JOURNAL of COMPUTER SCIENCES INSTITUTE

JCSI 27 (2023) 132–137
Received: 13 November 2022
Accepted: 7 December 2022

Comparison of shallow and deep learning methods of ECG signals classification for arrhythmia detection

Porównanie metod płytkiego i głębokiego uczenia do klasyfikacji sygnałów EKG zastosowanych do wykrywania arytmii

Dodon Turianto Nugrahadi*, Rudy Herteno, Dwi Kartini, Muhammad Haekal, Mohammad Reza Faisal

Department of Computer Science, Lambung Mangkurat University, Banjarbaru 70714, Indonesia

Abstract

The research aims to compare the classification performance of arrhythmia classification from the ECG signal dataset from the Massachusetts Institute of Technology–Beth Israel Hospital (MIT-BIH) database. This study uses shallow learning methods: Support Vector Machine, Naïve Bayes, and Random Forest. 1D Convolutional Neural Network (1D CNN), Long Short Term Memory (LSTM), and Gated Recurrent Unit (GRU) are deep learning methods that are used for the study. The models are tested on a dataset with 140 samples that are grouped into four class labels. Each sample has 2160 features. Those models are tested for classification performance. This research shows Random Forest and 1D CNN have the best performance.

Keywords: ECG signals; arrhythmia classification; shallow learning; deep learning

Streszczenie

Badanie ma na celu porównanie wydajności klasyfikacji arytmii na podstawie zestawu danych sygnału EKG z bazy danych Massachusetts Institute of Technology – Beth Israel Hospital (MIT-BIH). W pracy zastosowano następujące metody: Support Vector Machine, Naïve Bayes i Random Forest. Ponadto wykorzystano następujące metody głębokiego uczenia: 1D Convolutional Neural Network (1D CNN), Long Short Term Memory (LSTM) oraz Gated Recurrent Unit (GRU). Modele zostały przetestowane na zbiorze danych zawierającym 140 próbek pogrupowanych w cztery etykiety klas. Każda próbka zawierała 2160 cech. Przeprowadzone testy wydajności klasyfikacji wskazały, że Random Forest i 1D CNN wykazują najwyższą wydajność.

Słowa kluczowe: sygnałów EKG; klasyfikacja arytmii; płytkie uczenie; głębokie uczenie

*Corresponding author

Email address: dodonturianto@ulm.ac.id (D. T. Nugrahadi)

1. Introduction

Nowadays, signal classification methods play an essential role in a wide variety of life. Examples of the implementation of this method are for the classification of music genres and moods [1], the classification of machine conditions [2], and the classification of signals in the health field [3], [4].

One of the signal classification cases in the health sector is the electrocardiogram (ECG) signal classification. ECG is a signal that describes the electrical activity carried out by the heart and is very important in diagnosing heart rhythm disturbances caused by changes in electrical impulses in the heart. This disorder is also known as arrhythmia [4]. One way to help detect heart rhythm disturbances early on automatically can be done with the help of deep learning. This paper compares deep learning methods of arrhythmia classification using ECG signals.

Currently, many studies have been carried out that utilize machine learning or artificial intelligence to recognize ECG patterns and then automatically predict heart condition categories. Machine learning techniques based on shallow learning for heart disease classification are used, such as Random Forest [5], Support Vector Machine (SVM) [6], K-Nearest Neighbors (KNN) [7], Naïve Bayes [8]. The deep learning technique used for the classification of cases of heart disease is Convolutional Neural Network (CNN) [6], Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) [4][6].

In research on the classification of heart disease, there are two treatments for ECG data. The first treatment using raw ECG data was done by [9]. In this study, the selected raw ECG data were classified using the shallow learning type classification method, namely SVM. The second treatment is to analyze the raw ECG signal by transforming the signal. One of the transformation techniques used is the wavelet transform [10]. The raw ECG signal usually consists of different frequency components and even noise, making it difficult for classification algorithms to learn to determine patterns. So the wavelet transform is used to eliminate these things. However, humans recognize arrhythmic patterns from visualization of the raw ECG signal. In addition, the classification of raw ECG signals can lighten the computer's work because it does not perform signal transformation calculations.

Based on this explanation, a comparative study was conducted on classifying raw ECG signals with shallow and deep learning algorithms. The goal is to determine which shallow and deep learning algorithms can perform better classification. In addition, this study also aims to determine which classification performance is better between algorithms based on shallow learning and deep learning.

This report is divided into five section: 1) Introduction, 2) Dataset and method section explains the dataset and classification algorithms that are used, 3) Research implementation section explains the steps in implementation of this study, 4) Result, and the last section is 5) Conclusions.

2. Dataset and methods

2.1. Dataset

The raw ECG signal dataset used in the study is from research [9]. With details as can be seen in Tabel 1. The source of the research data came from the arrhythmia database of the Massachusetts Institute of Technology–Beth Israel Hospital (MIT-BIH). Six seconds of cardiac records resulted in 2160 data points.

Arrhythmias cause the heart to beat faster, slower, or irregularly. The condition causes symptoms such as feeling tired and pain in the chest. To detect arrhythmias, doctors use a heart record or electrocardiogram. An electrocardiogram (EKG) is a recording of heart activity obtained by attaching an electrode to the skin to capture the electric current generated by the heart. A series of heart activities recorded by the EKG can be used as an indicator of a heart rhythm disorder, which doctors or nurses can use to take appropriate action.

Table 1: Raw ECG dataset

No	Class Label	Number of features	Number of samples
1	Normal	2160	35
2	Atrial Fibrillation	2160	35
3	PVC Bigeminy	2160	35
4	Ventricular Tachycardia	2160	35
Total			140

A normal sample is regular heart rate ECG data, also known as normal sinus rhythm. A normal sample can be visualised in Figure 1, where the heart is beating in a regular sinus rhythm between 60 - 100 beats per minute (specifically 82 bpm).

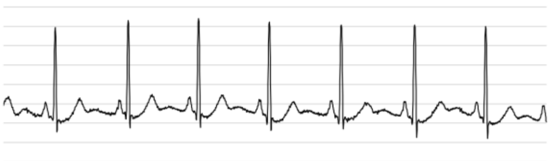

Figure 1: Normal.

Atrial Fibrillation is a type of arrhythmia. In this condition, a rapid heartbeat often causes poor blood flow. The heart's upper chambers (atria) beat out of coordination with the lower chambers (ventricles). These conditions may have no symptoms, but when they do they include palpitations, shortness of breath, and fatigue. Atrial Fibrillation visualization is shown in Figure 2.

Extra heartbeats known as premature ventricular contractions (PVCs) begin in one of the heart's two lower pumping chambers (ventricles). The regular heart rhythm is disrupted by this extra beat, which can sometimes make the chest feel like it is pounding or jumping up and down. This irregular heartbeat is also known as PVC Bigeminy, which is visualized in Figure 3.

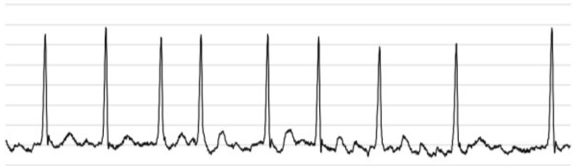

Figure 2: Atrial Fibrillation.

Ventricular Tachycardia is a condition in which the heart's lower chambers (ventricles) beat very fast. This irregular heartbeat ECG is shown in Figure 4.

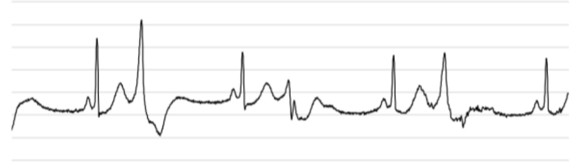

Figure 3: PVC Bigeminy.

This condition occurs because of a problem with the heart's electrical impulses. This condition can develop as a complication of a heart attack, or it can occur in people with certain conditions, such as valvular heart disease.

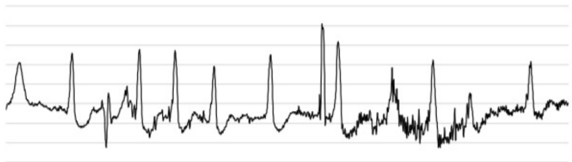

Figure 4: Ventricular Tachycardia.

2.2. Methods

There are two type learning that are used in this study: shallow learning and deep learning.

Shallow learning is a type of machine learning that learns from data with predetermined features extracted from features. In this study, shallow learning algorithms that are used are Support Vector Machine, Naïve Bayes, and Random Forest.

Vapnik introduced the Support Vector Machine (SVM) in 1992 as a machine learning method that works on the Structural Risk Minimization (SRM) principle, which aims to find a hyperplane to separate two classes in the input space. This method uses hypotheses as linear functions in feature space with high dimensions by implementing a learning bias derived from statistical learning theory. The level of accuracy in the model that the switching process will produce with

SVM depends on the kernel functions and parameters. [6].

The Naive Bayes algorithm is a classifier that uses the probabilistic and statistical methods proposed by the British scientist Thomas Bayes to predict future potentials based on past experiences. The main feature of Naïve Bayes is a solid assumption for the independence of each condition of an event. Calculations for comparison of new cases with old cases are indexed based on input parameters with mathematical calculations using probabilities that occur for cases considered similar (Bayesian model). The advantage of Naïve Bayes is that it does not require a lot of data to estimate the parameters needed for classification [8].

Random Forest has many decision trees, which can be hundreds or even thousands of trees, which predict individual classes. The Random Forest aims to form a single representative decision from multiple decision trees. The majority vote of the entire tree can be defined as the result of class predictions [5]. The structure of the Random Forest can be seen in Figure 5

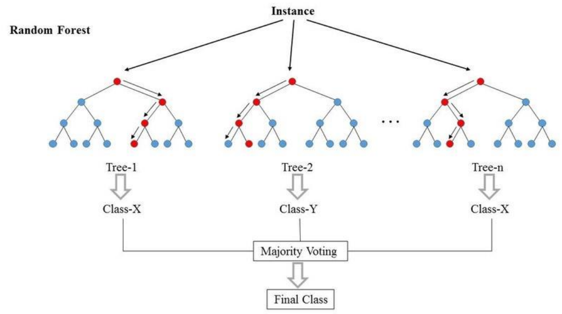

Figure 5: Structure of Random Forest [11].

In contrast to shallow learning, deep learning gets features from data from the feature extraction process, which is carried out automatically without human intervention. Deep learning algorithms automatically learn their features and weights, allowing deep learning to use the best features to get the best classification performance. Deep learning algorithms that are used in this study are 1D Convolutional Neural Network (1D CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU).

Wiesel and Hubel first carried out a Convolutional Neural Network (CNN) regarding the virtual cortex in the sense of sight. CNN is an architecture that can be trained and consists of several stages, inputs and outputs. CNN trains and tests each input data through a series of processes, namely the convolutional layer, followed by pooling to extract features from successive input data. The next step is the pooling process. In this process, the data is flattened and then entered into the fully connected-layer process to complete the classification process [12]. Usually, CNN is used to process input data in images as two-dimension data. CNN is also known as 2D CNN. Figure 6 shows an example of a 2D CNN architecture for 2D data.

In 2D CNN, the convolutional layer will produce two-dimensional data. CNN can also be used for one-dimensional data cases, for example, signals including ECG. For the case of one-dimensional data, CNN is known as 1D CNN. In 1D CNN, the convolutional layer will run in one dimension and produce one-dimensional data [14].

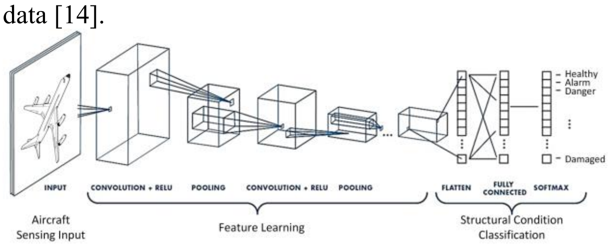

Figure 6: 2D CNN architecture [13].

Examples of 1D architecture for signal cases can be seen in Figure 7.

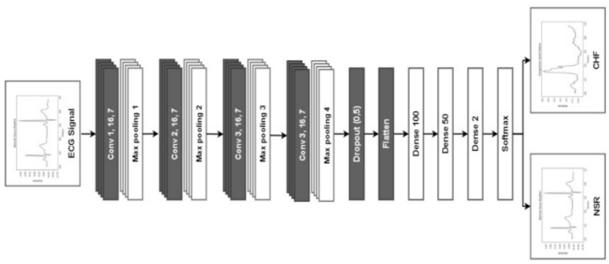

Figure 7: 1D CNN architecture [14].

Recurrent Neural Network (RNN) is an artificial neural network for processing sequential data such as speech recognition and language modelling. Long Short-Term Memory (LSTM) is a variation of RNN that was created to avoid the problem of long-term dependency on RNN [4]. LSTM can remember long-term information, and LSTM also has an iterative processing model like RNN. In the LSTM, there is a path that connects the old context to the new context or what is also known as the cell state, memory cell or memory path. The path makes the values in the old context easily linked to the new context if needed with minor modifications. The LSTM can delete or add information to memory lanes governed by the sigmoid function [6].

LSTM has four gates: forget gate, input gate, cell gate, and output gate. Forget gate is a gate that decides whether information should be removed or not from processing. The input Gate is a gate that decides to determine whether an input will be added to the cell gate memory. The cell gate is a gate that functions as a memory for a layer; this gate's ability is to remember long-term information. The output gate is a gate whose function is to decide what will be produced based on the input and cell gate.

Gated Recurrent Unit (GRU) is an architecture created in 2014 by Kyunghun Cho. GRU provides faster convergence, and the results can be compared with LSTM. GRU also has a smaller error value compared to LSTM. The purpose of the GRU is to make each recurrent unit so that it can capture every relationship (dependency) in different time scales carried out adaptively. The GRU has a component that regulates the flow of information called the gate, and the GRU here has two gates, namely, the reset gate and the update gate [15].

3. Research implementation

The steps in the implementation of the research can be seen in Figure 8.

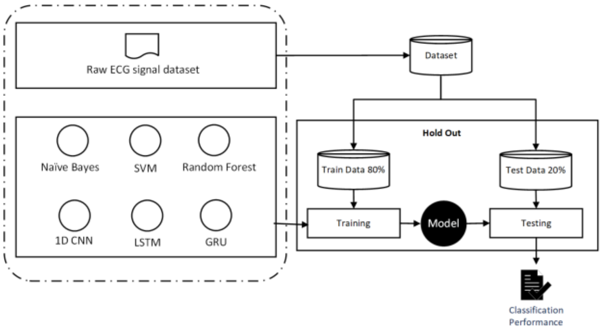

Figure 8: Steps of research implementation.

Based on the picture above, it can be seen that the raw ECG signal data is divided into 80% training data and 20% as test data. Furthermore, the training data will be used in the training process using the classification algorithm. The classification algorithms used in this study is as follows:

- Shallow learning-based classification algorithms are Naïve Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF).
- Classification algorithm based on deep learning, namely 1D Convolutional Neural Network (1D CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU).

From Figure 8, we conduct six experiments to make six classification models. Then each model will be tested using test data. So there are six performance classifications. The research questions raised in the Introduction section can be answered by comparing these performances.

Implementation of shallow learning algorithms using the Python programming language with the Scikit-Learn library. In Table 2 it can be seen the parameters used for each shallow learning algorithm.

Table 2: Shallow learning: algorithm's parameters

Algorithm	Parameters
NB	prior=none var_smoothing=1e-09
SVM	C=1.0 kernel=rbf degree=3 gamma=scale coef0=0.0
RF	n_estimators=100 criterion=gini max_depth=none min_sample_split=2 min_sample_leaf=1 mint_weight_fraction_leaf=0.0 max_features=sqrt bootstrap=True

The implementation of the deep learning's algorithm uses the Python programming language with the Keras and Tensorflow libraries. The architecture and parameters used in each algorithm can be seen in Table 3.

Table 3: Deep learning: algorithm's parameters

Algorithm	Parameters
1D CNN	1. Convolutional layers: Conv1D • Filter = 256 • Kernel size = 3 • Kernel = orthogonal • Bias = glorot_uniform • Regularizer = regularizers. L1L2(l1=1e-5, l2=1e-4) • activation = relu 2. BatchNormalization layers 3. Pooling layer: MaxPool1D 4. Convolutional layers: Conv1D • Filter = 128 • Kernel size = 3 • Kernel = orthogonal • Bias = glorot_uniform • Regularizer = regularizers. L1L2(l1=1e-5, l2=1e-4) • activation = relu 5. BatchNormalization layers 6. Pooling layer: MaxPool1D 7. Convolutional layers: Conv1D • Filter = 64 • Kernel size = 3 • Kernel = orthogonal • Bias = glorot_uniform • Regularizer = regularizers. L1L2(l1=1e-5, l2=1e-4) • activation = relu 8. BatchNormalization layers 9. Pooling layer: MaxPool1D 10. Convolutional layers: Conv1D • Filter = 64 • Kernel size = 3 • Kernel = orthogonal • Bias = glorot_uniform • Regularizer = regularizers. L1L2(l1=1e-5, l2=1e-4) • activation = relu 11. BatchNormalization layers 12. Pooling layer: MaxPool1D 13. Flatten layer 14. Dense layer • unit=64 • activation = relu 15. Dropout layer 16. Dense layer • unit=32

Algorithm	Parameters
	• activation = relu optimizer=Adam loss=categorical_crossentropy
GRU	1. CuDNNGRU layers • unit = 128 2. BatchNormalization layers 3. CuDNNGRU layers • unit = 64 4. BatchNormalization layers 5. CuDNNGRU layers • unit = 32 6. BatchNormalization layers 7. Dense layer • unit=32 • activation = relu 8. Dropout layer 9. Dense layer • unit=16 • activation = relu 10. Dense layer • unit=4 • activation = softmax optimizer=Adam loss=categorical_crossentropy
LSTM	1. CuDNNLSTM layers • unit = 128 2. BatchNormalization layers 3. CuDNNLSTM layers • unit = 64 4. BatchNormalization layers 5. CuDNNLSTM layers • unit = 32 6. BatchNormalization layers 7. Dense layer • unit=32 • activation = relu 8. Dropout layer 9. Dense layer • unit=16 • activation = relu 10. Dense layer • unit=4 • activation = softmax optimizer=Adam loss=categorical_crossentropy

4. Results

The results of the six experiments conducted in this study can be seen in Table 4. These results show that the highest accuracy for the shallow learning algorithm is 60%, namely by using the Random Forest algorithm. As for the deep learning algorithm, the 1D CNN algorithm works better than other deep learning algorithms with a classification performance of 60%.

Table 4: Classification performance of each model

Learning Type	Methods	Accuracy (%)
Shallow Learning	SVM	50
	NB	35
	RF	60
Deep Learning	1D CNN	60
	LSTM	45
	GRU	30

The classification performance comparison chart can be seen in Figure 9. As explained in section 2.1, Datasets, samples are high dimension data with 2160 features. The Introduction section also mentions that raw ECG data may contain noise and other different frequency components that can interfere with the classification algorithm for pattern recognition. Those problems can cause the low performance of classifications such as SVM, NB, LSTM and GRU.

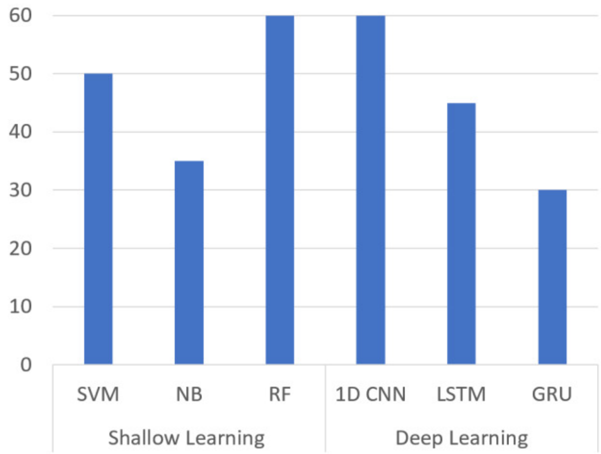

Figure 9: Comparison of classification performance.

Although the classification performance of RF and 1D CNN cannot be categorized as good, these two algorithms can work better than other algorithms. It is because of the characteristics of RF and 1D CNN.

The characteristic of RF is to create many decision trees that take features randomly. With these characteristics, RF can select the features that can provide the best classification performance. Therefore, the RF performance is better than other shallow learning classification algorithms.

The 1D CNN algorithm has a convolutional layer that functions as a filter which becomes the parameter that is updated in the learning process, and an optimum value is obtained that can provide the best classification performance. This is why the performance of the 1D CNN algorithm is better than the other deep learning algorithms used in this study.

5. Conclusions

From the results of this study, it can be concluded that the shallow learning type classification algorithm that can provide the best classification performance on raw ECG signals is Random Forest. In comparison, the 1D CNN algorithm is the best for deep learning classification algorithm types. Both of them produce the same classification performance, which is 60%, so the capabilities of the shallow and deep learning classification algorithms work equally well.

The following research will focus on optimizing the 1D CNN algorithm to get the classification performance of raw ECG signals to predict arrhythmias better. The way we are going to do this is to try out different 1D CNN architectures. In addition, hyperparameter tuning and other fine-tuning will be carried out to produce a better classification model.

6. Acknowledgements

In this research, the computer system's computation time was provided by the Data Science Lab of the Computer Science Department, Faculty of Mathematics and Natural Sciences, Lambung Mangkurat University. This work was supported by the Program Dosen Wajib Meneliti (PDWM) grant from PNBP Lambung Mangkurat University.

References

[1] A. Shakya, B. Gurung, M. S. Thapa, M. Rai, B. Joshi, Music classification based on genre and mood, International Conference On Computational Intelligence, Communications, and Business Analytics (2017) 168-183, https://doi.org/10.1007/978-981-10-6430-2_14.

[2] R. F. R. Junior, F. A. D. Almeida, G. F. Gomes, Fault classification in three-phase motors based on vibration signal analysis and artificial neural networks, Neural Computing and Applications 32(18) (2020) 15171-15189, https://doi.org/10.1007/s00521-020-04868-w.

[3] M. K. Delimayanti, B. Purnama, N. G. Nguyen, M. R. Faisal, K. R. Mahmudah, F. Indriani, M. Kubo, K. Satou, Classification of brainwaves for sleep stages by high-dimensional FFT features from EEG signals, Applied Sciences 10(5) (2020) 1797-1809, https://doi.org/10.3390/app10051797.

[4] Z. Ebrahimi, M. Loni, M. Daneshtalab, A. Gharehbaghi, A review on deep learning methods for ECG arrhythmia classification, Expert Systems with Applications X 7 (2020), https://doi.org/10.1016/j.eswax.2020.100033.

[5] M. Kropf, D. Hayn, G. Schreier, ECG classification based on time and frequency domain features using random forests. In 2017 Computing in Cardiology (CinC) (2017) 1-4, https://doi.org/10.22489/CinC.2017.168-168.

[6] O. M. A. Ali; S. W. Kareem, A. S. Mohammed, Evaluation of electrocardiogram signals classification using CNN, SVM, and LSTM algorithm: A review, In 2022 8th International Engineering Conference on Sustainable Technology and Development (IEC) (2022) 185-191, https://doi.org/10.1109/IEC54822.2022.9807511.

[7] C. Venkatesan, P. Karthigaikumar, R. J. M. T. Varatharajan, A novel LMS algorithm for ECG signal preprocessing and KNN classifier based abnormality detection, Multimedia Tools and Applications 77(8) (2018) 10365-10374, https://doi.org/10.1007/s11042-018-5762-6.

[8] B. Krithiga, P. Sabari, I. Jayasri, I. Anjali, Early detection of coronary heart disease by using naive bayes algorithm, In Journal of Physics: Conference Series Vol. 1717 No. 1 IOP Publishing (2021) 012-040, http://doi.org/10.1088/1742-6596/1717/1/012040.

[9] R. A. Cahya, C. Dewi, B. Rahayudi, Arrhythmia classification from electrocardiogram results using support vector machine with feature selection using genetic algorithms, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (2018) 1170 – 1178.

[10] T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, C. Ou, Automatic ECG classification using continuous wavelet transform and convolutional neural network, Entropy 23(1) (2021) 119-132, https://doi.org/10.3390/e23010119.

[11] S. I. Dimitriadis, D. Liparas, How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease, Neural Regeneration Research 13(6) (2018) 962-970, https://doi.org/10.4103%2F1673-5374.233433.

[12] M. Dovbnych, M. P. Wójcik, A comparison of conventional and deep learning methods of image classification, Journal of Computer Sciences Institute (2021) 303-308, https://doi.org/10.35784/jcsi.2727.

[13] I. Tabian, H. Fu, Z. S. Khodaei, A convolutional neural network for impact detection and characterization of complex composite structures, Sensors, 19(22) (2019) 4933-4958, https://doi.org/10.3390/s19224933.

[14] D. Li, J. Zhang, Q. Zhang, X. Wei, Classification of ECG signals based on 1D convolution neural network, 19th International Conference on e-Health Networking Applications and Services (Healthcom) (2017) 1-6, https://doi.org/10.1109/HealthCom.2017.8210784.

[15] H. M. Lynn, S. B. Pan, P. Kim, A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks, IEEE Access 7 (2019) 145395-145405, https://doi.org/10.1109/ACCESS.2019.2939947.