Article

A Study on Facial Expression Change Detection Using Machine Learning Methods with Feature Selection Technique

Sang-Ha Sung [1], Sangjin Kim [1,*] , Byung-Kwon Park [1], Do-Young Kang [2] , Sunhae Sul [3] , Jaehyun Jeong [4] and Sung-Phil Kim [4,*]

[1] Department of Management Information Systems, Dong-A University, Busan 49236, Korea; sangha@donga.ac.kr (S.-H.S.); bpark@dau.ac.kr (B.-K.P.)
[2] Department of Nuclear Medicine, Dong-A University, Busan 49236, Korea; dykang@dau.ac.kr
[3] Department of Psychology, Pusan National University, Busan 46241, Korea; ssul@pusan.ac.kr
[4] Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea; jh0410636@unist.ac.kr
* Correspondence: skim10@dau.ac.kr (S.K.); spkim@unist.ac.kr (S.-P.K.); Tel.: +82-05-200-7484 (S.K.)

Abstract: Along with the fourth industrial revolution, research in the biomedical engineering field is being actively conducted. Among these research fields, the brain–computer interface (BCI) research, which studies the direct interaction between the brain and external devices, is in the spotlight. However, in the case of electroencephalograph (EEG) data measured through BCI, there are a huge number of features, which can lead to many difficulties in analysis because of complex relationships between features. For this reason, research on BCIs using EEG data is often insufficient. Therefore, in this study, we develop the methodology for selecting features for a specific type of BCI that predicts whether a person correctly detects facial expression changes or not by classifying EEG-based features. We also investigate whether specific EEG features affect expression change detection. Various feature selection methods were used to check the influence of each feature on expression change detection, and the best combination was selected using several machine learning classification techniques. As a best result of the classification accuracy, 71% of accuracy was obtained with XGBoost using 52 features. EEG topography was confirmed using the selected major features, showing that the detection of changes in facial expression largely engages brain activity in the frontal regions.

Keywords: machine learning; classification; feature selection; BCI; EEG

1. Introduction

Along with the fourth industrial revolution, data-driven research in the biomedical engineering field has been actively progressing. The brain–computer interface (BCI) is one of the emerging topics in the field that studies the direct interaction between the brain and external devices [1]. With the development of devices capable of measuring neural activity, BCI research is actively progressing [2,3]. Through the BCI, one can understand the information represented in the brain signals and predict action only from the brain signals [4]. In particular, it is possible to measure the emotional state of the subject by capturing relevant brain signals with the BCI [5]. Emotion recognition has been used in various application fields as a method of grasping human emotion states through computer systems [6,7]. Emotions can be objectively classified through physiological signals such as blood pressure response, skin response, pupil reflex, and brain signals [5]. Among them, recognizing emotions using brain signals has recently attracted great attention [5]. Such systems often harness electrical activity of the brain measured by tens of electrodes placed on the scalp, which is termed as electroencephalography (EEG) [8]. Therefore, it is important to find major features in EEG data because the dimensionality of potential EEG features can be huge and the relationship between features is complex [9]. In many studies, to analyze EEG data for the BCI, the feature selection technique has been widely

used [10–12]. The feature selection technique can reduce the complexity of the model and help improve accuracy [10]. In addition, efforts to infer human emotional states by applying machine learning techniques are in progress [10,13]. For instance, recent studies have applied a machine learning technique to classify mental states (concentration and drowsiness) from EEG data with very high accuracy [12,14,15].

Today, machine learning algorithms are being used in various applications that deal with data [14]. Machine learning can be classified into supervised learning and unsupervised learning according to the type of data used for learning [15]. If the training data used for learning contain a label that means the correct answer, it is called supervised learning, and it is called unsupervised learning when learning without a label. In addition, machine learning is used as a model for classification when the presented label means an individual category, and is used as a regression model when the label is a continuous variable [15]. This study focuses on classification with supervised learning, where the label consists of the binary information of the success of detection of facial expression changes—e.g., '1' if a person correctly detects a facial expression change and '0' if the person incorrectly does it. Therefore, the EEG-based BCI was built in this study to predict whether or not a person correctly detects facial expression changes of others. Such a BCI would be able to help us to develop an intelligent system to evaluate social interactions of individuals and assist to improve one's empathic skills. We utilized machine learning methodology with feature selection technique for classifying EEG data. The purpose of this study is to derive the results of how the extracted main features affect the prediction of the correct detection of facial expression changes. Feature selection methods such as Fisher score, chi-square, mutual information, and Gini importance were used to examine the influence of each feature and to select the main features. In addition, random forest, decision tree, XGBoost, and support vector machine (SVM) were used to classify EEG into correct and incorrect facial expression change detection. The combination of the methodologies that showed the best classification accuracy was selected using the proposed feature selection technique and machine learning classification technique. Using the set of key features representing the highest classification accuracy, we examined specific brain areas important when people detect facial expressions of others.

The following sections of the paper describe the details of the research conducted. Section 2 presents the extensive algorithms and methodologies used in this study. Section 3 describes the data and experimental methods used. Section 4 describes the experimental results. Section 5 discusses the significance and limitations of this study, and Section 6 describes the conclusion.

2. Methods

In this section, we describe the feature selection methods and classification methods. Some feature selection methods and classification techniques are explained through detailed formulas to understand them. This session will help you understand how each technique will be used. With this, we used several feature selection methods and classification methods to find the best combination for BCI data analysis.

2.1. Feature Selection Methods

In building a classification model, it is a very important process to select features that affect the classification result [16]. In this study, popular feature selection methods were used, such as F-score, chi-square, and mutual information, which are univariate feature selection methods [16–18]. In addition, a feature selection technique using Gini importance was also used [19]. These variable selection methods rank variables in consideration of the influence between the target variable and the dependent variable. By using these feature selection methods, the complexity of building the model can be greatly reduced [20].

2.1.1. Fisher Score

The Fisher score (F-score) is one of the most popular feature selection methods [21]. F-score is a univariate selection method and selects the optimal features based on a statistical model [16]. It can be used mainly in linear models. The characteristics between heterogeneous classes are different, and the F-score increases as the characteristics between homogeneous classes are similar [22,23]. After statistically analyzing the relationship between the dependent variable and independent variable, the influence of each independent variable is derived with weight. F-score has the advantage of fast calculation, and it is easy to check the influence of each variable. The equation for calculating the F-score is as follows.

$$Fscore_i = \frac{|S_B|}{|S_W|} \tag{1}$$

$$S_B = \sum_{i=1}^{C} n_i (u - u_i)(u - u_i)^T \tag{2}$$

$$S_W = \sum_{i=1}^{C} \sum_{x \in c_i} (x - u_i)(x - u_i)^T \tag{3}$$

where the F-score is the ratio of between-class scatter (S_B) and within-class scatter (S_W), C is the total number of classes, and n_i is the number of samples belonging to c_i.

2.1.2. Chi-Square

Chi-square is also a popular feature selection method [16]. This technique statistically checks the relationship between each independent variable and class [17]. Since the chi-square test measures the dependence of each variable, it is easy to identify independent variables that are not related to class [24]. In other words, if there is no association between the dependent variable and the independent variable as a result of the statistical test, it is judged that the independent variable is not significant. The formula for chi-square is as follows:

$$Chi = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{4}$$

where O_i is the observed value of each class, E_i is the expected value of each class, and k is the number of classes.

2.1.3. Mutual Information

Mutual information (MI) is one of the univariate selection methods [18]. MI has the characteristic of using a nonparametric methodology. When the independent variable and the dependent variable are completely independent, the amount of information becomes 0, and when they are related, the amount of information increases. If there is an inverse relationship, the amount of information decreases. In other words, it is an indicator to judge how close the relationship between the variables is. The equation for calculating MI is as follows:

$$MI(A, B) = \frac{P(A \cap B)}{P(A) * P(B)} \tag{5}$$

In the feature selection technique, MI is used to identify the dependency between the dependent variable and the independent variable [25]. The higher the value, the higher the dependence and the influence [26].

2.1.4. Gini Importance

Gini importance is utilized to measure the importance of the feature [26]. It is calculated based on Gini impurity [19]. As the importance of a specific feature increases, the impurity of the corresponding node decreases significantly. Therefore, the lower the impurity, the higher the importance of the feature.

Gini impurity becomes 0 when the values classified through nodes are heterogeneous, and 1 in the opposite case. The following is the formula for Gini impurity:

$$Gini(A) = 1 - \sum_{k=1}^{m} p_k^2 \tag{6}$$

where p_k is the ratio of data belonging to k class, and m is the number of classes. Impurity after classification can be measured through the above formula. The closer the impurity is to 0, the higher the homogeneity. The importance of each node can be measured through the measured impurity. The formula for the importance of a node is as follows:

$$I(C_j) = w_j Gini(C_j) - w_{j_{left}} Gini\left(C_{j_{left}}\right) - w_{j_{right}} Gini\left(C_{j_{right}}\right) \tag{7}$$

where w_j means the ratio of the number of samples corresponding to node C_j among the total number of samples. That is, the weight impurity of the parent node is subtracted from the sum of the weight impurity of the child nodes. Node importance can be used to measure how much each feature has an impact on creating a tree and classifying samples.

2.2. Classifiers

Classification is a matter of predicting a given category of data. Classification is being used in many areas today and is steadily used in the research field relevant to BCI. Classification can be classified into binary classification and multiple classification according to the number of classes. In this study, the binary classification problem was the focus because the dataset was composed of two classes that indicate whether facial expressions were detected. The classifiers such as random forest, gradient boost, XGBoost, and SVM were used to classify whether or not facial expressions are detected.

2.2.1. Random Forest

Random forest is an ensemble machine learning model. It is used for both classification and regression, like decision trees [27].

The decision tree is one of the supervised learning methods to classify data through classification rules. The derived model is composed of a tree structure, which is easy to understand. However, it has a limitation of frequent overfitting of the training data.

Therefore, a random forest was used to solve the overfitting problem. Since it generalizes and uses the result values of several randomly generated decision trees, the overfitting problem is significantly reduced [28]. It uses a technique of bootstrap and aggregation (bagging) to generate a tree over a subset of the training data. It is advantageous for generalization because it aggregates the results of many trees through the voting technique.

2.2.2. Gradient Boost

Gradient boosting is a prediction model that can perform regression analysis or classification analysis, and is an algorithm belonging to the boosting family of ensemble methods. It is an algorithm that creates several weak learners to make strong learners [28]. The correct answer is predicted using the preceding tree, and the remaining residuals are predicted using the next tree [29]. By repeating this process, the model builds a number of trees (weak learners). These trees are combined to create a strong classifier. To improve the performance of each classifier, errors must be quantified using a loss function, and residuals should be reduced. Gradient boosting uses the algorithm of gradient descent in the process and induces learning in a direction in which the loss function can be minimized.

2.2.3. XGBoost

Since gradient boosting requires compute-intensive tasks, it is necessary to calculate it efficiently. XGBoost is a machine learning model created to compensate for the demerit of the gradient boosting [29]. It is similar to gradient boosting, but it controls the complexity of the tree by adjusting the loss function. Therefore, it has a faster operation speed compared

with the existing gradient boost [27]. In addition, it supplements the problem of overfitting by using the random subsampling technique of each individual tree.

2.2.4. SVM

SVM is one of the most popular machine learning models [30,31]. In this study, it was used as a binary classifier that classifies into two classes. It is a model that defines criteria for classification. These criteria are called the decision boundary, and the greater the number of features, the more complex it appears. The more accurate the decision boundary used to classify the data class, the higher the accuracy. However, it is difficult to accurately classify all data, so some outliers are ignored. The distance between the decision boundary and the support vector is called margin. By adjusting the margin, the overfitting problem can be reduced. It can obtain various types of decision boundaries by using several kernel options. That is, the result of the model changes a lot according to the adjustment of the parameter value.

3. Data Analysis

3.1. Data Collection

3.1.1. Participants

Seventy-five participants (49 males, 26 females), aged 19–30 years (mean 22.96 ± 2.4 years old), volunteered to take part in the experiment. All reported having normal or corrected-to-normal vision with reportedly no neurological disorder or psychiatric illness. Of those, 10 participants were excluded due to a problem in data acquisition, and 14 due to an insufficient number of samples in any of the two classes. This exclusion left 51 participants' data for subsequent analyses.

3.1.2. Task

The main purpose of the task was to measure the perceptual sensitivity of individuals to emotional changes in others' facial expression. The experiment was designed according to the previous study by Ha and Shim [32]. In every trial of the task, a fixation (white cross) was presented for 500 ms at the center of the screen. Then, an animated human face whose expression dynamically changed from neutral expression to emotional expression began to appear. Participants watched the animated face and reacted as quickly as possible by pressing the space bar on the keyboard when they recognized an emotion in the presented face (Figure 1a). The facial animation consisted of 26 facial expression images (created in FaceGen Modeller, Singular Inversions, www.facegen.com, accessed on 27 July 2021), from neutral to emotional ones, where each image was displayed for 300 ms—a full presentation of the animated face thus took 7800 ms. When participants pressed the button, the facial animation stopped, and the next response window was shown on the screen to ask which emotion participants recognized with numerical mapping of keys from 1 to 6: '1'= fear, '2' = sad, '3' = surprise, '4' = happy, '5' = angry, and '6' = disgust. Both male and female faces were presented, respectively. A trial for one of the six emotions for each gender was repeated 5 times, yielding a total of 60 trials in the task (6 emotions × 2 gender (male, female) × 5 repetitions) (Figure 1b).

3.1.3. EEG Acquisition and Processing

EEG signals were acquired using 31-channel wet Ag/Cl electrodes (anti-champ, Brain Products GmbH, Gilching, Germany) at a sampling rate of 500 Hz. The acquired EEG signals were band-pass-filtered with 1 Hz and 100 Hz cut-off frequencies using a finite impulse response (FIR) filter. The position of 31 electrodes was determined following the 10/20 international system: FP1, FPz, FP2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, and O2 (Figure 1c). Additional electrodes were attached to the left mastoid as a ground and the right mastoid as a reference. The impedance of all electrodes was kept below 10 kOhm.

(**a**)

(**b**)

(**c**)
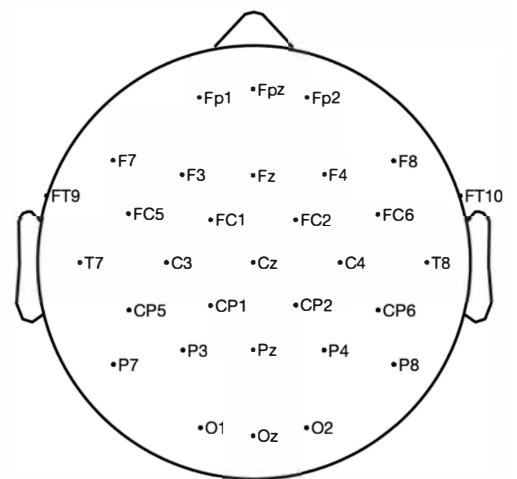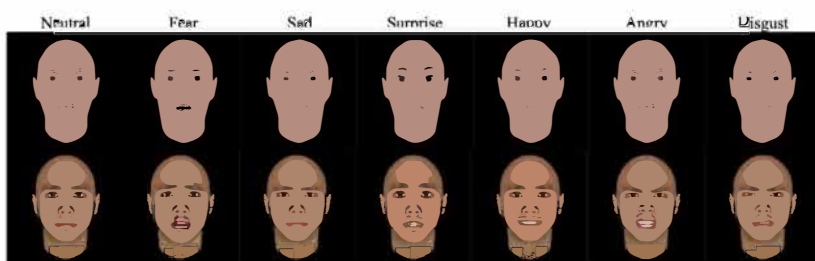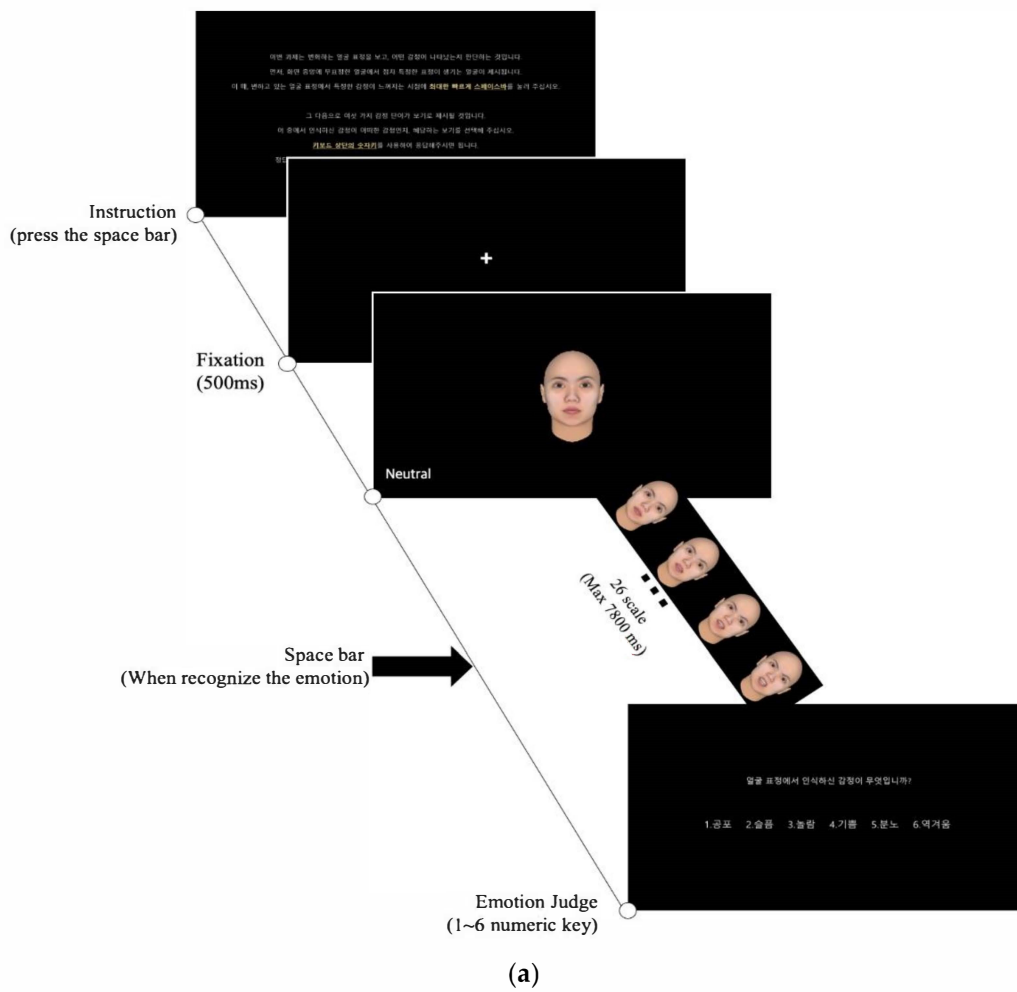
Figure 1. Figure about an experiment to acquire data. (**a**) Experiment protocol; (**b**) facial expression stimuli; the upper row is female, bottom is male. From left to right, the first is neutral face and in turn, the extreme expression of each emotion is shown; (**c**) EEG montage.

EEG preprocessing was conducted using the EEGLAB toolbox in MATLAB software (Mathwords, Inc., Natick, MA, USA) [33]. First, the EEG signals were band-pass-filtered again with a 58–62 Hz notch-filter to remove line noise using an FIR filter. Second, we elim-

inated noisy electrodes by investigating the correlation of a single channel with others [34]. All rejected channels were spherically interpolated to simplify subsequent analyses. After interpolation, the common average reference (CAR) method was applied for re-referencing. Independent component analysis (ICA) was then used to remove ocular and muscle artifacts. The artifact component was detected using the ICLabel plugin that classified EEG independent components (ICs) automatically [35]. By using this tool, those ICs with any of the following labels that showed the label probability ≥0.8 were rejected: 'muscle' and 'eye'.

An epoch of the EEG signal was extracted 1200 ms to 200 ms before response onset (i.e., time to press space bar) from each trial. The baseline period was set to be 0 ms to 500 ms after fixation onset. Epoch was standardized by the baseline signal's mean and standard deviation. Short-time Fourier transform (STFT) was applied to each epoch with a window length of 256 ms, and the overlap of 240 ms. The log-transformed power spectrum obtained by STFT in each window was subdivided into 6 frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), low-beta (13–20 Hz), high-beta (20–30 Hz), and gamma (30–50 Hz). The log-transformed power values within each band were averaged in each window, yielding a time–frequency data matrix with varying time size but fixed frequency size for each trial. After that, the information on the time window was averaged, leaving channel (31) and frequency (6) information for each trial.

The EEG dataset used in this study was aggregated from all EEG data of individual participants, where there were 50 trials of data on average for each participant. The aggregation yielded the EEG data of 2629 trials in total from 51 subjects. As described in the previous section, the analysis of EEG resulted in 186 features, including EEG spectral power values in 6 frequency bands at 31 EEG channels (i.e., electrodes) calculated in each trial. The size of the feature matrix was then 2629 × 186 (number of trials (samples) × number of features). Each trial (sample) in this data matrix was assigned the label of 1 if participants recognized facial expression correctly or 0 if they did incorrectly.

3.2. Feature Selection

By using the feature selection methods, the influence of each independent variable on the dependent variable was calculated, and sorting was performed in the order of the largest influence. In order to derive the model result according to the change in the number of accumulated variables, the accuracy of the model was measured by adding major variables one by one in order. Through the procedure, the accuracy that changes as individual variables are added can be measured, and the number of accumulated variables with the highest classification accuracy can be identified. In this study, the combination with the highest classification accuracy was selected as the main feature.

In addition to the method of adding a single variable in consideration of the interaction between the main variables, the method of selecting and combining the top 10 variables with high influence was additionally used. A combination was created and used from 10 variables selected through each feature selection method, and the combination variable with the highest classification accuracy was identified. However, when comparing the first feature selection method with the second feature selection method, the accuracy of the second method was low, so in this study, the combination created by gradually adding a major single variable was finally utilized.

3.3. Computing Environment

The computing environment used in this study is detailed in Table 1. We used Python language to construct feature selection methods and machine learning methods. The Pandas package and the Numpy package were mainly used for data preprocessing and preparation for analysis, and the Sklearn package was additionally used. The random value generated during the sampling process during the study was fixed to '101'. The train set and test set used a ratio of 7:3. In addition, if you use Google Colaboratory (Google

colab), you can conduct the same experiment as in this study without having to build a separate environment.

Table 1. H/W and library specification. Indicates hardware specifications and detailed Python library versions used in this study.

Type	Item	Specification
	CPU	Intel Core i7-8700
Hardware	GPU	GeForce GTX 1060 3 GB
	RAM	16 GB
	Pandas	1.1.3
Library	Numpy	1.19.2
	Scikit-learn	0.23.2
Language	Python	3.8.5

3.4. Experiment Diagram

Figure 2 shows the diagram for the experiment. The first stage, data acquisition stage, performs data collection and preprocessing through experiments. In the next step, the feature selection methodology is applied. The priority of features is checked through the feature selection methodology. Then, classification modeling is performed using machine learning methodologies. The extracted features are used for classification modeling. The performance of a machine learning model is checked through its accuracy. Then, the model with the highest classification accuracy is analyzed. In the final stage, we analyze how key features are located in scalp topography and which sub-bands are mainly used.
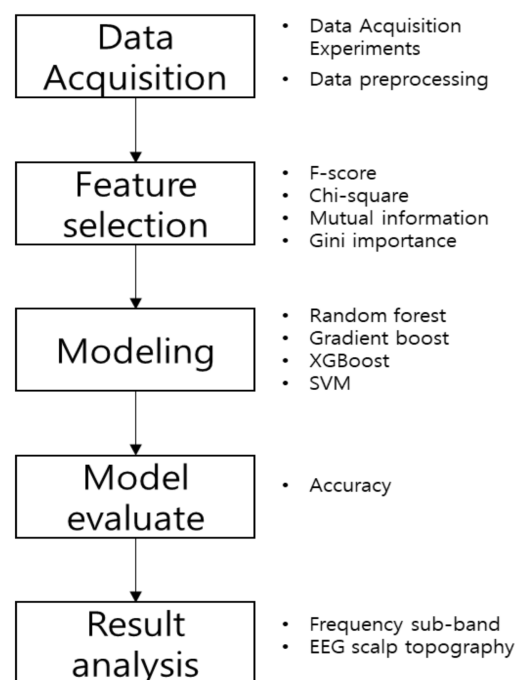
Figure 2. Diagram representing the experimental process. It represents the various methodologies and work content utilized at each stage.

3.5. Base Score

Base score is the classification accuracy of basic machine learning methods. By comparing the feature selection-based accuracy and the base score, we checked whether the feature selection methods affect the accuracy improvement. It was calculated using the aforementioned machine learning methods. After fixing the random seed, classes were classified

through each method. In this study, accuracy was utilized to evaluate the performance of the model. The expression for accuracy is as in Equation (8).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

True positive (**TP**) and true negative (**TN**) mean that the predicted result of the model and the actual value are the same or different. False positive (**FP**) and false negative (**FN**) mean the opposite of TP and TN. Accuracy is the number of correctly predicted data divided by the total number of data. The classification results are shown in Table 2.

Table 2. The accuracy of the ensemble model that used several models together was the highest.

Data	Random Forest	Gradient Boost	XGBoost	SVM	Ensemble
BCI_186	0.681	0.682	0.680	0.680	0.690

As a result of calculating the base score, the average accuracy of each model was about 68%. In this study, the standard score was set to 69%, and the improvement of the performance with the feature selection was tested.

4. Results

This section describes the results of the study. It describes the results of variable selection through feature selection, classification results for each model, and the results of BCI data analysis through major variables. The results of data classification are presented in detail in the form of a table, and EEG topography is also presented. This section describes the key variables selected and their regions of expression in the brain.

4.1. Result of Feature Selection

4.1.1. Result of Adding Influential Feature

Feature influence was measured through several feature selection methods. After sorting the features in the order of the measured influence, one variable from highest to lowest influence was cumulatively added one by one for each trial to derive the classification result. Both the training data and the test data were fixed using the same random seed value. The classification results from applying the feature selection methods are shown below in Table 3.

Table 3. The result of calculating the number of features showing the highest classification performance when classified while adding features one by one.

Feature Selection & Classification Model		F-Score	Chi-Square	Mutual Information	Gini Importance	Ensemble
Random Forest	Accuracy	69.4%	69.2%	69.2%	69.0%	70.0%
	The number of features	38	30	80	111	157
Gradient Boost	Accuracy	69.6%	69.0%	70.0%	69.2%	69.2%
	The number of features	34	41	8	7	129
XGBoost	Accuracy	70%	**71.0%**	69%	70.0%	70.0%
	The number of features	76	**52**	6	8	73
SVM	Accuracy	68.3%	68.3%	68.3%	68.3%	68.3%
	The number of features	1	1	1	1	1

Table 3 shows the results of machine learning techniques applied with the feature selection method. Most of the algorithms, except SVM, have improved classification accuracy through feature selection methods. In particular, the XGBoost model based on the top 52 features extracted through chi-square showed the highest performance with 71% classification accuracy. This is a result of using only about 36% of the total features, and

it can be seen that it not only improved accuracy but also improved the computational efficiency of the model.

4.1.2. Result of Feature Subset

All possible combinations were created using the top 10 features selected through feature selection methods. The accuracy of the classification model was calculated using a combined subset of features. As seen in the previous experiment, the dataset was fixed through random seeds. The classification results are shown in Table 4.

Table 4. The result of constructing a combination using 10 main features and extracting the feature with the highest classification performance.

Feature Selection and Classification Model		F-Score	Chi-Square	Mutual Information	Gini Importance	Ensemble
Random Forest	Accuracy	69.4%	68.3%	68.1%	70.0%	69.2%
	Subset of features	73, 153, 95, 25, 124, 127, 29	95, 25, 73, 10, 94, 79	11, 2, 90, 137, 180	152, 148, 64, 127, 126, 141, 151	84, 139, 69, 130, 93, 12, 63
Gradient Boost	Accuracy	69.4%	69.4%	69.4%	**70.3%**	69.2%
	Subset of features	10	10	23, 50, 2, 90, 137	**152, 153, 64, 27**	148, 69, 93, 12
XGBoost	Accuracy	69%	69.0%	70.0%	70.0%	69.4%
	Subset of features	73, 95, 29	124, 73, 10	50, 11, 2, 137	152, 153, 64, 27, 151	83, 130, 64
SVM	Accuracy	68.4%	68.4%	68.3%	68.4%	68.3%
	Subset of features	153, 95, 10	153, 95, 10	23	152, 153	84

Table 4 shows the results of machine learning methods using main feature subsets. Table 4 shows lower accuracy compared with Table 3. Most of the classification accuracy shown in Table 4 is similar to that of the base score. However, the accuracy of the combination of features using Gini importance tends to be somewhat high. In particular, when using gradient boost, it shows an accuracy of 70.3% through four features. It has a somewhat lower classification accuracy than that of the previous experiment but can bring a great effect in terms of computational efficiency.

From Tables 3 and 4, it can be seen that applying the feature selection method can have a great advantage in terms of accuracy and computational efficiency. In particular, when using chi-square and XGBoost, there was a 2% improvement of the accuracy compared to the base score. When Gini importance and gradient boost were used, a classification accuracy of 70.3% was obtained with four features.

4.2. Result of BCI Interaction

In the model with the highest performance (71%, chi-square and XGBoost), we counted the number of selected features for each EEG channel and frequency sub-bands to examine the spatial distribution of the selected features over the whole brain. We observed that the selected features were widely distributed over almost all EEG channels (26 out of 31 channels, Figure 3a), indicating that there were no specific EEG channels and brain areas that dominantly generated key features for classification. We additionally divided the entire channels into eight brain areas to compare the number of selected features among different brain areas: prefrontal (Fp), frontal (F), fronto-central (FC), temporal (T), central (C), centro-parietal (CP), parietal (P), and occipital (O) areas. Then, we calculated the average number of selected features per channel in each area as the number of channels varied across the areas. As a result, the fronto-central (FC) area exhibited the largest average number of selected channels, followed by the prefrontal (Fp) area showing the second largest (Fp: 2.3, F: 1.2, FC: 2.5, T: 1.75, C: 2, CP: 1.25, P: 1.4, O: 1.3). Accordingly, we found that while most brain areas provided EEG features useful for classification, the fronto-central and prefrontal areas contributed the most.

(a)

(b)
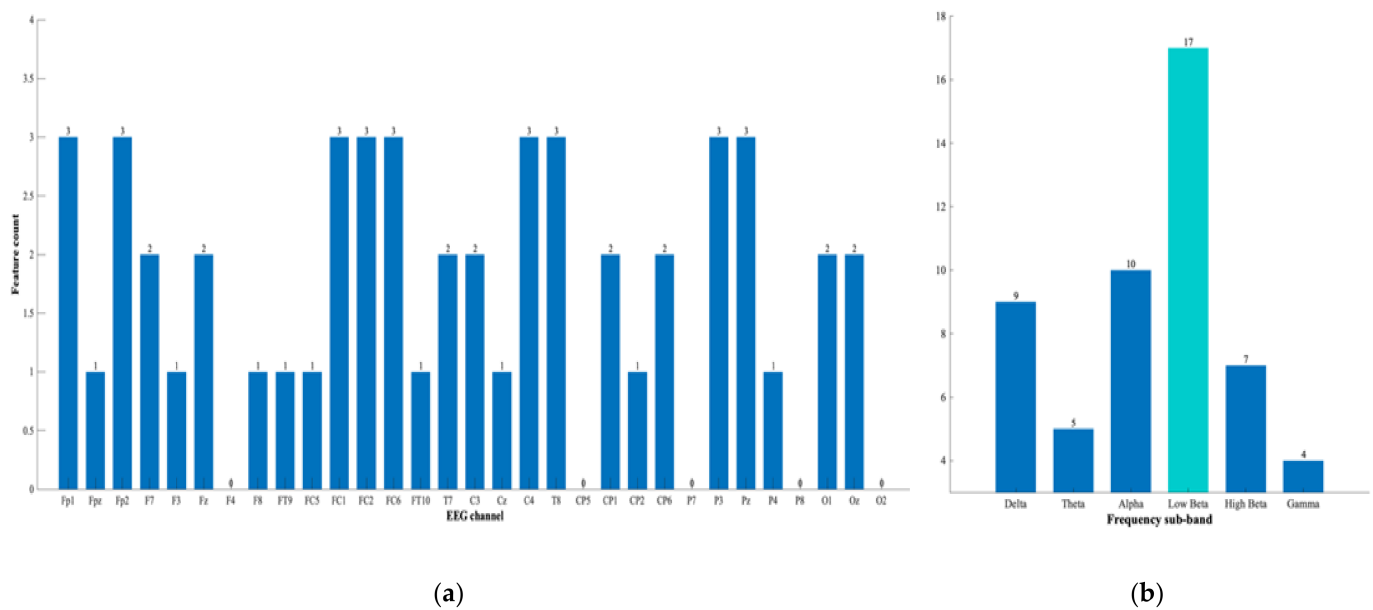
Figure 3. Feature count distribution. The features selected from the model with the highest accuracy (71%, chi-square and XGBoost) are counted for each dimension. (**a**) EEG channel. It is difficult to find the dominant spatial pattern; (**b**) Frequency sub-band. The low-beta, the frequency band with the highest feature count, is shown in light blue.

Next, we analyzed the spectral distribution of the number of selected features over frequency sub-bands given as follows: delta, theta, alpha, low-beta, high-beta, and gamma. The largest number of selected features was found in the low-beta band (Figure 3b). Low-beta (12–20 Hz) activity is known to be associated with increased mental states of high engagement, performance, and concentration [36,37], as well as with emotional stimulus processing and social interactions [38,39]. Therefore, the result here indicates that the modulation of low-beta activity may underlie the correct recognition of facial expression of others.

Figure 4 depicts the EEG scalp topography of chi-square selection importance rank score in the low-beta band. We calculated the importance rank score as shown in Equation (9) in the chi-square variable selection:

$$Importance\ rank = 1 - \left(\frac{ascending\ order}{feature\ size} \right) \qquad (9)$$

where an ascending selection order indicates the order in which a given feature was selected (from 1 to the number of features) and a feature size denotes the total number of features. Note that the importance rank score ranges from 0 to 1. For instance, if the chi-square variable selection results in the feature index in the following order: 153, 95, 10, etc. and the total number of selected features is 186, the importance rank score of the feature 153 is 0.9946 and that of the lastly selected feature is 0. The topography demonstrates that low-beta activity at the frontal and parietal areas tended to yield more important features for classification. This result could be related with the role of working memory (WM) in the recognition of facial expressions. It is known that loading into WM reduces the plasticity of facial expression recognition, leading to the false perception of facial stimuli [40]. As the prefrontal and parietal areas are primarily engaged in WM [41], our observation of the distribution of important features over fronto-parietal areas may reflect the operation of WM in the facial expression recognition. In addition, low-beta activity of parietal areas provides a substrate of WM that can synergistically integrate sensory information and executive commands from frontal areas [41]. Accordingly, our result of important low-beta features over fronto-parietal areas may indicate how well the visual information of facial expression is integrated with the function of emotional change detection in WM.
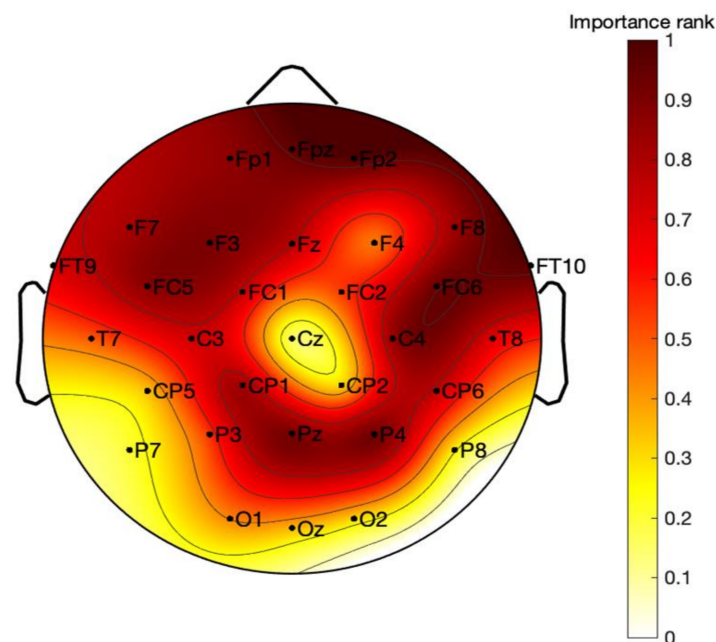
Figure 4. EEG scalp topography of chi-square selection importance rank score in low-beta band. The importance rank is calculated from Equation (9). For example, if low-beta and Fp2 is the 26th variable in ascending order, the importance rank is calculated with $1 - \left(\frac{26}{186}\right)$. The variable extracted first has the highest importance ranking score, resulting in a reddish color on the topography.

5. Discussion

The pattern of EEG spectral features selected in the study indicated that the fronto-parietal Low-Beta frequency bands appeared to contain useful information to predict whether a person recognized emotions of others correctly or incorrectly from dynamic facial expressions. A number of studies have related low-beta rhythms to emotion recognition from facial expressions. A study showed that affective pictures modulate event-related beta oscillations such that event-related synchronization (ERS) of beta oscillations vary with emotional valence, indicating that beta ERS distinguishes early bottom-up processing of visual, emotional stimuli [42]. In particular, these beta ERD processes may mark some of the deficiencies seen in autism spectrum disorder (ASD), which suffers from problems with social cognition and affective processing [43]. Considering facial expression processing in terms of decision-making, perceptual decision-making is based on sensory evidence, and hence prefers more definite sensory input to the brain. In perceptual decision-making, fronto-parietal beta oscillations during the stimulus processing are suggested as a marker of top-down attentional mechanisms that control the accumulation of decision-relevant sensory information [44]. Maksimenko et al. (2020) suggested that increased power of fronto-parietal beta oscillations is related to decision-making process by reflecting the processing of ambiguous stimuli [45].

Since the present study focused on a feasibility to predict whether or not a person recognizes emotions from facial expressions of others correctly using only the person's EEG, we did not investigate if the prediction performance varied with the type of emotions (e.g., happiness, sadness, and disgust). It is possible that some emotions are easier to recognize than others (e.g., happiness could be easier to recognize than surprise). We will investigate the effect of emotion type on the prediction performance and related neural activity patterns in subsequent studies. In addition, we predicted the correctness of emotion recognition using EEG signals acquired right before participants detected a change in facial expression (i.e., when participants pressed the key). However, it is unclear if the actual decision-making process about the type of emotion occurred in this period or not because participants answered the emotional type after they pressed the key. With an incorrect answer, it remains elusive whether the error of emotion recognition

occurred during sampling sensory evidence before detection or during maintaining sensory information in working memory after detection (note that facial expression disappeared after key pressing). Our next study with modified experimental paradigms that can help differentiate these processes will address this question.

There are several limitations that need further investigation in the follow-up studies. First, we only investigated spectral features in the current study. However, facial expression recognition may involve interactions between the brain regions, as evident in our topographic analysis where we observed important features over frontal and parietal regions. Therefore, it is plausible that large-scale networking between these two regions may indicate how well a person recognizes facial expression, and appropriate features for the networking would require analyses in the domain of brain connectivity. Connectivity features can be extracted from EEG using well-known methods, including phase synchrony or Granger causality [46]. Second, the classification performance reached above 70% using the optimally selected features. This study used unique data to explore biologically meaningful outcomes that affect brain signals. The accuracy presented in this study can be improved through additional algorithmic research. Moreover, better research results can be obtained by collecting more samples through additional research. Third, the class size was imbalanced in our data as the subjects could correctly recognize facial expressions more often than they showed incorrect recognition. Therefore, the class size of correct recognition was bigger than that of incorrect recognition. Addressing this imbalance issue by any state-of-the-art method (e.g., generative adversarial network) may help improve classifiers [47].

The EEG features and classification algorithms proposed in this study can be implemented in a BCI that allows us to infer how well a person recognizes others' emotional changes. Such a BCI system will be useful to evaluate social interactions of individuals. It will allow us to examine the BCI user's EEG patterns during social interactions and estimate the user's ability to understand others' facial expressions. Furthermore, this BCI system may be used to develop an intelligent system, presumably integrated with artificial intelligence (AI), that can assist the BCI user to improve empathic skills. The BCI may read EEG patterns during social interactions, evaluate EEG features related to emotion recognition, and provide appropriate feedbacks via AI to the user such that the user can recognize others' facial expressions better. The future studies will address this system development by integrating BCI and AI together.

6. Conclusions

In the study, we proposed an EEG-based classification method to predict whether a person correctly recognizes facial expressions of others using machine learning algorithms with feature selection methods. The proposed method used feature selection methodologies to extract major features that affect the recognition of facial expression changes, and classified whether correct recognition of facial expression changes are detected by applying various machine learning algorithms. As a result of experimenting using EEG data from 51 subjects, we showed XGBoost with chi-square exhibited the best performance with 52 major features compared to other models. From the selected features, a dominant trend appeared in the low-beta EEG frequency band. As a result of scalp topography mapping of 52 selected major variables, it was shown that the recognition of facial expression changes mainly engages the frontal brain areas. In future research, more objective results will be derived by predicting various emotional types and exploring other EEG feature domains.

Author Contributions: S.K.—research idea, formulation of research goals and objectives, guidance and consulting, and examine of calculation results. S.-H.S.—analysis of literature, analysis of experimental data, validation of model, and draft and final copy of the manuscript. B.-K.P.—consulting and literature analysis. D.-Y.K.—medical consulting and literature analysis. S.S.—experimental design and psychological consulting. J.J.—obtaining experimental data, analysis of literature, analysis of experimental data, and part of the manuscript draft. S.-P.K.—obtaining experimental data, analysis

of experimental data, guidance and consulting, and part of the manuscript draft. All authors have read and agreed to the published version of the manuscript.

References

1. Yang, B.; Li, H.; Wang, Q.; Zhang, Y. Subject-based feature extraction by using fisher WPD-CSP in braincomputer interfaces. *Comput. Methods Programs Biomed.* **2016**, *129*, 21–28. [CrossRef]
2. Pfurtscheller, G.; Müller-Putz, G.R.; Schlögl, A.; Graimann, B.; Scherer, R.; Leeb, R.; Brunner, C.; Keinrath, C.; Lee, F.; Townsend, G.; et al. 15 years of BCI research at Graz University of Technology: Current projects. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2006**, *14*, 205–210. [CrossRef] [PubMed]
3. Dornhege, G.; del RMillán, J.; Hinterberger, T.; McFarland, D.J.; Müller, K.R. *Toward Brain-Computer Interfacing*; MIT Press: Cambridge, UK, 2007; ISBN 9780262042444.
4. Daros, A.; Zakzanis, K.; Ruocco, A. Facial emotion recognition in borderline personality disorder. *Psychol. Med.* **2013**, *43*, 1953–1963. [CrossRef]
5. Doma, V.; Pirouz, M. A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. *J. Big Data* **2020**, *7*, 7–18. [CrossRef]
6. Pan, J.; Li, Y.; Wang, J. An EEG-based brain–computer interface for emotion recognition. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016. [CrossRef]
7. Pantic, M.; Rothkrantz, L.J. Automatic analysis of facial expressions: The state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1424–1445. [CrossRef]
8. Shariat, S.; Pavlovic, V.; Papathomas, T.; Braun, A.; Sinha, P. Sparse dictionary methods for EEG signal classifcation in face perception. In Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing, Kittila, Finland, 29 August–1 September 2010. [CrossRef]
9. Oskoei, M.A.; Gan, J.Q.; Hu, H. Adaptive Schemes Applied to Online SVM for BCI Data Classification. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2010. [CrossRef]
10. Isa, N.E.M.; Amir, A.; Ilyas, M.Z.; Razalli, M.S. Motor imagery classification in Brain computer interface (BCI) based on EEG signal by using machine learning technique. *Bull. Electr. Eng. Inform.* **2019**, *8*, 269–275. [CrossRef]
11. Ewan, S.N.; Philippa, J.K.; David, B.G.; Dean, R.; Freestone, A. Generalizable Brain-Computer Interface (BCI) Using Machine Learning for Feature Discovery. *PLoS ONE* **2015**, *10*, e0131328. [CrossRef]
12. Acı, Ç.İ.; Kaya, M.; Mishchenko, Y. Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods. *Expert Syst. Appl.* **2019**, *134*, 153–166. [CrossRef]
13. Abraham, A.; Pedregosa, F.; Eickenberg, M.; Gervais, P.; Mueller, A.; Kossaifi, J.; Gramfort, A.; Thirion, B.; Varoquaux, G. Machine Learning for Neuroimaging with Scikit-Learn. *Front. Neuroinform.* **2014**, *8*, 14. [CrossRef]
14. Tan, P.; Sa, W.; Yu, L. Applying Extreme Learning Machine to classification of EEG BCI. In Proceedings of the IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems (CYBER), Chengdu, China, 19–22 June 2016. [CrossRef]
15. Kaper, M.; Meinicke, P.; Grossekathoefer, U.; Lingner, T.; Ritter, H. BCI competition 2003-data set IIb: Support vector machines for the P300 speller paradigm. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1073–1076. [CrossRef]
16. Patil, A.R.; Chang, J.; Leung, M.; Kim, S. Analyzing high dimensional correlated data using feature ranking and classifiers. *Comput. Math. Biophys.* **2019**, *7*, 98–120. [CrossRef]
17. Thaseen, S.; Kumar, C.A.; Ahmad, A. Integrated Intrusion Detection Model Using Chi-Square Feature Selection and Ensemble of Classifiers. *Arab. J. Sci. Eng.* **2019**, *44*, 3357–3368. [CrossRef]
18. Hoque, N.; Bhattacharyya, D.; Kalita, J. MIFS-ND: A mutual information-based feature selection method. *Expert Syst. Appl.* **2014**, *41*, 6371–6385. [CrossRef]
19. Genuer, R.; Poggi, J.; Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **2010**, *31*, 2225–2236. [CrossRef]
20. Ding, S. Feature Selection based F-score and ACO Algorithm in Support Vector Machine. *Int. Symp. Knowl. Acquis. Model.* **2009**, *1*, 19–23. [CrossRef]

21. Jlmaz, E.Y. An Expert System Based on Fisher Score and LS-SVM for Cardiac Arrhythmia Diagnosis. *Comput. Math. Methods Med.* **2013**, *2013*, 849674. [CrossRef]
22. Gu, Q.; Li, Z.; Han, J. Generalized Fisher Score for Feature Selection. *arXiv* **2012**, arXiv:1202.3725.
23. Morison, K.; Wang, L.; Kundur, P. Power system security assessment. *IEEE Power Energy Mag.* **2004**, *2*, 30–39. [CrossRef]
24. Forman, G. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* **2003**, *3*, 1289–1305.
25. Ali, S.I. A Feature Subset Selection Method based on Conditional Mutual Information and Ant Colony Optimization. *Int. J. Comput. Appl.* **2012**, *60*, 5–10.
26. Jović, A.; Brkić, K.; Bogunović, N. A review of feature selection methods with applications. In Proceedings of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 25–29 May 2015. [CrossRef]
27. Anju, N.; Sharma. Survey of Boosting Algorithms for Big Data Applications. *Int. J. Eng. Res. Technol. (IJERT)* **2017**, *5*.
28. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A Comparative Analysis of XGBoost. *arXiv* **2019**, arXiv:1911.01914.
29. Rahman, S.; Irfan, M.; Raza, M.; Ghori, K.M.; Yaqoob, S.; Awais, M. Performance Analysis of Boosting Classifiers in Recognizing Activities of Daily Living. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1082. [CrossRef]
30. Wang, X.; Chen, J.; Wang, P.; Huang, Z. Infrared Human Face Auto Locating Based on SVM and A Smart Thermal Biometrics System. *Int. Conf. Intell. Syst. Des. Appl.* **2006**, *2*, 1066–1072. [CrossRef]
31. Ganapathiraju, A.; Hamaker, J.E.; Picone, J. Applications of support vector machines to speech recognition. *IEEE Trans. Signal Process.* **2004**, *52*, 2348–2355. [CrossRef]
32. Ha, H.; Shim, E.-J. Differences in Facial Emotion Recognitions According to Experiences of Childhood Maltreatment. *Korean Stud.* **2018**, *29*, 97–123. [CrossRef]
33. Delorme, A.; Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **2004**, *134*, 9–21. [CrossRef] [PubMed]
34. Kim, S.P. Preprocessing of EEG. *Comput. EEG Anal.* **2018**, 15–33. [CrossRef]
35. Pion-Tonachini, L.; Kreutz-Delgado, K.; Makeig, S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* **2019**, *198*, 181–197. [CrossRef] [PubMed]
36. Abhang, P.A.; Gawali, B.W.; Mehrotra, S.C. Technical aspects of brain rhythms and speech parameters. *Introd. EEG-Speech-Based Emot. Recognit.* **2016**, 51–79. [CrossRef]
37. Kropotov, J.D. *Quantitative EEG, Event-Related Potentials and Neurotherapy*; Academic Press: Cambridge, MA, USA, 2010. [CrossRef]
38. Schubring, D.; Schupp, H.T. Emotion and brain oscillations: High arousal is associated with decreases in alpha-and lower beta-band power. *Cereb. Cortex* **2021**, *31*, 1597–1608. [CrossRef]
39. Berntsen, M.B.; Cooper, N.R.; Romei, V. Emotional valence modulates low beta suppression and recognition of social interactions. *J. Psychophysiol.* **2019**, *34*, 235–245. [CrossRef]
40. Kostandov, E.A.; Cheremushkin, E.A.; Yakovenko, I.A.; Ashkinazi, M.L. The role of the context of cognitive activity in the recognition of facial emotional expressions. *Neurosci. Behav. Physiol.* **2012**, *42*, 293–301. [CrossRef]
41. Gelastopoulos, A.; Whittington, M.A.; Kopell, N.J. Parietal low beta rhythm provides a dynamical substrate for a working memory buffer. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 16613–16620. [CrossRef] [PubMed]
42. Zhang, W.; Lu, J.; Liu, X.; Fang, H.; Li, H.; Wang, D.; Shen, J. Event-related synchronization of delta and beta oscillations reflects developmental changes in the processing of affective pictures during adolescence. *Int. J. Psychophysiol.* **2013**, *90*, 334–340. [CrossRef]
43. Cooper, N.R.; Simpson, A.; Till, A.; Simmons, K.; Puzzo, I. Beta event-related desynchronization as an index of individual differences in processing human facial expression: Further investigations of autistic traits in typically developing adults. *Front. Hum. Neurosci.* **2013**, *7*, 159. [CrossRef] [PubMed]
44. Buschman, T.J.; Miller, E.K. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* **2007**, *315*, 1860–1862. [CrossRef]
45. Maksimenko, V.A.; Kuc, A.; Frolov, N.S.; Khramova, M.V.; Pisarchik, A.N.; Hramov, A.E. Dissociating Cognitive Processes During Ambiguous Information Processing in Perceptual Decision-Making. *Front. Behav. Neurosci.* **2020**, *14*, 95. [CrossRef] [PubMed]
46. Dauwels, J.; Vialatte, F.; Musha, T.; Cichocki, A. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *Neuroimage* **2010**, *49*, 668–693. [CrossRef]
47. Sampath, V.; Maurtua, I.; Aguilar Martín, J.J.; Gutierrez, A. A survey on generative adversarial networks for imbalance problems in computer vision tasks. *J. Big Data* **2021**, *8*, 27. [CrossRef] [PubMed]