2017 IEEE 10th International Workshop on Computational Intelligence and Applications
November 11-12, 2017, Hiroshima, Japan

MACD-Histogram-based Recurrence Plot: A New Representation for Time Series Classification

Keiichi Tamura
Graduate School of Information Sciences,
Hiroshima City University,
3-4-1, Ozuka-Higashi, Asa-Minami-Ku,
Hiroshima 731-3194, Japan
E-mail: ktamura@hiroshima-cu.ac.jp

Takumi Ichimura
Department of Management and Systems,
Prefecture University of Hiroshima,
1-1-71, Ujina-Higashi, Minami-ku,
Hiroshima 734-8559, Japan
ichimura@pu-hiroshima.ac.jp

***Abstract*—Time series classification is the most active research topics in time series data mining, because they cover a broad range of applications in many different domains. There are three important things that we need to consider in time series classification; representation, similarity measurement, and assignment strategy. Representation for time series is a technique that converts time series to feature vectors representing the characteristics of time series. Chaotic time series analysis have been well-studied; Moreover, recurrence plotting underlying chaos theory is one of the most robust the feature expression for time series. In this study, we propose new representation for the time series classification utilizing the recurrence plot technique. Moving average convergence divergence (MACD) histogram is the acceleration of time that represents the features of time series. Therefore, the proposed method is based on MACD histogram. In particular, a recurrence plot that is made from MACD histogram is called a MACD-Histogram-based recurrence plot (MHRP). The recurrence plot is referred to as a gray-scale image and we utilize stacked auto-encoders as a classier for MHRPs. To evaluate the performance of the classifier with the MHRP technique, we implemented it and conducted experiments using the UCR time series classification archive. The experimental results showed that the proposed classifier outperforms not only distance-based 1-NNs, but also our previous MACD-Histogram-based method.**

***Index Terms*—Time series classification, Time series mining, Recurrence plot, Chaotic time series analysis, MACD histogram**

I. Introduction

A time series is a sequence of measurements that are observed temporal changes in a phenomenon at regular intervals. Time series are now-ubiquitous data, because they cover the broad range application areas in natural science, such as bioinformatics, engineering, geoscience, astronomy, medicine, and agriculture [1], [2]. There are two types of time series: univariate and multivariate time series. In this study, we focus on univariate time series, which are a sequence of primitive items (e.g., real numbers, integer values, or symbols), including stock prices, biomedical signals, electrocardiogram values, earthquake waves, radio waves, and event streams. Time series mining, which includes forecasting, clustering, outlier detection, classification, frequent pattern and motif extraction, and visualization, has attracted numerous researchers and practitioners in many different application domains, because they have been basic techniques for the broad range applications handling time series [3], [4], [5], [6].

Since time series classification [7] has practical applications, it has been the most active research topics in time series mining. For example, similar biosignals are usually related to the same biological activity. If we can classify unclassified biosignals effectively, biological signals can be accurately recognized. There are three important things that we need to consider in time series classification; representation, similarity measurement, and assignment strategy. Representations for time series are techniques that convert a time series to a feature vector representing the characteristics of time series. Chaotic time series analysis [8] have been well-studied; moreover, recurrence plotting underlying chaos theory is one of the most robust representation for time series. A recurrence plot of a time series is a square matrix in which the value of each element in the matrix corresponding to those times at which a state of a dynamical system recurs [9].

In this study, we propose a new representation for the time series classification utilizing the recurrence plot technique and a new classifier for recurrence plots. Moving average convergence divergence (MACD) histogram [10] is the acceleration of time that represents the latent features of time series. In our previous work [11], [12], a high-level symbolic representation for time series involving MACD histogram showed good classification performance; therefore, this motivated us to develop a new recurrence plot. In particular, a recurrence plot of a time series is created from MACD histogram, not from original values in the time series. This recurrence plot is called a MACD-Histogram-based recurrence plot (MHRP). To classify time series, the proposed classifier utilizes stacked auto-encoders [13], because a recurrence plot is a square matrix, which is referred to as a gray-scale image.

To evaluate the proposed classifier, we implemented the classifier and experiments were conducted by using the UCR time series classification archive [14]. The experimental results show that the proposed classifier is good classification performance compared with our previous methods. The rest of the paper is organized as follows. In Section II, related work is described briefly. In Section III, some preliminaries for the proposed classifier are explained. In Section IV, MHRP

978-1-5386-0469-4/17/$31.00

and the proposed classifier are proposed. In Section V, the experimental results are shown and we discuss the method's performance. We conclude the paper in Section VI.

II. Related Work

There are three major approaches to time series classification: distance-based, feature-based, and model-based [7] approaches. The distance-based approaches define distance functions, measures the distance between time series, and classifies the time series by referring to the mutual distance. The feature-based approach convert time series to feature vectors and classifies time series according to the similarities between feature vectors. In the model-based approach, we are trying to apply statistical model analysis to time series classification.

Distance-based approaches have been thoroughly studied, and many studies have reported that 1-NN is the simplest yet the most stable algorithm. The early studies were based on the Euclidean distance; however, the Euclidean distance is not robust against small gaps between time series and its shape difference. To address this problem, the dynamic time warping (DTW) distance was proposed [15]. DTW improves the performance of time series classification dramatically. Ding et al. [3] reported that 1-NN with DTW, in general, performs well, and the difference between its performance and that of other subsequent distance metrics is small.

Shapelets [16], [17] are one of the most well-known techniques for feature-based and model-based approaches. Shapelets are segments of time series that identify class efficiently. They are extracted by evaluating the class prediction qualities of numerous candidates extracted from the series segments. Since SAX(Symbolic Aggregate approXimation) [18]was proposed, researchers have focused on the feature-based approach using SAX. SAX-VSM [19] is a state-of-the-art algorithm based on SAX and the "bag of words" model. Each class is represented by a feature vector and the feature vector is weighted by TF*IDF weighting. An unlabeled time series is assigned to a class in which the unlabeled time series has the highest feature score.

The proposed classifier is a kind of the feature-based approach. MHRPs generated from time series represent features of time series and MHRPs are classified instead of time series. The assignment strategy is based on neural networks. As attention to deep learning has increased, some researchers reviewed classification methods utilizing neural networks. Cui et al. [20] proposed multi-scale convolutional neural networks for time series classification. Wang et al. developed a representation for time series called GASF/GADF-MTF and proposed a classifier using deep convolutional neural network. Umeda [21] proposed a new time series representation extracted by the structure of the attractor using topological data analysis.

III. Preliminaries

In this section, some preliminaries for MHRP and the proposed classifier are explained.

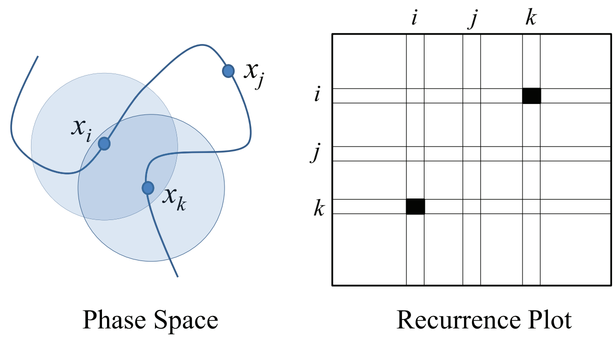

Fig. 1. Example of Recurrence Plot.

A. Recurrence Plot

A recurrence plot is a two-dimensional plotting figure, where if the distance $dist(x_i, x_j)$ between the data points x_i and x_j of data point series $X = <x_1, x_2, \cdots, x_n>$ in a phase space is smaller than a threshold r, the figure has a plotting point at the coordinates (i, j). The recurrence plot visualizes the latent features in time series by extracting sets of data points close in the phase space and plotting points on the two-dimensional space. Fig. 1 shows an example of a recurrence plot. In this figure, since data points x_i and x_k are located within a radius r, there are two plotting points at the coordinates (i, k) and (k, i), whereas there is no plotting points at the coordinates (j, k), (k, j), (i, j), and (j, i).

The recurrence plot technique is one of the most robust non-linear time series analysis based on chaotic theory. If a time series has cyclically pattern, the recurrence plot of the time series has cyclically patterns on it. Conversely, if a time series has aperiodic change, characteristic patterns that represents aperiodic change are shown in the recurrence plot of the time series. Since recurrence plots are visualized by gray-scale images and are easy to understand intuitively, we can sketch out background features that generate the same time series.

Let the i-th time series in a time series data set TS be $T_i = <t_{i,1}, t_{i,2}, \cdots, t_{i,n}>$. In this study, T_i is a simple time series, where each value is a primitive value such as a real number. Time delay coordinates are used for mapping time series to data points in the d-dimensional phase space. In the time delay coordinates space, values that are sampled every arbitrary delay time τ are extracted as a state value that is a data point in k-dimensional coordinate $at_j(d, \tau, T_i) = (t_{i,j}, t_{i,j+\tau}, t_{i,j+2\tau}, \cdots, t_{i,j+(d-1)\tau})$. The trajectory drawn by a sequence of extracted state values from a time series is called an attractor. The attractor is a sequence of extracted state values:

$$AT(d, \tau)[T_i] = < at_j(d, \tau, T_i) \mid j = 1, 2, \cdots > . \quad (1)$$

For example, given parameters $\tau = 2$ and $d = 2$, a series of coordinates $AT(2, 2)[T_i] = < (t_{i,1}, t_{i,3}), (t_{i,2}, t_{i,4}), \cdots, (t_{i,n-2}, t_{i,n}) >$ is a attractor of T_i in the 2-dimensional phase space.

For short, $AT(d, \tau)[T_i]$ is denoted by $AT[T_i]$. Let the recurrence plot of T_i be $RP(T_i)$, and (l, k)-element of $RP(T_i)$

is defined as:

$$PR(T_i)_{l,k} = \begin{cases} 1 & dist(AT[T_i]_l, AT[T_i]_k) < r, \\ 0 & (otherwise), \end{cases} \tag{2}$$

where $AT[T_i]_l$ and $AT[T_i]_k$ are the l-th and the k-th elements of $AT(d,\tau)[T_i]$, respectively. A parameter r is a threshold for determining approximation. The function $dist$ returns the Euclidean distance.

B. MACD Histogram

The MACD and its histogram are defined as the velocity and the acceleration of a time series. These criteria are used for the technical analysis of stock prices, which provides the indicator of stock trading. A series of stock prices is referred to as a time series; therefore, the chances of profiting from trading a stock can be determined by analyzing the time series of stock prices. In the theory, an object moves in a two-dimensional space and a time series is regarded as trajectories of its two-dimensional positions. Velocity and acceleration of the object are calculated using the observed changes in position.

The exponential moving average (EMA) is a type of weight moving average known as an exponentially weighted moving average. The definition of EMA for the t-th element of T_i is

$$\begin{aligned} EMA(ws)[T_i]_t &= \gamma \times t_{i,t} + (1-\gamma)EMA[T_i]_{t-1} \\ &= \sum_{k=0}^{ws}(\gamma(1-\gamma)^k t_{i,(t-k)}), \end{aligned} \tag{3}$$

where ws is the window size, and $\gamma = 2/(ws-1)$. The parameter ws is the length of the window. The weighting for each older data item decreases exponentially. Suppose that $t = k$. This implies that the average is calculated using the $(k-ws)$-th to the k-th element.

Let the EMA sequence of T_i under ws be $EMA(ws)[T_i]$. Given two EMA sequences $EMA(ws_1)[T_i]$ and $EMA(ws_2)[T_i]$, where $ws_1 < ws_2$, the difference between $EMA(ws_1)[T_i]_t$ and $EMA(ws_2)[T_i]_t$ is called the MACD:

$$\begin{aligned} &MACD(ws_1, ws_2)[T_i]_t = \\ &EMA(ws_1)[T_i]_t - EMA(ws_2)[T_i]_t,\ ws_1 < ws_2. \end{aligned} \tag{4}$$

The MACD is considered to be a derivative value of the EMA and is referred to as a velocity of time series. The EMA of the MACD, where the size of window is ws_3, is called the signal.

$$\begin{aligned} &SIG(ws_1, ws_2, ws_3)[T_i]_t = \\ &\qquad EMA(ws_3)[MACD(ws_1, ws_2)[T_i]]_t. \end{aligned} \tag{5}$$

The difference between the signal and the MACD is called the MACD histogram. The MACD histogram is a derivative value of the MACD and is regarded as the acceleration of the time series.

$$\begin{aligned} &HIST(ws_1, ws_2, ws_3)[T_i]_t = \\ &\quad MACD(ws_1, ws_2)[T_i]_t - SIG(ws_1, ws_2, ws_3)[T_i]_t \end{aligned} \tag{6}$$

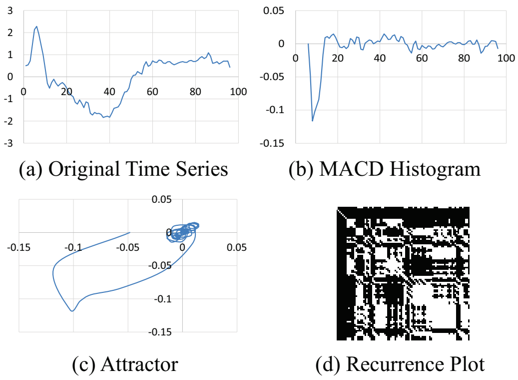

Fig. 2. Example of MHRP.

IV. Proposed Method

In this section, a new representation MHRP for time series classification and the proposed classifier are proposed.

A. MACD-Histogram-based Recurrence Plot

The MACD histogram of the i-th time series T_i is denoted by $HIST(ws_1, ws_2, ws_3)[T_i] = <ht_{i,1}, ht_{i,2}, \cdots, ht_{i,n}>$. It is, hereinafter, referred to as $HIST[T_i]$. Given $HIST[T_i]$ of T_i, the attractor $HAT(d,\tau)[HIST[T_i]]$ is defined as follows.

$$\begin{aligned} &HAT(d,\tau)[HIST[T_i]] = \\ &\qquad < hat_{i,t}(d,\tau,HIST[T_i]) \mid t = 1, 2, \cdots >, \end{aligned} \tag{7}$$

$$\begin{aligned} &hat_{i,t}(d,\tau,HIST[T_i]) = \\ &\qquad (ht_{i,t}, ht_{i,t+\tau}, ht_{i,t+2\tau}, \cdots, ht_{i,t+(d-1)\tau}) \end{aligned} \tag{8}$$

For short, $HAT(d,\tau)[HIST[T_i]]$ is denoted by $HAT[HIST[T_i]]$. The definition of MHRP of T_i $MHRP(T_i)(= RP(HIST[T_i])$ is as follows.

$$\begin{aligned} &MHRP(T_i)_{l,k} = \\ &\begin{cases} 1 & dist(HAT[HIST[T_i]]_l, HAT[HIST[T_i]]_k) < r \\ 0 & (otherwise) \end{cases} \end{aligned} \tag{9}$$

Fig. 2 shows an example of MHRP. A time series and its MACD histogram are shown in Fig. 2-(a) and Fig. 2-(b), respectively. In Fig. 2-(c) shows the attractor of the MACD histogram in the two-dimensional phase space. MHRP of the attractor extracted from the MACD histogram is shown in Fig. 2-(d). In our previous work [11], [12], we confirmed similar time series belonging to different classes are distinguished by using their MACD histograms through experiments. Therefore, MHRP has the has discrimination ability for time series classification.

B. Classifier using Stacked Auto-Encoders

The proposed classifier is based on stacked auto-encoders. An autoencoder is a kind of neural network used for unsupervised learning. The main propose of learning on the autoencoder is to learn a representation for a data set. Stacked auto-encoders are a neural network consisting of multiple layers of auto-encoders, with the output of each layer wired to the input of the continuous layer. Stacked auto-encoders with

TABLE I
DATA SETS

Name	Number of classes	Size of training set	Size of testing set	Time series Length
CBF	3	30	900	128
DistalPhalanxOuthlineAgeGroup	3	139	400	80
DistalPhalanxOuthlineCorrect	2	276	600	80
DistalPhalanxTW	6	139	400	80
ECG	2	100	100	96
ECG5000	5	500	4500	140
ECGFiveDays	2	23	861	136
ElectricDevices	7	8926	7711	96
Face(all)	14	560	1690	131
FacesUCR	14	200	2050	131
ItalyPowerDemand	2	67	1029	24
MedicalImages	10	381	760	99
MiddlePhalanxOuthlineAgeGroup	3	154	400	80
MiddlePhalanxOuthlineCorrect	2	291	600	80
MiddlePhalanxTW	6	154	399	80
MoteStrain	2	20	1252	84
PhalangesOuthlinesCorrect	2	1800	858	80
Plane	7	105	105	144
ProximalPhalanxOuthlineAgeGroup	3	400	205	80
ProximalPhalanxOuthlineCorrect	2	600	291	80
ProximalPhalanxTW	6	205	400	80
SonyAIBORobotSurface	2	20	601	70
SonyAIBORobotSurfaceII	2	27	953	65
SwedishLeaf	15	500	625	128
SyntheticControl	6	300	300	60
TwoLeadECG	2	23	1139	82
TwoPatterns	4	1000	4000	128

num_l hidden layers are usually built by using unsupervised layer-wise pre-training.

Fig. 3 shows the training process of the proposed classifier. Let the training data set $TRD = \{(TRD_1, cl_1), (TRD_2, cl_2), \cdots, (TRD_n, cl_n)\}$, where $cl_i \in CL$ is the class label of time series TRD_i. Given the number of layers num_l and a list of the number of hidden layer's units, the detail of the training process is as follows.

(1) For each time series TRD_l in TRD, the MACD histogram of TRD_l $HIST[TRD_l]$ is extracted.
(2) For each MACD histogram $HIST[TRD_l]$ is converted to $HAT(d,\tau)[HIST[TRD_l]] = HAT[HIST[TRD_l]]$, where d is the number of dimensions in the phase space, and τ is the delay parameter.
(3) For each $HAT[Hist[TRD_l]]$, MHRP of it $MHRP(TRD_l)$ is generated.
(4) The stacked auto-encoders with num_l hidden layers are built using unsupervised layer-wise pre-training.
(5) All $num_l + 1$ layers are combined together to form stacked auto-encoders with num_l hidden layers and a final layer that is softmax classifier layer capable of classifying time series. The fine tuning is conducted on the combined network using the supervised post-training.

C. Classification using SAEs

The process of classifying an un-labeled time series is as follows. Given a unlabed time series ULT and the classifier.

(1) The MACD histogram of the un-labeled time series $HIST[ULT]$ is extracted.
(2) $HIST[ULT]$ is converted to $HAT(d,\tau)[HIST[ULT]] = HAT[HIST[ULT]]$, where d is the number of dimensions in the phase space, and τ is the delay parameter.
(3) The MHRP of $HAT[HIST[ULT]]$ $MHRP(TRD_l)$ is generated.
(4) The classifier outputs classification probabilities, where $MHRP(ULT)$ is input data.

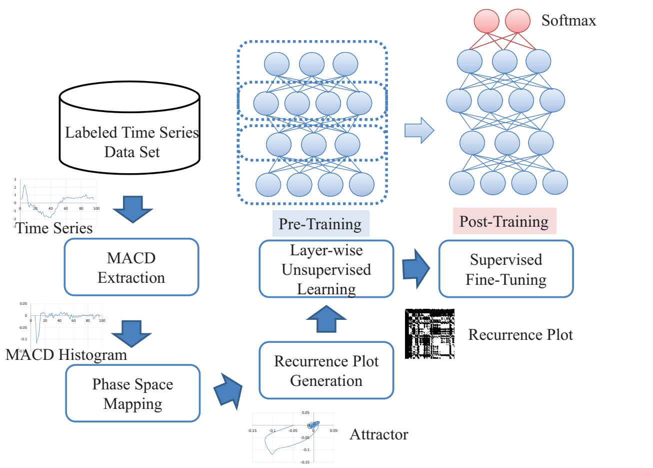

Fig. 3. Training Process of Proposed Method.

V. EXPERIMENTS

To evaluate the proposed classifier, we conducted experiments using the UCR time series classification archive [14], which is the largest available time series classification benchmark data set. This archive includes 85 types of labeled time series data sets with a variety of lengths, class numbers, and data sizes. Each data set is divided into two types of data sets: a training data set, and a test data set. In the experiments, we used 27 data sets in UCR Time Series Classification Archive, where the length of each time series is less that 150. Table I

TABLE II
ERROR RATES OF TRAINING DATA SETS AND TEST DATA SETS

	Error Rates of Training Data			Error Rates of TEST Data		
Name	(128, 64, 32)	(256, 128, 64)	(512, 256, 128)	(128, 64, 32)	(256, 128, 64)	(512, 256, 128)
CBF	0.000	0.000	0.000	0.367	0.383	0.329
DistalPhalanxOutlineAgeGroup	0.014	0.000	0.014	0.138	0.135	0.130
DistalPhalanxOutlineCorrect	0.029	0.007	0.011	0.158	0.163	0.167
DistalPhalanxTW	0.101	0.000	0.000	0.178	0.193	0.188
ECG	0.000	0.000	0.000	0.080	0.060	0.060
ECG5000	0.006	0.000	0.000	0.066	0.067	0.065
ECGFiveDays	0.000	0.000	0.000	0.004	0.000	0.001
ElectricDevices	0.304	0.305	0.246	0.368	0.388	0.366
Face(all)	0.016	0.000	0.000	0.255	0.272	0.266
FacesUCR	0.020	0.000	0.000	0.175	0.192	0.145
ItalyPowerDemand	0.000	0.000	0.000	0.029	0.045	0.029
MedicalImages	0.210	0.050	0.050	0.411	0.399	0.380
MiddlePhalanxOutlineAgeGroup	0.110	0.007	0.013	0.190	0.190	0.188
MiddlePhalanxOutlineCorrect	0.024	0.010	0.014	0.187	0.193	0.163
MiddlePhalanxTW	0.136	0.000	0.020	0.341	0.336	0.343
MoteStrain	0.000	0.000	0.000	0.121	0.130	0.120
PhalangesOutlinesCorrect	0.123	0.108	0.089	0.176	0.190	0.162
Plane	0.000	0.000	0.000	0.000	0.000	0.000
ProximalPhalanxOutlineAgeGroup	0.068	0.033	0.040	0.107	0.107	0.107
ProximalPhalanxOutlineCorrect	0.088	0.068	0.071	0.100	0.096	0.093
ProximalPhalanxTW	0.088	0.005	0.039	0.190	0.170	0.190
SonyAIBORobotSurface	0.000	0.000	0.000	0.185	0.260	0.161
SonyAIBORobotSurfaceII	0.000	0.000	0.000	0.215	0.219	0.195
SwedishLeaf	0.056	0.000	0.000	0.066	0.082	0.053
SyntheticControl	0.067	0.000	0.000	0.430	0.413	0.427
TwoLeadECG	0.000	0.000	0.000	0.006	0.008	0.005
TwoPatterns	0.011	0.000	0.000	0.636	0.619	0.619

shows 27 data sets. For each data set, error rates of classifying test data set were measured.

The least error rates were found by varying the following parameters: $\tau \in \{1, 2, 3\}$ and $r \in \{0.03, 0.025, 0.02, 0.015, 0.01, 0.005, 0.001\}$. In the experiments, we used three types of five-layer stacked auto-encoders. The numbers of units in hidden layers are $(128, 64, 32)$, $(256, 128, 64)$, and $(512, 256, 128)$. The number of epochs is 100 in both pre-training and post-training. The drop rate is 0.5, the optimizing algorithm is Adadelta, and the activation function in hidden layers is ReLU.

Table II shows the error rates of training data sets and test data sets when the numbers of units in hidden layers is changed to $(128, 64, 32)$, $(256, 128, 64)$, and $(512, 256, 128)$. The error rates show the lowest value obtained when changing the parameters. For training data sets, the error rates are almost zero, indicating that high classification performance is obtained. As for the test data sets, since there is difficulty in classifying the UCR Time Series Classification Archive itself, it is necessary to compare them with other methods, but not much change is seen even if the numbers of units in hidden layers changes.

Table III shows the results of comparing the proposed classifier with other methods. The proposed method was compared with four types of 1-NN classifiers: *EQ 1-NN*, *BWW DTW 1-NN*, *DTW 1-NN*, and *MHSAX*. Underlined values indicate the lowest error rate. The *EQ 1-NN* classifier utilizes the Euclidean distance and the *BWW DTW 1-NN*, and *DTW 1-NN* classifiers employ the DTW distance. *MHSAX* is 1-NN classifier that was proposed in our previous work [12]. The proposed method obtains the lowest error rates of the tested methods for 17 out of 27 data sets. The performance of three data sets, *CBF*, *SyntheticControl*, and *TwoPatterns* show poor performance compared with other methods. These three data sets include symmetric patterns belonging to different classes. To distinguish these kind of time series, MHRP needs to involve raw values of time series.

VI. CONCLUSION

In this study, we propose new representation for the time series classification utilizing the recurrence plot technique. The MACD histogram is the acceleration of time that represents the features of time series. Therefore, the proposed method is based on MACD histogram. In particular, a recurrence plot that is made from MACD histogram is called a MACD-Histogram-based recurrence plot (MHRP). The recurrence plot is referred to as a gray-scale image and we utilize stacked auto-encoders as a classier for MHRPs. To evaluate the proposed classifier, we implemented the classifier and experiments were conducted by using the UCR time series classification archive . The experimental results show that the proposed classifier is good classification performance compared with our previous MACD-Histogram-based method. In our future work, we are planing to utilize other deep network models to improve the classification performance and modify MHRP to distinguish symmetric patterns.

TABLE III
COMPARISONS OF ERROR RATES

	EQ 1-NN	*BWW DTW 1-NN*	*DTW 1-NN*	*MHSAX*	*MHRP*
CBF	0.148	0.004	**0.003**	0.04	0.329
DistalPhalanxOutlineAgeGroup	0.218	0.228	0.208	0.205	**0.13**
DistalPhalanxOutlineCorrect	0.248	0.232	0.232	0.231	**0.158**
DistalPhalanxTW	0.273	0.272	0.29	0.262	**0.178**
ECG	0.12	0.12	0.23	0.09	**0.06**
ECG5000	0.075	0.075	0.076	**0.064**	0.065
ECGFiveDays	0.203	0.203	0.232	0.105	**0**
ElectricDevices	0.45	0.376	0.399	**0.322**	0.366
Face(all)	0.286	0.192	0.192	**0.211**	0.255
FacesUCR	0.231	0.088	0.095	**0.039**	0.145
ItalyPowerDemand	0.045	0.045	0.05	0.048	**0.029**
MedicalImages	0.316	**0.253**	0.263	0.359	0.380
MiddlePhalanxOutlineAgeGroup	0.26	0.253	0.25	0.24	**0.188**
MiddlePhalanxOutlineCorrect	0.247	0.318	0.352	0.258	**0.163**
MiddlePhalanxTW	0.439	0.419	0.416	0.388	**0.336**
MoteStrain	0.121	0.134	0.165	**0.107**	0.120
PhalangesOutlinesCorrect	0.239	0.239	0.272	0.258	**0.162**
Plane	0.038	**0**	**0**	**0**	**0**
ProximalPhalanxOutlineAgeGroup	0.215	0.215	0.195	0.17	**0.107**
ProximalPhalanxOutlineCorrect	0.192	0.21	0.216	0.161	**0.093**
ProximalPhalanxTW	0.292	0.263	0.263	0.23	**0.17**
SonyAIBORobotSurface	0.305	0.305	0.275	0.251	**0.161**
SonyAIBORobotSurfaceII	0.141	0.141	0.169	**0.136**	0.195
SwedishLeaf	0.211	0.154	0.208	0.083	**0.053**
SyntheticControl	0.12	0.017	**0.007**	0.06	0.413
TwoLeadECG	0.253	0.132	0.096	0.034	**0.005**
TwoPatterns	0.090	0.002	**0.000**	0.002	0.619

ACKNOWLEDGMENT

This work was supported by the MIC/SCOPE #162308002 and Hiroshima City University Grant for Special Academic Research.

REFERENCES

[1] M. Last, A. Kandel, and H. Bunke, *Data Mining in Time Series Databases*, ser. Series in machine perception and artificial intelligence. World Scientific, 2004.

[2] P. Esling and C. Agon, "Time-series data mining," *ACM Comput. Surv.*, vol. 45, no. 1, pp. 12:1–12:34, Dec. 2012.

[3] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, "Querying and mining of time series data: Experimental comparison of representations and distance measures," *Proc. VLDB Endow.*, vol. 1, no. 2, pp. 1542–1552, Aug. 2008.

[4] S. Basu and M. Meckesheimer, "Automatic outlier detection for time series: An application to sensor data," *Knowl. Inf. Syst.*, vol. 11, no. 2, pp. 137–154, Feb. 2007.

[5] M. Shokoohi-Yekta, Y. Chen, B. Campana, B. Hu, J. Zakaria, and E. Keogh, "Discovery of meaningful rules in time series," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: ACM, 2015, pp. 1085–1094.

[6] J. Lin, E. Keogh, S. Lonardi, J. P. Lankford, and D. M. Nystrom, "Visually mining and monitoring massive time series," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '04. New York, NY, USA: ACM, 2004, pp. 460–469.

[7] Z. Xing, J. Pei, and E. Keogh, "A brief survey on sequence classification," *SIGKDD Explor. Newsl.*, vol. 12, no. 1, pp. 40–48, Nov. 2010.

[8] J. C. Sprott, *Chaos and Time-Series Analysis*. Oxford University Press, 2003.

[9] N. Marwan, "A historical review of recurrence plots," *The European Physical Journal Special Topics*, vol. 164, no. 1, pp. 3–12, Oct 2008.

[10] J. J. Murphy and J. J. Murphy, *Technical analysis of the financial markets*. Fishkill, N.Y.: New York Institute of Finance, 1999.

[11] K. Tamura, T. Sakai, and T. Ichimura, "Time series classification using macd-histogram-based sax and its performance evaluation," in *Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 2419–2424.

[12] K. Tamura and T. Ichimura, "Mhsax-based time series classification using local sequence alignment technique," in *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2017*, 2017, pp. 286–291.

[13] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Dec. 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1756006.1953039

[14] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: A survey and empirical demonstration," *Data Min. Knowl. Discov.*, vol. 7, no. 4, pp. 349–371, Oct. 2003.

[15] E. J. Keogh and M. J. Pazzani, "Scaling up dynamic time warping for datamining applications," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '00. New York, NY, USA: ACM, 2000, pp. 285–289.

[16] L. Ye and E. Keogh, "Time series shapelets: A new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '09. New York, NY, USA: ACM, 2009, pp. 947–956.

[17] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme, "Learning time-series shapelets," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '14. New York, NY, USA: ACM, 2014, pp. 392–401.

[18] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing sax: A novel symbolic representation of time series," *Data Min. Knowl. Discov.*, vol. 15, no. 2, pp. 107–144, Oct. 2007.

[19] P. Senin and S. Malinchik, "Sax-vsm: Interpretable time series classification using sax and vector space model," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, Dec 2013, pp. 1175–1180.

[20] Z. Cui, W. Chen, and Y. Chen, "Multi-scale convolutional neural networks for time series classification," *CoRR*, vol. abs/1603.06995, 2016. [Online]. Available: http://arxiv.org/abs/1603.06995

[21] Y. Umeda, "Time series classification via topological data analysis," *Transactions of the Japanese Society for Artificial Intelligence*, vol. 32, no. 3, pp. 1–12, 2017.