Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 20(2): 1716–1729.
DOI: 10.3934/mbe.2023078
Received: 05 September 2022
Revised: 17 October 2022
Accepted: 26 October 2022
Published: 04 November 2022

Research article

Bio-metric authentication with electrocardiogram (ECG) by considering variable signals

Hoon Ko[1,2,†], Kwangcheol Rim[3,†] and Jong Youl Hong[4,*]

[1] Research & Development Center, MetaiONE Inc., Business Incubation Center (#504), Chosun University, Gwangju 61452, Korea

[2] Instituto Superior de Engenharia do Porto (ISEP/IPP), Porto 4249-015, Portugal

[3] The Department of Mathematics, Chosun University, 309 pilmundae-ro, Gwangju 61452, Korea

[4] College of Culture & Sports, Korea University, Korea University Sejong Campus, Sejong City 30019, Korea

† The authors contributed equally to this work as the first author.

* **Correspondence:** Email: herr_hong@korea.ac.kr; Tel: +821076376437.

Abstract: The use of conventional bio-signals such as an electrocardiogram (ECG) for biometric authentication is vulnerable to a lack of verification of continuity of signals; this is because the system does not consider the change in signals caused by a change in the situation of a person, that is, conventional biological signals. Prediction technology based on tracking and analyzing new signals can overcome this shortcoming. However, since the biological signal data sets are massive, their utilization is crucial for higher accuracy. In this study, we defined a 10×10 matrix for 100 points based on the R-peak point and an array for the dimension of the signals. Furthermore, we defined the future predicted signals by analyzing the continuous points in each array of the matrices at the same point. As a result, the accuracy of user authentication was 91%.

Keywords: biometric authentication; electrocardiogram; variable signal; smart devices; wearable devices

1. Introduction

Generally, fingerprint, facial, voice and iris recognition have been used for static authentication and show high accuracy. However, they are vulnerable to theft [1]. In contrast, dynamic biometrics, such as electrocardiogram (ECG) and electromyography (EMG), are relatively difficult to steal. However, they exhibit poor correlations concerning the same users, existing signals, and new input signals. Therefore,

establishing a correlation between the existing and new signals will improve the existing biometric security systems. Many techniques can evaluate biometric accuracy. Although it is possible to verify the user authentication accuracy using a sequence, it is difficult to authenticate the same user based on new signals [2]. To illustrate, currently, Convolutional Neural Network (CNN) based algorithms form the mainstream algorithms that cut input signals into small pictures; however, upon receiving new signals, they cannot prove whether they are coming from the same person. This can be solved by proving the continuity of input signals with the previous signals [3]. In addition, since bio-signals report different values depending on the person's situation, the authentication systems should be able to predict signals that can change in the future. This study aims to authenticate bio-signals in real time using an ECG. A typical ECG cycle contains five points: P, Q, R, S and T. A normal person measures 60–100 ECG signals in a minute. R and S are easy to detect since they correspond to the highest and lowest points of the biometric signal; however, detecting P, Q and T is challenging. In a conventional ECG study, each point (R, S) and some intervals (with segments) are used to find individual features, such as the PR/RT interval, the PR/ST segment, and the QRS complex. Since the system must detect a special point from all signals in a real-time system, it is necessary to analyze the relationships between each measured signal. In this study, we measured a total of 100 points, namely, the R-peak and the points on the left [−50] and right [+49] of the R-peak; they were defined as the features centered on the R-peak [4]. Thereafter, we analyzed each feature of [−50, R, +49] from ECG signals and conducted a study to examine the distribution of the input of the point to the feature. The remainder of this study proceeds as follows. Section 2 defines the existing research and problems. Section 3 presents a model for analyzing the relationship between the proposed R-peak and the surrounding points, Section 4 presents the analysis and discussion of the experiment, and Section 5 concludes the study.

2. Related works

The biggest drawback of ECG is that it recognizes newly measured points differently from the existing measured and stored points [5, 6]. [5, 6] sought to explore current research which has been conducted on Recurrent Neural Networks (RNNs) in four areas such as biometric authentication, expression recognition, anomaly detection and applications. They reviewed each proposed method's methodology, object, output and benefits. For example, in a fingerprint recognition system, although fingerprints do not change from birth to growth or death, ECG provides varying output points depending on the measuring time, place, and emotion [7]. In [8], H. Ko et al. conducted a study to overcome the unstable measured ECG signature. They tried to find each pattern in the threshold to overcome the problem of bio-information about measurement. With the pattern from the threshold, they found the user's identification. To study the unique patterns, we measured the ECG for 120 seconds. As the first step, we divided them every 1–2 seconds, and then it used CNN to identify. As the result, the accuracy was 83.36%. [9] was an extension of [8]; Ko et al. studied the adjusted $(Q_i * S_i)$ for advanced personal identification. That study also analyzed the ECG signal flow in $(Q_i * S_i)$ and defined i as 50. Currently, big data systems pre-process and store raw data, use the same procedure to update big data after receiving and processing raw data continuously, and train models using the updated big data. This method is suitable for applications that require real-time processing, such as security systems for authorization/authentication. Furthermore, as mentioned above, since ECG signals have multiple changing factors such as time, place, and emotion, there is no way to

establish a relationship between previous signals and signals measured in real-time [10]. Recently, CNNs have been used in research related to authentication/authorization methods using bio-signals. CNN refers to converting 0–1999 points (one cycle of ECG), as displayed in Figure 2, into small pictures, grouping similar patterns and finding a pattern included in the pertinent pattern [11]. In other words, it is a multilayered feed-forward artificial neural network utilized for analyzing visual images. CNN is classified as a deep neural network in deep learning and is primarily applied in visual image analysis. Furthermore, a Space-Invariant Artificial Neural Network (SIANN) based on the shared-weight architecture and shift-invariant characteristics [11] has been applied to image and video recognition, recommendation systems, image classification, medical image analysis, natural language processing, etc. Studies have been conducted to evaluate the usage of biometric data within biometric systems considering different or other data sources of the same user and how two datasets can be connected to build or improve security systems [12], as it has been recognized as a shortcoming that should be solved in security modules using dynamic biological signals. As mentioned before, although many researchers have been studying authentication techniques using biometrics such as ECG and EMG, they still face a problem wherein the measured points are different, and the current results are not adaptable for a real-time system. This study proposes a model with various signals that analyzes continuing signals.

3. Proposed model with variable signals

3.1. Concepts

This study may have significance in that it is possible to predict the flow of ECG values to be generated in the future by continuously identifying the values of the left-right fixed section centering on the R peak of the ECG signal. In authentication using dynamic biometric information, measuring the correspondence between old information and newly measured biometric signals is a weakness from a security point of view. This is because, as mentioned above, the measured value of dynamic biometric information is different depending on the situation and condition whenever it is measured. This study's proposed method is analyzing the signal of an existing specific section and mathematically grasping each section's information, then analyzing the results of the newly measured information with the previously stored information. Indeed, the newly measured information is also continuously stored in the database and is used as information for a new signal to be measured in the future. Therefore, the primary method of this study is to continuously track the characteristics of the signal so that the characteristics of biometric information according to the user's situation are memorized in advance. Figure 1 shows how to measure with a wearable device.

3.2. Collection & Classification

Figure 2 shows the ECG signal measurement method used in this study, and the following explains how the dataset is constructed. Typically, an adult has a heart rate of 60–100 beats per min and records 10–16 heartbeats every 10 sec in an ECG test. In this study, we set the ECG experiment to create 2000 points for each cycle of ECG. In other words, 120,000–200,000 points were recorded per minute, and each point was entered as an input in a serial format [13].

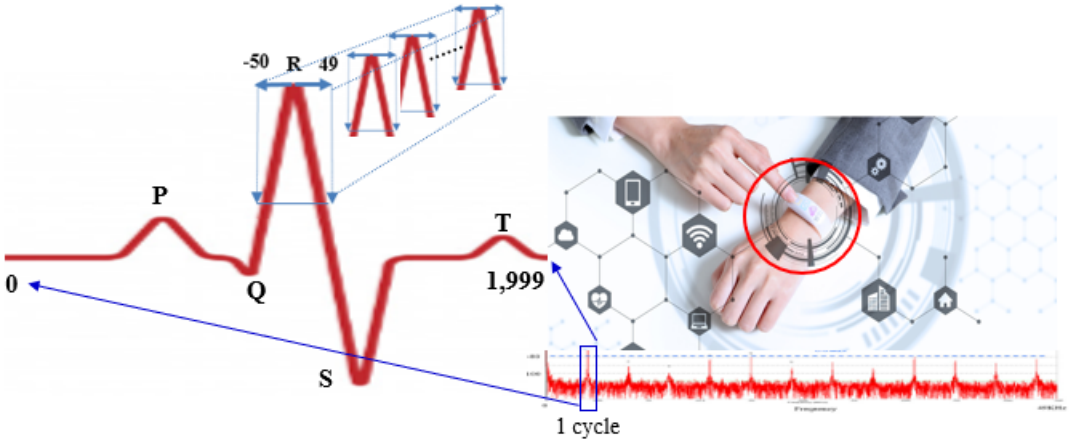

Figure 1. How to measure.

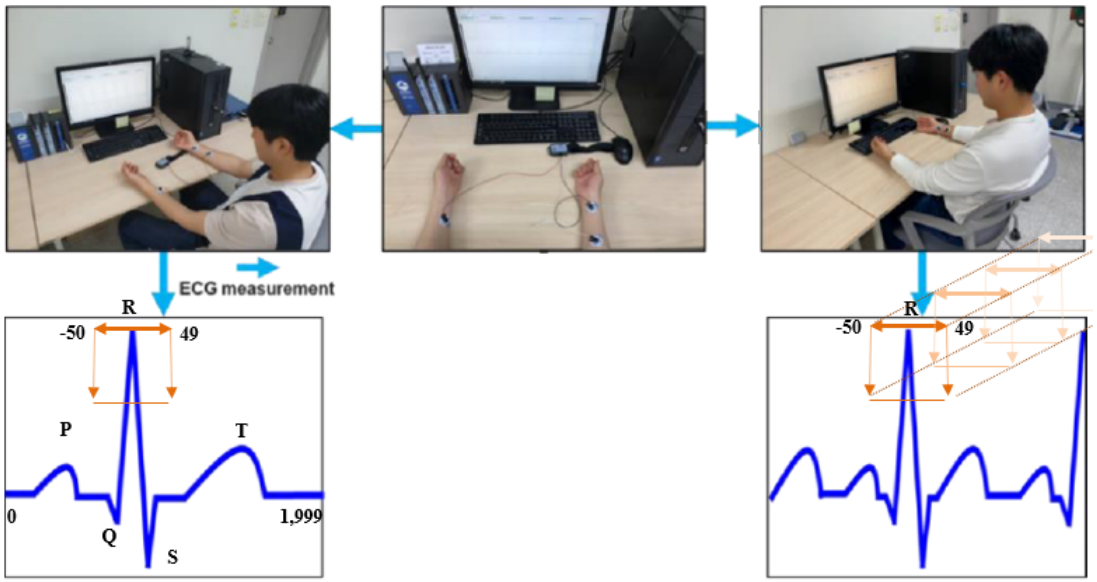

Figure 2. ECG collection in lead I between 0 and 1999 in one cycle [14].

An experiment to analyze the normalization method and its user recognition performance was performed using MATLAB R2019a on a personal PC with an Intel Core i7 processor. The numbers of subjects that participated in the ECG measurement for the experiment were 50 males and 35 females, ranging from 23 to 34; all of them were associated with Chosun University. The ECG measurement was performed once for each subject at 2-day intervals. It was measured for 60 s before and 120 s after exercise in a comfortable sitting posture, as shown in Figure 2. To measure the post-exercise ECG, the subject exercised with a stepper for 60s with attached electrodes. The post-exercise ECG was measured immediately after the exercise in a sitting posture. The ECG obtained from the measuring instrument MP160 was a lead-I ECG, which could be acquired from both wrists using wet electrodes at a sampling rate of 2000 Hz based on the international standard 12-lead electrocardiography. An experiment for evaluating the normalization method using the measured

ECG data was conducted to analyze the average similarity of the ECG [14]. One cycle constitutes a signal based on each cycle's first points [15]. The R-peak point is located at the highest position in the signal to detect the R-peak from each cycle. After finding the highest point, the model jumps by 50 points in the left direction and records the point in the 50^{th} column. Similarly, it moves 49 columns to the right of the R-point and saves the value in the 49^{th} column. This process is repeated continuously, and the resultant data obtained forms the dataset. Thereafter, this dataset, consisting of the 10×10 matrix, is defined for 100 points, that is, the points ranging from -50 on the left to 49 on the right of the R-peak, and is used to compose continuous matrices for the arrays.

3.3. Data representation procedure

When the representation with the group is configured in the manner displayed in Figure 3, which includes (a) measure, (b) matrix process and (c) n-dimensional array, it is possible to analyze the points of the same position in repeated cycles as continuous points. Groups [$a_{11}.group$, $a_{12}.group$, ..., $a_{nn}.group$] are formed for the same position in each matrix, and the mean, variance (σ^2), and standard deviation (σ) of the points belonging to each group are generated using the equations proposed below. The equations for finding the points of each group [$a_{ij}.group$]. The group are as follows. Herein, M represents the matrix, i and j represent the room's location, n is the maximum value of the location of entry, and N is used to calculate the mean.

$$mean(a_{ij}) = (M_1.a_{11} + M_2.a_{12} + \cdots + M_n.a_{ij})\frac{1}{N} \tag{3.1}$$

$$a_{ij}\sigma^2 = \frac{1}{N}\sum(mean\,(a_{ij}) - M_i.a_{ij})^2 \tag{3.2}$$

$$a_{ij}\sigma = \sqrt{\frac{1}{N}\sum(mean\,(a_{ij}) - M_i.a_{ij})^2} \tag{3.3}$$

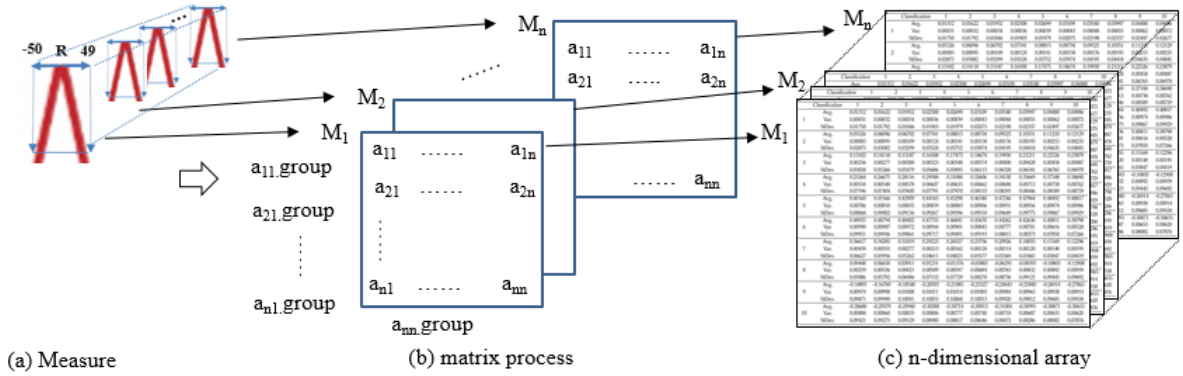

Figure 3. Data representation procedure with a group.

3.4. Matrix organization

The 10×10 matrix comprises 50 points from the left and 49 points from the right out of 128 R-peak signals. Table 1 provides the values to the matrix that summarizes the points obtained from Eqs

Table 1. Value of matrix. In this table, the values measured in a specific range left/right based on the R peak are organized into an array and a matrix, and the mean, variance, and standard deviation of the values at the same location are calculated.

	Type	1	2	3	4	5	6	7	8	9	10
	Avg.	0.01312	0.01622	0.01952	0.02308	0.02699	0.03109	0.03540	0.03997	0.04480	0.04986
1	Var.	0.00031	0.00032	0.00034	0.00036	0.00039	0.00043	0.00048	0.00055	0.00062	0.00072
	StDev.	0.01750	0.01792	0.01846	0.01903	0.01979	0.02075	0.02198	0.02337	0.02497	0.02677
	Avg.	0.05526	0.06096	0.06702	0.07341	0.08015	0.08734	0.09525	0.10351	0.11210	0.12129
2	Var.	0.00083	0.00095	0.00109	0.00124	0.00141	0.00158	0.00176	0.00195	0.00215	0.00235
	StDev.	0.02875	0.03082	0.03299	0.03528	0.03752	0.03974	0.04195	0.04418	0.04635	0.04845
	Avg.	0.13102	0.14118	0.15187	0.16308	0.17475	0.18674	0.19930	0.21211	0.22526	0.23879
3	Var.	0.00256	0.00277	0.00300	0.00323	0.00348	0.00374	0.00400	0.00428	0.00458	0.00487
	StDev.	0.05058	0.05266	0.05479	0.05686	0.05895	0.06113	0.06328	0.06541	0.06765	0.06978
	Avg.	0.25264	0.26675	0.28116	0.29588	0.31088	0.32606	0.34130	0.35669	0.37188	0.38690
4	Var.	0.00518	0.00548	0.00578	0.00607	0.00635	0.00662	0.00688	0.00713	0.00738	0.00762
	StDev.	0.07196	0.07404	0.07600	0.07791	0.07970	0.08133	0.08293	0.08446	0.08589	0.08729
	Avg.	0.40160	0.41566	0.42909	0.44165	0.45298	0.46340	0.47246	0.47964	0.48492	0.48817
5	Var.	0.00786	0.00810	0.00835	0.00859	0.00883	0.00906	0.00931	0.00956	0.00974	0.00986
	StDev.	0.08868	0.09002	0.09136	0.09267	0.09396	0.09518	0.09649	0.09775	0.09867	0.09929
	Avg.	0.48935	0.48794	0.48402	0.47755	0.46841	0.45670	0.44262	0.42636	0.40811	0.38798
6	Var.	0.00990	0.00987	0.00972	0.00944	0.00901	0.00845	0.00777	0.00701	0.00616	0.00528
	StDev.	0.09951	0.09936	0.09861	0.09717	0.09491	0.09193	0.08815	0.08373	0.07850	0.07266
	Avg.	0.36617	0.34285	0.31819	0.29223	0.26527	0.23756	0.20926	0.18055	0.15169	0.12296
7	Var.	0.00439	0.00355	0.00277	0.00213	0.00162	0.00128	0.00114	0.00120	0.00148	0.00195
	StDev.	0.06627	0.05956	0.05262	0.04611	0.04025	0.03577	0.03369	0.03465	0.03847	0.04419
	Avg.	0.09448	0.06650	0.03911	0.01231	-0.01376	-0.03883	-0.06291	-0.08593	-0.10803	-0.12908
8	Var.	0.00259	0.00336	0.00421	0.00509	0.00597	0.00684	0.00763	0.00832	0.00892	0.00939
	StDev.	0.05086	0.05792	0.06486	0.07132	0.07729	0.08270	0.08736	0.09123	0.09443	0.09692
	Avg.	-0.14893	-0.16769	-0.18548	-0.20301	-0.21885	-0.23327	-0.24643	-0.25840	-0.26914	-0.27865
9	Var.	0.00974	0.00998	0.01008	0.01011	0.01014	0.01003	0.00984	0.00963	0.00938	0.00914
	StDev.	0.09871	0.09990	0.10041	0.10055	0.10068	0.10013	0.09920	0.09812	0.09685	0.09558
	Avg.	-0.28680	-0.29379	-0.29940	-0.30388	-0.30714	-0.30913	-0.31004	-0.30991	-0.30871	-0.30655
10	Var.	0.00888	0.00860	0.00833	0.00806	0.00777	0.00748	0.00718	0.00687	0.00653	0.00620
	StDev.	0.09421	0.09275	0.09129	0.08980	0.08817	0.08646	0.08472	0.08286	0.08082	0.07876

(3.1)–(3.3), namely, mean *(Avg.)*, variance *(Var.)*, and standard deviation *(StDev)*, respectively. Table 1 only gives the mean, variance, and standard deviation according to each cycle of the ECG signal, but the bio-signal is continuously measured through the sensor. New values will be measured according to the health status, condition, etc. of the person being measured, but since the mean, variance, and standard deviation are continuously measured, the biometric information of the person changing is also calculated and is shown in Table 1. If these results continue, eventually, the user's future measurement values can be predicted. Accordingly, the predicted values are expected to be used in a dynamic biometric information authentication algorithm. Therefore, the results of Table 1 can be beneficial in the future.

3.5. Procedures

Analysis of variance (ANOVA) is used to compare variance values from the means (or arithmetic mean) of different groups. It is used in various scenarios to see if there is a difference between the means of different groups. In this study, the results summarized in Table 1 were used to measure the accuracy after comparison with newly measured values. As mentioned above, 2000 measurements were made for each cycle of the ECG signal. Generally, the heart rate of humans is between 60 and 100, so there were 120,000 to 200,000 measurements per minute, and the cost to process was very high. Therefore, it was carried out according to the following processing procedure. First, the procedure has to find the R-peak of each period and compose the values of fixed positions (±10 and ±50 in this study) into a matrix by referring to the procedure in Figure 3. Then, the values of the same position of each row and column, constituted based on the R-peak of each period, are arranged as consecutive input values. For example, the first value of the first period is stored in the space of (a_{11}) in the first array, and the first value in the second period is stored in the second array (a_{21}). Therefore, the period's first value is stored in each matrix's first space. After all, if consecutive values in a specific space of a dataset consisting of a matrices are analyzed through mathematical analysis, for example, the mean (3.1), variance (3.2), and standard deviation (3.3), it is expected that tracking and forecasting the (same position) value of the next period will be possible. After this procedure, Table 1 was finally created, and the stored values are continuously updated according to the analysis of the continuously measured ECG values. The reason for the update is to reduce the error of the bio-signal value that is changed whenever it is measured, which has been pointed out as an existing problem. For the input signal, the value generated by the above processing is compared with the values in Table 1 to measure the accuracy.

4. Analysis and discussion

In this section, we discuss the description of the dataset, the detailed accuracy results, and the results of the matrices [−10, R,+10] and [−50, R,+49].

4.1. Description of dataset

The mean, variance and standard deviation of the 100 points were obtained from the flow of points generated at 50 distance points on the left and 49 distance points on the right of the R-peak point of a person (the R-peak is the 51st). Furthermore, when new input points were entered, each point's mean, variance and standard deviation were substituted to determine whether the positions of the 100 points

are included in the training points [16]. Table 2 summarizes the analysis results using the collected and organized datasets [17, 18]. As a result, since the points located in each location of each matrix were stored in the same position as the next point, we could analyze the point of the same position of each signal. We used Weka version 3.8.5, a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rule mining and visualization to analyze the flow of these points. Every 100 points of each signal were configured as a Weka-compatible dataset. It supports several standard data mining tasks: more specifically, clustering, data pre-processing, regression, classification, feature selection and visualization. In Weka, the input is expected to be formatted following the attribute-relational file format (.arff extension). The techniques are predicated on the assumption that the data is available as one flat file or relation, where a fixed number of attributes describes the data points.

4.2. Detailed accuracy results

Table 2 summarizes the results of correct instances from the dataset. From a total of 189 instances, the proposed technique classified 172 instances correctly by 91.01% and 17 instances incorrectly by 8.99%. The mean absolute error (MAE) calculated using Eq (4.1) was 0.0102, and the root mean square error (RMSE) was 0.2999 [19, 20].

Table 2. Correct instances.

Type	Correct State	Accuracy
Correctly Classified Instances	172	91.01%
Incorrectly Classified Instances	17	8.99%
Kappa statistic	0.8002	
Mean absolute error (MAE)	0.0899	
Root mean squared error (RMSE)	0.2999	

$$MAE = \frac{1}{N}\sum \|y_1 - y_2\| \tag{4.1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum (y_1 - y_2)^2} \tag{4.2}$$

We can use Eqs (4.1) and (4.2) to get MAE and RMSE. Next, to get the *kappa* value, we need to get P_0 and P_e by using Eq (4.3) and (4.4).

$$P_0 = \frac{TP + FP}{TP + TN + FP + FN} \tag{4.3}$$

$$\begin{aligned} P_e = &[\frac{TP + TN}{TP + TN + FP + FN} \times \frac{TP + FP}{TP + TN + FP + FN}] \\ &+[\frac{FP + FN}{TP + TN + FP + FN} \times \frac{TN + FN}{TP + TN + FP + FN}] \end{aligned} \tag{4.4}$$

$$Kappa\ K = \frac{P_0 - P_e}{1 - P_e} \tag{4.5}$$

Table 3. Kappa agreement table.

Range	Contents
0	Less than chance agreement
0.01–0.20	Slight agreement
0.20–0.40	Fair agreement
0.40–0.60	Moderate agreement
0.60–0.80	Substantial agreement
0.80–0.99	Almost perfect agreement
1	Perfect agreement

The *kappa* value generated by Eq (4.5) was evaluated using Table 3. As per Table 3, the value of the generated Kappa, which was 0.8002, exhibited an *Almost perfect agreement* [21]. Table 4 presents the results determined for each class. TP denotes the true positive rate, which was determined [22, 23]. As per the results, when a person was classified as a user, the probability of that person being a user was 95.9%, whereas when a person was not determined as a user, the positive result was 82.4%. The false positive (FP) rate, which determines the number of incorrectly classified desired classes among those that do not belong to the desired class, was 17.6% [24]. The precision was 90.6%, and the recall was 95.9%.

Table 4. Detailed accuracy by class.

Item	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.959	0.176	0.906	0.959	0.932	0.803	0.891	0.895	User
	0.824	0.041	0.918	0.824	0.868	0.803	0.891	0.820	Not User
Weighted Avg.	0.910	0.128	0.910	0.910	0.909	0.803	0.891	0.868	

$$TP\ Rate = \frac{1}{TP + FN} \sum TP \tag{4.6}$$

$$FP\ Rate = \frac{1}{FP + TN} \sum FP \tag{4.7}$$

$$Precision = \sum \frac{TP}{TP + FP} \tag{4.8}$$

$$Precision = \sum \frac{TP}{TP + FN} \tag{4.9}$$

4.3. *Results of [−10,R,+10] and [−50,R,+49]*

We performed the analysis using 100 continuous points, that is, 50 points on the left and 49 points on the right based on the R-peak, for 2000 points in one cycle of ECG. Table 5 displays the results of analyzing whether a person was an existing or new user by comparing the range of the 100 continuous points and the range of the points extracted in more detail [25]. Table 5 displays the minimum, median, average and maximum points extracted using the points of two intervals, [−10, R,10] and [−50, R,49].

Table 5. The results of [−10,R,+10] and [−50,R,+49].

item	scope	minimum	median	average	maximum
Avg.	[−10, R, +10]	0.36617	0.45670	0.44880	0.48935
	[−50, R, +49]	-0.31004	0.12213	0.12245	0.48935
Var.	[−10, R, +10]	0.00439	0.00883	0.00839	0.00990
	[−50, R, +49]	0.00031	0.00612	0.00550	0.01014
Stdev.	[−10, R, +10]	0.06627	0.09396	0.09118	0.09951
	[−50, R, +49]	0.01750	0.07821	0.06934	0.10068

Figure 4 displays the graphs for the extracted values of the two intervals; it superimposes two figures and defines the extensive range of all the points. If the points calculated by specifically dividing, such as the minimum, median, average and maximum, are within the range of Figure 4 (right rectangle), then the person is a current user. If it is a different user, then at least one point deviated from the ranges of *Avg.*, *Var.* and *StDev*. As seen in Figure 4, each value was within the range of the corresponding value of Figure 4. Therefore, the resultant conclusion was that the two ranges were from the same user.

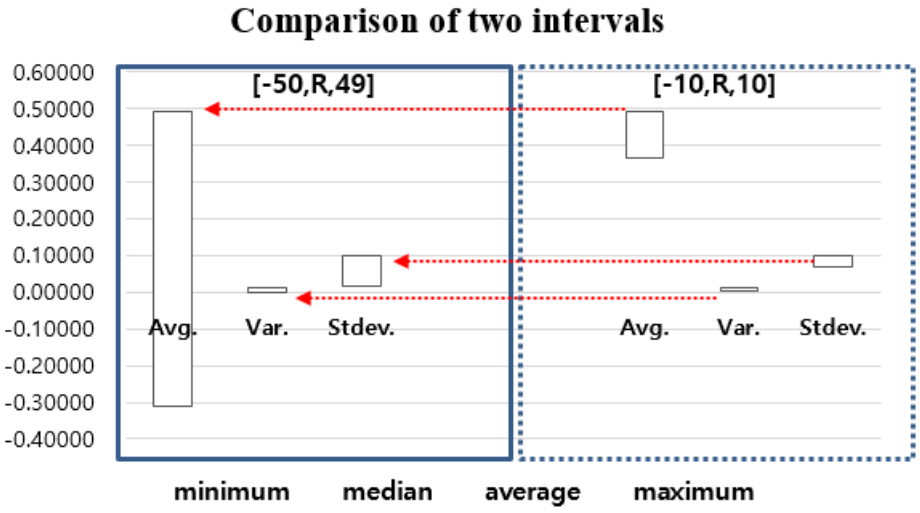

Figure 4. Comparison of two results.

4.4. *Discussion*

We addressed the current problem of bio-signal authentication. The biggest drawback of ECG is that it recognizes newly measured points differently from the existing measured and stored points, and

ECG provides varying output points depending on the measuring time, place and emotion. This method of biometric information measurement has difficulties in applying authentication. This is because, although an unchanging reference point is required for authentication, the reference point changes every time dynamic biometric information such as ECG is measured [26, 27]. In this study, measurements were made based on the R-peak value and the left and right values of fixed positions to solve the problem of the dynamic reference point of dynamic biometric information. Then, the period of the continuous signal was measured in the same way; and as the period was repeated, the value of the dynamic reference point was analyzed and applied to authentication. If the results of this study are applied to other dynamic bio-signals such as EMG, the application of a new authentication method using biometric information is expected. This study has remarkable significance in predicting user information that will change in the future by designating only a specific part of the user's biometric information and ECG and conducting follow-up studies. However, the current research is data that measures the information for general user status. It may be challenging to apply it to various emotional changes. Therefore, it is necessary to reinforce the ECG measurement contents according to emotion, which will be promoted in the future.

5. Conclusions

Biological signals are classified into static biometrics, such as those used in fingerprint, iris, facial, and vein recognition, and dynamic biometrics, such as ECG, Electroencephalography (EEG) and EMG. Multiple studies have been conducted on biosecurity using these static biological signals; however, with advancements in computing technology, attackers can easily expose them. Hence, many researchers are conducting studies using dynamic bio-signals that change every time they are measured. However, since these points provide a different value every time they are measured, applying them in security systems has been challenging. Recently, machine learning and deep learning have been applied to these bio-signals; however, they are not perfect solutions for recognizing new input bio-signals. To solve this problem, this study proposes a method of tracking the input dynamic bio-signals and analyzing their flow to predict future input points. Compared to conventional studies that track biometric information in the early stages, our study performed an analysis using the points on the left and right sides of the R-peak. To solve the inconvenience of authentication using CNN, the proposed algorithm merged coordinate data, rather than an image, into a column and authenticated the correspondence by simple analysis of variance. The accuracy of the proposed technique was obtained as 91.01%. This study intends to continuously update or manage the results generated by processing continuous signals with the same algorithm; it is expected that it will be effective for user authentication by predicting a signal to be input in the future. In addition, it is expected that signals that change according to a person's health status can also be predicted as a result of continuous measurement. In the future, we plan to conduct an expanded study by adding enhanced equations to the current study.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A2A01064089). This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded

by the Ministry of Education (No. 2021R1I1A3040361).

Conflict of interest

The authors declare no conflict of interest.

References

1. S. Tripathi, J. Murgi, K. R. S. Soni, R. Singh, A literature survey on multi model bio-metric system, *J. Comput. Technol.*, **10** (2021), 1–5.
2. W. Wu, S. Pirbhulal, G. Li, Adaptive computing-based biometric security for intelligent medical applications, *Neural Comput. Appl.*, **32** (2020), 11055–11064. https://doi.org/10.1007/s00521-018-3855-9
3. C. Che, P. Zhang, M. Zhu, Y. Qu, B. Jin, Constrained transformer network for ECG signal processing and arrhythmia classification, *BMC Med. Inf. Decis. Making*, **21** (2021), 184. https://doi.org/10.1186/s12911-021-01546-2
4. Z. Chen, P. Zhao, F. Li, T. T. Marquez-Lago, A. Leier, J. Revote, iLearn: an integrated platform and meta-learner for feature engineering, machine-learning analysis and modeling of DNA, RNA and protein sequence data, *Briefings Bioinf.*, **21** (2020), 1047–1057. https://doi.org/10.1093/bib/bbz041
5. J. M. Ackerson, R. Dave, N. Seliya, Applications of recurrent neural network for biometric authentication & anomaly detection, *Information*, **12** (2021), 272. https://doi.org/10.3390/info12070272
6. M. R. Ogiela, L. Ogiela, U. Ogiela, Biometric methods for advanced strategic data sharing protocols, in *IMIS 2015-The Ninth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2015)*, (2015), 179–183. https://doi.org/10.1109/IMIS.2015.29
7. S. K. Cherupally, S. Yin, D. Kadetotad, C. Bae, S. J. Kim, J. S. Seo, A smart hardware security engine combining entropy sources of ECG, HRV, and SRAM PUF for authentication and secret key generation, *IEEE J. Solid State Circuits*, **55** (2020), 2680–2690. https://doi.org/10.1109/JSSC.2020.3010705
8. H. Ko, S. B. Pan, L. Měsíček, Personal identification study for touchable devices with ECG, *Concurr. Comput. Pract. Exp.*, **32** (2020), e5169. https://doi.org/10.1002/cpe.5169
9. H. Ko, M. R. Ogiela, L. Ogiela, L. Mesicek, M. Lee, J. Choi, et al., ECG-based advanced personal identification study with adjusted (Qi * Si), *IEEE Access* , **7** (2019), 40078–40084. https://doi.org/10.1109/ACCESS.2019.2903575
10. A. Lee, Y. Kim, Photoplethysmography as a form of biometric authentication, in *2015 IEEE Sensors*, (2015). https://doi.org/10.1109/ICSENS.2015.7370629
11. Y. Li, Y. Pang, K. Wang, X. Li, Toward improving ECG biometric identification using cascaded convolutional neural networks, *Neurocomputing*, **391** (2020), 83–95. https://doi.org/10.1016/j.neucom.2020.01.019

12. S. Khan, S. Parkinson, L. Grant, N. Liu, S. McGuire, Biometric systems utilising health data from wearable devices: applications and future challenges in computer security, *ACM Comput. Surv.*, **53** (2020), 1–29. https://doi.org/10.1145/3400030

13. J. A. Jahanshahi, H. Danyali, M. S. Helfroush, Compressive sensing based the multi-channel ECG reconstruction in wireless body sensor networks, *Biomed. Signal Process. Control*, **61** (2020), 102047. https://doi.org/10.1016/j.bspc.2020.102047

14. G. Choi, H. Ko, W. Pedrycz, A. K. Singh, S. B. Pan, Recognition system using fusion normalization based on morphological features of post-exercise ECG for intelligent biometrics, *Sensors*, **20** (2020), 7130. https://doi.org/10.3390/s20247130

15. H. Choi, B. Lee, S. Yoon, Biometric authentication using noisy electrocardiograms acquired by mobile sensors, *IEEE Access*, **4** (2016), 1266–1273. https://doi.org/10.1109/ACCESS.2016.2548519

16. R. E. Wright, Logistic regression, *Am. Psychol. Assoc.*, **7** (1995), 217–244.

17. J. Pérez, J. Díaz, J. Garcia-Martin, B. Tabuenca, Systematic literature reviews in software engineering—Enhancement of the study selection process using Cohen's kappa statistic, *J. Syst. Software*, **168** (2020), 110657. https://doi.org/10.1016/j.jss.2020.110657

18. F. Thabtah, N. Abdelhamid, D. Peebles, A machine learning autism classification based on logistic regression analysis, *Health Inf. Sci. Syst.*, **7** (2019), 12. https://doi.org/10.1007/s13755-019-0073-5

19. A. Cano, B. Krawczyk, Kappa updated ensemble for drifting data stream mining, *Mach. Learn.*, **109** (2020), 175–218. https://doi.org/10.1007/s10994-019-05840-z

20. A. D. Raadt, M. J. Alexandra, R. J. Bosker, H. A. K. Kiers, Kappa coefficients for missing data, *Edu. Psychol. Meas.*, **79** (2019), 558–576. https://doi.org/10.1177/0013164418823249

21. M. J. Warrens, Kappa coefficients for dichotomous-nominal classifications, *Adv. Data Anal. Classif.*, **15** (2021), 193–208. https://doi.org/10.1007/s11634-020-00394-8

22. K. Itano, K. Ueki, T. Iizuka, T. Kuwatani, Geochemical discrimination of monazite source rock based on machine learning techniques and multinomial logistic regression analysis, *Geosciences*, **10** (2020), 63. https://doi.org/10.3390/geosciences10020063

23. N. Senaviratna, T. Cooray, Diagnosing multicollinearity of logistic regression model, *Asian J. Prob. Stat.*, **5** (2019), 1–9. https://doi.org/10.9734/ajpas/2019/v5i230132

24. M. D. Cock, R. Dowsley, A. C. A. Nascimento, D. Railsback, J. Shen, A. Todoki, High performance logistic regression for privacy-preserving genome analysis, *BMC Med. Genomics*, **14** (2021), 23. https://doi.org/10.1186/s12920-020-00869-9

25. K. Shah, H. Patel, D. Sanghvi, M. Shah, A comparative analysis of logistic regression, random forest and KNN models for the text classification, *Augmented Human Res.*, **32** (2020), 11055–11064. https://doi.org/10.1007/s41133-020-00032-0

26. L. Ogiela, M. R. Ogiela, U. Ogiela, Efficiency of strategic data sharing and management protocols, in *IMIS 2016-The 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2016)*, (2016), 198–201. https://doi.org/10.1109/IMIS.2016.119

27. M. R. Ogiela, L. Ogiela, U. Ogiela, Biometric Methods for Advanced Strategic Data Sharing Protocols, *IMIS 2015 - The Ninth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2015)*, (2015), 179–183. https://doi.org/10.1109/IMIS.2015.29

AIMS Press