Novel Model Architecture for EEG Emotion Classification

*Lars Rune Christensen , #Mohamed Ahmed Abdullah
*IT University of Copenhagen , #Sudan University of Science and Technology

Abstract — *Enhancing the communication between the human and the machine is the core purpose of the HCI field (Human Machine Interaction), Identifying human emotion is an important aspect of enhancing this communication. This work will identify 10 distinctive emotions, the emotions will be measured by Dominance, Valance, and Arousal. The classification will run five different models with respect to Riemannian Geometry to enhance the filtering. The EEG signal is being collected in a form of two sessions per subject, one session for training purposes the other for testing, then the data is filtered then the model was trained to give an accuracy between 50% to 70%. SAM assessment technique has been conducted to tag the data.*

Keywords — *EEG; Emotion Recognition; Emotion Detection; HMI; BCI, Riemannian Geometry, TangentSpace*

I. INTRODUCTION

One of the fastest and simplest and noninvasive techniques to record the brain waves is the EEG Electroencephalogram, we can easily grasp that is we compare it to other invasive techniques like fMRIfunctional Magnetic Resonance Imaging or NIRS near-infrared spectroscopy or MEG Magnetoencephalography. Therefor EEG is widely accepted especially in the field of HMI Human Machine Interaction and BCI Brain-Computer Interface field, as it's low budget and doesn't require medical professionals to operate it [17]. The BCI field is getting better as we are able to accurately extract the features from the EEG signals, also we should be able to distinguish and classify those signals accurately, as it might represent eye blink or emotional swings. The higher the accuracy of predicting the classes and the higher the number of classes that can be predicted, the better the BCI system is. As bigger this field gets as researchers will adopt it in different fields like health and clinical trials, but still there is no commercial adoption for the field as it's still lacking a variety in the number of classes that it can cover also lacking the concrete predictable accuracy, especially if we translate it to real-world scenario, as currently, we have to operate it under controlled environment in order to get acceptable results. We can see there is no major invention in the field since 20 years ago, the classification and the processing algorithm that wasproposed since the initiation of the field of EEG and the diagonalization methods such as CSP common spatial pattern, CCA canonical correlation analysis, ICA independent component analysis, with the multiple variations of each other's and possible combinations, we can clearly see that there is no major innovation in the field.[4]We can also say that the new method and techniques achieved only moderate improvement and do not increase reliability in a significant way.The multiple subfields of the BCI making the research to diverge and to be diffuse and not focused on a single issue, for example, MI motor-imagery, SSEP steady-state evoked potentials and P300, are treated with dedicated pre-processing, signal processing and classification tools they all intertangled and faces of the same issue.The classification techniques also suffered from this, it's also dispersed and fragmented. Now we have the "spatial filtering" approach and the"hard machine learning" approach. The "spatial filtering" techniques rely on increasing the SNRSignal to noise ratio, then a simple classification algorithm.The "hard machine-learning" approach has a good generalization across persons and across sessions, but it requires a big number of training data, in order to get a decent result. As well as it is computationally expensive.But we don't need big data for the "spatial filtering" approach, in fact, it's bad in the generalization capabilities but fast for training the models due to the simplicity of the model and the lower volume of training data and lower computational cost. But its hand tailored to the problem that it's trying to fix.In light of this situation, it has been stated that "the field would benefit from a new approach in research development that focuses on robust algorithm development". It has also been recommended to start regarding the pre-processing, feature extraction and classification not as isolated processes, but as a whole [12].

II. RELATED WORK

The aim ofstudy [8] is to help candidates control their emotions, by predicting the human emotions to create a close loop of playing music (as stimuli)then measure the emotion. This will help candidatesto increase awareness or enhance the brain functionality or be use it as a treatment to reach to a certain emotion, the accuracy of the model has been collected by a user's survey, the survey asking the candidates to develop a mental strategy then measure the level of achieving the targeted strategy. The emotion in this study is measured by the valance and arousal.

This study [11] is proposing a new Deep Learning Network DLN to predict the correlation between the input signals, three auto encoders have been used along with two softmax for classifying the valance and arousal, by using PCA on the initial input, the classification result of this model is 49.5% accurate, the model is identifying 3 levels of valance and arousal. The DLN model scored better than SVN and Bayes classifiers. This Study [13] is proposing a hyper deep learning model to link CNN and RNN (Convolutional Neural Network and Recurrent Neural Network)the CNN is able to find the correlation between input signals, and the RNN is able to learn long term dependencies, this model also has the potential of giving predictions not only for an entire trial but also for each time step, which is keen in real-time emotion monitoring scenarios.The DLN models are simple to setup and construct unlike the traditional models, DLN isdelegating the task of finding the best signal correlation to the layer of the neural network.This study [14] trying to help with Major Depressive Disorderit's a mental disorder can be detected from the amygdala region in the brain, In this study, an emotional up-regulating method has been conducted to healthy candidates by performing a set of tasks, then self-report the affective state.The tasks e.g. remember a happy memory and the accuracy will be measured by t-test the results from using the CSP was 72% from 11 sessions. In this study [15] the target is to spell out the words, the results can reach 70% after some training session per user. The modes being applied is LDA linear discriminant analysisand SVM support vector machines.

This study [16]is trying to classify four emotions (happy, sad, fear, and neutral), using 6 electrodes (FT7, FT8, T7, T8, TP7, and TP8) placed using the 10–20 system, the study used 75 film clip on a 15 subjects, each film lasted for 2 minutes, each 24 film clip is considered as a session, those sessions is only to reduce the fatigue effect on the signals and to keep the subjects focus, SAM Self-Assessment Manikin is used, the classification algorithm was SVM with a bimodal deep auto-encoder (BDAE), 85.11% has been achieved as an accuracy.

III. METHODOLOGY

Trails is a term resembling a short time of the EEG signal, the purpose of it is to simplify the signal analysis [2] in other words if we represent the EEG data in a table format where the columns are the channels and the rows are the actual values (signals) the trail will be a set of rows where the subject was looking at a particular image or doing some task.

A. Apparatus

EPOC Plus with 14 channels: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. with LSB = 0.51μV and 128 Hz with a resolution of 16 bits.

B. Setup

Tow healthy subjects participated in this study. The data will be collected during 4 sessions. The subject will be looking at the screen of a computer while sitting comfortably, they will be asked to minimize their eye blinking and stop completely any movement.

Session 1: a five second of countdownappears on the monitor then the participants will look at computer screen and watch images that is changing every 2.5 seconds, as per this study [3] the P300 ERP (Event Related Potential) will happen after 300 millisecondsso 2.5 second is anoptimumtime span as it's big enough to capture the whole emotion and small enough to notbore the subject so they won't get distracted. Multiple images will be shown to the subjects per session. The data from this session will be tagged by the image and the trails will be identified, the trails are the set of signals while the subject was looking at a certain image.

Session 2: Same as session 1 but this data will be used for testing the model. This data will be a little distorted because the subject will be familiar with the images and the emotions will be less this time due to the emotion memory.In order to tackle this and to reduce the emotion memory effect, the order of the images will be shuffled this time.

Session 3: now the same images will be presented and the subject has to select from 3 buttons describing the valance as (happy, natural, sad) and the arousal as 3 level and the dominance as 3 levels as well [1]. The maximum number of distinct emotions from this setup will be 20 distinct emotions.

Session 4: the same as session 3. The data from this session will be compared with session 3 to eliminated the conflicts e.g. if the subject select happy in session 3 and sad in session 4 this emotion will be discarded.

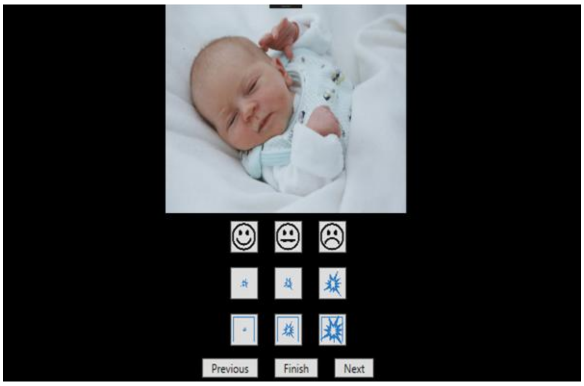

Fig.1 the emotions collecting tool buttons are valance, arousal and dominance for each row

The first subject watched 131 images then the data is tagged with the appropriate emotions, he has 5 distinct emotions from the range of possibility which is 20 possibilities. The second subject watched 60 images then the data was tagged as well, he has 10 distinct emotions tagged.

C. Model Generation

Emotion recognition is a classification problem and due to the SAM model now we can group the emotion by the three main criteria, dimensions Valance, Arousal, Dominance. We have multiple models we can use to detect the emotion. Rather than using only one model we are using the best model per subject, which means we will train all the models on the subject then we pick the best one. The first step is the channel selection.

Riemannian Geometry [2]:

This approach has been applied in the past on radar signal processing and image processing. If there is a reason for using the covariance matrix as a feature in a classifier the obvious choice will be to vectorize this covariance matrix in order to process the quantity as a vector and then we can use any vector-based classification algorithm so this approach will be the same as CSPcommon spatial pattern, like special filtering first then we can use any classification. The direct vectorization of special covariance matrix will not take into account the correlation between the vectors the coefficients of the symmetric and the symantec positive definite SPD, along with the vectorized covariance metrics do not follow the normal distribution so a large number of classification algorithm are less effective like e.g. linear discernment analysis LDA which is optimum for Gaussian distribution. The Riemannian method is talking into account the Riemannian Geometry at each point a scalar product can be defined in the associated tangent space.

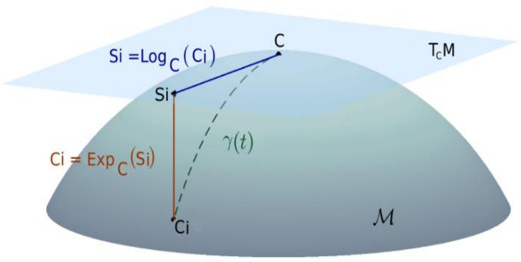

Fig 2. Manifold M and the corresponding local tangent space TCM at point C.

ElectrodeSelection [5]:

The channels selection is two types,first*subject based*: which is the best channel per subject or in other words we choose the channels that give us the best results per subject. The second type is *Application Based:* which choose the best channel for a particular action or stimuli. *Backward selection* and *forward selection*are including or excluding one channel at a time for best fit, but the computational power of this solution is exponential as well as it'sit's time-consuming. In this model, we choose to use the method proposed by [5] as it's using the Riemannian distance between two covariance matrices to choose the best channels.

The range of model can be described as follow:

1) xDAWN [9][10]:

   to provide the best SNR Signal to Noise Ration we choose to use [9], which is a filtering technique that is takin into account the signal and the ration, as opposed to older techniques like PCA (Principal component analysis) which takes into account the signal solely. Then the LogisticRegression will run on the filtered data from the previous step.

2) *Hankel [6]*

   The way that Hankel works is by building acovariance matrices by concatenating time delayed trial along the channel axis (Hankel matrices). It takes into account the temporal as well as spatial information. It's essential to reduce the size of the states. Again, we use the Logistic Regression model for classification.

3) *CSSP [7]*

   Common Spatio-Special Pattern, this approach is a combination of the previous ones, this technique will feed the well-known CSP to the Hankel covariance, then we will use the logistic regression as a posterior step.

This is considered as a controlled environment because of multiple factors, first, we didn't have the urge to clear the artefact e.g. eye blinking (movement) as we asked the subjects to try to reduce the eye blinking as much as possible, but in real life scenarios, this artefact has to be removed.

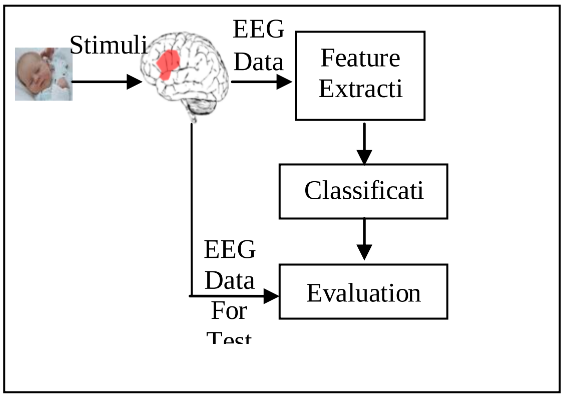

Fig.3 Describe the steps of EEG data collection and model generation

Also, we calculate when to show the images and then tag it with appropriate emotion this will create very clear lanes for the trails, thus the trails will be clearly defined due to the experiment setup and it will make the whole process easier. Compared to real-life scenarios where the trails are not clearly defined. In addition to the image set for example GAPEDit's awell-knowndataset and the contained images are tagged with proper emotions (arousal, valance and dominance) from a set of previous studies, so it's easy to collect the appropriate images that are almost guaranteed to trigger the targeted emotions, also the fact that it's open sourced and free to use makes it simpler and more convenient. Adding to that, the range of emotion that we choose can produce 20 different emotions maximum, but the standard SAM survey contains 5 items in each dimension which will result in 35 different emotions. In addition to that, the fact that all the sessions have been done in the same time window, while the mood of the subject is stillpersistent,and the subject is also protected from any distractions like an external sound, or sudden sounds or any nearby movement. All these percussions will not be present in a real-life scenario and will affect the quality of the signals and in turn the classification accuracy.

D. Evaluation

Predicting whatimagethe subject looked at, and predicting the emotion that the subject get when he looked at that image, is a totally different kinds of predictions. In other words,person(A) could see a picture and the triggered emotional response will be happy but the same image could be seen by person (B) and the triggered emotion is sadness, hence the SAM method. Also, the pictures are infinite, building our model on a specific set of pictures is not a good idea, as we will need to retrain it every time we want to change those images.

The first step will be to collect the EEG data and tagging it with the user responses (emotions), the responses that we collected it using SAM,next step will be fitting all the proposed models by the data that we collected it from session 1, step three will be choosing the best model per person depending on the best results. The evaluation is conducted from the data collected in session 2, where the subject is introduced to the same pictures but in a deferent order (the pictures are shuffled)then he tags it with his emotion at that moment, the shuffling is to help with the emotional memory and to increase the accuracy of our evaluation, the premise here is the subject will have the same emotional response if he saw the picture again.The accuracy is calculated based on the classes that have been predicted successfully, the classes have been produced by feeding the data from session two to the best model, which is selected per subject, the accuracy of the first subject is 42.8% and the second subject is 72%.This is because the second subject has less distinct emotions than the other subject. Another factor is the signal accuracy could be deferent between the subjects resulting in a difference in the accuracy.

IV. CONCLUSION AND DISCUSSION

Based on the controlled experiment setup of the study and the different models that have been adopted, the accuracy is ranging from 50-70% depending on the distinct emotion that the subject selected, as the range widens the accuracy shrinks. Giving the fact that the training data for the models was only one session, the models were able to produce a very height accuracy, but in the other hand the model generation is a very slow and expensive process, due to the multiple models being trained per subject, that can be tackled in the re-training phase by re-training only the most accurate model. Thelow number of subjects in this study makes it harder to see the value of other models,not all the models are useful for all the subjects. This architecture needs to be challenged more by using more classification algorithms and more data for more accuracy. Increasing the data might also mean increasing the number of sessions,to collect more data for training and only one session for testing purposes, this will increase the accuracy. Finally, we urge future studies to consult an ethical board or to collect written approval rather than a verbal one. Its relatively easy to notice a pattern and predict it, predicting emotions from EEG signals is relatively easy, but that doesn't mean we understand emotions, we just noticed a phenomenon then we tookthe best advantage of it. This field needs more input from the neuroscience and the medical field to flourish our understanding of emotions and to influence the machine learning models thus flourish the BCI and HMI fields.

V. FUTURE WORK

This architecture is fit for window-based analysis (Trials) and the window for the current setup is 2.5 second for online analysis this need to be reconsidered. Another thing could be enhanced could be in the model selection, for the current setup we select the best model per subject a better approach could be picking the highest probability across the models but now we need to consider the retraining because if we relied on all the models we have to retrain all of them every time which computationally expensive.

REFERENCES

[1] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," J. Behav. Ther. Exp. Psychiatry, vol. 25, no. 1, pp. 49–59, 1994.

[2] A. Barachantet al., "Classification of covariance matrices using a Riemannian-based kernel for BCI applications To cite this version : HAL Id : hal-00820475 Classification of covariance matrices using a Riemannian-based kernel for BCI applications," 2013.

[3] H. Xu and K. N. (Kostas) Plataniotis, "Affect recognition using EEG signal," IEEE 14th Int. Work. Multimed. Signal Process., pp. 299–304, 2012.

[4] K. J. Miller, G. Schalk, D. Hermes, J. G. Ojemann, and R. P. N. Rao, "Spontaneous Decoding of the Timing and Content of Human Object Perception from Cortical Surface Recordings Reveals Complementary Information in the Event-Related Potential and Broadband Spectral Change," PLoSComput. Biol., vol. 12, no. 1, pp. 1–20, 2016.

[5] A. Barachant, S. Bonnet, A. Barachant, S. Bonnet, C. Selection, and A. Barachant, "Channel Selection Procedure using Riemannian distance for BCI applications To cite this version : Channel Selection Procedure using Riemannian

distance for BCI applications," 2011.

[6] S. Lemm, B. Blankertz, G. Curio, and K. R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," IEEE Trans. Biomed. Eng., vol. 52, no. 9, pp. 1541–1548, 2005.

[7] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," Brain Topogr., vol. 2, no. 4, pp. 275–284, 1990.

[8] G. Cheng, "A closed-loop Brain-Computer Music Interface for continuous affective interaction," no. December, pp. 176–179, 2017.

[9] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: application to brain computer interface," Biomed Eng, IEEE Trans, vol. 56, pp. 1–9, 2009.

[10] B. Rivet, H. Cecotti, A. Souloumiac, E. Maby, and J. Mattout, "Theoretical analysis of XDAWN algorithm: Application to an efficient sensor selection in a P300 BCI BT - 19th European Signal Processing Conference, EUSIPCO 2011, August 29, 2011 - September 2, 2011," Eur. Signal Process. Conf., no. Eusipco, pp. 1382–1386, 2011.

[11] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation," Sci. World J., vol. 2014, 2014.

[12] M. Congedo, A. Barachant, and A. Andreev, "A New Generation of Brain-Computer Interface Based on Riemannian Geometry," arXivPrepr. arXiv1310.8115, vol. 33, no. 0, 2013.

[13] X. Li et al., "Emotion Recognition from Multi-Channel EEG Data through Convolutional Recurrent Neural Network," 2016 Ieee Int. Conf. Bioinforma. Biomed., pp. 352–359, 2016.

[14] H. Zhang et al., "A feasibility study of detecting brain signal in EEG during emotional self-regulation," pp. 184–187, 2017.

[15] S. I. Alzahrani, "P300 Wave Detection Using EmotivEpoc+ Headset: Effects of Matrix Size, Flash Duration, and Colors," p. 76f., 2016.

[16] W. L. Zheng, W. Liu, Y. Lu, B. L. Lu, and A. Cichocki, "EmotionMeter: A Multimodal Framework for Recognizing Human Emotions," IEEE Trans. Cybern., pp. 1–13, 2018.

[17] L. R. Christensen, M. A. Abdullah "EEG Emotion Detection Review," 2018 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol., pp. 1–7, 2018.