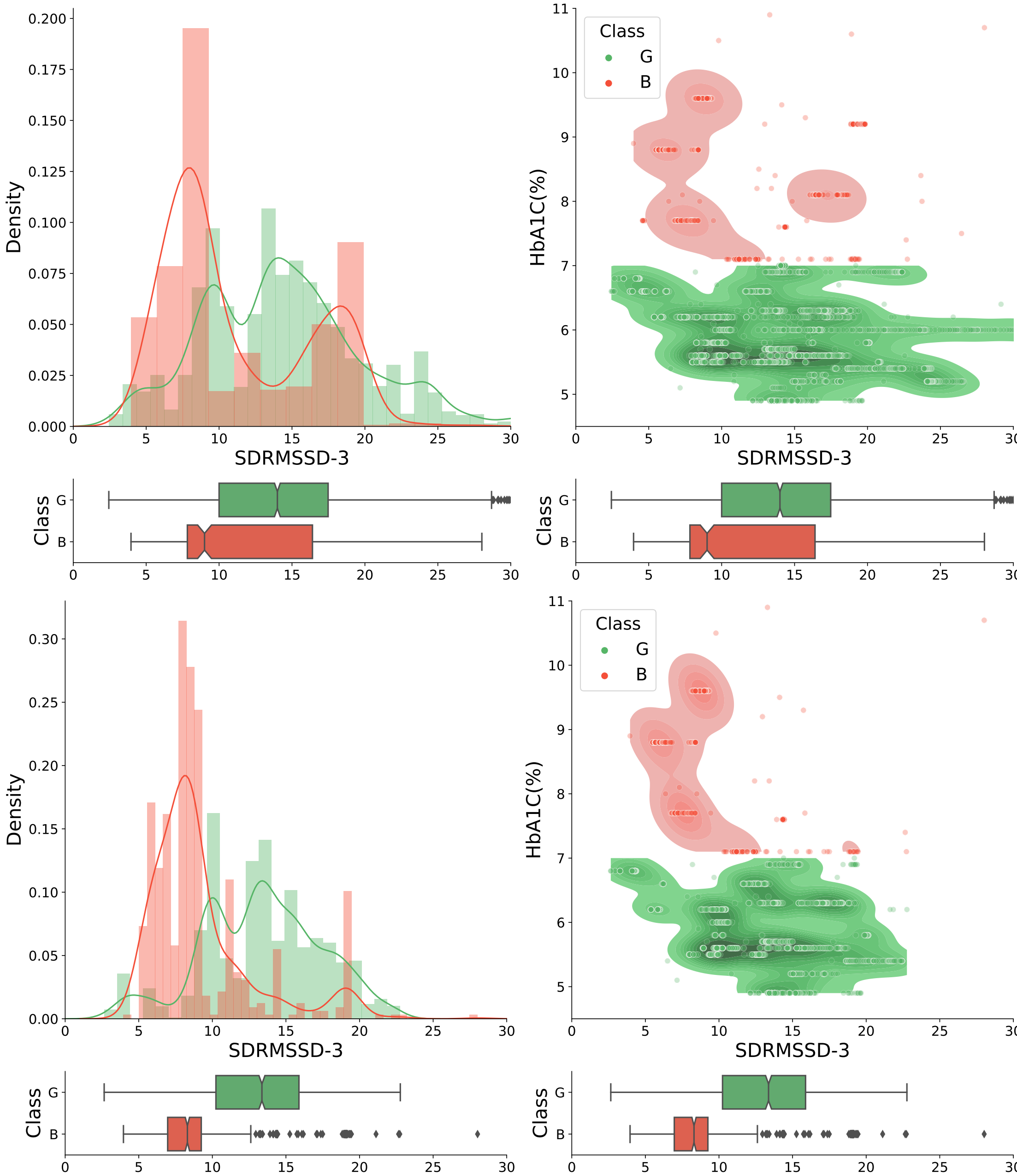

Fig. 3. Distribution plots for 24h SDRMSSD-3 of the original dataset (top) and after IQR sample removal (bottom)

86.52% and weighted F1 score of F1w = 86.71%) outscoring the other HRV parameters.

Hence, SDRMSSD-3 can be considered the single most valuable HRV parameter for distinguishing good from bad blood glucose regulation. Also, we concldued the longer the measurement period is, the better correlation to HbA1c and classification performance.

Promising trends of HbA1c in the data for the bad glucose regulation class were observed, and its correlations with and predictions from the HRV parameters will be the subject of a future study. In addition we plan to analyze which threshold values will provide optimal cut-off threshold-based univarate classification and start with implementation of machine and deep learning methods for developing classification and estimation measurement prototype measurement devices.

In addition, early conclusions show that prior knowledge of type 2 diabetes may produce better classification results and this will be part of our future work on blood glucose regulation classification.

References

[1] Innovation Dooel, "Glyco project - measure ECG and glucose levels with a small, non-invasive, wearable monitor," 2019, project partially funded by Fund of Innovations and Technical Development, North Macedonia. [Online]. Available: http://glyco.innovation.com.mk/

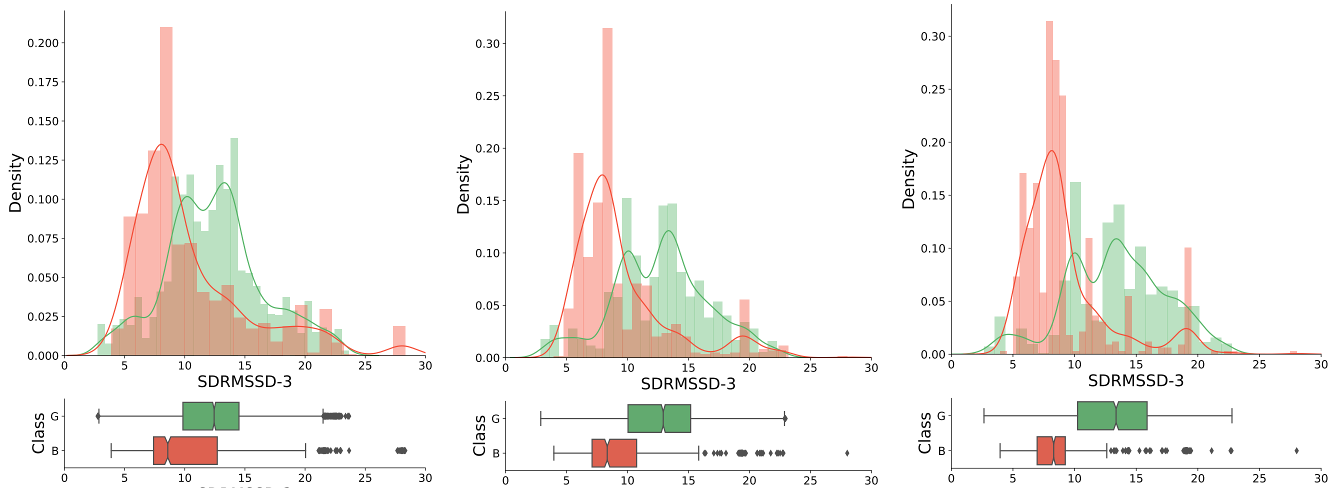

Fig. 4. Distribution plots for SDRMSSD-3 after IQR sample removal on datasets 16h (left), 20h (middle) and 24h (right)

TABLE III
PERFORMANCE DEPENDENCE APPLYING A SIMPLE THRESHOLD CUT OFF CLASSIFICATION FOR SDRMSSD-3 OVER THE ORIGINAL DATASET AND AFTER IQR-SAMPLE OUTILER REMOVAL

Original DS	ACC	F1M	F1w	F1	F0	PPV	SEN	SPC	NPV
24h	78.75%	68.56%	79.79%	50.66%	86.46%	44.65%	58.55%	83.38%	89.78%
20h	75.55%	65.62%	76.65%	47.16%	84.09%	41.77%	54.16%	80.94%	87.49%
16h	73.76%	62.89%	74.32%	42.80%	82.98%	40.37%	45.55%	81.52%	84.49%
DS-IQR-sample	ACC	F1M	F1w	F1	F0	PPV	SEN	SPC	NPV
24h	86.52%	82.23%	86.71%	73.49%	90.97%	70.57%	76.67%	89.70%	92.27%
20h	82.31%	75.56%	82.46%	62.72%	88.40%	61.15%	64.37%	87.70%	89.11%
16h	76.50%	69.44%	76.77%	54.76%	84.13%	52.90%	56.76%	83.10%	85.18%

[2] S. C. Malpas and T. J. Maling, "Heart-rate variability and cardiac autonomic function in diabetes," *Diabetes*, vol. 39, no. 10, pp. 1177–1181, 1990.

[3] M. Gusev, L. Poposka, G. Spasevski, M. Kostoska, B. Koteska, M. Simjanoska, N. Ackovska, A. Stojmenski, J. Tasic, and J. Trontelj, "Noninvasive glucose measurement using machine learning and neural network methods and correlation with heart rate variability," *Journal of Sensors*, vol. 2020, 2020.

[4] M. Gusev, E. Guseva, and L. Poposka, "Design of a non-invasive ECG-based glucose measurement system," in *MIPRO, Data Science and Biomedical Engineering, 43th Int. Conf., 2020, Proceedings*, 2020, pp. 321–326.

[5] R. D. Mahajan and B. Mishra, "Using glycated hemoglobin hba1c for diagnosis of diabetes mellitus: an indian perspective," *Int J Biol Med Res.*, vol. 2, no. 2, pp. 508–512, 2011.

[6] C. Coopmans, T. L. Zhou, R. M. Henry, J. Heijman, N. C. Schaper, A. Koster, M. T. Schram, C. J. van der Kallen, A. Wesselius, R. J. den Engelsman *et al.*, "Both prediabetes and type 2 diabetes are associated with lower heart rate variability: The maastricht study," *Diabetes care*, vol. 43, no. 5, pp. 1126–1133, 2020.

[7] S. Imthiaz, "A study of heart rate variability in offsprings of type 2 diabetes mellitus," Ph.D. dissertation, PSG Institute of Medical Sciences and Research, Coimbatore, 2019.

[8] E. B. Schroeder, L. E. Chambless, D. Liao, R. J. Prineas, G. W. Evans, W. D. Rosamond, and G. Heiss, "Diabetes, glucose, insulin, and heart rate variability: the atherosclerosis risk in communities (aric) study," *Diabetes care*, vol. 28, no. 3, pp. 668–674, 2005.

[9] H. Kudat, V. Akkaya, A. Sozen, S. Salman, S. Demirel, M. Ozcan, D. Atilgan, M. Yilmaz, and O. Guven, "Heart rate variability in diabetes patients," *Journal of international medical research*, vol. 34, no. 3, pp. 291–296, 2006.

[10] I. Vishinov, M. Gusev, M. Vavlukis, and L. Poposka, "Correlating short-term heart rate variability and instantaneous blood glucose measurements," in *2020 28th Telecommunications Forum (TELFOR)*. IEEE, 2020, pp. 1–4.