Correlation and Symmetrical Uncertainty-Based Feature Selection for Multivariate Time Series Classification

Ahmad Saikhu[1]*   Agus Zainal Arifin[1]   Chastine Fatichah[1]

[1]*Department of Informatics, Institut Teknologi Sepuluh Nopember, Indonesia*
* Corresponding author's Email: saikhu@if.its.ac.id

Abstract: Data relating to everyday phenomena can be recorded in the form of time series. Recently, modeling of time-series has become a topic of interest in data mining. Conducting an analysis of multivariate time series effectively is an essential task for decision-making activities in fields such as meteorology, medicine, and finance. Features selection is a key problem in the analysis of multivariate time series. Rainfall prediction, biomedical classification, pattern recognition, sensor network analysis and so on all have different input features. The problem is that these features have interdependencies and time-delay relationships. Currently, research on the selection of input features of these data still depends on whether they are linear or non-linear. In this paper, we propose a new integration strategy between Pearson Correlation and Symmetrical Uncertainty for relevant feature selection based on linear and non-linear relationships for multivariate time-series classification. We evaluated the goodness of fit of feature subsets using merit value. The meteorological data set was used to test the proposed method. The result showed that the method was able to reduce the number of features by 77.9% features and increase their merit value 2.25 times compared to no input features selection.

Keywords: Multivariate time series, Input selection, Irrelevant, Redundant, Linear and non-linear relationship, Merit value.

1. Introduction

Time-series data are familiarly used in diverse fields such as meteorology, medicine, finance, business, and sensor networking. Time-series modeling is required for prediction, clustering, and classification for decision-making. The data are taken from numerous observations at different time points, as in the case of multivariate time series (MTS). The MTS data often have a very high dimensionality. Prediction and classification of time series have been widely discussed in many fields and have recently attracted much attention in data mining research. Classifying a highly dimensional data poses a challenge because of the need to reduce the vast number of features.

Several MTS Classification (MTSC) methods have been developed, including distance-based classification to determine the distance between two sequences of time series [1]. A second method is a feature-based classification by transforming the data sequence to a feature set prior to classification based on features availability in the MTS. Meanwhile, another approach is model-based MTSC without class labels.

In this research, we analyzed MTS using a features-based approach. This approach was chosen because it has a simple and fast filter implementation. Feature selection, i.e. keeping the relevant features and removing redundant ones.

MTSC is very important in data mining, for example for season classification in the field of climatologic/meteorology, classification of normal/abnormal ECG/EEGs, or classification of product quality in manufacturing. Generally, MTSC problems are different from data mining classification. For example, the values of the attributes are ordered by time, there are tens or even hundreds of variables, and the length of data samples is different. Therefore, the traditional

approach of machine-learning for MTSC is problematic because all do to its tens/hundreds of features that could not be broken into univariate time series issues result in loss relationship between features [2].

There are several MTSC methods, among others instance-based classification [3] feature-based classification [4], and the model-based classification [5]. The focus of the first classification method is to predict the test sample based on similarity with training data. In general, MTS classifier uses 1-nearest-neighbour, which is computed Euclidean distance and dynamic time warping (DTW). Meanwhile, feature-based MTSC is based on a number of variable predictors. Model-based classification is a classification determined by the specific interaction patterns between object classes for the MTS data.

In MTS, there are commonly up to tens or hundreds of input features for classification, part of which are irrelevant and redundant. For computational efficiency, only relevant features should be selected. Redundancy among features can occur for either linear relationships, non-linear relationships, or both. Hence, the necessity of characteristics identification of the relationships between features to ensure that the selected features are relevant and free of redundancy.

There are several methods for selecting relevant input features and removing redundant features in classification. Some are statistical based, such as correlation methods, chi-square and relief-F [6–8]. Another method is based on mutual information (MI) by utilizing information gain and entropy [9–11]. Statistical methods identify the relevance and redundancy of a feature based on the linear relationships between features, while MI identifies the relevance and redundancy of a feature based on its non-linear relationships.

One research implemented a feature selection method called Class Separability Feature Selection on a MTS using trace-based class separability criterion [12]. Another research proposed a feature selection method by using supervised classification within an information-theoretic framework [13], whereby MI is exploited for measuring the statistical relationships between a subset of features and the class labels of the samples.

Two variables x and y have a linear relationship if any change in the value of x tends to be followed by y with a constant ratio [13, 14]. If the linear relationship x and y is plotted in a scatter diagram, the result will be a straight-line pattern. The relationships between input features and the relationships between input features and class labels

can also be non-linear, so this condition is called 'non-linear relevant'.

On the other hand, a relationship between variables x and y is non-linear if a change in the value of x tends to be followed by y at a non-constant ratio [10, 15]. The relevance and redundancy of these two types of relationships must be identified before feature selection for MTSC.

Several researches investigated feature-based MTSC, for example by utilizing the feature subset selection for pre-processing and applying data mining techniques for classification [16, 17]. In [4], a number of research feature-based MTSC methods to reduce the dimensions of MTS data with a different length of the period are compared.

In [18] it is successfully demonstrated that the classification of feature-based MTS is faster than other methods. In [19] this is confirmed, where pre-processing was used to select relevant features and remove redundant features before classification. However, MTSC is a complex problem, because the predictors can have many features, there can be many relationships between features, and they depend on the time variable.

In fact, feature selection based on non-linear relationships does not always successfully identify relationships between features that are linear [20]. Conversely, feature selection based on linear relationships can fail to identify non-linear relationships. In this study, we used two approaches for selecting relevant features and removing redundant features. The main idea of this approach is to identify good features based on both linear and non-linear relationships. For linear relationships, the Pearson Correlation (PC) was used for relevant feature selection and redundant feature removal. For non-linear relationships, Symmetrical Uncertainty (SU) was applied. Both methods are needed since ones cannot replace the other. Therefore, both methods need to be integrated in order to obtain complete predictors, both those which have linear and non-linear patterns.

Hence, in this paper, we propose a new strategy for selection of relevant input features from MTS based on linear as well as non-linear relationships using PC and SU namely Pearson Correlation and Symmetrical Uncertainty Filter (PCSUF) to produce a new features subset. These feature subsets are measured by the goodness of fit, which is called merit value.

The PCSUF method provides the improvement of the performance of the MTS model. The PCSUF method was applied to the rainfall classification and obtains the good merit-value, therefore, it can be applied for other MTS data. We also compare the

PCSUF method to the Correlation-based Feature Selection (CFS) method.

The rest of this paper is organized as follows. Section 2 introduces time series and relevant/redundant features and discuss our methods, the algorithm used and case a study. Section 3 discusses the results and their analysis. Conclusions are drawn in section 4.

2. The material and methods

This section describes time series theory, methods for selecting relevant features and removing redundant features, the data used for the case study, and merit value.

2.1 Time series

A time series is a time-based sequences of data objects that have relationships with each other, written as: $X_i(t)$; $[i = 1, 2, 3,.., n, t = 1, 2, 3, .., m]$. If the value of $n > 2$, known as MTS, while $n = 1$ is called a univariate time series (UTS). MTSC predicts the values of the attribute (class) targets based on a number of attribute predictors of the observed values of each temporal attribute (sequential) with the same time interval. The attributes can be inter-dependent or dependent between on the time variable.

In the analysis of time series data, another important element that must be included is the time-delay. This factor occurs between variables or between variables and class labels. For MTSC, it is necessary to formulate the input of variable time-delay using the Cross-Correlation Function (CCF) technique as expressed in Eq. (1) [21], where $r_{xy}(T)$ is the correlation between variable x and variable y at time period T and σ_{xy} is covariance(x,y), $\sigma_{xx}(0)$ is variance(x) and $\sigma_{yy}(0)$ is variance(y).

$$r_{xy}(T) = \frac{\sigma_{xy}(T)}{\sqrt{\sigma_{xx}(0)}\sqrt{\sigma_{yy}(0)}} \qquad (1)$$

If we have the relationship between two time-series, y_t and x_t, we can also identify the relationship between the series y_t to past lags of the series x_{t+h} using a sample CCF. x_{t+h} is variable x with a lag of h. If x_{t+h} correlates significantly with response variable y_t, then it can be a predictor for y. Lags h can be either positive or negative where a negative value of h represents the value of variable x in the previous h period, which will be correlated with the value y at time t.

The CCF value gives the correlation between x_{t-2} and y_t. When one or more x_{t+h}, with negative h, are predictors of y_t, it is sometimes said that x leads y. We can test the hypothesis that the two series x_t and y_t are cross-correlated by comparing the sample CCF $r_{xy}(T)$ with their approximate standard error:

$$1/\sqrt{n-k} \qquad (2)$$

2.2 Relevant and redundant features

We need a definition for features that are originally assumed relevant predictors for the class label. In [22], the feature set is divided into three subsets of features that are mutually exclusive, i.e. strongly relevant, weakly relevant, and irrelevant features. The first subset must exist and be maintained as a predictor of optimal features. It cannot be removed because this would affect the outcome of the prediction of the class labels. The second subset is less necessary as a predictor feature, except in certain conditions. The last subset is not required as a predictor feature in any condition. An optimal predictor is composed of all the strongly relevant features, most of the weakly relevant features and none of the irrelevant features.

Creating the relevant feature subset still faces another problem, namely redundancy between features. This occurs because of a significant correlation between two features.

A formal definition of redundant features is required in order to identify and design a method to remove these redundant features. Features that are weakly relevant should be separated into features that are redundant and those that are not redundant. Therefore, we need an efficient way to obtain optimal features predictors (see Fig. 1 (c) and (d)).

We evaluate relevant and irrelevant features in a subset as a linear relationship using Pearson's Correlation (PC) as expressed in Eq. (3) [23], where ρ_{xy} is a correlation between variable x and variable y. ρ_{xy} declared significant if its value is higher than a critical value in the PC table with a certain degree of freedom (DOF) and significant level (SL).

$$\rho_{xy} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \qquad (3)$$

For evaluating a non-linear relationship between variables and between variables and class labels, we use symmetrical uncertainty (SU). SU is formulated
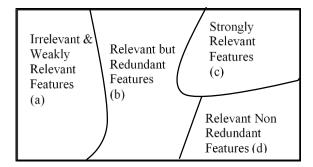
Figure. 1 Irrelevant, relevant and redundant features

based on entropy and information gain If we have variable *x*, then the entropy is defined as expressed in Eqs. (4) and (5) [24], where $P(x_i)$ is the prior probabilities for all values of *X*, and $P(x_i/y_i)$ is the posterior probability of *X* given the values of *Y*.

$$H(X) = \sum_i P(x_i) log(P(x_i)) \qquad (4)$$

$$H(X|Y) = -\sum_j P(y_j) \sum_i P(x_i|y_j) log\left(P(x_i|y_j)\right) \qquad (5)$$

The entropy of variable *X* after observing the values of another variable *Y* is defined as expressed in Eq. (6). The amount by which the entropy of *X* decreases provides additional information about *X* provided by *Y*.

$$IG(X|Y) = H(X) - H(X|Y) \qquad (6)$$

However, information gain is biased in favor of features with more values. Furthermore, feature values have to be normalized to ensure that they are comparable and have the same effect. Therefore, we choose SU (Eq. (7)), which is suitable for discrete data. If the data is continuous, it is necessary to execute the process of discretization.

$$SU(X,Y) = 2 \left[\frac{IG(X|Y)}{H(X) + H(Y)} \right] \qquad (7)$$

SU(X, Y) is declared significant if its value is higher than a critical value from the Chi-Square table with a certain DOF and SL.

In this research, we used merit-value to measure the performance of PCSUF feature selection. There are several previous studies using merit-value to measure the optimality of CFS performance, such as drought modeling [25] that achieve merit-value 0.5-0.9 with selected features from 24 to 15, epileptic classification [26] and forest-fire modeling [27] that

achieve the high performances of the constructed models.

The advantages of using merit-value measure are that execution is faster than other methods, especially on large numbers of features and independent of the classifier [28, 29]. Merit-value is calculated by Eq. (8) where if the merit value is higher, then the feature subset is better [30].

$$r_{z,C} = \frac{k\bar{r}_{z,i}}{\sqrt{k+k(k-1)\bar{r}_{i,i}}} \qquad (8)$$

$r_{z,C}$ is the merit value between an input feature and a class label. $\bar{r}_{z,i}$ is the average correlation among inter-correlations, that is the correlation between predictors and class labels and $\bar{r}_{i,i}$ is the average correlation among intra-correlations, that is the correlation among predictors, and k is the number of features.

2.3 Algorithm

The proposed method for relevance and redundancy analysis can be implemented by using the PCSUF) algorithm. The PCSUF algorithm is a
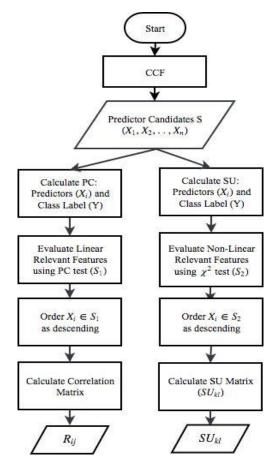

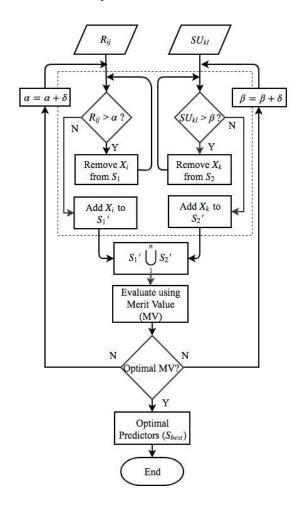Figure. 2 Selection of relevant features

Figure. 3 Removal of redundant features

further development of the FCBF algorithm [22] and ECMBF algorithm [31]. Figs. 2 and 3 show that it consists of three phases:

a. Select the relevant features subset based on linear and non-linear relationships between features and class labels (Fig. 2).
b. Remove the subset of redundant features based on linear relationships (left-side of Fig. 3) and remove the redundant non-linear relationships (right-side of Fig. 3).
c. Try various combinations of values for α and β to get the S best ones (Fig. 3).

2.4 Case study: climate data

Rainfall modeling is an important problem that conducted in many fields, such as agriculture, water resources management, hydrology and facility maintenance and control. Several previous works applying numerical, statistical and soft computing approaches were unable to provide accurate models [32].

An important component in identifying patterns is a measurement of climate variables, i.e. air temperature, relative humidity, air pressure, water evaporation, wind speed, sea surface temperature, and duration of solar radiation [32]. Research related to climate modeling and prediction of rainfall, especially for the region of Indonesia, still use partial climatic variables that do not completely describe the relationships between climate factors.

In this study, we add global variables and take into account their relationship with time-lag. The addition of input variables makes the model more comprehensive but inefficient in the computing process. Therefore input variable selection by eliminating redundant predictors must be done [33].

Local observation data were obtained from weather observation station class I "Juanda" of the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG), located in East Java (Coordinates: S 07°22'38", E 112°22'38", WMO ID: 96935). This weather station observes an area of 30 km^2 in Surabaya and surroundings.

The Local features recorded by this station are daily rainfall, relative humidity (RH), minimum/ mean/maximum temperature, and solar radiation (SR). The daily observation period covers 11 years, January 1, 2006 until December 31, 2016. Fig. 4 shows a time series plot for average minimum temperature /month, average solar radiation and humidity.

Global variable data (climate indices) was obtained from The Royal Netherlands Meteorological Institute (KNMI), available online (http://climexp.knmi.nl). It Includes 6 variables: Nino3, Nino4, Nino34, sea-surface temperature (SST), air pressure (AP) and Southern Oscillation Index (SOI). The monthly observation period cover are 11 years, January 2006 until December 2016.

Nino34 stands for SST anomalies in the Equatorial Pacific 5°S to 5°N latitude and 170°W to 120°W longitude; Nino3 stands for SST anomalies in the Pacific over the region 5°S to 5°N latitude, 150°W to 90°W longitude; Nino4 stands for SST anomalies in the Pacific(150°W-160°E and 5°N-5°S); SOI stands for differences in surface air pressure between Tahiti and Darwin; SST stands for anomalies in The Java Sea and anomalies in the Indian Ocean south of Java.

Because we wanted to evaluate the relationships between features and class labels in a monthly period, we first had to aggregate daily data to monthly data. In order to get monthly rainfall, we accumulated the daily data for each month. For monthly maximum/minimum temperature, we chose a maximum/minimum value of each month. For monthly mean temperature, humidity and solar radiation, we averaged the daily data for each month.
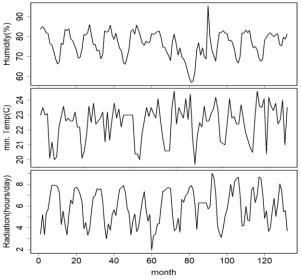
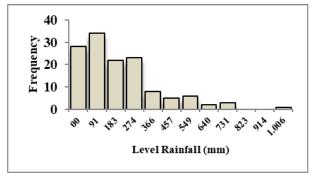Figure. 3 Min. temperature, humidity and solar radiation, Jan 2006-Dec 2016

Figure. 5 Histogram of monthly rainfall in Surabaya

Class labels were created by discretization of monthly rainfall. If the monthly rainfall value was between 0 and 100 mm, the class label is 0 (no rain/low rainfall rate) and if the rainfall value was > 100 mm, the class label is 1 [34]. Fig. 5 explains the characteristics of monthly rainfall.

3. Result and discussion

Fig. 6 shows the input feature selection process for the meteorological data set. As the first step, for the identification and generation of time-lag features, we used CCF on the local/global variables and the rainfall class label. Fig. 7 shows the CCF correlogram between rainfall class and minimum temperature (min_T). This figure shows that CCF produced 8 time-delay features for min_T that were significantly correlated with rainfall because the cross-correlations were higher than UCL/LCL (0.1/-0.1). We took only positive lags from CCF correlogram.
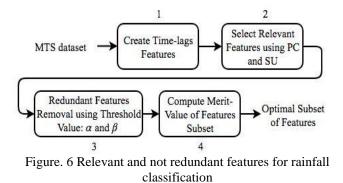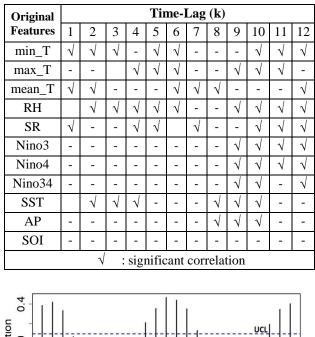

Figure. 6 Relevant and not redundant features for rainfall classification

Table 1. Significant time lag features

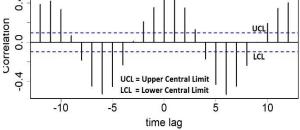| Original Features	Time-Lag (k)											
	1	2	3	4	5	6	7	8	9	10	11	12
min_T	√	√	√	-	√	√	-	-	-	√	√	√
max_T	-	-	√	√	√	-	-	√	√	√	-
mean_T	√	√	-	-	-	√	√	√	-	-	-	√
RH		√	√	√	√	√	-	-	√	√	√	√
SR	√	-	-	√	√		√	-	-	√	√	√
Nino3	-	-	-	-	-	-	-	-	√	√	√	√
Nino4	-	-	-	-	-	-	-	-	√	√	√	√
Nino34	-	-	-	-	-	-	-	-	√	√	-	√
SST		√	√	√	-	-	-	√	√	√	-	-
AP	-	-	-	-	-	-	-	√	√	√	-	-
SOI	-	-	-	-	-	-	-	-	-	-	-	-
√ : significant correlation												

Figure. 7 CCF of rainfall and minimum temperature

All results of CCF for time-delay features with a significant correlation are shown in Table 1. There were 57 significant time-delay features so that the total number of input variables was 68 features, including five local features and six global features as predictors.

In the second step, we selected the linear and non-linear relevant input features using PCSUF with SL = 0.05 and DOF = 130 in the PC table so that the significant correlation was $\rho_{xy} > 0.178$. We got 48 linear relevant features. We also identified the non-linear features that were relevant by using SU.

Table 2. Merit value and feature number

Methods	Merit Value (Features Number)			
	All Features	PC	SU	PCSUF
Original Features	0.295 (11)	-	-	-
No Feature Selection	0.296 (68)	-	-	-
Select Relevant	0.650 (49)	0.653 (48)	0.457 (17)	-
Select Relevant & Redundant Removal	-	0.341 (8)	0.667 (9)	**0.666 (15)**

The significant features were found for SU > 0.103 with DOF = 2 and SL= 0.05 at χ^2 table. We got 32 linear relevant features, 17 non-linear relevant features and 16 were intersection linear and non-linear relevant features, as shown in Fig. 8.

In the third step, we removed the redundant features using PCSUF by trying out various α and β values from α = 0.2; β = 0.2 to α = 0.7; β = 0.7 with δ = 0.1. Furthermore, the goodness of fit of the results of relevant feature selection and redundant feature removal were measured using merit value.

In the fourth step, to evaluate the efficiency of the selection mechanism of the input variables, we measured the merit value. That measurement was used for two conditions: with feature selection and without feature selection for classification of rainfall. Table 2 presents the results of measuring the merit value and the number of selected features.

We tested the new strategy on six different scenarios, i.e. computing the merit value by using the 11 original features as the first scenario and without feature selection as the second scenario. The third scenario used relevant feature selection without redundant feature removal. The fourth scenario involved feature selection and redundant feature removal only with the PC method for different α values.

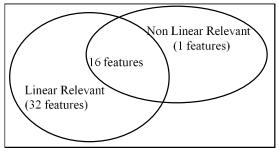Meanwhile, in the fifth scenario, we used feature

Figure. 8 Relevant features for rainfall classification
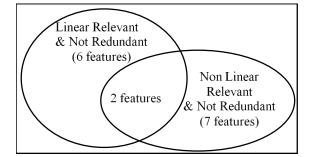

Figure. 9 Relevant and not redundant features for rainfall classification

selection and redundant feature removal for non-linear relationships by using the SU method with various β values. In the last scenario used feature selection and redundant feature removal with PCSUF. The best results for the PCSUF scenario and the five other scenarios are shown in Table 2.

When no feature selection was applied (first and second scenario), then the resulting feature set was not effective, i.e. the merit value was low.

Using the third scenario, the resulting feature set was quite effective, with merit value increased twice, but not efficient enough because the number of input features was reduced by 29% from the original number of input features. The resulting feature set of the third scenario was insufficiently effective, with the merit value lower than in the second scenario.

In our case study, the PCSUF method produces selected features from 68 to 15 as shown Fig. 9 with merit-value 0.666 compared to the CFS method [25] obtain selected features from 68 to 8 with merit-value 0.341. Based on the literature [25], the acceptable of merit-values is 0.5-0.9. Therefore, PCSUF method obtains better merit-value than CFS method.

4. Conclusions

The process of identifying relationship patterns between input features and class labels or among input features is preferred before MTSC modeling. This is done in order to determine the appropriate method for feature selection, either PC, SU or PCSUF. Due to the PC method cannot fully identify the patterns of non-linear relationships, while the SU method cannot fully identify to the patterns of linear relationships.

A new integration strategy, PCSUF, for selecting relevant features based on linear and non-linear relationships was successfully implemented. This method produced a number of efficient input features in a MTSC problem. In a case study using climatological MTS data from "Juanda" weather

station, PCSUF managed to reduce the number of input features to 22.1%. Using only 15 input features out of 68 features, its merit value was increased 2.25 times.

Compared with no elimination of redundant features and feature selection based on linear relationships, PCSUF was more efficient in terms of the number of input features and more effective in terms of merit value.

Comparing PCSUF with SU, particularly for effectiveness, the results were relatively similar, i.e. they both detected all relationship patterns between features and class labels for both linear and non-linear relationships.

For the time-delay feature, CCF was still used, which assumes a linear time dependence. In future research, we aim to develop non-linear time dependencies to create the time-delay feature. Next, we need general goodness of fit to measure optimal input feature selection applying to both linear and non-linear associations. This is because the merit value only measures the results of feature selection based on linear relationships.

Acknowledgments

References

[1] K. Vasimalla, "A Survey on Time Series Data Mining", *International Journal of Innovative Research in Computer and Communication Engineering,* Vol. 2, No. 5, pp. 170–179, 2014.

[2] T. Górecki and M. Łuczak, "Multivariate time series classification with parametric derivative dynamic time warping", *Expert System with Application*, Vol. 42, No.5, pp. 2305–2312, 2015.

[3] J. Zhao, S. Member, and L. Itti, "Descriptors with Hybrid Sampling", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 3, 2016.

[4] B. D. Fulcher and N. S. Jones, "Highly comparative feature-based time-series classification", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 26, No. 12, pp. 3026–3037, 2014.

[5] A. Kotsifakos and P. Papapetrou, "Model-based Time Series Classification", In: *Proc. of The Thirteenth International Symposium on Intelligent Data Analysis*, pp. 1–12, 2014

[6] W. Yuan, D. Guan, L. Shen, and H. Pan, "An Empirical Study of Filter-based Feature Selection Algorithms Using Noisy Training Data", In: *Proc. of 4th IEEE International Conference on Information Science and Technology*, pp.210-212, 2014.

[7] B. Tang, H. He, P.M. Baggenstoss and S. Kay, "A Bayesian Classification Approach Using Class-Specific Features for Text Categorization", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 6, pp. 1602–1606, 2016.

[8] J. Tang, S. Alelyani, and H. Liu, "Feature Selection for Classification: A Review, in Feature selection for classification: A review. In: Data Classification: Algorithms and Applications", *Book Section*, CRC Press, pp. 37–64, 2014.

[9] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data", *Knowledge and Information Systems*, Vol. 34, No. 3, pp. 483–519, 2013.

[10] R. J. May, H. R. Maier, G. C. Dandy, and T. M. K. G. Fernando, "Non-linear variable selection for artificial neural networks using partial mutual information", *Environmental Modelling & Software*, Vol. 23, pp. 1312–1326, 2008.

[11] M. N. S. Zainudin, N. Sulaiman, N. Mustapha, and T. Perumal, "Feature Selection Optimization using Hybrid Relief-f with Self-adaptive Differential Evolution", *International Journal of Intelligent Engineering and Systems*, Vol. 10, No. 2, pp. 21–29, 2017.

[12] M. Han and X. Liu, "Feature selection techniques with class separability for multivariate time series", *Neurocomputing*, Vol. 110, pp. 29–34, 2013.

[13] B. Bonev, F. Escolano, D. Giorgi, and S. Biasotti, "Information-theoretic selection of high-dimensional spectral features for structural recognition", *Computer Vision and Image Understanding*, Vol. 117, No. 3, pp. 214–228, 2013.

[14] E. Y. Arini, R. Hidayat, and A. Faqih, "Rainfall Simulation Using RegCM4 Model in Kalimantan during El Nino Southern Oscillation", *Procedia Environmental Science*, Vol. 24, pp. 70–86, 2015.

[15] J. H. Faghmous and V. Kumar, "Spatio-Temporal Data Mining for Climate Data : Advances, Challenges, and Opportunities", *Data Mining and Knowledge Discovery for Big Data,* Vol. 1*,* pp. 83–116, 2013.

[16] Y. Wang, T. Zheng, Y. Zhao, J. Jiang, Y. Wang, L. Guo, and P. Wang, "Monthly water quality forecasting and uncertainty assessment via

bootstrapped wavelet neural networks under missing data for Harbin, China", *Environmental Science and Pollution Research*, Vol. 20, No. 12, pp. 8909–8923, 2013.

[17] H. Yoon, K. Yang, and C. Shahabi, "Feature subset selection and feature ranking for multivariate time series", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 9, pp. 1186–1198, 2005.

[18] E. T. Mustafa Gokce Baydogan and G. Runger, "A Bag-of-Features Framework to Classify Time Series", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 11, pp. 1–7, 2013.

[19] B. Chakraborty, "Feature Selection and Classification Techniques for Multivariate Time Series", In: *Proc. of Innovative Computing, Information and Control, International Conference*, pp. 2–5, 2007.

[20] A. Schepen, Q. J. Wang, and D. Robertson, "Evidence for using lagged climate indices to forecast Australian seasonal rainfall", *Journal of Climate*, Vol. 25, No. 4, pp. 1230–1246, 2012.

[21] M. Wegener, "Spatio Temporal Data Mining", *Handbook of Regional Science*, pp. 741–758, 2014.

[22] L. Yu and H. Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy", *Journal of Machine Learning Research*, Vol. 5, pp. 1205–1224, 2004.

[23] A. El Ibrahimi and A. Baali, "Application of neural modeling and the SPI index for the prediction of weather drought in the Saïss Plain (Northern Morocco)", *International Journal of Intelligent Engineering and Systems*, Vol. 10, No. 5, pp. 1–10, 2017.

[24] S. S. Shreem, S. Abdullah, and M. Z. A. Nazri, "Hybrid feature selection algorithm using symmetrical uncertainty and a harmony search algorithm", *International Journal of System Science*, Vol. 47, No. 6, pp. 1312–1329, 2016.

[25] G. B. Demisse, T. Tadesse, and Y. Bayissa, "Data Mining Attribute Selection Approach for Drought Modelling : A Case Study for Greater Horn of Africa", *International Journal of Data Mining and Knowledge Management Process*, Vol. 7, No. 4, pp. 01-16, 2017.

[26] Md. Mursalin, Y. Zhang, Y. Chen, and N. V Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier", *Neurocomputing,* Vol. 241, pp. 204–214, 2017.

[27] N. Ngoc Thach, B.T. Ngo Dang, X.C. Pham, H.T. Nguyen, H.T. Bui, N.D. Hoang, and D.T.

Bui, "Spatial pattern assessment of tropical forest fire danger at Thuan Chau area (Vietnam) using GIS-based advanced machine learning algorithms: A comparative study", *Ecological Informatics*, Vol. 46, No. May, pp. 74–85, 2018.

[28] M. Hall, "Correlation-based Feature Selection for Machine Learning", *Thesis*, Hamilton, New Zealand: University of Waikato, Department of Computer Science, 1999.

[29] B. Kumari and T. Swarnkar, "Feature Subset Selection in Large Dimensionality Micro array using Wrapper Method : A Review", *International Journal of Computer Science and Information Technology*, Vol. 2, No. 2, pp. 792–798, 2011.

[30] M. K. Ebrahimpour and M. Eftekhari, "Ensemble of feature selection methods: A hesitant fuzzy sets approach", *Applied Soft Computing Journal*, Vol. 50, pp. 300–312, 2017.

[31] S. Y. Jiang and L. X. Wang, "Efficient feature selection based on correlation measure between continuous and discrete features", *Information Process Letters*, Vol. 116, No. 2, pp. 203–215, 2016.

[32] J. Sánchez-Monedero, S. Salcedo-Sanz, P. A. Gutiérrez, C. Casanova-Mateo, and C. Hervás-Martínez, "Simultaneous modelling of rainfall occurrence and amount using a hierarchical nominal-ordinal support vector classifier", *Engineering Application of Artificial Intelligence*, Vol. 34, pp. 199–207, 2014.

[33] J. Abbot and J. Marohasy, "Using lagged and forecast climate indices with artificial intelligence to predict monthly rainfall in the brisbane catchment, Queensland, Australia", *International Journal of Sustainable Development and Planning*, Vol. 10, No. 1, pp. 29–41, 2015.

[34] A. Pribadi, P. Wongwises, U. Humphries, A. Limsakul, and A. Wangwongchai, "Diurnal Rainfall Variation over Three Rainfall Regions within Indonesia Based on Ten Years of TRMM Data", *Journal of Suistainable Energy & Environment*, Vol. 3, pp. 81–86, 2012.