2024 18th International Conference on Automatic Face and Gesture Recognition (FG)

A Spectro-Statistical Approach for Emotion Identification from EEG Signals

Lownish Rai Sookha[1], Gulshan Sharma[1], M. A. Ganaie[1], Abhinav Dhall[2,1]
[1] Indian Institute of Technology Ropar, India
[2] Flinders University, Australia

***Abstract*— Automatic identification of emotions is important in human-centered computing. It allows machines to better understand user emotions. Identifying emotions via neural sensing techniques such as electroencephalogram (EEG) is a promising approach. In this paper, we aim to identify the emotions class from EEG signals. We frame emotion identification as a classification task and apply spectral and statistical encoders to extract the relevant features. We validate our approach on EmoNeuroDB dataset. Our method outperforms the EmoNeuroDB baseline, achieving a 42.10% increase in class prediction accuracy.**

I. Introduction & Background

Emotion is a psycho-physiological phenomenon which plays an important role in shaping human interactions. It is closely linked to an individual's personality, mood, temperament, and motivation [1]. One key way in which emotion influences human interaction is through its impact on communication. Emotions affect how messages are conveyed, perceived, and interpreted during social exchanges. Emotions can be expressed through verbal cues such as voice tone, choice of words, as well as through nonverbal cues like facial expressions, body language, and gestures [2].

In human-machine interactions, the *automatic* identification of emotions allows the machine to understand the user's emotional state [3]. This enables the machine to be more *empathetic* in its interactions with humans [4]. Consider a scenario where a virtual assistant detects sadness in a user's voice. Rather than processing commands mechanically, the assistant can adjust its responses to offer a human-like empathetic interaction. This improves the user experience, making interactions with machines more natural and satisfying.

User emotions can be identified with various sensing modalities such as facial expressions, speech patterns, textual content, and physiological signals [5]. Facial expressions refer to the different arrangements of facial muscles, which can be systematically analyzed to determine a user's emotional states [6]. Speech-based emotion identification relies on recognizing the connection between acoustic features and emotions [7]. Changes in features like pitch, intensity, and prosody in speech can indicate various emotions. For example, when someone is sad, their speech may have a lower pitch, less variation in pitch, and a slower speech rate. Text-based emotion identification involves determining the sentiment conveyed in a text. This involves analyzing syntactic patterns and linguistic style to identify the underlying emotions [8].

However, the aforementioned sensing modalities may not always reflect a person's true emotional state. Some users may have atypical facial and vocal expressions, while others may intentionally mask their emotions. On the other hand, physiological sensing captures users' *implicit signals*, which are difficult to manipulate. It tracks changes in parameters like heart rate, skin conductance, and brain activity, providing insights into user's emotions and arousal levels.

Analyzing brain activity in response to emotional stimuli can be conducted through methods like functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). These methods allow to observe which areas of the brain are activated when experiencing specific emotions. EEG is non-invasive method, which provides precise measurements of electrical activity in the brain [9]. It has a high temporal resolution, allowing to capture rapid changes in brain activity of a user in real-time [10]. However, analyzing EEG signals can be challenging due to noise interference and limited spatial resolution. These signals are recorded via multiple channels, which often introduces internal and external noise. This noise complicates the recognition of emotional responses.

In recent years, conventional machine learning and deep learning based approaches have been applied to identify emotions from the EEG signals [11]. These approaches have explored spatio-temporal aspects [12], and fusion of EEG with visual and audio modalities [13]. In machine learning approaches, there are typically two stages: feature extraction and classifier selection [12]. EEG features can be extracted by signal processing techniques such as Fourier transform or wavelet analysis, as well as statistical measures like mean, variance, or entropy calculations [14], [15]. These extracted features serve as inputs to the chosen classifier, which then learns patterns and makes predictions [16]. On the other hand, deep learning based approaches allow for the automatic learning of feature representations from raw EEG signals. It can capture complex relationships and dependencies within the signal, leading to improved emotion identification [17].

In this regard, we propose a hybrid approach where emotion classification is performed by integrating feature representations obtained from machine learning and deep learning-based classifiers. We compute the weighted sum of these feature representations and then apply the argmax operation to determine the emotion class with the highest confidence level. Our empirical finding on EmoNeuroDB dataset suggests that performing hybrid fusion leads to

979-8-3503-9494-8/24/$31.00

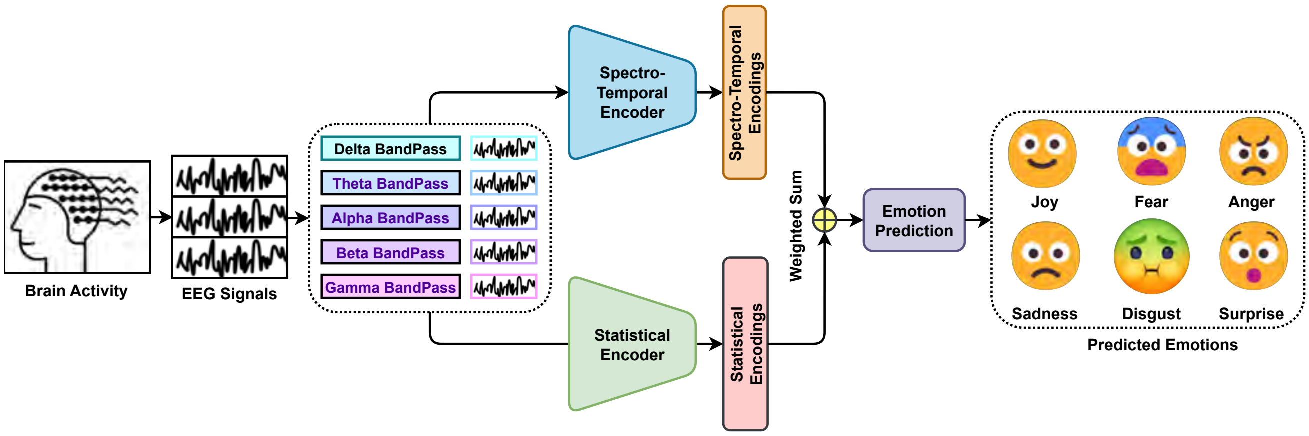


Fig. 1. **Proposed Approach**: EEG siganls are input into a 2nd order butter-worth bandpass filters to extract δ, θ, α, β, and γ bands. These filtered signals are then input into spectro-temporal and statistical encoders, which are responsible for learning relevant features. These encoders then outputs probabilities for each emotion class. To make the final prediction, we calculate a weighted sum of these class probabilities and apply argmax operation to determine the predicted emotion class.

improved emotion classification compared to using either approach independently.

II. Method

A. Problem Formulation

We formulate emotion identification as a classification task and the objective is to classify the EEG signals into one of six distinct emotion classes.

B. Dataset

The EmoNeuroDB [18] dataset includes multi-channel EEG recordings from 40 participants. During the recordings, participants are presented with an avatar in a virtual reality environment. The avatar displays six different emotions: fear, joy, anger, sadness, disgust, and surprise as stimuli. Participants are asked to mimic the avatar's facial expressions and their brain activity is recorded via DSI-24 wireless dry electrode EEG headset. 40 participants, including 16 females, with ages ranging from 19 to 57 years participated in the study. Each recording lasted for 15 seconds and was sampled at a rate of 300 Hz.

The dataset is split into three subject-independent subsets: training, validation, and testing, with no overlap between them. The training set contains 360 EEG recording samples and validation and test sets have 180 samples each. Each set has a distinct subjects, with 3 EEG recordings for each emotion. The dataset is perfectly balanced for each class. For more information please refer to the [18].

C. Feature Extraction

To extract meaningful features from EEG signals, we first divide the EEG signal into δ (1-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-30 Hz), and γ ($\geq$30 Hz) bands via applying a second-order Butterworth bandpass filter [19], [20]. Afterwards, we apply spectro-temporal and statistical encoders. Input to these encoders are topographic maps and statistical features, respectively. The output from these encoders are class-wise probabilities from the classification algorithms.

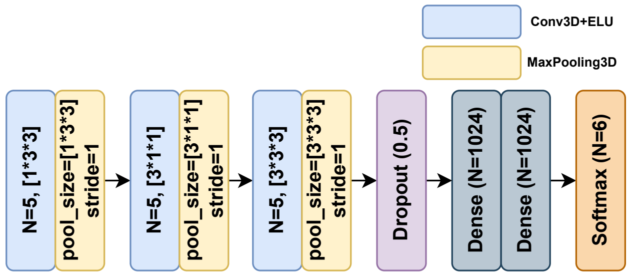


Fig. 2. 3D-CNN architecture designed to extract latent features from the topographic maps. Refer to Section II for details.

1) Spectro-temporal Encoder: Spectro-temporal encoding are generated in two stages: In first stage we generate topographic maps, which captures both the spectral and spatial information of the EEG signal [20]. These maps are generated at 0.5-second intervals and have a resolution of 64$*$64 pixels. Afterwards, we apply 3D-CNN to learn latent features from the topographic maps. The architecture of the 3D-CNN consists of three stacks of convolution with a max-pooling layer, followed by two fully connected layers. The output of this encoder comprises class-wise probabilities extracted from the softmax layer of the 3D-CNN. Figure 2 illustrates the 3D-CNN architecture.

2) Statistical Encoder: Input to this encoder is statistical features of the EEG signals. These feature includes Permutation Entropy, Spectral Entropy, Singular Value Decomposition Entropy, Sample Entropy, Approximate Entropy, Katz Fractal Dimension, Petrosian Fractal Dimension, and Higuchi Fractal Dimension [21], [22]. After computing these features, we apply following classification algorithms.

- Random Forest (RF) [23], is a ensemble of decision trees, where each tree is trained on a randomly selected subset of the training data. The final prediction is the majority voting among individual trees.

TABLE I

COMPARISON OF CLASS-WISE ACCURACY ON THE VALIDATION SET FOR DIFFERENT ENCODERS.

Methods	Fear	Joy	Anger	Sadness	Disgust	Surprise	Mean Accuracy
Baseline Method [18]	0.10	**0.53**	0.10	0.10	0.07	**0.27**	0.19
Spectro-temporal Encoder	0.10	0.10	0.10	0.43	0.27	0.07	0.18
SVM based Statistical Encoder	0.00	0.00	0.37	0.03	**0.70**	0.10	0.20
LDA based Statistical Encoder	0.30	0.20	0.27	0.47	0.10	0.20	0.26
RF based Statistical Encoder	0.33	0.17	0.4	0.23	0.13	0.17	0.24
RF based Statistical Encoder (Top 50% features)	0.10	0.17	**0.43**	**0.57**	0.13	0.17	0.26
Proposed Approach	**0.33**	0.43	0.23	0.20	0.33	0.10	**0.27**

- Linear Discriminant Analysis (LDA) [24], is a generative algorithm, which constructs a linear decision boundary between classes in the feature space.
- Support Vector Machine (SVM) [25], is a discriminative classifier, which finds the optimal decision boundary such that there is maximal distance between decision boundary and support vectors.

D. Classification Method

Once we obtain the class-wise probabilities from Spectro-temporal Encoder and Statistical Encoder, we combine these probabilities using a weighted sum. The argmax operation is then applied to the weighted sum to predict the emotion class (refer to Figure 1). We test various weight values, ranging from 0.1 to 0.9, and report the results in Table I on optimal weight value.

E. Model Training & Validation Procedure

When training the Spectro-temporal Encoder, we apply Adam optimizer with a learning rate of 0.001 and categorical cross-entropy for loss computation. We set the batch size to 32 and maximum number of epochs at 30.

While training the RF based Statistical Encoder, the top 50% of the relevant features are used to train similar RF encoder to achieve better performance. The RF based Encoder has been trained using the following parameters: *50* trees, *gini* criterion and a minimum of *2* samples required to split a node.

To evaluate the classification performance, we apply 5-fold cross-validation. Within each fold, we select the best-performing model for the final evaluation. This selection process helps us avoid choosing a model converged at poor local minima [26]. The experiments are conducted on an Nvidia A100 GPU, which is equipped with 40 GB of GPU memory.

III. RESULTS & DISCUSSION

A. Overall Results

We observe that weighted sum of class-wise probabilities from spectral and statistical encoders produces the best results. While the baseline method achieves an average accuracy of 0.19, our approach achieves 0.27 on the validation set. On the test set, the baseline accuracy is at 0.19, whereas our method achieves 0.24, which is *42.10%* increase in class prediction accuracy.

We compare different models via mean accuracy metric. Table I displays the overall results of different encoder types.

1) While the spectro-temporal encoder has lower accuracy compared to the baseline method, its predictions improves when combined with the RF based statistical encoder.
2) The various statistical encoders tested have outperformed the baseline method. However, SVM and LDA based statistical encoders are biased towards certain emotions.
3) The proposed method, which combines spectro-temporal and RF based statistical encodings with a weighted approach, achieves the highest overall average accuracy.

B. Class-wise Results

For statistical encoder, we experiment with different classifiers. We observe that LDA based statistical encoder has better overall accuracy. However its class-wise results are more biased towards *Sadness* class (refer to Figure 3(c)). We observe a similar trend with SVM based statistical encoder, which is biased towards the *Disgust* class (refer to Figure 3(b)).

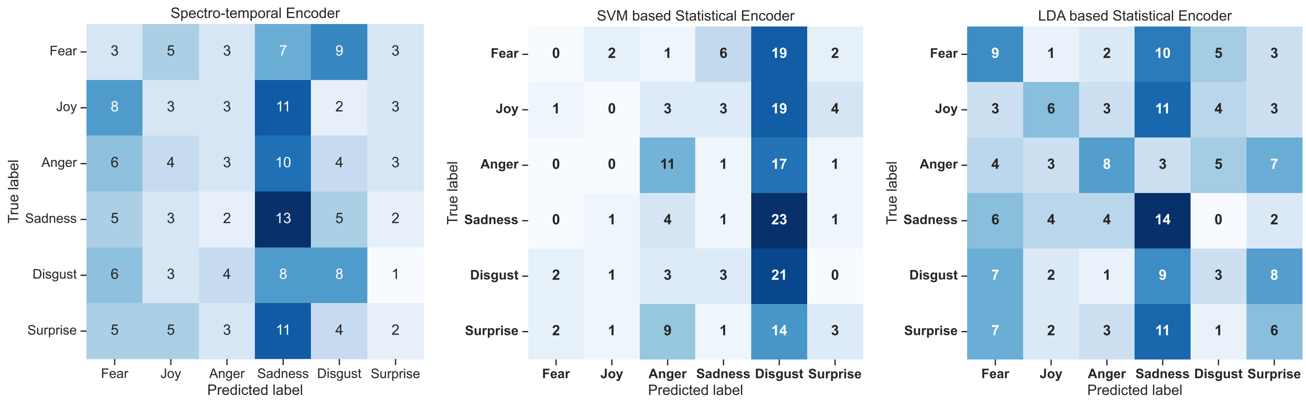
In case of RF based statistical encoder, which is inferior to LDA based statistical encoder in term of overall accuracy, we observe that class-wise accuracy is better distributed. There is less confusion among the classes, and it shows higher accuracy in identifying *Fear* and *Sadness* class. However, many instances of *Joy* are classified as *Fear*. *Surprise* remains the least accurately classified emotion. We also observe that upon selecting the top 50% of the features via Gini index, there is an increase in accuracy (refer to Table I) and reduced confusion among the classes (refer to Figure 3(e)).

Our proposed method, demonstrates an improvement in both overall prediction accuracy and individual emotion accuracies, notably on *Joy*, *Anger* and *Sadness*. However, many of the *Surprise* instances are still misclassified as *Anger*. Another noteworthy observation is that the individual encoders tend to favour a particular emotion. Our proposed fusion method maintains a better ratio across the various classes.

Figure 3 shows class-wise accuracy. A significant amount of the *Surprise* signals have been misclassified as *Fear*. This can might be due to the spontaneity present in both emotions which have led the features to appear relatively simpler.

TABLE II
THE PRECISION, RECALL, AND F1 SCORES (MACRO WEIGHTED) FOR DIFFERENT ENCODERS ON THE VALIDATION SET.

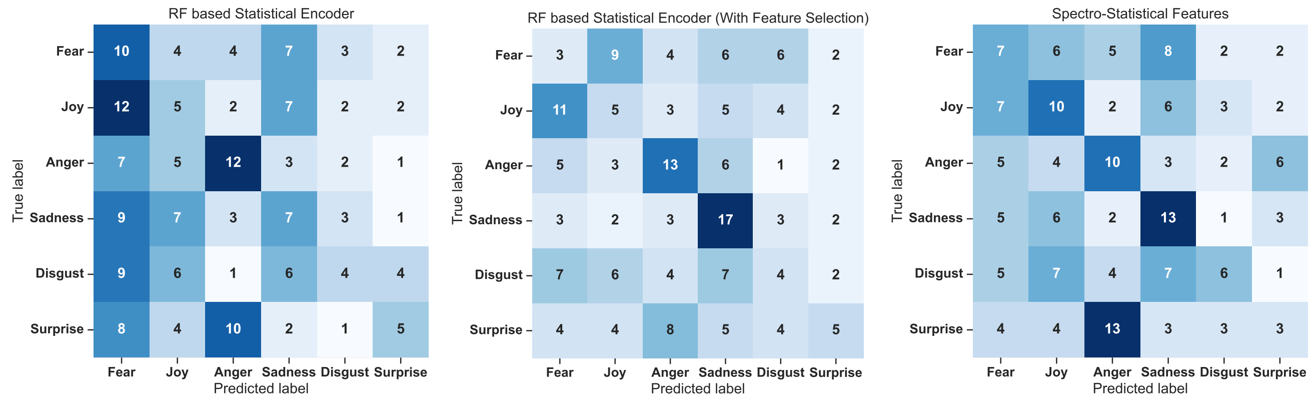
Methods	Precision	Recall	F1-Score
Spectro-temporal Encoder	0.17	0.18	0.16
SVM based Statistical Encoder	0.15	0.20	0.14
LDA based Statistical Encoder	0.26	0.26	0.25
RF based Statistical Encoder	0.26	0.24	0.24
RF based Statistical Encoder (Top 50% features)	0.25	0.26	0.25
Proposed Approach	**0.27**	**0.27**	**0.26**



(a) Spectro-temporal Encoder. Predictions are biased towards *Sadness* class.

(b) SVM based Statiscal Encoder. Predictions are biased towards *Disgust* class.

(c) LDA based Statistical Encoder. Predictions are biased towards *Sadness* class.



(d) RF based Statistical Encoder. Predictions are biased towards *Fear* class.

(e) RF based Statistical Encoder (Top 50% features). The Anger and Sadness classes demonstrate better performance.

(f) Proposed Approach. *Surprise* class is often mis-classified as *Anger*.

Fig. 3. Confusion matrices depicting the performance of different encoders on the validation set.

C. Proposed Approach on Validation & Test Set

While comparing the proposed approach on validation and test set, we observe some variations in the performance across individual classes. There has not been much changes in the *Fear* and *Anger* classes, but we have observed significant differences in accuracy for *Joy*, *Disgust*, and *Surprise* class. *Disgust* performed very poor on the test set, while class performed well and nullifying the effect. The performance of *Disgust* class is notably poor on the test set, whereas *Surprise* class performed well, effectively neutralizing the overall impact. Table III presents a comparison of accuracies between the validation and test sets. We additionally compute precision, recall, and F1 scores for various encoders on the validation set. In Table II, we present the macro weighted scores, revealing that our proposed method outperforms the different encoders.

Furthermore, we observe performance variations between the validation and test set (refer to Table III). The variation may be due to the high inter-participant variability. As validation and test set include non-overlapping participants data.

IV. CONCLUSION

In this paper, we present a spectro-temporal and statistical embedding fusion approach for the identification of emotions from EEG signals. Emotions are elicited through a virtual avatar presented to participants, and the identification task

TABLE III

CLASS-WISE RESULTS OF PROPOSED APPROACH.

Emotion Class	Validation Set	Test Set
Fear	0.33	0.33
Joy	0.43	0.30
Anger	0.23	0.20
Sadness	0.20	0.30
Disgust	0.33	0.03
Surprise	0.10	0.30
Mean Accuracy	0.27	0.24

is framed as a classification problem. Our approach involves calculating a weighted sum of prediction probabilities generated by spectro-temporal and statistical encoders, resulting in a robust solution for emotion identification. As part of future work, we propose incorporating data augmentation strategies, especially synthetic EEG data generated via generative algorithms [10] to improve the robustness and generalization. In addition, we suggest to explore feature selection techniques to minimize redundant features. This will help in improving interpretability of the proposed approach.

REFERENCES

[1] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2011.

[2] Paul Ekman and Wallace V Friesen. *Unmasking the face: A guide to recognizing emotions from facial clues*, volume 10. Ishk, 2003.

[3] Ramanathan Subramanian, Julia Wache, Mojtaba Khomami Abadi, Radu L. Vieriu, Stefan Winkler, and Nicu Sebe. Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing*, 9(2):147–160, 2018.

[4] Tengfei Song, Wenming Zheng, Peng Song, and Zhen Cui. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*, 11(3):532–541, 2020.

[5] Naveed Ahmed, Zaher Al Aghbari, and Shini Girija. A systematic survey on multimodal emotion recognition using learning algorithms. *Intelligent Systems with Applications*, 17:200171, 2023.

[6] Shan Li and Weihong Deng. Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 13(3):1195–1215, 2022.

[7] Siddique Latif, Rajib Rana, Sara Khalifa, Raja Jurdak, Junaid Qadir, and Björn Schuller. Survey of deep representation learning for speech emotion recognition. *IEEE Transactions on Affective Computing*, 14(2):1634–1654, 2023.

[8] Jiawen Deng and Fuji Ren. A survey of textual emotion recognition and its challenges. *IEEE Transactions on Affective Computing*, 14(1):49–67, 2023.

[9] Gulshan Sharma. Physiological sensing for media perception & activity recognition. In *Proceedings of the 2022 International Conference on Multimodal Interaction*, ICMI '22, page 696–700, New York, NY, USA, 2022. Association for Computing Machinery.

[10] Gulshan Sharma, Abhinav Dhall, and Ramanathan Subramanian. MEDiC: Mitigating EEG data scarcity via class-conditioned diffusion model. In *Deep Generative Models for Health Workshop NeurIPS 2023*, 2023.

[11] K Kannadasan, Jainendra Shukla, Sridevi Veerasingam, B Shameedha Begum, and N Ramasubramanian. An EEG-based computational model for decoding emotional intelligence, personality, and emotions. *IEEE Transactions on Instrumentation and Measurement*, 2023.

[12] Zhongjie Li, Gaoyan Zhang, Longbiao Wang, Jianguo Wei, and Jianwu Dang. Emotion recognition using spatial-temporal EEG features through convolutional graph attention network. *Journal of Neural Engineering*, 20(1):016046, 2023.

[13] Tzyy-Ping Jung, Terrence J Sejnowski, et al. Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing. *IEEE Transactions on Affective Computing*, 13(1):96–107, 2019.

[14] Kalidasan Kannadasan, Sridevi Veerasingam, B Shameedha Begum, and Natarajan Ramasubramanian. An EEG-based subject-independent emotion recognition model using a differential-evolution-based feature selection algorithm. *Knowledge and Information Systems*, 65(1):341–377, 2023.

[15] Anurag Tiwari and Amrita Chaturvedi. A multiclass eeg signal classification model using spatial feature extraction and xgboost algorithm. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4169–4175, 2019.

[16] Pankaj Pandey, Gulshan Sharma, Krishna. P. Miyapuram, Ramanathan Subramanian, and Derek Lomas. Music identification using brain responses to initial snippets. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1246–1250, 2022.

[17] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

[18] Emoneurodb challenge. https://voxellab.pl/EmoNeuroDB/. [Accessed 21-12-2023].

[19] Ivan W Selesnick and C Sidney Burrus. Generalized digital butterworth filter design. *IEEE Transactions on signal processing*, 46(6):1688–1694, 1998.

[20] Gulshan Sharma, Pankaj Pandey, Ramanathan Subramanian, Krishna Prasad Miyapuram, and Abhinav Dhall. Neural encoding of songs is modulated by their enjoyment. In *Proceedings of the 2022 International Conference on Multimodal Interaction*, ICMI '22, page 414–419, New York, NY, USA, 2022. Association for Computing Machinery.

[21] Rab Nawaz, Kit Hwa Cheah, Humaira Nisar, and Vooi Voon Yap. Comparison of different feature extraction methods for eeg-based emotion recognition. *Biocybernetics and Biomedical Engineering*, 40(3):910–926, 2020.

[22] Xiang Li, Dawei Song, and Bin Hu. Exploring EEG features in cross-subject emotion recognition. *Frontiers in neuroscience*, 12:294333, 2018.

[23] Islam A. Fouad. A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms. *Ain Shams Engineering Journal*, 14(3):101895, 2023.

[24] Rukhsar Shaik, Hema Kumar Goru, Harshitha Nadakuditi, Gayathri Vasamsetti, and Balakrishna Velagapudi. A Comparative Study of Naive Bayes, LDA and Gradient Boosting Classifiers for Epileptic Seizure Detection Using Discrete Wavelet Transform. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pages 1–7, 2023.

[25] J. Ruiz de Miras, A.J. Ibáñez-Molina, M.F. Soriano, and S. Iglesias-Parro. Schizophrenia classification using machine learning on resting state EEG signal. *Biomedical Signal Processing and Control*, 79:104233, 2023.

[26] Kenji Kawaguchi. Deep learning without poor local minima. *Advances in neural information processing systems*, 29, 2016.