HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 4 (2024)

Towards Better Mental Health: Rapid Screening for Mood Disorder Using Electroencephalogram and Facial Expression Paired Data with Weakly Supervised Contrastive Learning

Grace Kim[1] and Ju Young Lee[#]

[1]Phillips Academy Andover, USA
[#]Advisor

ABSTRACT

Depression has become a growing crisis among teenagers in recent years, with rates of suicide and depression rising annually. Despite the increasing need for more accessible resources to combat this mental health emergency, most available mental health resources remain either too expensive or inaccurate to effectively help those affected by the surge in mental health crises. The method proposed in this paper aims to combat the mental health epidemic by using Electroencephalogram and Facial Expression data to train a multi-modal system that can analyze emotion to detect early signs of depression and suicidal ideations. The proposed system is composed of two modules: an emotion classification module and a mood disorder screening module. The proposed emotion classification module takes both Electroencephalogram data and facial images as inputs and classifies the individual's emotional categories. After predicting the emotional status, the mood disorder screening module detects abnormalities in the individual's mental health condition by comparing the results with those of other individuals. Through extensive experiments, I have demonstrated that the proposed system achieves an accuracy of 95.7% on the Electroencephalogram and Fical Expression dataset which proves its feasibility for real-world application.

Introduction

Over the past decade, mental health has become an increasingly prevalent issue among teenagers, growing to become a national health crisis. According to the Centers for Disease Control and Prevention, from 2000 to 2019, the percentage of high school students reporting symptoms of depression increased by 40%, and between 2007 and 2018, suicide rates among the 10-24 age group increased by 57%. In 2020 alone, it is estimated that there were 6,600 deaths by suicide among the 10-24 age group (Office of the Surgeon General, 2021).

Unfortunately, many teenagers who need help are not receiving it at crucial times, due to the high costs and other issues with therapy such as misdiagnoses. Thus as environmental stressors continue to mount, these individuals are left with no means of mitigating the psychological effects of such pressures. Previous attempts to increase the availability of resources, such as suicide hotlines or online mental health services, which have been in service since the 2000s, have not been effective in curbing this epidemic (Rauch, 2017).

Traditional mood disorder screening methods rely on survey checklists, such as the Patient Health Questionnaire (PHQ-9) and the Beck Depression Inventory (BDI). However, these methods to identify depression symptoms have lacked efficiency, accessibility, and accuracy. To solve this problem, this paper introduces a novel machine learning-based depression screening system that integrates facial expression classification and Electroencephalogram encoding techniques.

The rest of the paper is structured as follows: Chapter 2 provides background knowledge to better understand the proposed approach. Chapter 3 explains detailed information about the proposed mood disorder screening system. Chapter 4 demonstrates the effectiveness of the proposed approach through extensive experiments. Finally, Chapter 5 summarizes the paper.

Related Work

Facial Expression Classification

Facial expression classification involves using technology to recognize and interpret non-verbal cues that convey emotions. As illustrated in Figure 1, different emotions produce distinct physical changes in a person's face. This technology determines an individual's current mood by analyzing facial movements.

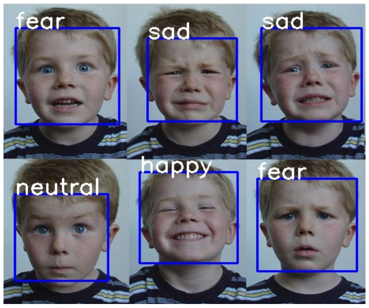


Figure 1. Facial expression recognition (Ashish 2020)

Facial expression recognition system is frequently developed using Convolutional Neural Network (CNN) which have demonstrated remarkable performance in various computer vision tasks. For instance, Akhand et al. proposed the use of deep CNN and the use of the transfer learning technique to improve the accuracy of facial recognition (Akhand et al. 2021). Another instance is Liu et al. 2020 where they utilized CNN to detect facial expressions and proved that CNN could predict emotional face value as accurately as the human volunteers (Liu et al. 2020).

Facial expressions are powerful indicators of emotional states, and there is a notable correlation between facial information and depression. Research shows that individuals with depression often display specific facial expressions and reduced emotional expressiveness, which can be detected through advanced facial recognition technology (Hu et al. 2023).

Electroencephalogram

Electroencephalography (EEG) is a non-invasive technique used to measure the electrical activity of the brain. It requires placing electrodes on the scalp to detect the electrical signals produced by neurons. Figure 2 illustrates the EEG signal capturing device and the obtained signal. EEG signals are widely used to diagnose and monitor conditions such as epilepsy, sleep disorders, encephalopathies, and brain death.

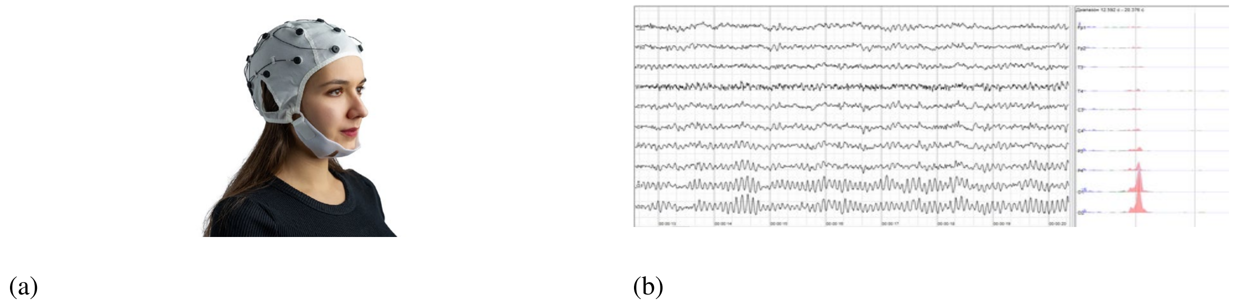


Figure 2. Electroencephalogram device and obtained signal
(a): Electroencephalogram device (OpenBCI 2021) and (b): Electroencephalogram signal (Wikipedia 2023)

EEG waveforms are often represented and measured in microvolts which reflect the strength of the electrical activities of the brain, as shown in Figure 2 (b). Due to its high temporal resolution, EEG is highly capable of detecting rapid changes in brain activity, making it ideal for real-time monitoring. It is also relatively low cost which enables the developed system to be more affordable compared to other neuroimaging techniques like magnetic resonance imaging or positron emission tomography.

Recent studies have discovered a high correlation between EEG signals and mood disorders such as anxiety and depression (de Aguiar Neto and Rosa 2019). Inspired by these findings, this study proposes a mood disorder screening utilizing EEG signals and the aforementioned facial images.

Proposed Mood Disorder Screening System

The purpose of the proposed mood disorder screening system is to diagnose and detect depression with enhanced accuracy and efficiency. To achieve this, the system integrates data from both an Electroencephalography (EEG) recorder and facial image analysis, as depicted in Figure 3. After collecting data from these sources, it is processed through encoders based on Convolutional Neural Networks (CNNs). These encoders generate embedding vectors, which are then compared mathematically to vectors representing the same emotional state.

This chapter is organized as follows: Section 3.1 explains how EEG and facial image data are utilized to create more accurate emotion-related representations. Section 3.2 discusses the application of transfer learning to enhance mood disorder screening through emotion classification. Finally, Section 3.3 explores the mood disorder screening process. In this post-processing step, the predicted emotion is compared with the emotions of other individuals or with the same individual's past emotions in response to similar emotional stimuli to determine irregularities.

Weakly Supervised Contrastive Learning

Figure 3 illustrates the proposed weakly supervised contrastive learning approach for training both the EEG encoder and the facial image encoder. The EEG encoder processes a set of recorded EEG signals, each representing one of five different emotions, and converts them into EEG feature vector $\boldsymbol{E}_{k}$, where $\boldsymbol{k}$ ranges from 1 to 5. Similarly, the facial image encoder processes a set of facial images, also representing different emotions, to generate image feature vectors $\boldsymbol{I}_{k}$.

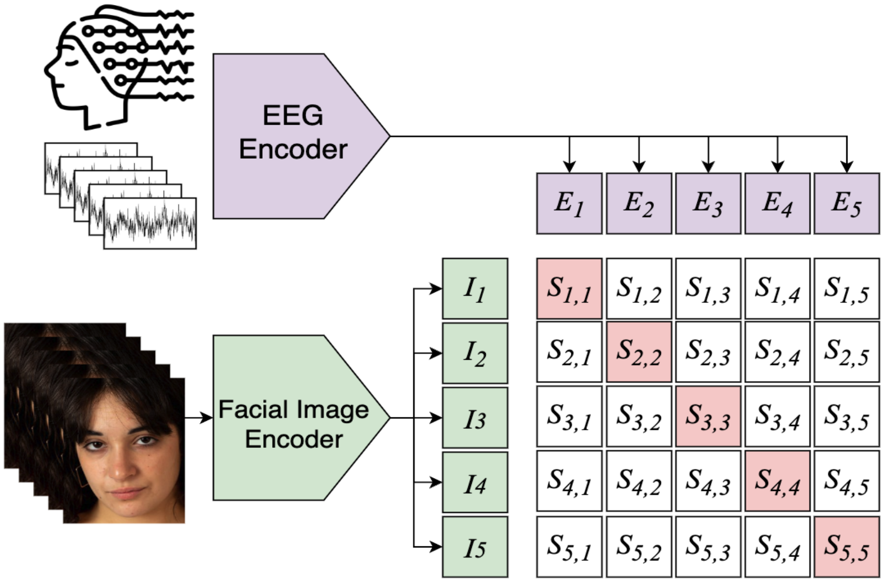

EEG Encoder
E_1
E_2
E_3
E_4
E_5
Facial Image Encoder
I_1
I_2
$I3$
$I4$
$I5$
$S_{1,1}$ $S_{1,2}$ $S_{1,3}$ $S_{1,4}$ $S_{1,5}$
$S_{2,1}$ $S_{2,2}$ $S_{2,3}$ $S_{2,4}$ $S_{2,5}$
$S_{3,1}$ $S_{3,2}$ $S_{3,3}$ $S_{3,4}$ $S_{3,5}$
$S_{4,1}$ $S_{4,2}$ $S_{4,3}$ $S_{4,4}$ $S_{4,5}$
$S_{5,1}$ $S_{5,2}$ $S_{5,3}$ $S_{5,4}$ $S_{5,5}$

Figure 3. Proposed weakly supervised contrastive learning for multimodal mood disorder screening

The resulting vectors should be mathematically similar to one another if stemmed from the same emotion. Therefore, to measure the similarities between the Facial and EGG vectors, they are inputted into the Similarity Function shown in Equation 1.

Equation 1: Similarity Function

$$S_{E_i,I_j} = \frac{E_i \circ I_j}{|E_i| \times |I_j|}$$

Where, $\boldsymbol{S_{A,B}}$ denotes the similarity score between two feature vectors and $\boldsymbol{E}$ and $\boldsymbol{I}$ represent the data from EEG and facial expressions. The feature vectors that are the most similar to one another will have a value of 1 and the most different will result in a value of -1. Then once the feature vectors are produced, they are put through a classifier that categorizes the vectors into emotions. The mathematical process of predicting the possibility of the vectors fitting each emotion category is explained in Equation 2.

Equation 2: Softmax Function

$$P_{i,j} = \frac{e^{S_{i,j}}}{\sum_{k=1}^{5} e^{S_{i,k}}}$$

Where $P_{i,j}$ denotes the predicted probability of emotion categories from EEG and facial image. The variable $\boldsymbol{i}$ and $\boldsymbol{j}$ represents data from EEG and facial encoders. Finally, to measure the amount of loss of the prediction of the proposed system, the output probability is fed into a cross-entropy loss function shown in equation 3.

Equation 3: Cross Entropy Loss Function

$$L = -log_{e^{P_{i,j}}}$$

The $\boldsymbol{L}$ here represents the loss which is the difference between the prediction given by the CNN and its corresponding ground truth. The measured loss values range from 0 to infinite. Where 0 signifies an accordance with the prediction and truth..

Emotion Classification with EEG and Facial Image

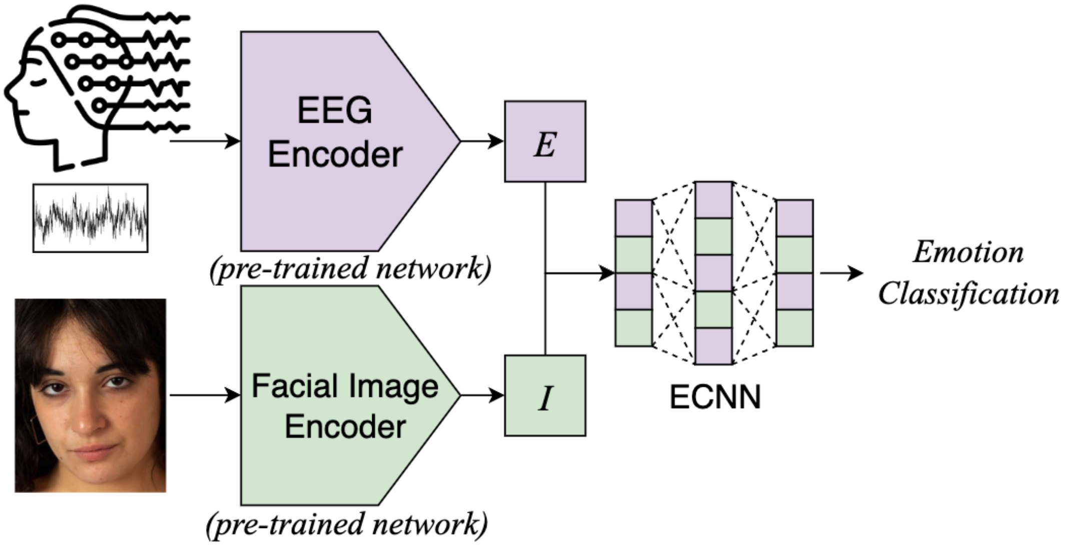


Figure 4. Transfer learning for emotion classification using both eeg and facial image

Figure 4 illustrates the transfer learning phase utilizing the pre-trained network from section 3.1. Both pre-trained facial image encoder and EEG encoder extract emotion-related features. These features are then fed into the Emotion Classification Neural Network (ECNN) to predict the emotional state of the individual based on their EEG and facial image data.

Since both encoder networks are pre-trained to find better representations of emotion-related features, the proposed transfer learning not only converges faster than training from scratch but also yields more accurate results. The detailed effectiveness of this approach will be explained in Chapter 4 with extensive experimental results.

Mood Disorder Screening

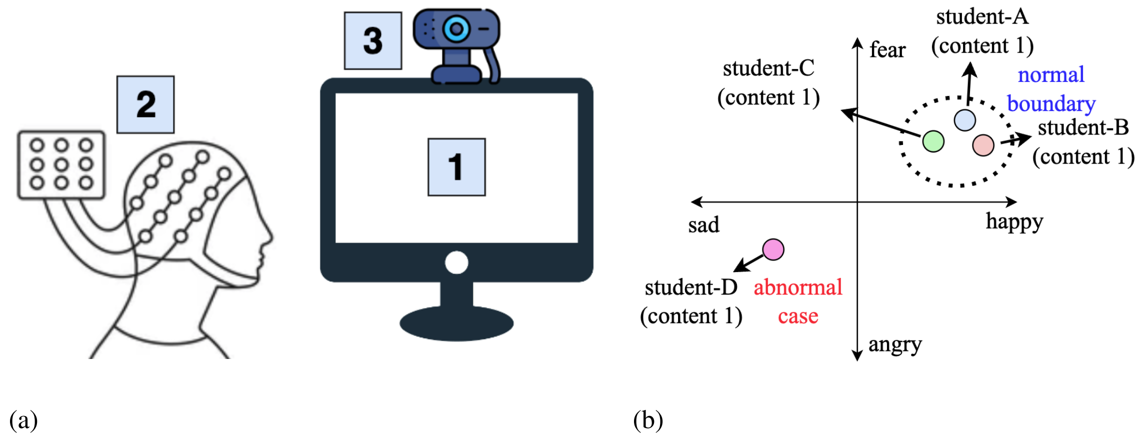


Figure 5. (a): Obtaining EEG and Facial Expression Data from individuals (1: video designed to elicit specific emotions, 2: webcam, and 3: EEG device) and (b): Emotional sates visualization graph depicting the difference between normal group and individual who might have mood disorders.

In this chapter, I introduce a mood disorder screening process utilizing predicted emotional states predicted from the proposed emotion classification network. Figure 5 (a) illustrates the setup for obtaining facial images and EEG signals from individuals. The subjects watch videos designed to elicit specific emotions on a monitor. During this process, an EEG device and a webcam capture the individuals' emotion-related facial images and EEG signals. These data are processed by the proposed emotion classification network and collected from various individuals.

Figure 5 (b) demonstrates how the predicted emotional states are used to screen individuals who might have mood disorders. It is assumed that an individual with a mood disorder would exhibit different emotional reactions to the same content compared to others, who generally display similar emotions. This scientific approach can effectively identify individuals who may need mental health care.

Experimental Results

This chapter provides comprehensive experimental results demonstrating the effectiveness of the proposed approach. First, I introduce the data used in this paper along with four evaluation metrics commonly used in classification tasks. Then, I demonstrate a performance comparison with state-of-the-art convolutional neural networks architectures. Finally, I conducted an ablation study to examine the effectiveness of the proposed weakly supervised contrastive learning.

Dataset

To train and evaluate the proposed system, two types of datasets are used: facial expression (Kaggle 2024) and EEG datasets (Liu et al. 2021). The facial expression dataset is comprised of 31,337 samples. 80% of the dataset is used to train the proposed system while the remaining 20% is used for testing. Figure 6 illustrates the category distribution of each emotion occupied in the facial expression dataset. While Figure 7 shows example images that are included in each emotions' section.

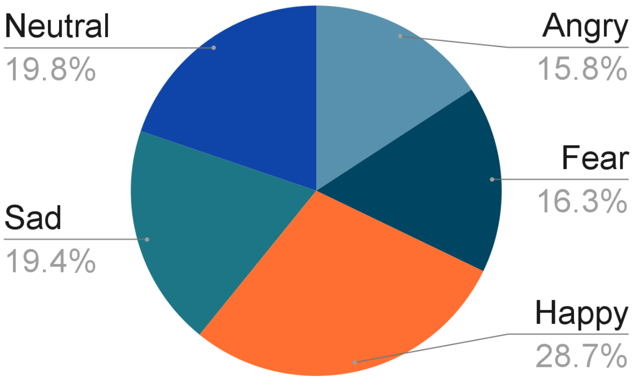


Figure 6. Category distribution of the dataset used in this paper

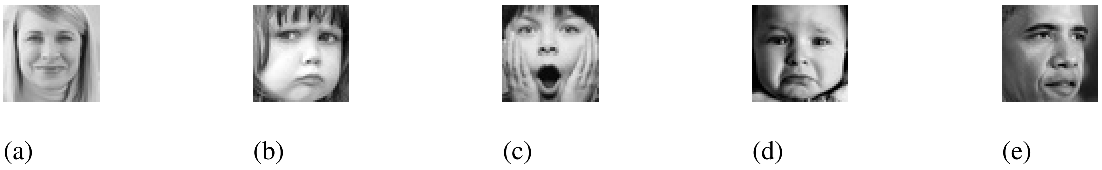

(a) (b) (c) (d) (e)

Figure 7. Five emotions in Face Expression Dataset (Kaggle 2024)
(a): happy, (b): angry, (c): fear, (d): sad, and (e): neutral

For the EEG dataset, twenty participants are shown videos ranging 2 to 4 minutes intentionally crafted to elicit specific emotion (from the same categories as the facial expression dataset). As the participants watched the videos, the EEG encoder collected the brain wave data. Each EEG sample is recorded at a sampling rate of 200 Hertz and 4,000 samples are collected for each emotion. Similar to the collection of facial expression data, 80% of EEG data are used for training and the remaining 20% are used for testing.

Inference Metrics

In order to assess the reliability of the proposed method, I utilized four evaluation metrics: accuracy, recall, precision, and F1-score. The first metric, shown in equation 4, is Accuracy which looks at the overall exactness in which the model can classify the data. The metric looks at the number of accurate predictions (true positives and negatives) that the model produced and divides that by the total number of data there were.

Equation 4: Accuracy

$$\text{Accuracy} = \frac{\mathit{True\ Positive + True\ Negative}}{\mathit{Total}}$$

The second metric is Recall, shown in equation 5. This metric looks at how many of the actual (true) positives the model identified as positive. It does so by dividing the number of true positives found by the sum of true positives and false negatives (the total number of actual positives).

Equation 5: Recall

$$\text{Recall} = \frac{\mathit{True\ Positive}}{\mathit{True\ Positive + False\ Negative}}$$

Precision, shown in equation 6, is slightly different from recall in that it looks at how much of the results labeled as positives by the model are true positives. Precision achieves this by dividing the true positives by the total number of data labeled as positive (both the true and false positives).

Equation 6: Precision

$$\text{Precision} = \frac{\mathit{True\ Positive}}{\mathit{True\ Postive + False\ Positive}}$$

The final metric is the F1-score, depicted in equation 7, which evaluates for the harmonic mean between precision and recall. This means that the F1 score evaluates the ability of the model to be able to classify the data as positive and be accurate when labeling the data as positive. Therefore a high F1 score correlates to a highly accurate model.

Equation 7: F1-score

$$\text{F1 Score} = 2 \times \frac{\mathit{precision} \times \mathit{recall}}{\mathit{precision} + \mathit{recall}}$$

Evaluation

Table 1. Evaluation comparison for four different CNN architectures

Architecture	Accuracy	Recall	Precision	F1-Score
VGG-19 (Simonyan et al. 2014)	0.9323	0.9404	0.9271	0.9337
MobileNetV2 (Sandler et al. 2018)	0.9354	0.9437	0.9271	0.9353
ResNet-50 (He et al. 2016)	0.9436	0.9519	0.9349	0.9433
DenseNet-121 (Huang et al. 2017)	0.9483	0.9549	0.9314	0.9466
+VGG-19	0.9392	0.9478	0.9332	0.9395
+MobileNetV2	0.9432	0.9488	0.9459	0.9409
+ResNet-50	0.9569	0.9626	0.9489	0.9542
+DenseNet-121	0.9571	0.9656	0.9489	0.9572

(+ marked denotes the model trained with the proposed weakly supervised contrastive learning.)

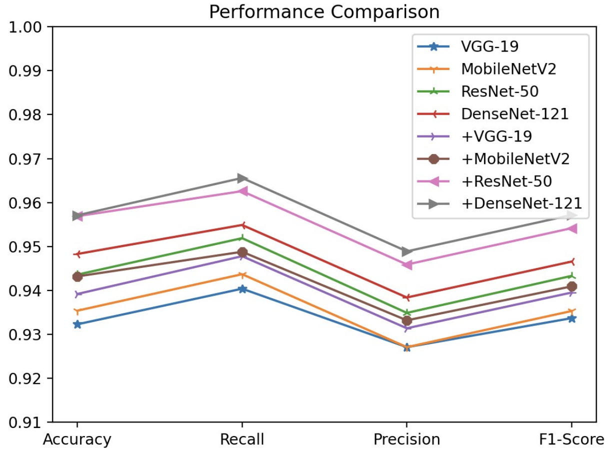


Figure 8. Evaluation comparison for four different CNN architectures

Once the CNNs have been trained with the proposed and previous method, their accuracies are evaluated using the four equations detailed in 4.2. The results of that evaluation are depicted in Table 1. Four types of CNN were used in this experiment, as can be seen in Table 1:VGG-19, MobileNetV2, ResNet-50, and DenseNet-121 (written in order of most shallow to deepest in comparison to one another). The results from the assessment showed that the CNN trained with the proposed method was the most accurate in comparison to the CNN trained using the previous method. This comparison is shown in figure 8 and 9 where all the CNN architecture have higher accuracy ratings compared to its counterparts.

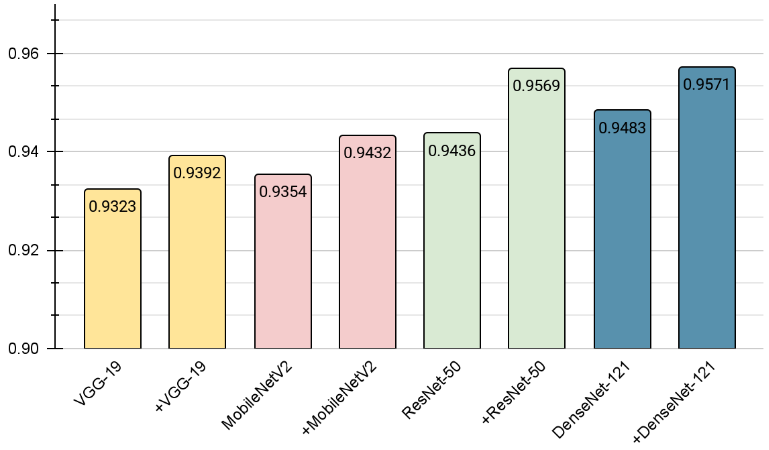
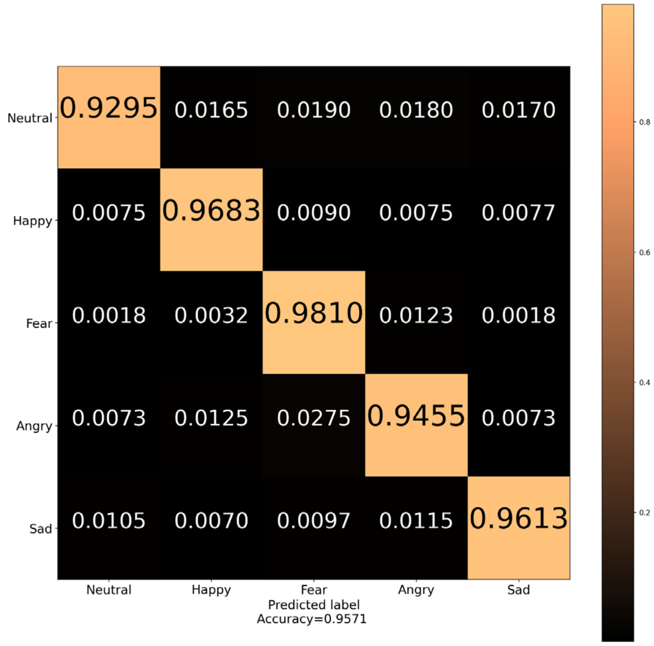


Figure 9. Accuracy improved for all four comparison architectures when the proposed approach was applied

More specifically, Denset-121 and ResNet-50 trained using the proposed method had the highest accuracy percentage with all of its scores exceeding 0.948.



	Neutral	Happy	Fear	Angry	Sad
Neutral	0.9295	0.0165	0.0190	0.0180	0.0170
Happy	0.0075	0.9683	0.0090	0.0075	0.0077
Fear	0.0018	0.0032	0.9810	0.0123	0.0018
Angry	0.0073	0.0125	0.0275	0.9455	0.0073
Sad	0.0105	0.0070	0.0097	0.0115	0.9613

Figure 10. Confusion matrix of the proposed method

Figure 10 depicts the results from DenseNet-121. The diagonal components represent the true positives of each emotion while the vertical component represents the emotion that the CNN classified the emotion as. According to the figure, Fear had the highest accuracy with a 98.1% accuracy. It is believed that this is the case since fear is one of the strongest emotions which in turn causes a strong physical reaction. Thus, giving more indicators for CNN to use and classify fear. On the other hand neutral was classified with the lowest accuracy with a 92.95% accuracy. This is believed to have happened since neutral is a very placid emotion with not as many physical indicators. Also, every person may express a ‘neutral’ feeling differently which hinders the CNN’s ability to classify it with the same precision as the other emotions.

Conclusion

In this research, I proposed a mood disorder system using electroencephalogram and facial expression paired data with weakly supervised contrastive learning. The proposed approach achieved an accuracy of 95.71% in classifying the emotional state of individuals based on their electroencephalogram and facial data. The ablation study clearly demonstrates that the proposed method contributed to improved accuracy. Additionally, I proposed a post-processing method for screening mood disorders based on the predicted emotional states. For my future work, I plan to implement the the proposed system into public areas such as schools to create an easily accessible mental health monitoring systems. I expect that the proposed system will help detect emotionally distressed individuals and screen those in need of assistance to allow them easier and faster access to treatment.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Akhand, M. A. H., Roy, S., Siddique, N., Kamal, M. A. S., & Shimamura, T. (2021). Facial emotion recognition using transfer learning in the deep CNN. Electronics, 10(9), 1036.

Ashish. (2020, May 31). “Facial Expressions Recognition using Keras Live Project- 2nd Part Testing”: YouTube. https://www.youtube.com/watch?app=desktop&v=XfTSfG47_q0

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. IEEE Computational intelligence magazine, 9(2), 48-57.

de Aguiar Neto, F. S., & Rosa, J. L. G. (2019). Depression biomarkers using non-invasive EEG: A review. Neuroscience & Biobehavioral Reviews, 105, 83-93.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://doi.org/10.48550/arXiv.1512.03385

Hu, B., Tao, Y., & Yang, M. (2023). Detecting depression based on facial cues elicited by emotional stimuli in video. Computers in Biology and Medicine, 165, 107457.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708). https://doi.org/10.48550/arXiv.1608.06993

Kaggle. (2024, Jun 26). “*Face expression recognition dataset*”: Kaggle. https://www.kaggle.com/datasets/jonathanoheix/face-expression-recognition-dataset

Magdin, M., Koprda, Š., & Tuček, D. (2022). Case Study Comparing the Accuracy Classification of Emotion and Analysis of Sentiment Using IBM Natural Language Understanding and Emotnizer Applications. DIVAI 2022.

Liu, D., Liu, B., Lin, T., Liu, G., Yang, G., Qi, D., ... & Lin, H. (2022). Measuring depression severity based on facial expression and body movement using deep convolutional neural network. Frontiers in Psychiatry, 13, 1017064.

Liu, S., Li, D., Gao, Q., & Song, Y. (2020, November). Facial emotion recognition based on cnn. In 2020 Chinese Automation Congress (CAC) (pp. 398-403). IEEE.

Office of the Surgeon General. (2021). Protecting youth mental health: the US surgeon general's advisory.

OpenBCI. (2021, May 20). "EEG ELECTRODE CAP KIT": OpenBCI https://shop.openbci.com/products/openbci-eeg-electrocap

Panchal, N. (2024, Feb 6). "*Recent Trends in Mental Health and Substance Use Concerns Among Adolescents*": KFF.

Rauch, J. (2017, Jul 24). "*The History of Online Therapy*":Talkspace https://www.talkspace.com/blog/history-online-therapy/#:~:text=Those%20who%20define%20online%20therapy,frequently%20discussed%20mental%20health%20issues

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520). https://doi.org/10.48550/arXiv.1801.04381

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Siniscalchi, K. A., Broome, M. E., Fish, J., Ventimiglia, J., Thompson, J., Roy, P., ... & Trivedi, M. (2020). Depression screening and measurement-based care in primary care. Journal of primary care & community health, 11, 2150132720931261.

Statista. (2023, Aug 3). "*Percentage of U.S. youth who experienced mental health challenges regularly as of 2023, by type*": Statista. https://www.statista.com/statistics/1412704/mental-health-challenges-among-us-youth-by-type/

Weisenburger, R. L., Mullarkey, M. C., Labrada, J., Labrousse, D., Yang, M. Y., MacPherson, A. H., ... & Beevers, C. G. (2024). Conversational assessment using artificial intelligence is as clinically useful as depression scales and preferred by users. Journal of Affective Disorders, 351, 489-498.

Wikipedia. (2023, Oct 23). "*Electroencephalography*": Wikipedia https://en.wikipedia.org/wiki/Electroencephalography