Received: June 26, 2021. Revised: October 21, 2021. 42

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Rules Extraction of Relevance Vector Machine for Predicting Negative Emotions from EEG Signals

Adhi Dharma Wibawa[1,2]* Evi Septiana Pane[3] Diah Risqiwati[1,4] Mauridhi Hery Purnomo[1,2]

[1]*Dept. of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[2]*Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[3]*Industrial Training and Education of Surabaya, Ministry of Industry, Indonesia*
[4]*Informatics Engineering, Universitas Muhammadiyah Malang, Indonesia*
* Corresponding author's Email: adhiosa@te.its.ac.id

Abstract: Many studies have reported that patients who are experiencing long-term negative emotions have higher risk of having health deterioration. Therefore, recognition of negative emotions from Electroencephalography (EEG) signals is crucial for monitoring patient conditions. In EEG emotion recognition, clinicians tend to need a clear explanation regarding the rules behind the EEG emotion classification process. Most of the EEG emotions classification use Support Vector Machine (SVM) causing lack of probabilistic prediction which can trigger longer computation time. To address the limitation, we applied Relevance Vector Machine (RVM) with Bayesian inference algorithm to calculate the probability of predicted output. Similar to SVM, RVM was unable to provide transparent rules behind its classification. Therefore, this study attempts to extract rules from RVM by implementing Random Forest algorithm to the relevance vectors. We extract the average energy spectrum in each frequency band as the leading feature of emotions in EEG. Through the resulted rules of RVM_RF, we found that negative emotions of EEG were determined by the average energy spectrum of delta band at fronto-central electrodes (FCZ>23.683, FC4>24.812), theta band at frontal electrode (F5>23.683), and alpha band in parietal-occipital electrode (PO8>20.212). From the evaluation on three sessions of data measurement, it shows that the proposed approach of RVM_RF can predict the negative emotions of EEG with the higher average accuracy of 85.33% and average precision rate of 0.933 compared with other rule-based methods such as RVM_CN2 Rule, RVM_C4.5 Tree, and SVM. All in all, this proposed approach has demonstrated the possibility to identify negative emotions from EEG signal using rules extraction from sparse learning method, RVM.

Keywords: Sparse-model, Negative emotions, Ensemble learning, Interpretable rules.

1. Introduction

People can encounter a variety of negative emotions in daily life such as hatred, anger, sadness, or fear due to many types of causes. Prolonged negative emotions can lead to many negative impacts such as immune alteration [1] or worsening other chronic diseases like Diabetes mellitus, Hypertension, or Cardiac disease [2]. For example, people who experience persistent negative emotions tend to increase the risk of having plaques hardening in the arteries, i.e. atherosclerosis [3]. For those reasons, a recognition of negative emotion is an important aspect of medication process of the patient with chronic diseases. Emotion recognition can be done by identifying the pattern that goes along with negative emotions from physical and physiological measures.

In the last decades, studies on emotion recognition have been heavily relying on physiological signals. Koelstra compare six modalities of physiological signals with responses to emotional video clip stimuli. The results show that EEG signals have significant correlation with the participants rating [4]. Compared to the emotional facial-expression recognition, emotion processing in

the brain appears early with an interval of approximately 180 milliseconds before emotional facial processing arises [5]. Therefore, the motivation for performing emotion recognition based on EEG signals is logical.

There have been many approaches in studies of EEG emotion recognition recently. These approaches range from features extraction [6, 7], features and/or electrodes selection [8], and classification using machine learning [9-11]. In the latter approach, many studies perform a black-box method like SVM. SVM is non-linear classifier model with given labelled training data. It works by finding a separating hyper-plane which discriminates between the class labels of data sample. SVM can handle most of the non-linear classification problems in various field. Nevertheless, it also has several drawbacks, notably the lack of probabilistic prediction output and the growing samples of Support Vectors (SVs) that linearly relate to the increased number of training data [12]. To overcome those limitations, Tipping [12] extends the idea of SVM by introducing a prior distribution on weight from Bayesian inference which is called RVM. Compared to SVM, RVM provides a sparser model with lower decision function while keeping its classification performance. In RVM, the sparse model comes from estimating hyperparameters that constraint the posterior distribution of weights [13]. Therefore, the corresponding posterior distribution of weights is set to zero when these hyperparameters approach infinity. Meanwhile, the rest of data samples with non-zero weights determine the decision functions of the RVM classification model. These remaining data are called Relevance Vectors (RVs). Therefore, by using RVs from RVM model, it is expected to gain an in-depth understanding behind the prediction making on data samples. Since its appearance, RVM has been implemented in several studies on EEG such as in motor imagery [14, 15], mental fatigue detection [16], and many more.

Although SVM and RVM produce a generalized model in various areas of application. Both methods lack the ability to generate transparent explanation behind their prediction result in a human comprehensible form [17]. Classification model able to deliver a set of human-understandable rule for prediction is an advantage for typical application such as health monitoring. Therefore, a physician can gain more interpretation from the rules learned during the classification [18]. A known method for rules extraction is based on Sequential Covering Approach (SCA). SCA has the ability to construct IF – THEN rules from predefined label of the training data. Rules based method and decision tree are two of the most appealing method in SCA. Both methods, generally provide interpretability behind decision prediction, but, in practice, they often suffer with low comprehensibility and accuracy. By expanding the strategy of SCA, Breiman proposes an ensemble learning from Tree classifier called Random Forest (RF) [19]. RF is an ensemble method that utilizes several tree algorithms as a learning model. RF could address the difficulty of single binary tree model to fit with the complex models. By extracting the rules from a few numbers of RVs, the rules obtained using RF are obviously smaller than those from RF. Therefore, the rules obtained from RVM_RF is comprehensible.

Based on those advantages, this study uses RF rule induction technique to extract rules for predicting negative emotions from EEG signals. In our proposed approach, the RVs from the best model of RVM are generated. Then, the RVs and its original label are used as an artificial training data for generating rules from RF rule induction. Compared to the previous study as shown in Table 1, RVM_RF is the first approach for rules extraction from RVM, which has never been tested on predicting negative study pattern from EEG signals. Specifically, the RVM is used for emotion classification, where RF was applied in the relevance vectors result to provide transparent explanation behind the RVM prediction model. So that the experts can obtain more interpretations of the rules learned during the classification. In addition, the RVM has been proven to have less execution time than SVM [16]. The proposed approach is using an open dataset on EEG emotion [20]. In the dataset, a set of video clips were used as stimuli to evoke participant in three class of emotions namely positive, negative, and neutral.

The main objective of this study is to extract a human-understandable rules to predict the negative emotions of EEG signals from RVM classification. Therefore, the two major contributions of this study are the following: presenting a framework of rule extraction from RVM and providing rules for predicting negative emotions from EEG signals. In addition to the major contributions, this paper also investigates electrodes which are significant for emotion recognition task within the dataset. These findings are also notable for developing a framework of EEG emotion recognition due to many interests in implementing fewer electrodes on real-time EEG emotion recognition.

The rest of the paper is organized as follow, the basics of this research are based on the related research described in Section 2. Dataset material and proposed methods introduced in Section 3. Section 4 presents experiment result about electrode selection and classification evaluation. Discussion the

Table 1. Summary of related works

Studies	Remarks	Classifier and accuracy result
[11]	A fused feature extraction method used to improve the precision of EEG-based emotion recognition.	SVM = 89.17% RVM = 91.18%
[21]	Develop a hybrid model (SVM-LOA) for epileptic seizure detection with SVM and Lion Optimization Algorithm (LOA) to optimize the SVM parameters.	SVM = 80.05% SVM-LOA = 96.78%
[22]	The Recursive partition- Parallel Random Forest (R-PRF) has been used to improve the performance of the medical data classification.	SVM-linier = 86.93% RF = 86.57% Decision tree = 78.9% R-PRF = 92.01%
[23]	The proposed rule based modified Convolutional neural network-Global Vectors (RCNN-GloVe) and rule-based modified Support Vector Machine-Global Vectors (RSVM-GloVe) were developed for classifying the twitter complex sentences.	RCNN-GloVe = 92.02% & 88.93% RSVM-GloVe = 87.91% & 85.02%
[16]	Claimed that RVM has less execution time than SVM during the EEG-based mental fatigue detection using cognitive test.	RVM = 92.6% & 93.7% SVM = 81.8% & 73.5%
[14]	It suggested different approach for EEG signal analysis other than the classic SVM	RVM with kernel: - Polynomial = 63.1% - Gaussian = 62.6% - Chaos = 63.6% SVM with polynomial kernel = 64.5%
[15]	Proposed a novel hybrid kernel function RVM by combining the Gaussian and the Polynomial to classify multi-task motor imagery EEG	RVM with kernel: - Polynomial = 81.18% - Gaussian = 81.34% - Hybrid = 82.50%

possibility of rules extraction to predict negative emotion in Section 5, followed by conclusions of this study in Section 6.

2. Related works

An end-user may need explanations (rules) when the purpose is to seek the generalized decision behind the prediction models. The motivation of rule extraction from RVM carries over from earlier knowledge in rules extraction from SVM. In particular, RVM has an advantage related to sparse property which uses significantly fewer basis functions, i.e. RVs, compared to the SVM. This minimum number of RVs in the resulted model is the representative of the class separating data. Therefore, the minimum number of RVs can be optimized to produce comprehensible rules that explains "how" a decision is made by the RVM model. With the emerging of RVM implementation in many areas of the field and significant development of the kernel type, it is required to learn patterns from RVM decision function that provides human-comprehensible explanations behind the model. This ability advances the RVM application for some fields that requires transparency of the decision.

Rules extraction from RVM follows the step of rules extraction from SVM. According to the review on SVM rules extraction, there are three common schemes in rules extraction of SVM: decompositional, pedagogical, and combination of earlier approach [17]. The decompositional approach takes the SVs and separating hyperplane from SVM model as an input for the rule extraction algorithm. For instances, Christy and Shyamala propose rule-based modified SVM using Fuzzy Rule Based Systems (FRBS) [24] to refine the rules; Han suggest the ensemble learning approach to generate rules from SVs as an artificial data input [25]; Fu proposed rules extraction for a non-linear SVMs called RulExSVM, and define the hyper-rectangles region (to determine rules) by using intersection of separating hyper-planes with each SVs which bind the corner of the region [26]. On the contrary, the pedagogical approach uses the SVM model as a black-box method. Method in this group performs SVM as a main learner of the training data. It uses the prediction results of SVM model as artificial data. These data are then used to train in a rules based algorithm (i.e. association rules or decision tree) to generate the corresponding rules, such as SVM with PCPAR (SVM_PCPAR) [27] and SVM with Decision Tree (SVM_DT) [28]. The last approach combines pedagogical and decompositional strategies. This approach employs SVs and separating hyperplanes to extract rules from the synthetic data based on SVs. In implementation, Barakat and Diederich suggest an eclectic rules-extraction from SVM [29]. The eclectic method

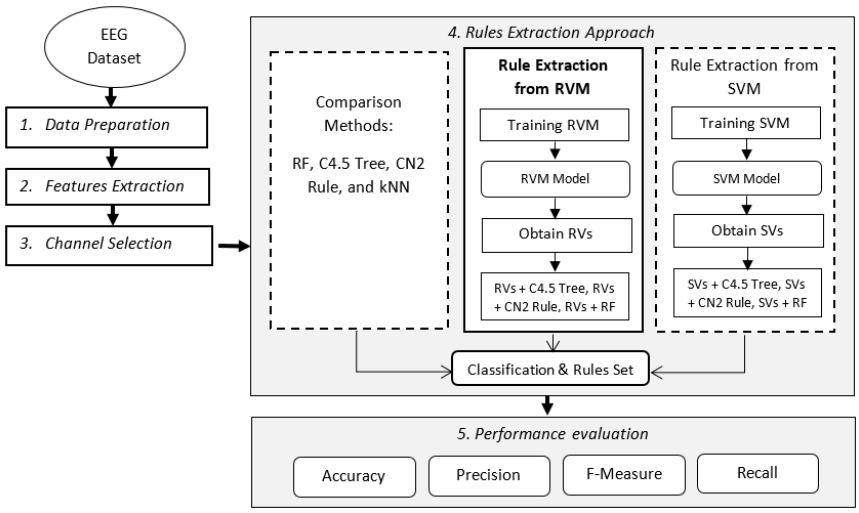

Figure. 1 Proposed rules extraction for negative emotions recognition from EEG

utilizes the knowledge obtained by the SVM and represented in its SVs as well as the synthetic data predicted with them. Mona and Mohamed develop a hybrid model with SVM and Lion Optimization Algorithm (LOA) for epileptic seizure detection [21]. The LOA based approach is used to optimize the SVM parameters and give better result than the SVM alone. Among those three approaches of rules extraction from SVM, the decompositional techniques appear to produce less number of rules which tend to have a good generalization performance compared with the original SVM form. Therefore, this method is suitable for application that requires comprehensible rules, like in our study.

3. Material and methods

In this study, we propose a novel approach for extracting rules to predict negative emotion from EEG using RVM and RF algorithm. This method utilizes the resulting RVs from best RVM model classification to generate rules by RF algorithm. This rule is then applied to predict negative emotions from EEG records. As comparison method, the SVM, K-Nearest Neighbour (kNN), RF, C4.5 Tree, and CN2 Rule are also employed to prove the motive behind the rule extraction from RVM. The overall methodology proposed in this work comprises of five tasks as shown in Fig. 1.

3.1 Dataset preparation

In this paper we used open EEG dataset of emotion, i.e. SEED dataset [20]. It contains 45 records of EEG data from 15 participants in three sessions of measurement within certain time interval (1st, 2nd, and 3rd weeks). In each session, the

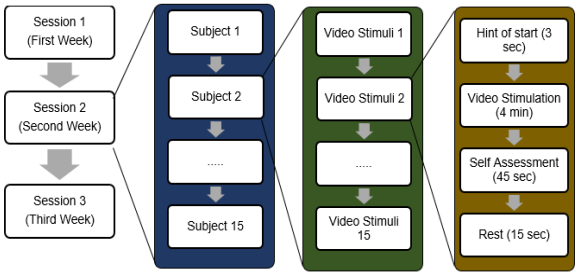

Figure. 2 The data acquisition process of SEED dataset

participants were stimulated by 5 videos for negative, positive, and neutral emotions. The data measurement process of SEED dataset was depicted in Fig. 2. In this study, we assumed that the neutral emotion is as the baseline condition. Therefore, we only included negative and positive emotion for further processing. Then, the total instances that were used in the next process were 15 subjects x 3 sessions x 10 videos (450 instances). Each video clip duration was about 4 minutes. After the video stimulation, every participant was asked to fill their emotional response regarding the video clip. The EEG signals were captured by using ESI NeuroScan System with 62-electrodes location according to the 10-20 international system (as shown in Fig. 3).

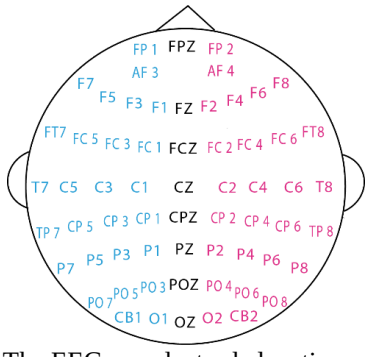

Figure. 3 The EEG cap electrode location used in ESI Neuroscan System [20]

The data was first down sampled to 200 Hz sampling rate. To eliminate noise and artefact such as eye blinking and other muscular remarks, the band pass filter from 0.3 to 50Hz were performed. After performing the noise removal process, the data was decomposed into five common EEG bands; there are gamma (30 – 50 Hz), beta (13-30 Hz), theta (4 – 8Hz), alpha (8-13Hz) and delta (0-4 Hz) bands using Short Term Fourier Transform (STFT) of 256 points with 1 second non-overlapping Hamming window.

3.2 Average energy spectrum features

The EEG signals contain a number of hidden information of brain function. The feature extraction methods were implemented to capture reliable hints from the EEG data in order to recognize and classify emotion. The extracted features were used later in classification process. In this study, the remarkable features extraction of differential entropy (DE) is applied [20]. The DE is known to be equal to the average energy spectrum. It follows the general idea of Shannon entropy which uses entropy to measure the complexity of continuous random variables. Since, the EEG data has a higher portion of low-frequency energy over high frequency. Therefore, the average energy spectrum features can be used to distinguish the pattern of low and high power of each EEG frequency band. It is defined as [6]:

$$h(X) = -\int_{-\infty}^{\infty} X\, log(X) dx = \frac{1}{2} log(2\pi e\sigma^2) \quad (1)$$

$$X = N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

For the feature extraction, we calculated average energy spectrum for each frequency band from all electrodes.

3.3 Electrode selection

With a large number of electrodes in the dataset (e.g. 62-electrodes), it is obvious that a useful selection method is required [30]. The purpose of this selection is to remove the irrelevant electrodes for emotion recognition. Therefore, this leads to minimize the computation time in further process. In our previous study [31], we employ stepwise discriminant analysis to select which electrode has high significance value of discrimination. The Wilk's lambda (Λ) score was used to tests how well each electrode feature contributes to the classification model. Each electrode feature is evaluated by putting it into the model and then taking it out by generating a Λ score. The Λ score is calculated as follows:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E}|+|\mathbf{H}|} \quad (3)$$

Where **H** is the Sums of Squares and Cross Product (SSCP) matrix and **E** is the error of SSCP matrix. The SSCP matrix is often used in the multivariate analysis of variance to store information about variability.

3.4 RVM classification

RVM is based on the principle of sparse learning algorithm with a prior distribution on weight that generates a sparse solution. It is first introduced by Tipping for regression and classification problem [12]. RVM has an identical function with SVM. In general, the output function y(x) for the input space $(x_i)_{i=1}^N$ and the target vector $t = [t_1, t_2, \dots, t_N]$ are defined as:

$$y(x, w) = \sum_{i=1}^{N} w_i K(x, x_i) + w_0 \quad (4)$$

Where N is number of instances,$w_i$is the weight, and $K(.)$ is the kernel function. Thus for $t_n = y(x_n; w) + \varepsilon_n$, the noise ε_n is assumed as Gaussian distribution with zero mean and variance σ^2, then:

$$p(t|w) = N(t|y(x; w), \sigma^2) \quad (5)$$

The basis function is the kernel function $\phi_i(x) = K(x, x_i)$. Since the basis function does not necessary meet Mercer's condition, then it is not required to be positive definite. For classification problem, RVM uses Bernoulli distribution to construct probability density function, given as:

$$p(t|w) = \prod_{i=1}^{N} y_i^{t_i} (1 - y_i)^{1-t_i} \quad (6)$$

Where weight $w = (w_0, \dots, w_N)^T$, target class $t = (t_1, \dots, t_N)^T$, and y_i are the sigmoid function of $\sigma\{y(x_i; w)\}.\sigma(y)$ with $\sigma(y) = 1/(1 + e^{-y})$. The weight w is constrained by Eq. (7) to ensure the generalization ability.

$$p(w|\alpha) = \prod_{i=1}^{N} N(w_i|0, \alpha_i^{-1}) \quad (7)$$

Where α is the hyper-parameter that determines the prior distribution of w value. Then, the Bayes' rule is used to obtain the posterior probability density function of $p(w|t, \alpha)$ expressed as:

$$p(w|t, \alpha) = \frac{p(t|W)p(w|\alpha)}{\int p(t|W)p(W|\alpha)dw} = \frac{g(w)}{p(t|\alpha)} \quad (8)$$

The weight w is assumed to be a Gaussian distribution. Therefore, we can obtain analytical formulation since all the probability density functions are Gaussian. The expression for the posterior probability density function equation over the weight is:

$$p(w|t,\alpha) \sim N(\mu, \Sigma) \quad (9)$$

Where the mean $\mu = \Sigma\Phi^T\mathbf{B}_t$with $\mathbf{B}$ are a $(N+1)\times(N+1)$ diagonal matrix with diagonal element and the covariance $\Sigma = (\mathbf{A} + \Phi^T\Phi)^{-1}$ with $\mathbf{A} = diag(\alpha_0, \alpha_1, \ldots, \alpha_n)$. Therefore, the learning of RVM is the search for the hyper parameter α and w. The sparse model is achieved when α values approach infinity, then the weight w approaches zero. To obtain solution for $(w|t,\alpha)$, Laplace method with Gaussian distribution is used. By taking the logarithm of $g(w)$, we approximate the solution using Hessian matrix that is given as:

$$\mathbf{H} = \nabla\nabla \log g(w)\,|_{w_{\mathrm{MP}}} = \Phi^T\mathbf{B}\Phi + \mathbf{A} \quad (10)$$

The values of w_{MP} and α_i are achieved iteratively as follows:

$$w_{\mathrm{MP}}^{new} = w_{\mathrm{MP}}^{old} - \left(\mathbf{H}^{old}\right)^{-1}\nabla \log g(w)|_{w_{\mathrm{MP}}} \quad (11)$$

$$\alpha_i^{new} = \frac{1-\alpha_i^{old}\Sigma_{ii}}{\mu_i^2} = \frac{\gamma_i}{\mu_i^2} \quad (12)$$

The procedure of RVM learning and prediction is shown in the following steps [32]:

1) Choose a suitable kernel type to map the dataset into high dimensional space and create the design matrix Φ.
2) Initialize starting values of α to calculate **A**,
3) Choose initial value of w to estimate w_{MP},
4) Fix and update the values of w and α using Eq. 11 and Eq. 12,
5) Repeat step (2)-(4) until meeting the convergence criteria, and
6) Predict the new data with the estimated model and obtain performance measures.

The proposed rules extraction from RVM in this study is given in Figure. 4. Firstly, the training data are applied to construct an RVM model with acceptable accuracy by performing 10-fold cross-validation. Then the RVs output from best fold of RVM model is used to generate rules with RF. Different with proposed scheme by Han [25], we use RVs with their original label as an artificial data input for RF. The reason behind this strategy is the assumption that the RVs samples have high probability, therefore, it is selected as the model representation. In the last stage, these rules results are evaluated by the testing data.

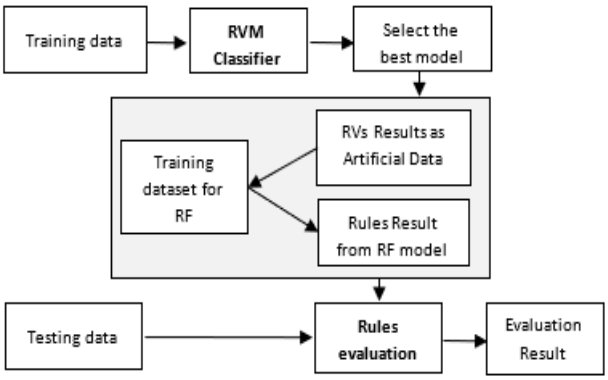

Figure. 4 The steps of proposed RVM_RF

3.5 Random Forest for rules extraction

Random Forest is one of the known algorithms in ensemble learning family. RF is working by combining several binary tree models to learn and predict the target-class label. Similar with the rules-based algorithm, each binary tree in RF also has ability to generate the form of IF-THEN rules. Therefore, it has the advantage of easy interpretability for human and machine. A random forest is a tree-structured based classifier, which is given as [19]:

$$\{\, h_k = h(x, \theta_k),\ k = 1, \ldots, N \,\} \quad (13)$$

Therefore, each tree gives a vote for typical class appearing in instances *x*. The basic principle of RF uses Bootstrap Aggregating (Bagging) and Random Features Selection (RFS). The randomness is injected by growing each tree through both principles. Bagging utilizes *N* classifier by generating addition data in the learning process. Random sampling technique produces the additional data with substitution from initial dataset. In this way, some observations may be reiterated in each new training dataset. However, any instances have equal probability to be appended in a new dataset. Meanwhile, RFS technique works by randomly choosing *k* features among *M* features in the dataset. Then, it builds the best splitting rule-based from these features. As for the prediction, aggregation of *N* resulting classifier is given as:

$$h(x) = \underset{yi}{\operatorname{argmax}} \sum_{k=1}^{N} I(h_k(x) = y_i) \quad (14)$$

The RF model can handle a large number of input features, such EEG data without a prior feature

selection process. It can be used to give estimation of which features are important for classification.

3.6 Performance evaluation

The rules were obtained from the RVs generated from the 10% training data set. The proportion of 10% data training was chosen because RVM tends to have a good performance with a smaller samples of training data [33]. We validate the rules on three sessions of data measurement. By validating on three sessions of measurement, we want to see whether the resulted rules are consistent over different time of data measurement.

Several measures such as accuracy, precision, recall, and F-measure are used to show the rules performance. Accuracy criterion computes the correctness of predicting samples towards the total number of samples in classification. Among several criteria in performance measure, we highly concern the precision criterion because it indicates how precise the rules could predict the number of relevant negative emotion of EEG samples correctly. This criterion is good to determine the accurate prediction of EEG sample originally labelled as negative emotions. It is given by the proportion of correct prediction number of negative emotion samples divided by the total number of predicted samples of negative emotion class. The recall criterion is a number of actual negative emotion samples that the model predicts as negative emotion samples divided by the total number of actual samples on negative emotion class. Meanwhile, the F-measure criterion shows the balance among recall and precision criteria. The accuracy, precision, recall, and F-measure criteria are given by Eqs. (15) to (18).

$$\text{Accuracy} = \frac{\text{Number of all correct predicted samples}}{\text{Number of all samples}} \quad (15)$$

$$\text{Precision} = \frac{\text{Number of correct predicted samples as negative emotion}}{\text{Number of all predicted samples as negative emotion class}} \quad (16)$$

$$\text{Recall} = \frac{\text{Number of correct predicted samples as negative emotion}}{\text{Number of actual samples of negative emotion class}} \quad (17)$$

$$\text{F} - \text{measure} = 2\ \frac{\text{Precision} \,.\, \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

4. Experimental results

Extracting rules for predicting negative emotions from RVM is the main intention of this study. In the first step, the EEG pre-processing stage resulted in average energy spectrum features extract from 5 bands of frequency in 62 electrodes. So, the dimension size output from the feature extraction process is 310 features. Then, the features extracted result was fed to the electrode selection process.

4.1 Electrode selection

From our previous work [31], we investigated the most optimum electrodes for emotion recognition using the Wilk's Lambda from five different frequency bands. The three optimal electrodes in each frequency band is shown in Table 2. The selection was according to their lambda score (> 0.5). Consequently, the number of features for classification were 15, 3 electrodes × 5 frequency bands.

In addition to Table 2, Fig. 5 maps the location of selected electrodes in each frequency band. Through Fig. 5, we see that the optimum three electrodes from each frequency band are spread in several areas of brain, which are in central electrodes (CZ, C3, CB2),

Table 2. Three optimum electrodes in each frequency band resulted from Wilk's Lambda

Frequency Band (Electrode / Lambda Score)				
Delta (δ)	Theta (θ)	Alpha (α)	Beta (β)	Gamma (γ)
FCZ / 0.843	CZ / 0.871	C3 / 0.769	FC1 / 0.752	PO5 / 0.774
CB2 / 0.747	F5 / 0.792	PO8 / 0.549	F6 / 0.598	CZ / 0.667
FC4 / 0.698	P1 / 0.732	TP8 / 0.448	CZ / 0.533	P8 / 0.587

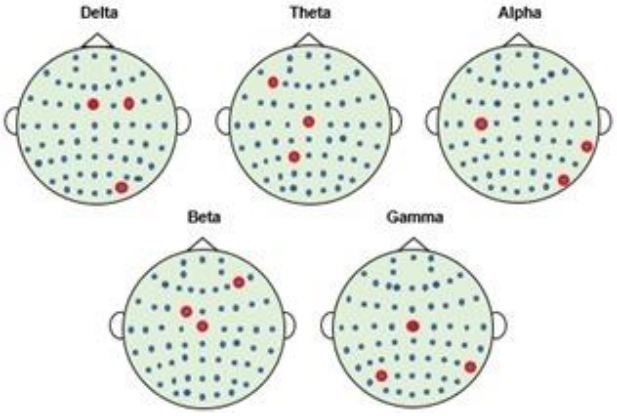

Figure. 5 Location of selected optimal channels according to the SDA results

fronto-central electrodes (FCz, FC4, FC1, F5, F6), parietal-occipital electrodes (PO8, PO5, P8), and temporal-parietal (TP8).

4.2 Classification evaluation

Predicting negative emotions from EEG signal using the extracted rules from RVM is the main intention of this study. The performance criteria such as accuracy, precision, recall, and F-measure were employed to measure the quality of the rules extracted from the proposed and compared methods including rules extraction from SVM with C4.5 Tree, CN2 Rule, and RF. The parameter details and implementation package of each algorithm during the experiment are given in Table 3. From the three performance criteria used in this study, we emphasize on precision of negative emotion class as the main point of analysis. The average results on 10-fold cross-validation on the train set are shown Table 4. Through Table 4, the RVM shows the highest precision rate and accuracy from the 10-fold cross-validation of the training set data. Besides, the standard deviation (SD) of accuracy and precision in RVM model is the lowest among other algorithms. This means that the accuracy among fold models tends to be close to the mean accuracy. Therefore, the rules from RVM tends to be more advantageous compared to other algorithms. To prove the sparse ability of RVM over SVM, we present the different numbers of RVs and SVs in each fold in Fig. 6. Clearly seen from Fig. 6, the number of RVs in all folds (mean and SD 13±1.77) are much less than the number of SVs (mean and SD 34±1.43).

Table 3. Detail parameters used in different methods

Classification	Detail of Parameters	Implementation Algorithm
RVM	Gaussian kernel and optimal kernel scale = 3	Tipping Bayesian Sparse v.2
SVM	Gaussian kernel, γ = 0.125, and *C*-value = 2	LIBSVM
C4.5 Tree	Min. number of instances in leaves = 2 and max. split of subset = 5	C4.5 (Orange v3.18)
CN2 Rule	Evaluation measure = entropy, max. rules length = 5, and ordered rule sequence	CN2 Rule (Orange v3.18)
RF	Number of tree = 3 and max. depth of each tree = 3	Orange v3.18
kNN	k = 5	Orange v3.18

Table 4. Average results of 10-fold cross-validation for negative emotion class

Classification Method	Accuracy % (mean ± SD)	Precision rate (mean ± SD)	Recall rate (mean ± SD)	F-Measure score (mean ± SD)
RVM	**95.33 ± 4.50**	**0.907 ± 0.19**	0.876 ± 0.15	**0.877 ± 0.17**
SVM	83.00 ± 16.19	0.870 ± 0.26	0.863 ± 0.21	0.863 ± 0.18
C4.5 Tree	76.85 ± 23.49	0.783 ± 0.24	0.854 ± 0.27	0.767 ± 0.23
CN2 Rule	80.55 ± 15.02	0.889 ± 0.18	0.740 ± 0.25	0.817 ± 0.14
RF	79.17 ± 26.13	0.845 ± 0.22	**0.937 ± 0.17**	0.787 ± 0.25
kNN	83.33 ± 18.83	0.895 ± 0.19	0.708 ± 0.32	0.846 ± 0.18

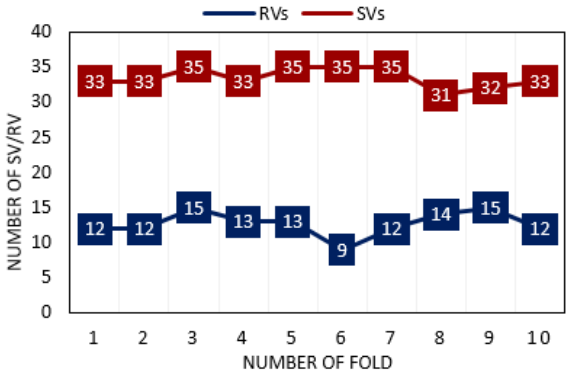

Figure. 6 Comparison of RVs and SVs number in each training fold

Through running the 10-fold cross-validation on training data, we found that the model from fold 9 achieved the best precision rate. The RVs and SVs with its original labels are then taken as artificial data for rules extraction phase using RF and other comparing models. The motivation for including SVs for rules extraction is to show comparison of performance achievements with the proposed approach.

After obtaining the RVs and SVs from the best fold model of RVM and SVM training, we further extracted the rules from the RVs and SVs using rule induction methods. These methods include the RF, CN2 Rule, and C4.5 Tree. The evaluation was performed on three sessions of data measurement according to Figure. 2. The performance results from the RVM_RF and other comparing methods are shown in Fig. 7. From Fig. 7 (a), (b), and (d), the RVM_RF obtained the highest average performance from three sessions of data measurement in terms of accuracy 85.33%, precision rate 0.933, and F-measure score 0.852.

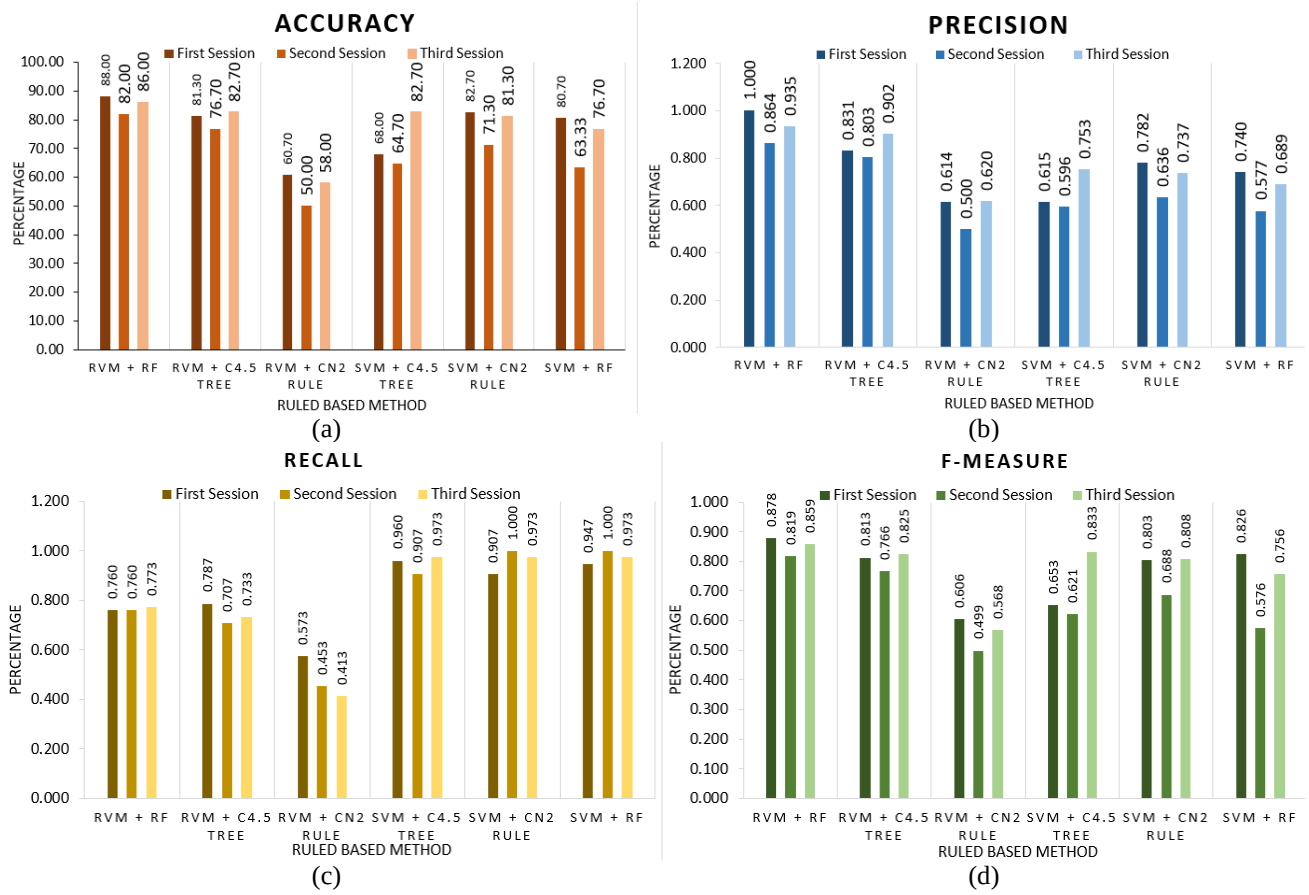

Figure. 7 Results of three sessions of data measurement for negative emotion class

Meanwhile, the least average performance is yielded by RVM_CN2Rule with accuracy of 56%, precision rate of 0.578, and F-measure rate of 0.557. Among three rule induction methods combined with RVM, the RF performed better compared with CN2 Rule and C4.5 Tree. As we can see in Fig. 7 (a), (b), and (d), the rules result from RVM_RF obtained better performance on predicted negative emotions than rules extraction from SVM. Although, the RVM_RF has a lower recall than rules extraction from SVM, but it leads in accuracy, precision, and F-measure score in total six methods.

4.3 Sample rules

For further exploration of the performance of RVM_RF for rules extraction, we compared a number of rules of the RVM_RF method with the other eight methods to predict negative emotions from EEG signal. The result is shown in Fig. 8, as we see, the RVM_RF generated least number of rules compared to the other methods. RVM_RF obtained similar result compared to RVM_C4.5Tree.

The sample of extracted rules from RVM_RF is shown in Fig. 9. The sample rules in Fig. 9 show that negative emotions of EEG are determined by average energy spectrum in alpha, theta, and delta frequency bands. Moreover, the negative emotions are found at fronto-central electrodes (FCZ, FC4), frontal electrode (F5), and parietal-occipital electrode (PO8) areas of brain.

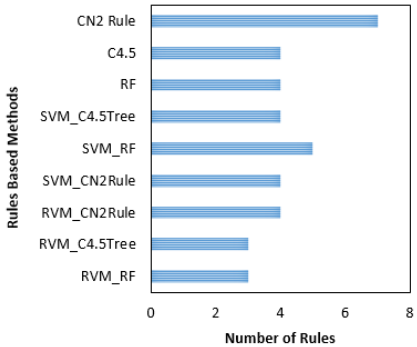

Figure. 8 Comparison of number of rules with proposed and another rules induction methods

Rule 1: **IF** average energy spectrum of delta band at FCZ
AND average energy spectrum of theta band at F5 > 23.683
THEN EEG sample is Negative Emotion
Rule 2: **IF** average energy spectrum of delta band at FC4 > 24.812
THEN EEG sample is Negative Emotion
Rule 3: **IF** average energy spectrum of alpha band at PO8 > 20.212
THEN EEG sample is Negative Emotion
ELSE EEG sample is Positive Emotion

Figure. 9 Sample of rules extracted using RVM_RF method

5. Discussion

This study demonstrates the possibility of rules extraction to predict negative emotions from EEG signal. In the first step, we used fewer electrode which optimal for emotion recognition [31]. Through the results of Wilk's Lambda score in Table 2, they indicate that the features of electrode from delta and theta frequency bands are more significant among other frequency bands for emotion recognition. From Fig. 5, we see that most of the electrodes selected in each frequency band are spread several areas of the brain. This result of electrode selection proves the contribution of certain brain areas for emotion recognition.

The results on classification of RVM and SVM training model show that the RVM generated less number of relevance vectors than SVM (see Fig. 6). This finding proves that RVM classification has sparser ability than SVM. Consequently, the testing computation time of RVM is less required than SVM. This comparison result of RVs and SVs number is confirmed well by earlier finding shown in [12]. The classification results on the 10-fold training data show that the RVM obtained the highest precision and F-measure compared to other methods. This evidence supports our motivation to extract the rules from RVM classification.

The evaluation of extracting rules to the three sessions of data measurement shows that RVM_RF obtained the highest average accuracy, precision rate, and F-measure score in all data sessions. Consequently, from Fig. 7(a) and (b), it shows that RVM_RF has better accuracy and precision compared with rules extraction from SVM (SVM_RF, SVM_C4.5Tree, and SVM_CN2Rule). This verifies that our proposed rule-extraction method has better learning ability from RVs than SVs to predict negative emotions. Through Fig. 7 (a) and (b), the rules extraction result from RVM_RF also has higher accuracy and precision compared with RVM_C4.5Tree and RVM_CN2Rule. This is because RF uses an ensemble of several Tree models with bagging method that could generate a strong learner which has more complexity and flexibility than a single model. Certainly, from Fig. 7 (a)-(c), the RVM_RF has higher accuracy and precision but lower recall than rules extraction from SVM. However, the rules of RVM_RF is generated only by RVs, which has smaller size than SVs (see Fig. 6), so the complexity and number of rules are more decreased than rules extraction from SVM (see Fig. 8). Overall, our rules evaluation on the testing data proves the possibility of using RVM_RF to predict negative emotions from EEG signal.

Additionally, we also present comparison of number of rules obtained from all comparing methods in Fig. 8 and the sample rules from the RVM_RF in Fig. 9. According to the rules number resulted in Fig. 7, it is shown that RVM_RF resulted the least number of rules which is also similar with RVM_Tree. The larger rules set may make the learning patterns more transparent but the comprehensibility of the rules is adversely affected. Nevertheless, the RVM_RF still obtained better precision result of evaluation than RVM_Tree. This means that RVM_RF does not only generate a few comprehensible rules, but also with high precision of prediction. The sample rules shown is confirmed well by the earlier study on brainwaves-emotion correlation analysis studies [34-36] in term of frequency band.

When we look at the sample rules, we found that average energy spectrum from delta frequency bands determines largely the negative emotions. Although, the literature stated that delta frequency band is usually associated with a deep sleep condition. This result is also confirmed well by the study of Balconi and Vanutelli [34] that showed a significant increase of variance of negative emotions in delta frequency band. Other than delta frequency band, we can see from the sample rules that theta and alpha frequency band are also considered as important bands for emotion recognition. These rules agree with the result obtained by Lee and Hsieh [35]. From the analysis on theta and alpha frequency band, a negative emotion shows a significantly higher correlation than positive emotions, particularly at the parietal and occipital sites such as PO8. Other than that, the EEG alpha power from sad emotion has the largest variance among other emotion. Moreover, in our result on electrode selection in Table 2, we can see that features from electrodes in delta and theta frequency bands have high score of Wilk's Lambda (the Wilk's lambda score indicates greater discriminatory ability of the features). These results are in accordance with the sample rules obtained in Fig. 9. Another study on brainwaves analysis of emotion detection by Wan Ismail [36] also confirms our rules. He finds that the anger emotion shows very clear reaction in theta band, while the sad emotion shows significant reaction at the delta and theta bands.

In order to show the difference of our proposed approach with previous works, we summarize several studies on the same public dataset in Table 5. Compared to other studies, our approach is advantageous in number of features included for classification since it reports the lowest number of electrodes. Consequently, the less amount of electrodes in emotion recognition will reduce the

Table 5. Comparison with previous studies on the same public dataset

Authors	Number of Features (Electrodes)	Method	Performance result	Provide rules of classification
Zheng and Lu [20]	20 (4 electrodes × 5 frequency band)	Deep belief Neural Network for classification	86.08 % of accuracy with different training and testing data	No
Zheng [38]	310 (62 electrodes × 5 frequency band)	Graph Regularized Extreme Learning Machine for classification	91.07 % of accuracy using 5-fold cross-validation scheme	No
W. Zheng [8]	20 (4 electrodes × 5 frequency band)	Group sparse correlation analysis for optimal electrode selection	80.20 % of accuracy using leave-one-trial-out cross-validation strategy of evaluation	
Chai [39]	310 (62 electrodes × 5 frequency band)	Subspace Alignment Auto-Encoder for reducing dimension discrepancy	81.81 % of accuracy from session-to session evaluation	No
Our proposed approach	15 (3 electrodes × 5 frequency band)	RVM and RF for classification and rules extraction	0.933 precision rate of predicting only negative emotions 85.33% of average accuracy from three sessions of data measurement	Yes

time setup for EEG headset and features extraction process [37]. According to the proposed approach categories in Table 5, the studies in general propose a method for classification, electrode selection, and features reduction. Instead of only performing emotion classification by RVM, our study also gives additional result in form of interpretable explanation for predicting negative emotions.

Although all of the studies in Table 5 use the same dataset and features extraction methods, comparing performance result within these studies is difficult to perform. Because each study has different strategies in validating their proposed method. However, the performance accuracy average achieved by our study is still comparable with other previous works. Besides, we use the fewest number of features to gain acceptable performance in EEG emotion recognition compared with other studies. From Table 5, it is clearly seen that our proposed approach is the only study that provides explanation behind the emotion classification, particularly of negative emotions. This shows that this study contributes to the higher precision rate on prediction of negative emotions and providing interpretable rules useful for end-users.

6. Conclusion

This study attempts to provide human-transparent rules to predict negative emotions from brain signals by combining the RVM and RF method. Specifically, the RVM is used for emotion classification, where RF was applied in the relevance vectors result to provide transparent explanation behind the RVM prediction model. In addition, we performed electrode selection using Wilk's Lambda score due to the large numbers of electrodes included in the dataset, which will avoid a high computation time.

The training result proves that the RVM model is much sparser than SVM. Through the rules evaluation result, our proposed method has higher precision and accuracy in predicting negative emotion than SVM, meaning that the relevance vectors provide better learning data than support vectors for extracting the rules. In terms of the rules extracted, our proposed method generates the least number of rules with high quality of precision rate compared with other methods. The RVM_RF obtained the highest average performance from three sessions of data measurement in terms of accuracy 85.33%, precision rate 0.933, and F-measure score 0.852. The sample rules showed that the average energy spectrum from delta, theta and alpha frequency bands mainly identifies negative emotions of EEG. This rule is confirmed well by the previous finding in brainwaves analysis on emotion recognition. With the fewest number of features that only 15 (3 electrodes × 5 frequency band) can gain acceptable performance in EEG emotion recognition compared with other studies. This shows that this study contributes to the higher precision rate on prediction of negative emotions and providing interpretable rules useful for end-users. Furthermore, these rules result will provide enhanced opportunity for timely prediction in a real-time emotion recognition, which might reduce the emergence of adverse effects from experiencing prolonged negative emotions.

Further improvement might put emphasis on utilizing α and *w* hyperparameters together with RVs for extracting rules through the finding of decision boundary in form of hyper-rectangles. Moreover, the performance criterion here is restricted to just classification performance. In the future, we suggest to include several rules quality measures such as fidelity, comprehensibility, and consistency.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Author Contributions

Conceptualization and methodology, Adhi Dharma, Evi S. Pane and Mauridhi Hery P; software, Evi S. Pane; validation, Adhi Dharma, Evi S. Pane, and Mauridhi Hery P; formal analysis, Adhi Dharma; investigation, Adhi Dharma and Evi S. Pane; resources, Evi S. Pane; data curation, Evi S. Pane; writing—original draft preparation, Evi S. Pane; writing—review and editing, Adhi Dharma, Evi S. Pane, and Diah Risqiwati; visualization, Evi S. Pane; supervision, Mauridhi Hery P; project administration, Adhi Dharma; funding acquisition, Adhi Dharma.

Acknowledgments

We thankful to Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan - LPDP), Ministry of Finance, The Republic of Indonesia that made this research paper possible through financial support.

References

[1] K. B. Koh, "Emotion, Interventions, and Immunity", *Springer New York*, pp. 299-315, 2013.

[2] S. D. Pressman and S. Cohen, "Does positive affect influence health?", Vol. 131, No. 6, pp. 925-971, 2005.

[3] G. Peter J., A. L. Marsland, D. C. H. Kuan, B. L. Schirda, J. R. Jennings, L. K. Sheu, A. R. Hariri, J. J. Gross, and S. B. Manuck, "An Inflammatory Pathway Links Atherosclerotic Cardiovascular Disease Risk to Neural Activity Evoked by the Cognitive Regulation of Emotion", Vol. 75, No. 9, pp. 738-745, 2014.

[4] Koelstra, Sander, C. Mühl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; Using physiological signals", *IEEE Trans. Affect. Comput.*, Vol. 3, No. 1, pp. 18-31, 2012.

[5] D. Maurer, R. L. Grand, and C. J. Mondloch, "The many faces of configural processing", Vol. 6, No. 6, pp. 255-260, 2002.

[6] R. N. Duan, J. Y. Zhu, and B. L. Lu, "Differential entropy feature for EEG-based emotion classification", *Int. IEEE/EMBS Conf. Neural Eng. NER*, pp. 81-84, 2013.

[7] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion Recognition From {EEG} Using Higher Order Crossings", Vol. 14, No. 2, pp. 186-197, 2010.

[8] W. Zheng, "Multichannel {EEG}-Based Emotion Recognition via Group Sparse Canonical Correlation Analysis", Vol. 9, No. 3, pp. 281-290, 2017.

[9] J. Atkinson and D. Campos, "Improving {BCI}-based emotion recognition by combining {EEG} feature selection and kernel classifiers", Vol. 47, pp. 35-41, 2016.

[10] P. Li, W. Jiang, and F. Su, "Single-channel {EEG}-based mental fatigue detection based on deep belief network", 2016.

[11] Q. Gao, C. H. Wang, Z. Wang, X. L. Song, E. Z. Dong, and Y. Song, "EEG based emotion recognition using fusion feature extraction method", *Multimedia Tools and Applications*, Vol. 79, No. 37-38. pp. 27057-27074, 2020.

[12] M. E. Tipping, "The relevance vector machine", *Adv. Neural Inf. Process. Syst.*, Vol. 12, pp. 653–658, 2000.

[13] M.-W. Mak, "Lecture Notes on Relevance Vector Machines", *The Hong Kong Polytechnic University*, 2016. [Online]. Available: http://www.eie.polyu.edu.hk/~mwmak/papers/RVM.pdf.

[14] E. Dong, G. Zhu, C. Chen, J. Tong, Y. Jiao, and S. Du, "Introducing chaos behavior to kernel relevance vector machine (RVM) for four-class EEG classification", *PLoS ONE*, Vol. 13, No. 6. 2018.

[15] E. Dong, K. Zhou, J. Tong, and S. Du, "A novel hybrid kernel function relevance vector machine for multi-task motor imagery EEG classification", *Biomed. Signal Process. Control*, Vol. 60, p. 101991, 2020.

[16] A. Setiawan, A. D. Wibawa, E. S. Pane, and M. H. Purnomo, "EEG-based mental fatigue detection using cognitive tests and RVM classification", In: *Proc. of 2019 Int. Conf. Artif. Intell. Inf. Technol. ICAIIT 2019*, pp. 180-185, 2019.

[17] N. Barakat and A. P. Bradley, "Rule extraction from support vector machines: A review", *Neurocomputing*, Vol. 74, No. 1–3, pp. 178–190, 2010.

[18] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective", *Artif. Intell. Med.*, Vol. 23, No. 1, pp. 89–109, 2001.

[19] L. Breiman, "Random Forests", *Mach. Learn.*, Vol. 45, No. 1, pp. 5-32, 2001.

[20] W. L. Zheng and B. L. Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks", *IEEE Trans. Auton. Ment. Dev.*, Vol. 7, No. 3, pp. 162-175, 2015.

[21] M. A. S. Ali and M. A. Elfattah, "A hybrid model (SVM-LOA) for epileptic seizure detection in long-term EEG records using machine learning techniques", *Int. J. Intell. Eng. Syst.*, Vol. 11, No. 5, pp. 162-172, 2018.

[22] Y. Cooli and C. Mahesh, "Recursive Parallel Partition Random Forest for Medical Disease Classification", *Int. J. Intell. Eng. Syst.*, Vol. 14, No. 5, pp. 112-120, 2021.

[23] C. Daniel and S. Loganathan, "A Comparison of Machine Learning and Deep Learning Methods with Rule Based Features for Mixed Emotion Analysis", *Int. J. Intell. Eng. Syst.*, Vol. 14, No. 1, pp. 42-53, 2021.

[24] A. C. F. Chaves, M. M. B. R. Vellasco, and R. Tanscheit, "Fuzzy rule extraction from support vector machines", In: *Proc. of Fifth International Conference on Hybrid Intelligent Systems (HIS'05)*, p. 6, 2005.

[25] L. Han, S. Luo, J. Yu, L. Pan, and S. Chen, "Rule Extraction From Support Vector Machines Using Ensemble Learning Approach: An Application for Diagnosis of Diabetes", *IEEE J. Biomed. Heal. Informatics*, Vol. 19, No. 2, pp. 728–734, 2015.

[26] X. Fu, C. Ong, S. Keerthi, G. G. Hung, and L. Goh, "Extracting the knowledge embedded in support vector machines", In: *Prof. of 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, pp. 291–296, 2004.

[27] J. He, H. Hu, B. Chen, P. C. Tai, R. Harrison, and Y. Pan, "Rule Extraction from SVM for Protein Structure Prediction", *Springer Berlin Heidelberg*, pp. 227–252, 2008.

[28] J. He, H. Hu, R. Harrison, P. Tai, and Y. Pan, "Transmembrane segments prediction and understanding using support vector machine and decision tree", *Expert Syst. Appl.*, Vol. 30, No. 1, pp. 64–72, 2006.

[29] N. Barakat and J. Diederich, "Eclectic rule-extraction from support vector machines", *Int. J. Comput. Intell.*, Vol. 2, No. 1, pp. 59-62, 2005.

[30] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing", *EURASIP J. Adv. Signal Process.*, Vol. 2015, No. 1, p. 66, 2015.

[31] E. S. Pane, A. D. Wibawa, and M. H. Pumomo, "Channel Selection of EEG Emotion Recognition using Stepwise Discriminant Analysis", In: *Proc. of 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM)*, pp. 14–19, 2018.

[32] T. Fletcher, "Relevance Vector Machines Explained", *Tech. Rep. - Univ. Coll. London*, pp. 1-9, 2010.

[33] B. Ribeiro and C. Silva, "RVM Ensemble for Text Classification", *Int. J. Comput. Intell. Res.*, Vol. 3, No. 1, 2007.

[34] M. Balconi and E. Vanutelli, "Empathy in Negative and Positive Interpersonal Interactions. What is the Relationship Between Central (EEG, fNIRS) and Peripheral (Autonomic) Neurophysiological Responses?", *Adv. Cogn. Psychol.*, Vol. 13, No. 1, pp. 105–120, 2017.

[35] Y.-Y. Lee and S. Hsieh, "Classifying Different Emotional States by Means of EEG-Based Functional Connectivity Patterns", *PLoS One*, Vol. 9, No. 4, p. e95415, Apr. 2014.

[36] W. O. A. S. Wan Ismail, M. Hanif, S. B. Mohamed, N. Hamzah, and Z. I. Rizman, "Human Emotion Detection via Brain Waves Study by Using Electroencephalogram (EEG)", *Int. J. Adv. Sci. Eng. Inf. Technol.*, Vol. 6, No. 6, p. 1005, 2016.

[37] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-Time EEG-Based Emotion Recognition and Its Applications", *Springer Berlin Heidelberg*, pp. 256–277, 2011.

[38] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying Stable Patterns over Time for Emotion Recognition from EEG", *IEEE Trans. Affect. Comput.*, Vol. 10, No. 3, pp. 417–429, 2019.

[39] X. Chai, Q. Wang, Y. Zhao, X. Liu, O. Bai, and Y. Li, "Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition", *Comput. Biol. Med.*, Vol. 79, pp. 205–214, 2016.