Gender Classification of EEG Signals using a Motif Attribute Classification Ensemble

Jean Li
Dept. Information Sci.
University of Otago
Dunedin, New Zealand
jeanli1231@gmail.com

Jeremiah D. Deng
Dept. Information Sci.
University of Otago
Dunedin, New Zealand
jeremiah.deng@otago.ac.nz

Dirk De Ridder
Dept. Surgical Sci.
University of Otago
Dunedin, New Zealand
dirk.deridder@otago.ac.nz

Divya Adhia
Dept. Surgical Sci.
University of Otago
Dunedin, New Zealand
divya.adhia@otago.ac.nz

***Abstract*—Effective analysis of EEG signals remains a challenging task. So far, the analysis and conditioning of EEG have largely remained gender-neutral. This paper explores the evidence of gender effects on EEG signals and confirms the generality of these effects by achieving successful gender prediction through EEG signals. Specifically, we propose a novel statistical feature representation that captures the gender discrepancy, and design a customized classification ensemble framework to overcome the non-stationary characteristics in EEG signals, utilizing findings obtained through several machine learning techniques including clustering, visualization, and metric learning. Apart from gender differentiation, the age effect on EEG gender patterns is also revealed.**

***Index Terms*—EEG, gender, age, clustering, classification, ensemble**

I. Introduction

Electroencephalography (EEG) is a measurement of multi-channel potentials that reflects the electrical activity of the human brain. The study of the brain's electrical activities through the EEG records is one of the most important tools for the diagnosis of neurological diseases, such as epilepsy, brain tumors, head injury, sleep disorders and dementia [1]. Compared with other non-stationary time series data, EEG signals demonstrate a high noise-to-information ratio. The signals can be hugely affected by a series of artifacts, i.e., EEG characteristics that differ from signals generated by activities in the brain [2]. Common artifacts include eye movements, jaw tension, and muscle contractions. Also, EEG signals are highly individual-specific. The EEG associated with seizure onset in one patient may closely resemble a benign pattern within the EEG of another patient [3]. This brings challenges to effective cross-subject pattern identification.

Research into gender influences is imperative to fully understand a host of brain disorders with gender differences in their incidence and/or nature [4]. Brains of men and women are different, both genetically determined and epigenetically adjusting to environmental factors [5]. This is partly mediated via a gender-specific release of hormones as well as differential involvement of the immune system, with microglia masculinizing and T cells feminizing brains. This, in turn, leads to a gender-specific difference in the prevalence of brain disorders: males are more prone to brain disorders arising during development such as autism (5x), conduct disorder (3x), ADHD (3x), schizophrenia (1.5x), dyslexia (3x), stuttering (2.3x) and Tourette's syndrome (4x), whereas females have more adult-onset brain disorders such as major depression (x2), anxiety (2x), panic disorders (2.5x), OCD (1.5x), PTSD (2x), bulimia (4x), migraine (3x), MS (2x), myasthenia gravis (4x) and Alzheimer's disease (2x) [5]. To unravel the underlying brain mechanisms behind these clinical differences it may be of relevance to understand the underlying differences in electrical brain activity between men and women. Detection of EEG gender patterns may enable a better understanding of the gender differences in the development of certain neuropsychiatric diseases, responses to treatments [6], and the bio-psycho-social-behavioral roles; which will, in turn, enable the development of novel targeted gender-specific interventions. We, therefore, embarked on a study using machine learning to extract differences between the two sexes using electroencephalographic data from a database of normal healthy people.

This study focuses on using machine learning techniques to explore evidence of gender effects on EEG signals and attempt a novel classification model to predict the gender of a subject through their EEG signals.

This paper is organized as follows. In Section II, we review related work and discuss some common approaches to represent EEG signals. Section III introduces the data preprocessing procedures and presents the EEG analysis outcome, justifying our modeling approach. Then in Section IV, we introduce our four-layer customized gender classification model and present the results on gender classification. We conclude the paper in Section V.

II. Related work

Previous studies have found some evidence of the gender effects on EEG. In [7], the EEG signals of 80 individuals between the ages of 8 and 12 years were analyzed. Differences between genders were found in this study group, with males having less theta but more alpha signals than females. Females were also found to have a developmental lag in the EEG compared to males. Another study [8] also reported sex differences in EEG asymmetry during self-generated cognitive and affective tasks. The study in [9] investigated the effects of age

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

and gender on sleep EEG power spectral density of individuals of age ranged from 20–60 years. The average power density within the 4-second epochs was calculated. It was found that females show significantly higher spectral power density in some power bands than males. Though significant effects of age on sleep EEG spectral power density were found, the study did not find any interaction between age and gender. However, the above studies did not validate the findings on external test individuals, therefore, the generality of the findings may be limited.

Although EEG signal classification has been widely explored for different purposes, such as disease diagnosis, only till recently, have a few studies [10] [11] attempted EEG-based gender classification. For classification purposes in general, the EEG signals are typically preprocessed by applying both band-pass filters and spatial filters before feature extraction. The most common types of features to present EEG signals are frequency band power features and time point features. Band power features represent the average energy level of EEG signals within a certain frequency range over a given time window called an epoch. Band power features need to be extracted respectively in each channel. Timepoint features are a concatenation of EEG signals from all channels, and they are typically used for event-related potentials classifications [12]. Due to the non-stationarity of EEG signals, band power features should be extracted from a reasonably small epoch. For example, in [3], the band power features are extracted within a sliding window with a length of 2 seconds. Spatial filters were also applied in other studies for EEG feature extraction. These can be obtained in a supervised manner, such as Common Spatial Patterns (CSP). CSP projects the signals into another matrix space that maximizes the distance between 2 classes. Reference [13] discusses the effectiveness of this approach and has proven it to be useful. Spatial filters can also be obtained through an unsupervised way such as Independent Component Analysis (ICA). In addition to the above, other EEG representing methods are also studied, these include sparse representation and deep learning. The sparse representation-based classification (SRC) method has shown a robust classification performance [14]. Deep learning, in which the features and the classifier are jointly learned directly from the EEG signals. The convolutional neural networks and restricted Boltzmann machines are the two most popular deep learning methods for EEG-based Brain-Computer Interfaces (BCIs) studies [12].

III. EEG DATA ANALYSIS

In this section, we will introduce the EEG dataset used for this study, and present the analysis of the overall dataset. Preliminary evidence of gender and age effects on EEG signals will be presented. Our analysis also suggests that individual-wise gender classification is more plausible than gender classification based on single epochs.

A. Dataset description

The dataset used for this study consists of a cleaned resting-state EEG streams set of 130 healthy individuals with a sampling rate of 128 Hz. The raw EEG signal was collected through a standard Mitsar amplifier with 19 channels (Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2). A high-pass filter of 0.15 Hz and a low-pass filter of 200 Hz were applied, followed by a careful inspection that manually removes the artifacts signal Sections include eye movements, jaw tension, teeth clenching, etc. The studied population consists of 85 females and 45 males age between 25 to 55 years. The distribution regarding the age and gender of the population can be found in Table I.

TABLE I
AGE AND GENDER OF THE SAMPLE SET

Age (years)	***Females***	***Males***
25-35	17	11
36-45	34	14
46-55	34	20
Sum	85	45

Due to the imbalanced sample numbers of male and female groups, data points from the minority group in training sets will be over-sampled for any classification.

B. Band power features

For feature extraction, we employ the band power features. These are promoted as the "gold standard" for many BCI applications involving the detection of mental states or emotions [12]. Our previous experiments showed that it outperforms CSP in visualization or classification tasks.

Since EEG signals cannot be segmented into physiologically relevant units, the conventional approach of segmenting the EEG streams into epochs according to time interval is adopted in this study. For each channel, the energy level falling within certain passbands was measured over each epoch. There are five major brain waves distinguished by their different frequency ranges. These frequency bands from low to high frequencies, respectively, are typically categorized in specific bands as 0.5–4 Hz (delta), 4–8 Hz (theta), 8–13 Hz (alpha), 13–30 Hz (beta) and $>$30 Hz (gamma) [1].

The length of the epochs determines the frequency range it covers. Longer epochs can capture a lower frequency range. As we are not clear in which frequency would the gender motif (if it exists) appears, a longer epoch is preferred in this aspect. However, EEG signals are highly non-stationary, if the epochs are too long, it may fail to capture the brain actively. To balance between the ability to capture the lower frequency range and the non-stationary nature of EEG signals, the epoch length of 4 seconds (frequency resolution of 0.25Hz) and 8 seconds (frequency resolution of 0.125Hz) are considered.

A preliminary epoch-wise gender classification through a multi-layer perceptron neural network model was performed to compare the effects of different epochs and band power features. As represented in Table II, different segmentation

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

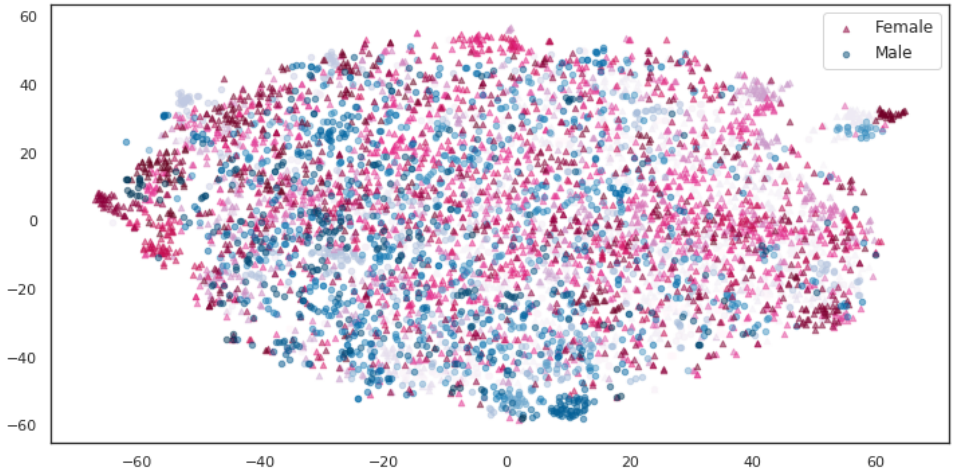

Fig. 1. Visualization of epoch features (normalized band power) from either a Male (blue dot) or Female (red triangle) subject. Darker shades indicate the corresponding subjects are older.

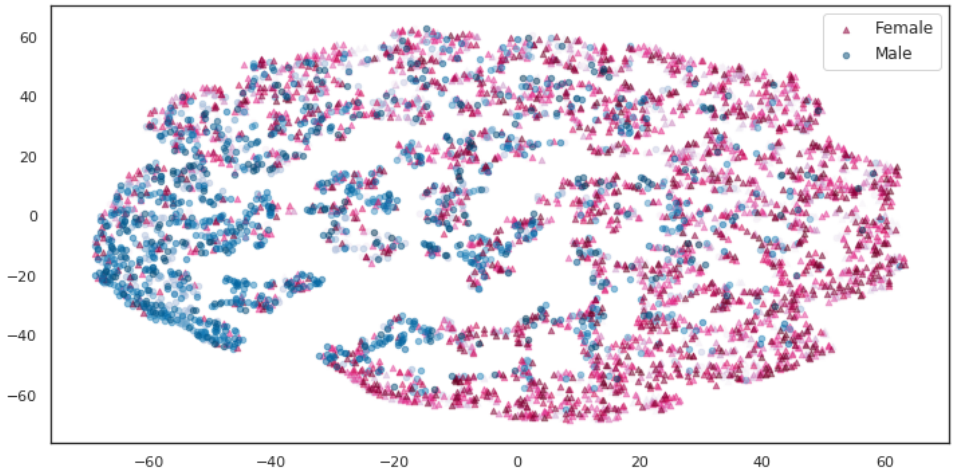

Fig. 2. Visualization of epoch features using clustering information. Darker shades indicate the corresponding subjects are older.

TABLE II
COMPARISON OF DIFFERENT EPOCH SEGMENTATION SCHEMES.

Epoch details	**test score mean(std)**	
	trained with 4 band power features[a]	***trained with 5 band power features***[b]
4 second epochs with 2 second sliding intervals	0.61(0.05)	0.63(0.05)
8-second epochs with 2 second sliding intervals	0.63(0.05)	0.63(0.06)
8-second epochs with 4-second sliding intervals	0.64(0.05)	0.63(0.06)

[a]4 band power include: delta, theta, alpha and beta
[b]5 band power include: delta, theta, alpha, beta and gamma

schemes of EEG signals with different band power range (4 band power: delta, theta, alpha, and beta; 5 band power: delta, theta, alpha, beta, and gamma) result in very similar test scores. Hence for computational considerations, the scheme using half-overlapping 8-sec epochs (with 4-sec sliding intervals) is adopted. This gives us 76 band power features (4 band power each channel, and there are 19 channels) for each epoch data point. The following analysis will be based on this epoch segmentation strategy.

C. Visualization of epoch data

The poor performance of the brute-force epoch-wise classification motivates us to visualize the epoch feature data and see if there exist any gender "motifs" for these EEG epochs. Fig. 1 presents the epoch-wise t-SNE visualization of epoch data represented by the normalized band power features. The t-SNE algorithm visualizes high-dimensional data by embedding the data points to a low-dimensional space, using gradient descent to reduce the Kullback-Leibler divergence between the original data and the projected data [17]. Each data point in Fig. 1 represents one epoch in the 76-D feature space. A red triangle represents this epoch is from a female, a blue circle represents

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

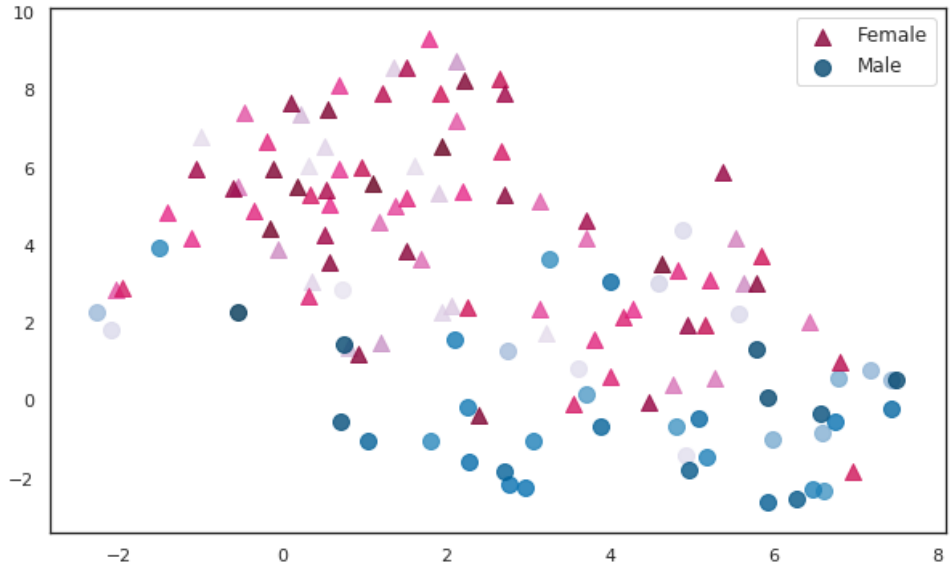

Fig. 3. Individual-wise visualization using the histogram vector of matched cluster centres.

the epoch is from a male, and a darker shade symbol represents the epoch is from an older individual. The bottom edge of the plot is mainly female epochs. A slight age difference among females is shown, as the epochs of older females (dark red symbols) tend to group around the edges. However, no clear discrepancy between genders is presented in this plot. This suggests that classification on epochs directly would be rather ineffective.

An intuition to circumvent this issue is to introduce a clustering procedure for the input EEG band power features. With clustering, each epoch is assigned to a certain cluster in an unsupervised manner based on mutual similarities. Each cluster represents a specific EEG *prototype*. It is then expected that the EEG epochs from a subject may demonstrate some patterns for one gender and others for another. If so, the gender of an individual may be determined by the distribution of these prototypes over all epochs. Hence by introducing the clustering procedure, we can avoid putting a gender label on each epoch, but still be able to label the individuals.

Before clustering, we resort to metric learning to customize the similarity measurement between epoch features. Suppose there are two epoch feature vectors $\mathbf{x}_i$ and $\mathbf{x}_j$. If both of them are of the same gender, we denote the 2-tuple $(\mathbf{x}_i, \mathbf{x}_j) \in S$; otherwise $(\mathbf{x}_i, \mathbf{x}_j) \in D$. The distance metric between $\mathbf{x}_i$ and $\mathbf{x}_j$ is defined as a generalized Mahalanobis distance:

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T A (\mathbf{x}_i - \mathbf{x}_j)}, \tag{1}$$

where matrix A is to be optimized by

$$\begin{aligned} &\min_{\mathbf{M} \in \mathbb{S}_+^d} \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in S} d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \\ &\text{s.t.} \quad \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in D} d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \geq 1 \end{aligned} \tag{2}$$

The metric-learn package[1] is used to handle the metric learning task for clustering.

For clustering, the feature vector $\mathbf{x}$ is first transformed to $\mathbf{y} = A^{1/2}\mathbf{x}$, and then the K-means algorithm is applied to generate the cluster prototypes from the transformed features. Suppose we have K clusters formed, with K cluster centers $\mathbf{c}_i$, $i = 1, 2, \cdots, k$. We can now produce another visualization of the epoch data. Specifically, a clustering-based epoch representation is generated by the epoch feature's Euclidean distance to the cluster centers.

Fig. 2 presents the epoch-wise t-SNE plot using clustering information. A red triangle represents this epoch is from a female, a blue circle represents the epoch is from a male, and a darker shade symbol represents the epoch is from an older individual. Each epoch is represented by its distance to the cluster centers. Compared with the plot using band power features directly (Fig. 1), the new plot demonstrates a much-improved discrepancy between males and females. Details on extracting clustered features and selecting the number of clusters will be discussed in Section IV.

D. Visualization of individuals

While epoch-wise gender motifs cannot be found, we suspect that gender information may be hidden in the statistical distribution of the motifs generated by clustering. Hence for each subject, we construct a *motif attribute vector*, in the form of the histogram of winning clusters across all epochs.

Suppose all the epoch features of subject i form a set $S_i = \{\mathbf{y}_{ij}, j = 1, \cdots, N\}$ (N is the number of epochs). The motif attribute vector is defined as

$$\mathbf{m}_i = h(\text{argmin}_k \|\mathbf{c}_k - \mathbf{y}_{ij}\|, \forall \mathbf{y}_{ij} \in S_i), \tag{3}$$

where $h(.)$ stands for the histogram operation.

Therefore, for subject-wise visualization, each individual-wise data point can be represented by the histogram vector

[1]URL http://contrib.scikit-learn.org/metric-learn/index.html

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

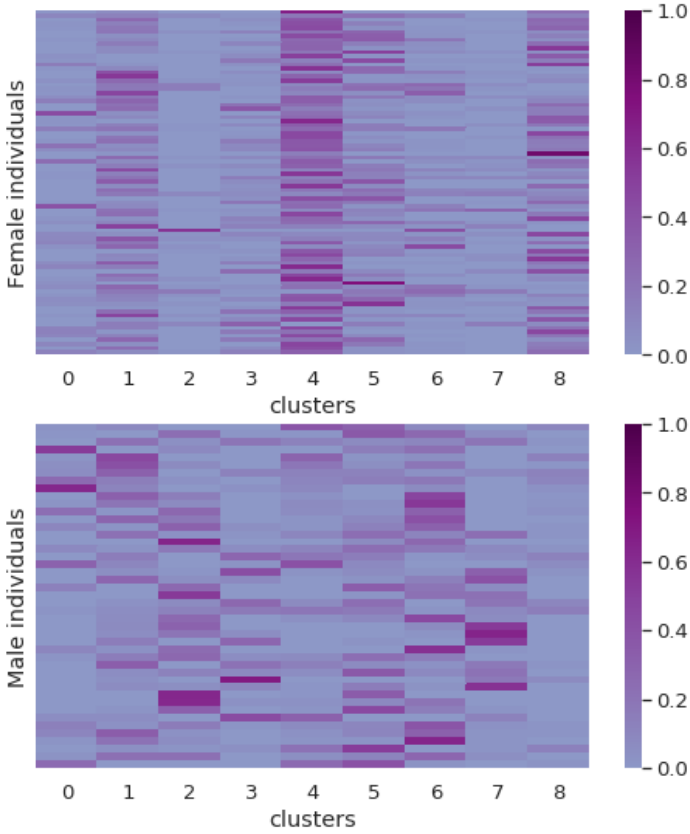

Fig. 4. Heat map of individual's epochs distribution

$\mathbf{m}_i$. Fig. 3 illustrates the t-SNE plot of individual-wise data points after clustering. Individual-wise gender discrepancy is clearly observed in this plot. This suggests the feasibility of individual-wise classification.

We now present more visualization evidence to support our motif attribute design. Depending on how frequent an individual's epochs hit the clusters, we can generate a *heat map* for male and female subjects, as shown in Fig. 4, which demonstrates the epoch distribution of the individuals when all the epochs are categorized into 9 clusters. Each row in the map represents the epochs distribution of one individual. Darker shades represent a higher percentage in this cluster. The individual is sorted by age. On the map, the age of individuals is increasing from top to bottom. For female individuals in general, a considerably high proportion of their EEG epochs are in cluster 4, followed by a relatively high proportion in clusters 1 and 8. Except for a few outliers, females rarely have a high percentage of their epochs distributed in clusters 2 and 7. Compared to female individuals, less consistency among male individuals is shown on the heat map, except they all have a small proportion of EEG epochs in cluster 8. Not all males manifest a strong proportion of their EEG epochs in clusters 2 or 7. These patterns tend to be the least female orientated. A significant number of younger male individuals (upper part of the second heat map) show a high proportion of their EEG epochs in cluster 1, which is a pattern generally possessed by females.

IV. Gender classification

Now that the motif attribute vectors can be effectively visualized and display good separability between genders, we proceed to construct a multi-stage Motif Attributes Classification Ensemble (MACE) framework, as presented in Algorithm 1.

Algorithm 1 MACE

1: Segment EEG time series data into 8-sec epochs with 4-sec sliding intervals. And Extract band power features from each epoch.
2: Apply weakly supervised metric learning on the epochs; *vid.* Eq.(2).
3: Apply K-means clustering and assign cluster labels to each epoch.
4: Construct motif attribute vectors for each individual; *vid.* Eq.(3).
5: Train multiple classifiers on the motif attribute vectors.
6: Hard vote ensemble of the classifiers and output final classification.

A. Framework

The computational framework consists of four procedures.

First is feature extraction, in which epochs are extracted from the time series EEG signal. The data of each individual contains 19 time series data from 19 channels. Artifacts were manually removed from the EEG streams. The time series data from each channel are segmented into 8-second epochs with 4-second sliding intervals. Then the energy levels from 4 band range, i.e. delta, theta, alpha, and beta, were extracted by computing the absolute power through approximating the area under the curve. The epochs are then scaled to have unit norm. After extraction, each epoch is represented by 76 features (4 band power each channel times 19 channels). The 130 individuals in the dataset have an average of 44 epochs.

The second stage is clustering. This is introduced to address the issue of gender motifs possibly not existing in every epoch. It also transforms epoch-wise data points into attribute vectors for subject-wise classification. This allows for individuals to have gender-neutral epochs.

Weakly supervised metric learning [16] is first applied to the training set to enhance the performance of K-means clustering. Random pairs of epochs with their similarity information (same gender or not) are generated to learn a distance metric that respects these relationships. Both training and testing sets are then transformed over this learned metric. The transformed epoch points are then assigned to a certain cluster through K-means. We then construct the individual-wise features by calculating the percentage of one's epochs in each cluster. For example, if there are 3 clusters in total, and an individual has 10 epochs in cluster 1, 30 epochs in cluster 2 and no epochs in cluster 3, this individual would have a data point with the value vector of [0.25, 0.75, 0]. The number of clusters decides the

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Fig. 5. Test scores of different number of clusters

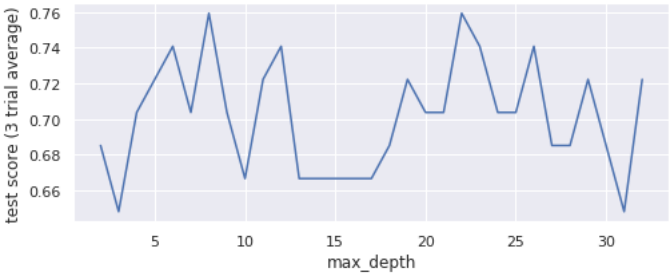

Fig. 6. Test scores of random forest with different maximum depth
Note: The test score is the average of a 3-trial test for a random stratum, the size of which is 10% of the full data set, and it shares the same distribution of age and gender with the full set.

number of different EEG patterns we attempt to extract from the dataset.

To determine the optimal number of clusters, we considered the average test scores. We examined the test scores with three different classifiers. The algorithms are multi-layer perceptron neural network (MLP), support vector machine (SVM) and random forest (RF). Fig. 5 demonstrate the average test score of a 10-trial test for a random stratum, the size of which is 10% of the full data set, and it shares the same distribution of age and gender with the full set. The performance of the classifiers varies. When the number of clusters is 9, all three classifiers generate high average test scores. When there are 24 clusters, though the performance of one classifier is relatively poor, the average test scores of the other two both reached their highest points.

Our experiment shows that the performance of the classifiers is largely improved with the introduction of clustering-based motif attributes. Detailed comparison of subject-wise accuracy when with and without the clustering stage will be demonstrated in Section IV-B.

The next stage of the framework is to train classifiers on individual-wise data points. Three classifiers are trained separately: MLP, SVM, and RF. In each trial, the minority class in the training set is oversampled to avoid the influence of the imbalanced dataset. Random forest reaches the best performance when the maximum depth of the individual trees is set to 8 and 22 as shown in Fig. 6.

Due to the performance variance of the three classifiers, a simple ensemble layer is introduced aiming to improve the performance and stability of classification. Specifically, a hard vote among the three classifiers is taken, and the final prediction is decided by majority voting.

B. Classification results

In this section, we will present the experiment results of the proposed MACE framework. The model performance will be evaluated by two metrics: the prediction accuracy and the area under the ROC curve (AUC). The effects of the clustering-based motif attributes and the ensemble on classification performance will also be presented.

- Internal accuracy

Table III demonstrate the internal test score. All the data were used for training and testing. In the table, 'F' represents Females, 'M' represents Males, and the numerical value is the range of age in this group

TABLE III
TEST SCORES BY GROUP

Classifiers	Accuracy by gender/age groups						Overall accuracy
	M25 -35	M36 -45	M46 -55	F25 -35	F36 -45	F46 -55	
MLP	0.73	0.86	0.95	0.94	0.94	0.91	0.91
SVM	0.73	0.79	0.85	0.94	0.76	0.74	0.79
RF8	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Ensemble	0.82	0.86	0.95	1.00	0.94	0.91	0.92

- Overall test accuracy

The analysis in Section IV-A suggests that the best number of clusters is likely to be 9 or 24. Therefore, we experimented with our customized framework with a cluster number of 9 and 24 respectively.

TABLE IV
10-FOLD INDIVIDUAL-WISE CROSS VALIDATION SCORES

Cluster number	Overall accuracy			Female accuracy	Male accuracy
	mean	*median*	*std*		
9 clusters	0.73	0.73	0.12	0.74	0.71
24 clusters	0.72	0.73	0.10	0.75	0.65

Note: the results are the 10-fold CV score of the voted prediction

The overall subject-wise classification accuracy is calculated by taking the average test score of 10 trials. In each trial, a 10% test sample was randomly selected, and the other 90% is used as a training set. As shown in Table IV, when the number of clusters is 9, both the mean and median test scores are 0.73, and the standard deviation is 0.12. The average accuracy of the female group (0.74) is higher than that of the male group (0.71). In the model with a cluster number of 24, the median is 0.73, and the mean score is 0.72 with a standard deviation of 0.1. The

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

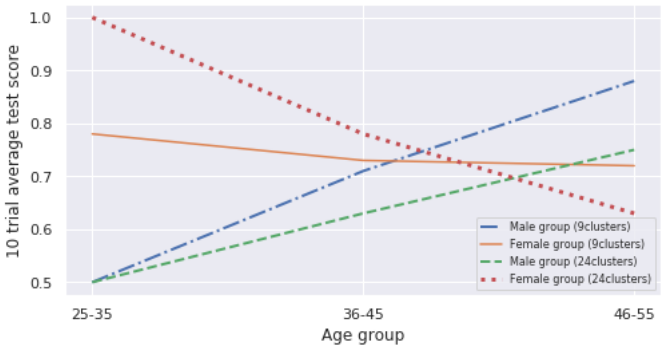

Fig. 7. Test scores of different gender/age groups

discrepancy between the accuracy of females and males is more prominent with 24 clusters. The average accuracy of the female group is 0.75, whereas it is 0.65 of the male group.

- Test accuracy by gender/age groups

To investigate the possible age effects on gender classification, We divide the tested individuals into 6 subgroups and calculate the accuracy within each group. These groups are females of an age between 25 to 35 years (F25-35), females of an age between 36 to 45 years (F36-45), females of an age between 46 to 55 years (F46-55), males of an age between 25 to 35 years (F25-35), males of an age between 36 to 45 years (F36-45) and males of an age between 46 to 55 years (F46-55). Table IX demonstrates the results of the voted prediction accuracy by subgroup.

TABLE V
TEST SCORES BY GROUP

Model[a] with	Accuracy by gender/age groups[b]					
	M25-35	M36-45	M46-55	F25-35	F36-45	F46-55
9 cluster	0.50	0.71	0.88	0.78	0.73	0.72
24 clusters	0.50	0.63	0.75	1.00	0.78	0.63

[a]the results are of the MACE framework with two different clustering procedure
[b]F=Females, M=Males, numerical value is the range of age in this group

The results by group suggest that the prediction accuracy rates for females are generally higher than that for males. Both of the models with a cluster number of 9 and 24 have an accuracy of only 0.5 for group M25-35. Within male groups, the accuracy rate of the models increases with individuals' age increasing. Compared to younger males. the EEG signals of older males are easier to differentiate from the EEG signals of females. This finding suggests that the EEG signals from older males have more male gender identity than those from younger males. On the contrary, within female groups, the accuracy rate of the models decreases with individuals' age increasing. The trend is more prominent in the model with 24 clusters than that in the model with 9 clusters. The interaction between age and the accuracy of gender classification is demonstrated in Fig 7.

Opposite to some previous studies [9], our study demonstrates some interesting interactions between age and gender.

- Ablation on clustering and motif attribute formation

To show the effect of the clustering and motif attribute formation on classification results, we conduct an ablation study where classification models are fed directly with features (without going through clustering and motif attribute formation). We trained the same classifiers (i.e. MLP, SVM, and RF) directly on the standardized epoch band power features. The 10-fold epoch-wise cross-validation scores are in Table VI.

TABLE VI
10-FOLD EPOCH-WISE CROSS VALIDATION SCORES

Classifiers	Overall accuracy		
	mean	*median*	*std*
MLP	0.62	0.61	0.05
SVM	0.60	0.61	0.06
RF8[a]	0.63	0.65	0.08

[a]RF8: random forest with maximum depth of 8 for each tree

To transform the epoch-wise prediction into individual-wise prediction, we use the percentage of the epoch-wise predictions for each individual as their individual-wise probability prediction. For example, if an individual has 40 epochs, 10 of which are categorized as male and 30 of which are categorized as females, then the chance of this individual being a male is 25%, and being a female is 75%. To evaluate the individual-wise classification performance, we calculated the AUC for 10 trials. Each trial uses 10% randomly generated samples as the test set and the other 90% as the training set. As shown in Table VII, for the neural network classifier, the mean AUC value of the 10 trials is 0.72 with a standard deviation of 0.2. For support vector machine classifier, the mean AUC is 0.66 and the standard deviation is 0.17. And the random forest classifier has a mean AUC of 0.72 with a standard deviation of 0.2.

TABLE VII
AUC OF 10 TRIALS (WITHOUT CLUSTERING)

Classifiers	AUC		
	mean	*median*	*std*
MLP	0.72	0.71	0.20
SVM	0.66	0.64	0.17
RF8[a]	0.72	0.73	0.20

[a]RF8: random forest with maximum depth of 8 for each tree

TABLE VIII
AUC OF 10 TRIALS (AFTER CLUSTERING)

Classifiers	AUC		
	mean	*median*	*std*
MLP	0.81	0.80	0.09
RF8	0.78	0.80	0.15

[a]RF8: random forest with maximum depth of 8 for each tree

Since the classifiers of our framework are trained directly on individual-wise clustered features, and support vector

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

machine does not directly provide probability estimates. Multi-layer perceptron and random forest classifiers are used for the AUC estimation after clustering because these two classifiers can estimate the probability more straightforwardly. Table VIII demonstrates the AUC of the classifiers after clustering. The 10 trials are run in the same way as in the test before clustering. For the neural network classifier, the mean AUC is increased from 0.72 to 0.81 after applying the clustering procedure, and the standard deviation dropped from 0.2 to 0.09. For the random forest classifier, the mean AUC is increased from 0.72 to 0.78, and the standard deviation decreased from 0.2 to 0.15. In comparison, the classifiers trained on clustered individual-wise features show better and more stable performance.

- The effect of the classifier ensemble procedure

We compared the 10-fold cross-validation scores of neural network, support vector machine, random forest, and the classifier ensemble. As shown in Table IX, the ensemble model keeps the same mean and variance with the best performed individually trained classifier. Also, the median of the classifier ensemble is better than all the other three classifiers individually. This demonstrates the effectiveness of the ensemble layer.

TABLE IX
10-FOLD CROSS VALIDATION SCORES(9 CLUSTERS)

Classifiers	Overall accuracy			Female accuracy	Male accuracy
	mean	*median*	*std*		
MLP	0.73	0.69	0.12	0.76	0.69
SVM	0.69	0.69	0.11	0.67	0.73
RF8[a]	0.72	0.69	0.12	0.76	0.67
Ensemble	0.73	0.73	0.12	0.74	0.71

[a]RF8: random forest with maximum depth of 8 for each tree

V. CONCLUSIONS

In this paper, we investigate the potential gender differences in resting state EEG signals. Gender differences have been found with females showing stronger gender effect: the classification accuracy is generally higher in the female groups than that in the male groups. Our study also found an interaction between gender and age. The EEG signal of older males shows a stronger gender effect than that of younger males, and the EEG signal of older females shows a weaker gender effect than that of younger females.

It is also found that epoch-wise gender discrimination is implausible and there are no clear gender "motifs" as revealed by clustering analysis. Nevertheless, by utilizing histogram information of a subject's epoch data, the motif attribute vectors manage to capture potential gender differences, and the MACE framework manages to perform subject-wise gender classification with the help of a classifier ensemble. However, a less satisfactory test accuracy resides in the younger male group. If we were allowed to explore further, we would like to look into this issue.

For future work, we would like to investigate the applicability of our algorithm in a wider age range and experiment with EEG signals obtained under different clinical stimuli.

REFERENCES

[1] S. Siuly, Y. Li, and Y. Zhang, EEG Signal Analysis and Classification. Cham: Springer International Publishing, 2016.

[2] S. Vanneste, J. Song, D. D. Ridder, Thalamocortical dysrhythmia detected by machine learning, NATURE COMMUNICATIONS 9:1103, 2018.

[3] A. Shoeb, J Guttag, Application of Machine Learning To Epileptic Seizure Detection. Appearing in Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 2010.

[4] L. Cahill, Why sex matters for neuroscience. Nat Rev Neurosci, vol. 7, no. 6, pp. 477–484, Jun. 2006.

[5] M. McCarthy, B. Nugent, and K. Lenz, Neuroimmunology and neuroepigenetics in the establishment of sex differences in the brain. Nat Rev Neurosci 18, 471–484, 2017.

[6] Arns, M. et al. EEG alpha asymmetry as a gender-specific predictor of outcome to acute treatment with different antidepressant medications in the randomized iSPOT-D study. Clin. Neurophysiol. 127, 509–519, 2016.

[7] A. R. Clarke, R. J. Barry, R. McCarthy and M. Selikowitz, Age and sex effects in the EEG: development of the normal child, Clinical Neurophysiology Volume 112, Issue 5, Pages 806-814, May 2001.

[8] R. J. Davidson, G. E. Schwartz, E. Pugash and E. Bromfield, Sex differences in patterns of EEG asymmetry, Biological Psychology Volume 4, Issue 2, Pages 119-137, June 1976.

[9] Carrier, J., Land, S., Buysse, D.J., Kupfer, D.J. and Monk, T.H., The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old). Psychophysiology, 38: 232-242. 2003.

[10] [1]M. J. A. M. van Putten, S. Olbrich, and M. Arns, Predicting sex from brain rhythms with deep learning, Sci Rep, vol. 8, no. 1, p. 3069, Dec. 2018.

[11] B. Kaur, D. Singh, and P. P. Roy, Age and gender classification using braincomputer interface, Neural Comput & Applic, vol. 31, no. 10, pp. 58875900, Oct. 2019.

[12] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, et al.. A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces: A 10-year Update. Journal of Neural Engineering, IOP Publishing, 15 (3), pp.55, 2018.

[13] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, Klaus-Robert Müller. Optimizing Spatial Filters for Robust EEG Single-Trial Analysis. IEEE Signal Processing Magazine 25(1), 41-56, 2008.

[14] Younghak Shin, Seungchan Lee, Minkyu Ahn, Hohyun Cho, Sung Chan Jun, Heung-No Lee, Noise robustness analysis of sparse representation based classification method for non-stationary EEG signal classification, Biomedical Signal Processing and Control, Volume 21, Pages 8-18, 2015.

[15] He X, Cai D, Niyogi P. Laplacian score for feature selection. InAdvances in neural information processing systems, pp. 507-514, 2006.

[16] Eric P. Xing, Andrew Y. Ng, Michael I. Jordan, and Stuart Russell. Distance metric learning, with application to clustering with side-information. In Proceedings of the 15th International Conference on Neural Information Processing Systems (NIPS'02). MIT Press, Cambridge, MA, USA, 521–528, 2002.

[17] L. van der Maaten and G.E. Hinton, Visualizing Data Using t-SNE, J. Machine Learning Research, vol. 9, pp. 2579-2605, 2008.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE