A Machine Learning Approach for Sex and Age Classification of Paediatric EEGs

Lan Wei[1], John C McHugh[2] and Catherine Mooney[1]

***Abstract*— Electroencephalography (EEG) is an important investigation of childhood seizures and other brain disorders. Expert visual analysis of EEGs can estimate subjects' age based on the presence of particular maturational features. The sex of a child, however, cannot be determined by visual inspection. In this study, we explored sex and age differences in the EEGs of 351 healthy male and female children aged between 6 and 10 years. We developed machine learning-based methods to classify the sex and age of healthy children from their EEGs. This preliminary study based on small EEG numbers demonstrates the potential for machine learning in helping with age determination in healthy children. This may be useful in distinguishing developmentally normal from developmentally delayed children. The model performed poorly for estimation of biological sex. However, we achieved 66.67% accuracy in age prediction allowing a 1 year error, on the test set.**

I. Introduction

The human brain undergoes significant maturational change during childhood, which is paralleled by changes within the electroencephalogram (EEG). Changes are particularly marked within the first year, after which more gradual evolution occurs into early childhood and onwards to adolescence. Appreciation of age-specific normative EEG patterns is clinically important as abnormalities of brain development can be associated with slowing of background activities as well as other patterns of abnormal activity.

Medical reporting of EEG is primarily based on visual inspection by qualified experts in Clinical Neurophysiology with a combined appreciation of age-related neurological disease and knowledge of age-related EEG patterns. Research on machine learning-based EEG feature estimation for age and sex determination is limited.

An automated EEG-based sex and age classification method would be a powerful tool for brain development researchers and the research to date, although limited, has shown some good results [1]. Kaushik *et al.* [2] developed a deep BLSTM-LSTM network model to predict sex and age from 60 EEG recordings with an age range from 6-55 years old, which achieved an accuracy of 93.7% for age classification and 97.5% for sex classification. Similarly, Nguyen *et al.* [3], [4] and Kaur *et al.* [5] developed methods with accuracy ranging from 88.3% - 97.5% and 93.8% to 97.7% for age and sex, respectively. However, all the previously mentioned methods suffer from some serious drawbacks. They used a very wide age range (from 6 - 69) and predicted the age into bins e.g. 6-10, 12-15, 19-34; 35-54; and 55-69. Moreover, Kaushik *et al.* and Nguyen *et al.* [2], [4] divided their EEGs into epochs, then trained and tested their models on EEG epochs that were likely to be from the same person. This has the potential to lead to overfitting and there is no guarantee that their models will perform well on independent tests sets.

In this study, we analyse the features of the EEGs, which could assist researchers in exploring the differences between male and female children and brain development. We developed machine learning-based methods to classify the sex and age of healthy children's EEG recordings (n=351). We took care to separate the children into either the training or test set to ensure that the test set was independent (the training and test set without overlap). For age prediction, we predicted the specific age, in years, of the child rather than an age range.

II. Materials and Methodology

A. CHI EEG Dataset

Ethical approval was granted from the Medical Research Ethics Committee of Our Lady's Children's Hospital Crumlin, Dublin, Ireland (GEN/617/17). EEGs from 351 healthy children and 301 children with epilepsy aged 6 to 10 were used in this study. 80% (n=282) of the healthy EEGs of each age were used to train the method for age classification, and the remaining 20% (n=69) were used for testing. The same training and test set were used to develop the sex classification method. Table I shows the number of EEG recordings used for training and testing.

TABLE I
NUMBER OF EEG RECORDINGS USED TO TRAIN AND TEST THE SEX AND AGE CLASSIFICATION METHODS.

	Sex	Age 6	Age 7	Age 8	Age 9	Age 10
Train	Female (n=107)	28	19	26	16	18
	Male (n=175)	36	43	34	36	26
Test	Female (n=25)	6	5	6	4	4
	Male (n=44)	9	11	9	9	6

B. Data Pre-processing

The CHI EEG dataset was sampled at different sampling frequencies of 200Hz, 256Hz, and 500Hz. Therefore, the EEG signal was resampled to 256Hz. A notch filter (50Hz)

*This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the NeuroInsight Marie Skłodowska-Curie grant agreement No. 101034252. We acknowledge the Research IT HPC Service at University College Dublin for providing computational facilities and support that contributed to the research results reported in this paper.

[1] FutureNeuro SFI Research Centre, School of Computer Science, University College Dublin, Dublin, Ireland lan.wei@ucd.ie, catherine.mooney@ucd.ie

[2] Our Ladys Children's Hospital, Crumlin, Dublin, Ireland

was used to remove power line interference, and DC offset was removed from the EEGs. Singular Spectrum Analysis (SSA) was used to remove artefacts with a window length of 4. SSA is a subspace-based technique for time series analysis [6], which comprises four basic steps: 1) embedding, 2) singular value decomposition, 3) grouping and 4) reconstruction. SSA is widely used for removing artefacts such as motion artefact [7], electromyogram (EMG), electrooculogram (EOG), and electrocardiogram (ECG) artefacts [8].

C. Channel Selection

EEG recording is a highly complex process, and the unique channel configuration of each EEG or clinical site needs to be adapted [9]. Channels from frontal, central, parietal, occipital and temporal regions are used according to longitudinal bipolar derivations [10] (Channels FP1-F7, F7-T3, T3-T5, T5-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F8, F8-T4, T4-T6, T6-O2, FP2-F4, F4-C4, C4-P4, P4-O2). In this study, channels FP1-F7, FP1-F3, F7-T3 and T3-T5 had the best performance on the training set (10-fold cross-validation). Therefore, we used these four channels to develop the methods.

D. Feature Estimation

In this work, time and frequency domain features were measured. Discrete wavelet transform was used to filter the signals within the frequency bands (delta, theta, alpha, beta and gamma) of interest [9]; 5s epochs with 2.5s overlap were used to develop 23 features for each channel. These features are as follows.

Standard features (9): The mean, standard deviation, signal envelope, kurtosis, skewness, complexity, mobility, variance and fractal dimension of pre-processed absolute amplitude were calculated in the time domain.

Features in sub-frequency bands (11): Wavelet decomposition was used to extract the different rhythm activities, with delta (0-4Hz), theta (4-8Hz), alpha (8-16Hz), beta (16-32Hz), and gamma (32-64Hz) [11]. The relative and absolute band power of these rhythm activities were estimated. The absolute band power of the EEG amplitude was also used in this study.

Frequency domain features (3): The mel-frequency cepstral coefficients (MFCCs) [3], power spectral density (psd), and amplitude modulation spectrogram (AM spectrogram) were extracted as features for developing the sex and age classification methods.

E. Data Balancing

In this study, the number of female children (n=107) in the training set was less than the number of male children (n=175). This leads to class imbalance, which makes it challenging to train machine learning models. Therefore, Synthetic Minority Over-sampling Technique (SMOTE) [12] was used to balance the data in the training set (k=5) for sex classification. For age classification, the number of the different ages are relatively balanced in the training sets (Age 6: n=64; Age 7: n=62; Age 8: n=60; Age 9: n=52; Age 10: n=44;). Therefore, the data balancing technique is not used.

F. Method Development

The random forest algorithm was implemented within the Python 3 environment. The parameters n-estimators and max-depth were optimised based on the performance of the training set (10-fold cross-validation) for the sex and age classification in the EEG recordings. The n-estimators value was tested from 100 to 300, and the max-depth was tested from 5 to 15. The best performance on the training and test set was achieved when n-estimators = 100, max-depth = 6 for the sex classification, n-estimators = 100, and max-depth = 5 for the age classification.

G. Performance Evaluation

Accuracy, precision, recall, and F1 score were used to evaluate the sex classification model. Accuracy, macro-averaged precision, macro-averaged recall, macro-averaged F1 score and mean absolute percentage error (MAPE) were used for age classification.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - P_t}{A_t}\right| \quad (1)$$

MAPE: Mean absolute percentage error; n: Number of times the summation iteration happens; A_t: Actual value; P_t: Predicted value;

III. Results

A. Feature Analysis

Figure 1 compares relative theta and alpha between male and female children at the age of 6 on channels FP1-F7, FP1-F3, F7-T3, and T3-T5. It is clear that female children at the age of 6 have higher relative theta and lower relative alpha compared with male children. In addition, Mann-Whitney U Test was calculated for the relative theta and relative alpha for male and female, with p-value less than 0.05.

Figure 4 shows the change in the power of different frequency bands with sex and age. Figure 4 shows that the power of different frequency bands (theta, alpha, beta and gamma) have different development trends with increasing age in male and female children.

B. Sex Classification

Table II shows the performance of the sex classification method. Sex classification, in this case, is a binary class classification method, which achieved an accuracy of 55.07% on the sex test set.

TABLE II
PERFORMANCE OF THE SEX CLASSIFICATION METHOD ON TRAINING AND TEST SET.

	Accuracy	Precision	Recall	F1 score
Train	73.76%	75.45%	76.68%	73.65%
Test	55.07%	52.57%	52.68%	52.52%

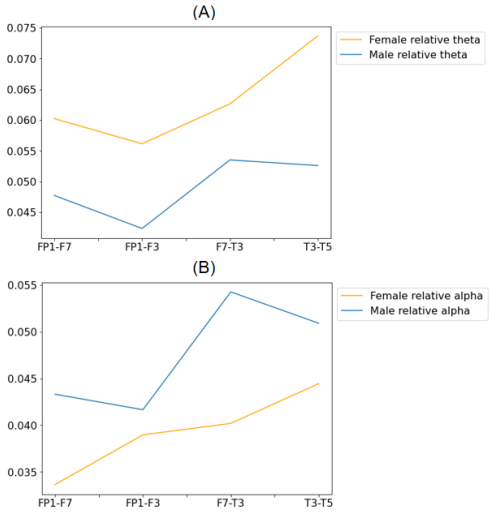

Fig. 1. Comparison of relative theta and alpha between male and female children at the age of 6 on channel FP1-F7, FP1-F3, F7-T3, T3-T5. The blue lines represent the changes in male children's relative theta and alpha bands. The orange lines represent the changes in female children's relative theta and alpha bands.

C. Age Classification

Table III shows the performance of the age classification method. Here, age is predicted into five classes, which achieved an accuracy of 30.43% on the age test set. As age is not in fact a categorical value, we also show results as a confusion matrix (Figure 2) where we tolerated a one year error. In this case we obtained 66.67% accuracy on the test set (Table III).

TABLE III

PERFORMANCE OF THE AGE CLASSIFICATION METHOD ON TRAINING AND TEST SET (1-Y-ERR: TOLERATE WITH ONE YEAR ERROR).

	Accuracy	Precision	Recall	F1 score	MAPE
Train	60.99%	60.40%	60.28%	60.99%	9.16%
Test	30.43%	29.97%	29.99%	29.69%	16.03%
Test 1-y-err	66.67%	65.14%	64.93%	64.96%	11.28%

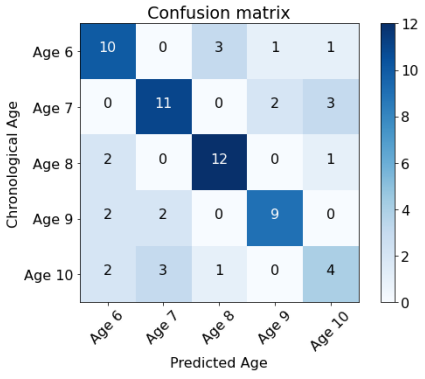

Fig. 2. Matrix of developed age classification method with one year error on the age test set.

Figure 3 presents a subanalysis for healthy children and children with epilepsy. The slope of the predicted age for healthy children is 0.51. The slope of the predicted age for children with epilepsy is 0.16. Moreover, the developed age classification model predicted that the age of children with epilepsy was significantly lower than that of healthy children after the age of 8 years. In addition, the Wilcoxon signed rank test of chronological age and predicted age was evaluated. The p-value for healthy children is 0.714. And the p-value of children with epilepsy is 0.0013, which is lower than the significance level (0.05).

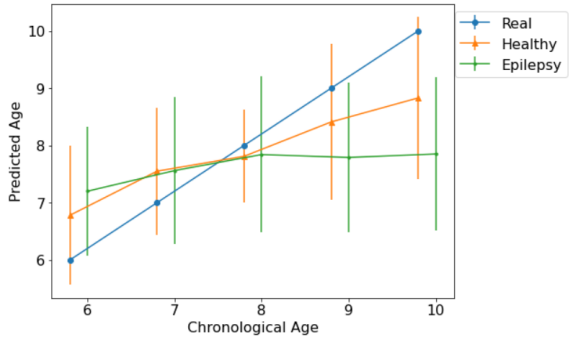

Fig. 3. Error bars for age classification for healthy children, children with epilepsy. The blue line indicates the chronological age. The orange and green lines shows the trend for age prediction for healthy children and children with epilepsy, respectively.

IV. DISCUSSION

Previous research showed that female children at the age of 6 have higher relative theta and lower relative alpha compared with male children [4], which we also observed in this study (Channels FP1-F7, FP1-F3, F7-T3 and T3-T5, see Figure 1). Mann-Whitney U Test was calculated for the relative theta and relative alpha for male and female, with p-value less than 0.05. Results show that the relative theta and alpha for male and female children have a statistically significant difference. Figure 4 shows that the power in the different frequency bands (theta, alpha, beta, and gamma) has significantly increased from 6-7 and then reduced from 7 to 10 on the left frontal area (FP1–F7, FP2-F3, F7–T3) for male children. For female children, the power in the different frequency bands (theta, alpha, beta, and gamma) significantly decreases from 6-7 and 8-9 years old, trending upward from 7 to 8 and 9-10 on the left front area. In addition, it is clear that the qualitatively EEG visual review is a reduction in frontal amplitudes with increasing maturation. There is an unusual peak in the male data around the age of 7 years and female data around the age of 8 years. Figure 3 presents a subanalysis for children with epilepsy. The slope of predicted age for healthy children (slope=0.51) is higher than that of children with epilepsy (slope=0.16). In addition, statistically significant differences were found in chronological age and predicted age for children with epilepsy ($p < .05$, Wilcoxon signed rank test). However, there were no statistically significant differences in chronological age and predicted age for healthy children ($p > .05$, Wilcoxon signed rank test). This behavior implies that the predicted chronological age in children with epilepsy diverged from healthy subjects.

Previous studies [1], [2], [4], [3] developed sex and age classification methods based on using EEGs from children

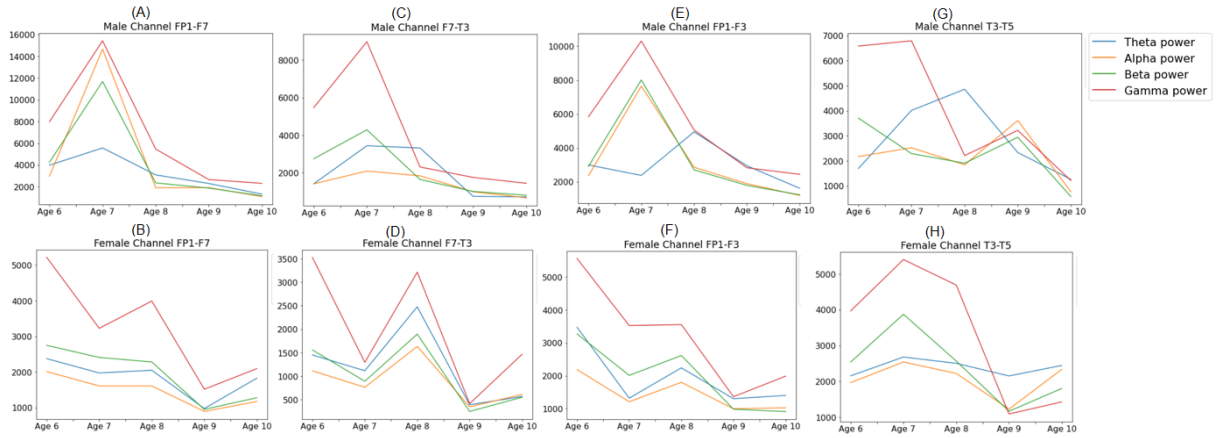

Fig. 4. The change of the power of the different frequency bands with age (from 6 to 10 years old) for male (on the left) and female (on the right) children in channels FP1-F7, FP1-F3, F7-T3 and T3-T5. The blue, orange, green and red lines indicate the change in the power of the theta, alpha, beta and gamma bands, respectively.

aged 6 up to adults age 69, and predict the age into a small number of bins with wide age ranges, for example, young, middle aged and elderly. In addition, some researchers [2], [4] divided their EEG dataset into epochs and trained and tested on epochs which may come from the EEG of the same person. This can lead to overfitting, and the performance of the classifier on people not included in the training data is unknown. For example, when we trained and tested our models using epochs from the same children in both the training and test set, for comparison, we achieved an accuracy of 98.87% and 99.07% for sex and age classification, respectively. However, this is clearly overfitting and the models will not perform well on children who were not included in the training data i.e. the models will not generalise. In our training and testing we took care to separate the children into either the training or test set to ensure that the test set was independent, which reduced the overfitting problem.

A limitation of the current study is that our models performed poorly for prediction of biological sex, with an accuracy of 55.07% on the test set. In future work, we would like to investigate what features are important for sex classification for children. In addition, because machine learning is a "black box" approach, clinicians may have difficulty trusting machine learning-based models [13]. In future work, we will use explainable AI (XAI) techniques to help gain users' trust in machine learning-based models.

V. Conclusions

In this study, we have developed a sex and age classification models for children's EEG data. The models developed in this work can predict the specific age rather than an age range. Moreover, the models were tested on a test set that was independent from the training set. Our model achieved an accuracy of 55.07% and 30.43% for sex and age classification, respectively, and 66.67% accuracy in age prediction allowing a 1 year error. Moreover, we analysed the features of EEGs that could assist researchers in exploring the differences between male and female children, and in the analysis of brain development. Our results demonstrate that there are sex and age differences in the EEG recordings of children aged 6-10 years. In addition, a subanalysis was done for children with epilepsy and found that predicted chronological age diverged from healthy subjects.

References

[1] O. Al Zoubi, C. Ki Wong, R. T. Kuplicki, H.-w. Yeh, A. Mayeli, H. Refai, M. Paulus, J. Bodurka, Predicting age from brain EEG signals—A machine learning approach, Frontiers in aging neuroscience 10 (2018) 184.

[2] P. Kaushik, A. Gupta, P. P. Roy, D. P. Dogra, EEG-based age and gender prediction using deep BLSTM-LSTM network model, IEEE Sensors Journal 19 (7) (2018) 2634–2641.

[3] P. Nguyen, D. Tran, T. Vo, X. Huang, W. Ma, D. Phung, EEG-based age and gender recognition using tensor decomposition and speech features, in: International conference on neural information processing, Springer, 2013, pp. 632–639.

[4] P. Nguyen, D. Tran, X. Huang, W. Ma, Age and gender classification using EEG paralinguistic features, in: 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2013, pp. 1295–1298.

[5] B. Kaur, D. Singh, P. P. Roy, Age and gender classification using brain–computer interface, Neural Computing and Applications 31 (10) (2019) 5887–5900.

[6] H. Hassani, Singular spectrum analysis: methodology and comparison.

[7] A. K. Maddirala, R. A. Shaik, Motion artifact removal from single channel electroencephalogram signals using singular spectrum analysis, Biomedical Signal Processing and Control 30 (2016) 79–85.

[8] J. Cheng, L. Li, C. Li, Y. Liu, A. Liu, R. Qian, X. Chen, Remove diverse artifacts simultaneously from a single-channel EEG based on SSA and ICA: A semi-simulated study, IEEE Access 7 (2019) 60276–60289.

[9] L. Wei, C. Mooney, Epileptic seizure detection in clinical EEGs using an XGBoost-based method, in: 2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), IEEE, 2020, pp. 1–6.

[10] A. Polo, P. Narvaez, C. Robles Algarín, Implementation of a cost-effective didactic prototype for the acquisition of biomedical signals, Electronics 7 (5) (2018) 77.

[11] L. Wei, H. Boutouil, R. R. Gerbatin, O. Mamad, M. Heiland, C. R. Reschke, F. Del Gallo, P. F. Fabene, D. C. Henshall, M. Lowery, et al., Detection of spontaneous seizures in EEGs in multiple experimental mouse models of epilepsy, Journal of Neural Engineering 18 (5) (2021) 056060.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, Journal of artificial intelligence research 16 (2002) 321–357.

[13] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, C. Mooney, Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: a systematic review, Applied Sciences 11 (11) (2021) 5088.