Optimizing Heartbeat Classification using Bézier Interpolation

M. Gusev*, N. Petrovski*, L. Tonkovikj*
* Sts Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, North Macedonia
E-mail: marjan.gushev@finki.ukim.mk, {nikola.petrovski1, lucija.tonkovikj}@students.finki.ukim.mk

***Abstract*—Analog to digital conversion of electrocardiograms depends on the sampling frequency influencing the determination of a proper heartbeat location and precision of further digital processing. We set a research question to find the optimal number of interpolation points to reduce the mistakes in the similarity check of heartbeats and classify ventricular beats. In addition, another research question aims at finding the optimal number of interpolation points applying Bézier interpolation to reveal the optimal performance/cost ratio. Final research question is to find the sampling frequency that will reveal optimal performance in classification of ventricular beats. The experiments evaluate all neighbouring pairs of heartbeats from the standard benchmark MIT-BIH arrhythmia dataset resampled to a 125 Hz sampling frequency. The results show that even one more interpolation point, which corresponds to a sampling frequency of 250 Hz, will increase the performance versus the original 360 Hz sampling frequency. At the same time, the optimal is interpolation with additional five or seven points corresponding to 750 Hz, and 1000 Hz respectively. We found that a threshold value of 34 reveals the optimal performance to conclude a change between ventricular heartbeats and others, even in a 10-bit precision of the analog-digital conversion. The processing time and performance/cost-benefit analysis show that one interpolation point is the most beneficial.**

***Keywords*—Interpolation, ECG, upsampling**

I. Introduction

Detection of ventricular beats and arrhythmia disorders motivates researchers to develop new effective and efficient algorithms. The best available tool in their effort is to analyze the measured electrical heart activity represented by an electrocardiogram (ECG). Most of the approaches for automated detection are based on pattern matching and signal processing algorithms, or developed machine learning (ML) models. However, all these approaches depend on the quality to represent the ECG in the digital form, especially considering the requirement for efficient transfer and storage of data. On one side, efficiency requires a smaller amount of data to be transferred and processed, and on the other side, detection precision requires more data. The analog-to-digital conversion (ADC) unit is responsible to create samples with a specific size determined by the bit resolution. Samples are created on a predefined time interval according to the sampling frequency.

Each heartbeat can be determined by the R-peak in the ECG (the extreme in the QRS complex representing a heartbeat in the ECG), and the sampling frequency affects its real-time stamp (location). Most research analyzed the impact of the sampling frequency on the detection of the R peaks, compression techniques for ECG signals, or calculation of Heart Rate Variability (HRV) focusing on reconstructing the missing data within ECG sampling. In our recent paper [1], we address the problem of classifying a heartbeat type by performing a similarity check on two heartbeats. After correcting the proper time location with Bézier interpolation [2], we perform the similarity check of neighbouring heartbeats and evaluate the performance versus different threshold values to distinguish between a class of ventricular (V) and class (NonV) of other heartbeat types. A ventricular beat is usually determined by a wider width of the QRS complex, meaning that the heart has not followed the normal heartbeat activation procedure. In this paper, we continue this research aiming at finding the optimal sampling frequency and perform efficient transfer from the end-user devices (sensor and smartphone or other IoT device) allowing effective algorithms. We address two research questions: to 1) find the threshold value that reveals optimal performance, and 2) find the number of points for applying Bézier interpolation which reveal the best performance/cost ratio. We conduct experiments with two methods for Bézier interpolation: Method I as a standard construction of a four control points Cubic curve [3]; and Method II as a modified construction using a centripetal Catmull-Rom curve [4], [5].

The benchmark ECG dataset is the MIT-BIH ECG benchmark with 30-minute ECG measurements of 44 patients (excluding four measurements for patients with paced rhythm out of 48 available in the dataset). The original measurements are sampled at a 360 Hz sampling frequency (assumed as default), and in our experiments, the measurements were downsampled to a 125 Hz sampling frequency, which generates 2.88 fewer data to be transferred. The test cases included the construction of Bézier curves by adding 1, 3, 5, 7, and 15 extra samples between a pair of two analyzed samples, which corresponds to a sampling frequency of 250 Hz, 500 Hz, 750 Hz, 1000 Hz, and 2000 Hz. Finally, each test case is evaluated according to achieved sensitivity, positive predictive value, and F1 score to detect a V beat, avoiding accuracy due to the imbalance of the V class.

II. Related work

This research can potentially reveal the lowest sampling frequency and bit resolution the ADC unit should use for a sufficient performance of detecting the features in the ECG signal. The conclusions in the analyzed literature

for QRS detection show that sampling frequencies below 100 Hz will result in decreased performance [7], with a large number of wrong QRS detections for sampling frequencies below 200 Hz [8], while [9] concludes negligible performance discrepancies for sampling frequencies above 125 Hz having 120 Hz as a cut-off for sufficient performance. Recommendations suggest sampling frequencies of 128 Hz [7], or between 250 - 500 Hz [8].

Analyzing the bit resolution, researchers have concluded that 8-bit resolution is unacceptable, satisfactory results are obtained over 10 bits, and 12 bits are recommended ADC value [10].

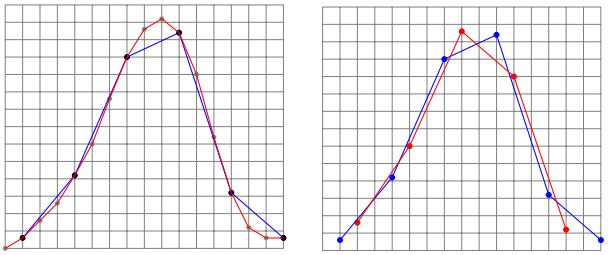
The impact of the sampling frequency on heart rate variability has been reported by several research papers [11]–[15]. Our earlier paper [1] contains a review of related work on the use of Bézier interpolation to improve ECG interpretation, the recovery of ECG missing parts with techniques tested on audio signals [16], or denoising techniques [17], developing a model of an ECG heartbeat [18], or analyzing ECG morphological features [19], [20], concluding that this method was not used to check the similarity of two heartbeats and reveal a tool to detect V versus NonV heartbeats.

III. Methods

Since the ECG is sampled at a particular sampling frequency, we do not know if the real R peak will be represented by the samples. Therefore, we aim at finding the real location of the R peak, and we need to search in the region between two points found to be extreme values. The main idea to apply this procedure is to transfer as little data as possible without losing precision performance and generate a digital presentation for a smooth curve and find a much better location of the time stamp.

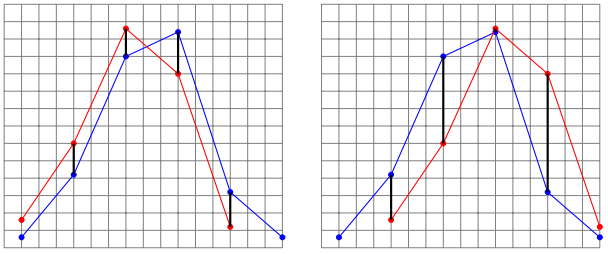
Experiments are performed on MIT BIH Arrhythmia ECG database (MITDB) [21] resampled to a sampling frequency of 125Hz and 10-bit resolution using the Mode-Median-Bucket downsampling method [6]. Each file of ECG samples is accompanied by an annotation file specifying the sample with a heartbeat. Heartbeats were annotated by two independent experts and a third one was engaged in the case of discrepancy. A heartbeat with wider QRS is classified as a V beat if it is annotated as a V (ventricular), E (escape), or F (fusion of ventricular and normal beat) type. All other beat types are classified as NonV class. The problem of detecting a wider QRS in the ECG is a complex task since the baseline fluctuates and detection of Q and S points may not be accurate, knowing that the ECG is usually measured on a real human where the muscle noise corrupts the ECG signal. This is why we apply the similarity check to find how much two heartbeats differ.

Fig. 1a) illustrates the difference between digital presentations sampled at different sampling frequencies (120 and 360 Hz), and Fig. 1b) shows differences in the digital presentations that depend on the location of the sample



a) different sampling freq. b) different start offsets

Fig. 1: QRS digital presentation: a) blue line sampled at 120 Hz and red line sampled at 360 Hz; b) both lines sampled at 120 Hz with different sample starts.



a) misaligned b) aligned to the R peak

Fig. 2: Similarity check of the same heartbeat sampled at 120 Hz with different sample starts - alignment according to the improper location of the R peak.

start. In both cases of the lower sampling frequency, the R peak can not be properly determined.

Similarity check between two heartbeats (blue and red lines in Fig. 2) is evaluated by a sum of Euclidean distances between corresponding samples in the digital representations (black lines in Fig. 2). In the case of a lower sampling frequency, the R peak will not be located on a proper location and a similarity check will generate large values even if the same heartbeat is checked with a different sampling start point.

In this paper, Bézier interpolation was applied to generate more samples to keep the ECG line smooth and determine the proper location of the R peak.

A. Method I

The first interpolation method follows the standard Cubic Bézier curve construction using four control points (P_1, P_2, P_3 and P_4, such that P_2 and P_3 do not have to be on the curve itself). Each segment S_i is defined by its starting point p_i and its ending point p_{i+1}. The segment S_i can be constructed using a Cubic Bézier curve, such that $P_1 = p_i$ and $P_4 = p_{i+1}$. The final graph is constructed by merging all of the segments together. However, the control points P_2 and P_3 remain to be found for each segment.

In our method, we use the algorithm [3] that starts with the Cubic Bézier curve $B(t)$ equation (1), where t

represents how far $B(t)$ is from p_i to p_{i+1}.

$$\begin{aligned} B(t) =&(1-t)^3 \cdot P_1 + 3t(1-t)^2 \cdot P_2 \\ &+ 3t^2(1-t) \cdot P_3 + t^3 \cdot P_4 \end{aligned} \tag{1}$$

Each segment S_i is constructed using a Cubic Bézier curve, such that $P_1 = p_i$, $P_2 = a_i$, $P_3 = b_i$ and $P_4 = p_{i+1}$, as represented in (2).

$$\begin{aligned} S_i(t) = (1-t)^3 \cdot p_i + 3t(1-t)^2 \cdot a_i \\ +3t^2(1-t) \cdot b_i + t^3 \cdot p_{i+1}, \quad 0 \leq i < 20 \end{aligned} \tag{2}$$

The transition between segment S_i and S_{i+1} is smooth around the point p_{i+1}, meaning that the first and second derivatives of S_i and S_{i+1} have to be equal in point p_{i+1}. (3) is obtained assuming that the segment S_i is in point p_{i+1} for $t = 1$, while the segment S_{i+1} is in the same point for $t = 0$.

$$\begin{aligned} S_i'(t=1) = S_{i+1}'(t=0), \quad i = 0, 1, \ldots, 19 \\ S_i''(t=1) = S_{i+1}''(t=0), \quad i = 0, 1, \ldots, 19 \end{aligned} \tag{3}$$

To find all missing control points (a_i and b_i) for all n segments, we need a system of $2n$ linear equations. However, only $2(n-1)$ can be obtained from (3). Therefore, the following (arbitrary) boundary conditions are imposed:

$$\begin{aligned} S_0''(t=0) = 0 \\ S_{n-1}''(t=1) = 0 \end{aligned} \tag{4}$$

(5) and (6) respectively present the first and second derivative of (2).

$$\begin{aligned} S'(t) =&3[-(1-t)^2 \cdot p_i + (1-3t)(1-t) \cdot a_i \\ &+ t(2-3t) \cdot b_i + t^2 \cdot p_{i+1}] \end{aligned} \tag{5}$$

$$\begin{aligned} S''(t) =&6[(1-t) \cdot p_i + (3t-2) \cdot a_i \\ &+ (1-3t) \cdot b_i + t \cdot p_{i+1}] \end{aligned} \tag{6}$$

(7) and (8) are constructed substituting (5) and (6) in (3) and (4).

$$\begin{aligned} a_{i+1} + b_i = 2 \cdot p_{i+1} \\ a_i + 2a_{i+1} = 2b_i + b_{i+1} \\ p_0 - 2a_0 + b_0 = 0 \\ a_{n-1} - 2b_{n-1} + p_n = 0 \end{aligned} \tag{7}$$

$$\begin{cases} a_{i+1} + b_i = 2 \cdot p_{i+1}, & i = 0, 1, \ldots, n-1 \\ a_i + 2a_{i+1} = 2b_i + b_{i+1}, & i = 0, 1, \ldots, n-1 \\ p_0 - 2a_0 + b_0 = 0 \\ a_{n-1} - 2bn - 1 + p_n = 0 \end{cases} \tag{8}$$

The matrix form of the system is given in (9) for easier computation.

$$\begin{bmatrix} 2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 4 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 4 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 4 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 & 4 & 1 \\ 0 & \cdots & 0 & 0 & 0 & 2 & 7 \end{bmatrix} \times \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{17} \\ a_{18} \\ a_{19} \end{bmatrix} = \begin{bmatrix} P_0 + 2 \cdot P_1 \\ 2(2 \cdot P_1 + P_2) \\ 2(2 \cdot P_2 + P_3) \\ \vdots \\ 2(2 \cdot P_{17} + P_{18}) \\ 2(2 \cdot P_{18} + P_{19}) \\ 8 \cdot P_{19} + P_{20} \end{bmatrix} \tag{9}$$

All a_i control points are obtained by solving (9), and all missing b_i points can easily be calculated by (10).

$$\begin{cases} b_i + 2 \cdot p_{i+1} - a_{i+1}, & i = 0, 1, \ldots, 18 \\ b_{19} = \frac{a_{19} + p_{20}}{2} \end{cases} \tag{10}$$

B. Method II

The second interpolation method is a modified Cubic Bézier curve construction, which converts a centripetal Catmull-Rom curve [5] to a Cubic Bézier form. It uses 4 consecutive points on the ECG graph (P_0, P_1, P_2, P_3) to construct the curve between the points P_1 and P_2 by generating two additional control points T_1 and T_2.

To process a given heartbeat we start by extracting 23 points $(p_i, i \in [0, 22])$, which consist of: the peak (p_{11}), 10 points before the peak $(p_i, i \in [1, 10])$, 10 points after the peak $(p_i, i \in [12, 21])$ and 2 points before and after the heartbeat (p_0 and p_{22}). Similarly to the first method, the heartbeat is divided into 20 segments $(S_i, i \in [1, 20])$, where S_i is the segment from point p_i to point p_{i+1}. The two additional points (p_0 and p_{22}), are only used to assist in the generation of segments S_1 and S_{20}.

Interpolation of a given segment S_i starts off by selecting the 4 consecutive points (P_0, P_1, P_2, P_3) needed for the generation of the curve running from p_i to p_{i+1}. The points used for the generation of the segment S_i are assigned in the following way: $P_0 = p_{i-1}$, $P_1 = p_i$, $P_2 = p_{i+1}$ and $P_3 = p_{i+2}$. Following the assignment of the points, we must generate the required control points T_1 and T_2. In this paper, we use a fascinating proposal for the creation of these points given by Cem Yuksel, Scott Schaefer and John Keyser in [5] and additionally implemented by [4], based on (11), where the four points (P_0, P_1, P_2, P_3) are the previously assigned points and the values d_1, d_2, d_3 are described such that $d_i = |P_i - P_{i+1}|^\alpha = \sqrt{|P_i - P_{i+1}|}$,

and $\alpha = \frac{1}{2}$ due to the Catmull-Rom curve being centripetal.

$$T_1 = \frac{d_1^2 P_2 - d_2^2 P_0 + (2d_1^2 + 3d_1 d_2 + d_2^2) P_1}{3d_1(d_1 + d_2)}$$
$$T_2 = \frac{d_3^2 P_1 - d_2^2 P_3 + (2d_3^2 + 3d_3 d_2 + d_2^2) P_2}{3d_3(d_3 + d_2)} \quad (11)$$

Following the generation of the control points T_1 and T_2, we will use the points $\{P_1, T_1, T_2, P_2\}$ in a Cubic Bézier curve equation, denoted by $B(t)$ (for $t \in [0, 1]$), given in (12) to generate the interpolated points for a given segment S_i.

$$B(t) = (1-t)^3 \cdot P_1 + 3t(1-t)^2 \cdot T_1 + 3t^2(1-t) \cdot T_2 + t^3 \cdot P_2 \quad (12)$$

C. Experiment testbed

A threshold value for the calculation of the similarity check is used to detect whether a beat belongs to the same class as the previous one. We test each pair of successive neighbouring heartbeats in MITDB and calculate the similarity coefficient as a sum of Euclidean distances between corresponding 21 samples around the R peak.

The following experiments are specified according to the applied method:

- *R360* - raw original ECG benchmark data sampled at 360 Hz,
- *R125* - data resampled at a sampling frequency of 125 Hz, and
- *B1,B3,B5,B7,B15* - data sampled at 125 Hz adding 1, 3, 5, 7, and 15 points applying Bézier interpolation.

Both methods of Bézier interpolation are tested.

D. Evaluation method

The MITDB ECG benchmark contains 100733 heartbeats with a large imbalance of the V beats versus NonV class. We evaluate relevant statistical measures: sensitivity (SEN) as a ratio of correctly detected out of all V beats, positive predictive value (PPV) as a portion of correctly detected out of all V detections, specificity (SPC) as a ratio between correctly detected NonV beats out of all NonV beats, and negative predictive value (NPV) as a portion of correctly detected NonV beats out of all NonV detections. However, due to the class imbalance, we focus on F1 score as a relevant performance measure.

An improvement factor (expressed in %) is calculated as a ratio between the achieved performance of applied Bézier interpolation method and the benchmark with data sampled at 360 Hz and resampled at 125 Hz. The performance/cost ratio aims at finding the experiment that revealed the highest performance versus cost. Also, we calculate the gain as a ratio between the increased performance (improvement factor) and increased cost, both expressed in %. If the GAIN is more than 1, then the experiment results with higher performance versus the cost.

TABLE I: Average execution times for Bézier interpolation to process 21 samples around R peak.

Exp.	Method I	Method II
B1	1.41 ms	0.25 ms
B3	2.23 ms	0.32 ms
B5	2.92 ms	0.39 ms
B7	3.66 ms	0.46 ms
B15	6.62 ms	0.75 ms

TABLE II: Similarity check performance (in %) for optimal threshold values Thr applying Bézier curves.

Method I	Thr	SEN	SPC	PPV	NPV	F1
360	-	89.21	96.87	82.24	98.22	85.58
125	-	87.29	96.70	81.15	97.91	84.11
B1	34.0	89.50	96.87	82.30	98.27	85.75
B3	34.0	89.74	96.95	82.71	98.31	86.08
B5	34.0	89.35	97.06	83.19	98.25	86.16
B7	33.0	88.84	97.17	83.61	98.17	86.15
B15	33.0	89.38	97.06	83.19	98.25	86.18
Method II	Thr	SEN	SPC	PPV	NPV	F1
360	-	89.21	96.86	82.22	98.22	85.58
125	-	87.29	96.70	81.13	97.91	84.10
B1	34.0	89.45	96.87	82.31	98.26	85.73
B3	34.0	89.56	96.96	82.75	98.28	86.02
B5	34.5	89.17	97.08	83.22	98.22	86.09
B7	34.5	89.20	97.07	83.18	98.22	86.09
B15	34.5	89.22	97.07	83.22	98.23	86.12

IV. Evaluation of Results

We conducted seven experiments of calculating the similarity check of successive neighbouring heartbeats. Bézier interpolation was applied in the last five experiments with both Method I and Method II.

Table I presents the average execution times for interpolating 1, 3, 5, 7 and 15 points for each detected heartbeat in the analyzed dataset. Optimal threshold values and corresponding performance results are presented in Table II.

In our earlier paper we concluded that both methods revealed almost equal performance (Table II) although Method II was much faster than Method I (Table I). In this paper, we focus on optimization and determine the threshold value which differentiates the similarity check coefficient if the beat has changed the class was found to be approximately 34 with slight differences for different experiments.

Fig. 3 presents details on optimal threshold values and the corresponding performance measures. Adding one point for Bézier interpolation (experiment B1), the threshold value of 34 yields the maximum F1 score of 85.7%. The results for different values of interpolated points per section are very similar. When interpolating three, five or seven points per section (Experiments B3 and B5 and B7), the maximum value for the F1 score was slightly increased to 86%, with the ideal threshold being 34 in both cases.

Additionally, Experiment B1 Method I results (Fig. 4) present the performance measures SEN, SPC, PPV, NPV

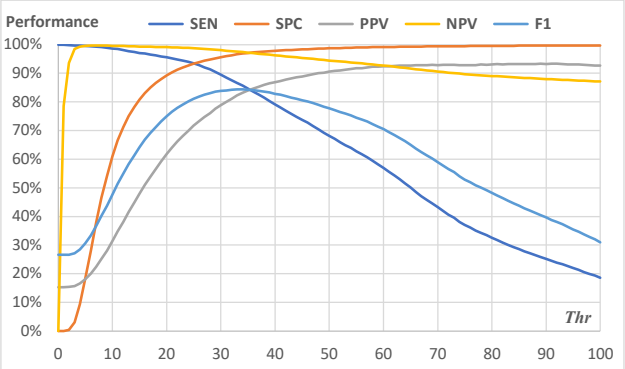


Fig. 3: Performance (in %) versus threshold values for Bézier interpolation (experiment B1 Method I)

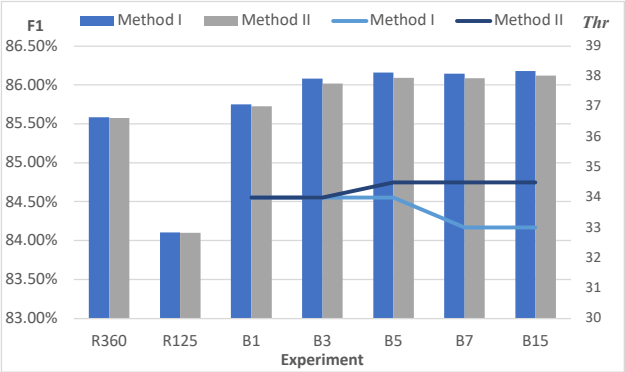


Fig. 4: Comparing optimal threshold values and corresponding F1 Scores of both Bézier interpolation methods

and F1. The lower threshold is the higher values of SEN and lower values of PPV, and the opposite, as threshold values get higher, SEN decreases and PPV increases. The optimal value is obtained for $Thr = 34$. The same trends and behaviour are observed for performance measure values in other experiments with both methods of Bézier interpolation.

Table III presents the improvement factor in % comparing the achieved F1 score versus the original benchmark (with a sampling frequency of 360 Hz), and also versus data obtained with downsampling to 2.88 times lower frequency (125 Hz). As expected, downsampling showed a decrease in performance by a factor of -1.73. However, we found that even adding one point with Bézier interpolation (experiment B1) achieves a slightly better performance (0.2%). This is a very interesting result, meaning that instead of transferring almost three times more data, we apply rather simple calculations and obtain better performance, which is a good lead toward ECG compression techniques.

Table IV presents the performance/cost ratio, revealing the optimal to be B1 experiment (adding one point) with method II (we get the highest performance for the smallest cost).

TABLE III: Improvement factor (in %) for similarity check with optimal threshold value when applying Bézier interpolation (Methods I and II)

	IF versus 360		**IF versus 125**	
Exp.	Method I	Method II	Method I	Method II
360	/	/	1.76	1.76
125	-1.73	-1.73	/	/
B1	0.20	0.18	1.96	1.94
B3	0.58	0.52	2.35	2.29
B5	0.67	0.60	2.44	2.37
B7	0.66	0.60	2.43	2.36
B15	0.69	0.63	2.46	2.40

TABLE IV: Performance/cost ratio for similarity check with optimal threshold value when applying Bézier interpolation (Methods I and II)

Exp.	Method I	Method II
B1	0.608	3.429
B3	0.386	2.688
B5	0.295	2.208
B7	0.235	1.871
B15	0.130	1.148

Adding more points with Bézier interpolation (experiments B3, B5, B7 or B15) achieves better performance, and the performance trend saturates after three added points. Table V presents the increased F1 score values (expressed in %) and also increased cost (measured as an average response time for calculations applying the corresponding Bézier interpolation method).

We observe that GAIN of 0.64 is a very small value to increase the number of interpolated points from one to three, and this value is even smaller for a larger number of points. This suggests that adding even one point with Bézier interpolation will increase the performance, with the highest performance/cost ratio, and the higher performance achieved with more points is small and is not cost-effective.

The recommendations of the American Heart Association' [22] suggest that the minimum sampling frequency for a digital ECG recording should be 500 Hz.

However, high sampling rates imply increased network throughput and processing requirements demanding computers with larger memories and faster processors. The results from the presented experiments show that a lower sampling rate would retain adequate information for ECG representation without introducing significant errors in the results.

Pizzuti et al. [11] conclude that a sampling frequency of 500 Hz gives redundant information and recommend 250 Hz without loss of statistical accuracy for ECG measurements. Kwon et al. [15] show that down-sampling to 500 or 250 Hz resulted in excellent concordance, while down-sampling to 100 Hz produced acceptable results for time-domain HRV analysis, but not for frequency-domain analysis, and down-sampling to 50 Hz proved to be unacceptable. The overall conclusion that a 250-Hz

TABLE V: GAIN versus B1 for similarity check with optimal threshold value when applying Bézier interpolation (Methods I and II)

	Method I			Method II		
Exp.	inc F1	inc cost	GAIN	inc F1	inc cost	GAIN
B3	20.06	58.16	0.35	17.87	28.00	0.64
B5	24.74	107.09	0.23	22.34	56.00	0.40
B7	23.99	159.57	0.15	21.88	84.00	0.26
B15	25.93	369.50	0.07	23.92	200.00	0.12

sampling frequency is acceptable for HRV analysis is also confirmed by our approach.

Pizzuti et al. [11] show a significant decrease in differences among the amplitude measurements between a sampling rate of 500 Hz and 125 Hz. down to 65% for the detection of the S or 83% for the detection of R in the QRS wave.

V. Conclusion

We have conducted experiments to determine the optimal threshold value for similarity check and detection (classification) of ventricular beats versus normal beats when applying Bézier interpolation. The results confirmed that a threshold value of 34 is the optimal value. Our experiments confirmed that the best performance/cost ratio is obtained by adding one more interpolation point. An ECG digital representation sampled at 125 Hz reveals sufficient information, and adding one more point (upsampling to 250 Hz) with Bézier interpolation reveals optimal performance.

Downsampling from 360 Hz to 125 Hz has decreased the similarity check performance by 1.73%. Interestingly adding one sample by Bézier interpolation, on downsampled data from 360 Hz to 125 Hz improves the similarity check by 0.2%, that is, it finds R peak more precise than the original 360 Hz sampled data. We can conclude it is good to realize sampling at a lower frequency (such as 125 Hz), transfer a smaller amount of data, add only one point to obtain a smooth curve with Bézier interpolation and then process data.

Although the highest improvement is revealed for adding 15 samples (corresponding to a sampling frequency of 2000 Hz), we observe a saturated performance increase after adding three more samples, which can be concluded that after sampling the ECG with a sampling frequency of more than 500 Hz will not be beneficial.

This research can potentially ensure that 125 Hz is the lowest sampling frequency and the 10-bit bit resolution of the ADC unit should provide sufficient performance for detecting the features in the ECG signal. The added value is running Bézier interpolation and adding one more point for upsampling that will reveal increased performance in detecting specific features in the ECG, such as classification of ventricular beats.

References

[1] M. Gusev, L. Tonkovikj, and N. Petrovski, "Effect of bézier interpolation on similarity of heartbeats in electrocardiograms," in *2022 30th Telecommunications Forum (TELFOR)*, pp. 1–4, IEEE, 2022.

[2] M. E. Mortenson, *Mathematics for computer graphics applications*. Industrial Press Inc., 1999.

[3] O. Aflak, "Bézier interpolation," 2020. Last accessed online on 15.09.2022 at: https://towardsdatascience.com/b%C3%A9zier-interpolation-8033e9a262c2.

[4] A. Joshi, "Constructing a cubic bezier that passes through four points," 2019. Last accessed online on 23.09.2022 at: https://apoorvaj.io/cubic-bezier-through-four-points/.

[5] C. Yuksel, S. Schaefer, and J. Keyser, "Parameterization and applications of catmull–rom curves," *Computer-Aided Design*, vol. 43, no. 7, pp. 747–755, 2011.

[6] S. Steinarsson, *Downsampling time series for visual representation*. PhD thesis, 2013.

[7] G. G. Berntson, K. S. Quigley, J. F. Jang, and S. T. Boysen, "An approach to artifact identification: Application to heart period data," *Psychophysiology*, vol. 27, no. 5, pp. 586–598, 1990.

[8] M. Malik, "Task force of the european society of cardiology and the north american society of pacing and electrophysiology. heart rate variability. standards of measurement, physiological interpretation, and clinical use," *Eur Heart J.*, vol. 17, pp. 354–381, 1996.

[9] R. J. Ellis, B. Zhu, J. Koenig, J. F. Thayer, and Y. Wang, "A careful look at ECG sampling frequency and r-peak interpolation on short-term measures of heart rate variability," *Physiological measurement*, vol. 36, no. 9, p. 1827, 2015.

[10] E. Ajdaraga and M. Gusev, "Analysis of sampling frequency and resolution in ecg signals," in *2017 25th Telecommunication Forum (TELFOR)*, pp. 1–4, IEEE, 2017.

[11] G. Pizzuti, S. Cifaldi, and G. Nolfe, "Digital sampling rate and ecg analysis," *Journal of biomedical engineering*, vol. 7, no. 3, pp. 247–250, 1985.

[12] L. Hejjel and E. Roth, "What is the adequate sampling interval of the ecg signal for heart rate variability analysis in the time domain?," *Physiological measurement*, vol. 25, no. 6, p. 1405, 2004.

[13] M. Gusev, S. Tudjarski, and A. Anagelevska, "Sampling rate impact on heart rate variability," in *2022 30th Telecommunications Forum (TELFOR)*, pp. 1–4, IEEE, 2022.

[14] S. Mahdiani, V. Jeyhani, M. Peltokangas, and A. Vehkaoja, "Is 50 hz high enough ecg sampling frequency for accurate hrv analysis?," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5948–5951, IEEE, 2015.

[15] O. Kwon, J. Jeong, H. B. Kim, I. H. Kwon, S. Y. Park, J. E. Kim, and Y. Choi, "Electrocardiogram sampling frequency range acceptable for heart rate variability analysis," *Healthcare informatics research*, vol. 24, no. 3, pp. 198–206, 2018.

[16] A. Prieto-Guerrero, C. Mailhes, and F. Castanié, "Lost sample recovering of ecg signals in e-health applications," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 31–34, IEEE, 2007.

[17] S. M. M. Islam, M. S. Farid, and M. A. Kiber, "Denoising ecg signal using adaptive filter algorithms and cubic spline interpolation for regaining missing data points of ecg in telecardiology system," *International Journal of Enhanced Research in Science, Technology and Engineering*, vol. 4, no. 12, 2015.

[18] R. Soontornvorn, H. Fujioka, V. Chutchavong, and K. Janchitrapongvej, "Modeling ecg waveform using optimal smoothing bézier-bernstein curves," in *TENCON 2017-2017 IEEE Region 10 Conference*, pp. 1235–1238, IEEE, 2017.

[19] E. K. Roonizi and R. Sameni, "Morphological modeling of cardiac signals based on signal decomposition," *Computers in biology and medicine*, vol. 43, no. 10, pp. 1453–1461, 2013.

[20] R. Chaturvedi and Y. Yadav, "Analysis of ecg signal by polynomial approximation," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, no. 5, pp. 1029–1033, 2014.

[21] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[22] H. Pipberger, G. Baule, A. Berson, S. Briller, D. Geselowitz, L. Horan, and O. Schmitt, "Recommendations for standardization of instruments in electrocardiography and vectorcardiography," *IEEE Trans. Biomed. Eng.*, no. 1, pp. 60–68, 1967.