Signal Processing and Emotion Detection using EEG signal Data with CNN and GCN

Imran Hossain Imon
Undergraduate Student
BRAC University
Dhaka, Bangladesh
hossainimran.imzz@gmail.com

***Abstract*—In this modern era of artificial intelligence working with brain signals played revolutionary roles, contributing to a wide range of analytics and recognition, thus Brain-Computer Interface (BCI) is a very important field of research. Similarly, this research focuses on recognizing emotion through brain signals collected. I have collected brain signals from two datasets (Dreamer and Gamemo) and through various feature extraction methods like FFT, Welch, Hjorth, and statistical data I have collected features which are then trained using a convolutional neural network and Graph Neural Network model to detect four types of emotion (boredom, fear, calm anger). With this model, an average of 90% of accuracy prediction was obtained, which shows a better classification than many other research studies.**

***Index Terms*—Brain-computer interface, BCI, Graph Neural Network, GCN, deep learning, signal processing, emotion detection**

I. INTRODUCTION

In this dynamic world of artificial intelligence, advancement in industrial automation along with robots and lifestyle is changing rapidly. One of these advances plays an important role in the field of Brain-Computer Interface (BCI). Among numerous research going into BCI emotion detection/classification has been trending rapidly, using heart rate, eye movement, brain signals, etc. Using sensors like portable EEG machines, ECG machines, or even devices like phones and watches, people have been collecting data on emotions and recording and training various kind of machine learning algorithms to detect emotions. However, we can only detect emotions with sensory machines and pre-trained models to classify in human interactions. Still, less research has been done to recreate the brain signal depending on the detected signals.

II. RELATED WORK

The research proposed in [4] approaches an entropy-based feature extraction for emotion recognition using an EEG machine of 62 channels. The data was collected from people shown multiple clips and EEG signals were recorded, from which alpha, beta, theta, and gamma signals were extracted, then authors proposed a Differential Entropy of feature extraction from signal along with other features using entropy like Differential asymmetry, DASM; and rational asymmetry, RASM, rather than traditional energy spectrum. After that linear dynamic system was used for feature smoothing followed by principle component analysis and minimal-redundancy-maximal-relevance (MRMR) for dimension reduction. Finally, machine learning models were trained in support vector machine and k nearest neighbors, and compared results to support their claim of detection of maximum 84.22% accuracy on their models.

In the paper, [5] the author has described the properties and advantages of 1D convolutional neural network (CNN) in signal processing and detection along with four different cases where 1D CNN has outperformed many other paradigms. The authors stated that unlike 2D CNN working on matrices 1D CNN is more computationally optimized and works better in feature extraction in signal also later stated cases like detecting early symptoms of heart arrhythmia, structural damages, and electrical disruption using 1D CNN.

Another paper worked with emotion recognition with Graph Convolutional Broad Network, where the authors stacked combinations of graph convolutional layers and regular convolutional layers (CNN) for better feature extraction and methods like PSD, FFT, DE, etc. Then board learning system (BLS) to enhance the features extracted. Finally, the model is run on Seed and Dreamer EEG dataset and comparative analysis was shown on the emotion recognition on various machine learning models SVM, GraphSLDA, GSCCA, DGCNN, GCB-net (SR), GCB-net (BLS). GCB-net(BLS) showed maximum accuracy, thus confirming their proposal [6].

A novel Dynamical Graph Convolutional Neural Networks (DGCNN) based approach was taken for emotion recognition, where the authors showed classification accuracy of a maximum of 90.2% using features like PSD, DE, DASM, RASM on five different frequency bands (delta, theta, alpha, beta and gamma). DGCNN outperformed many other machine learning or graph-based models as it uses non-linear CNN and graph convolutions are able to find better intrinsic properties within the signals [7].

It is shown how to recognize emotional feelings from photographs of people's faces. What are the most often employed techniques and tactics for dealing with the problem? What function do CNNs play as a global pattern in the face expression detection challenge? A thorough assessment of the literature was undertaken to choose 51 articles that matched

Identify applicable funding agency here. If none, delete this.

the inclusion and exclusion criteria [9].

I have found highly related research to mine. That research used the DREAMER dataset and their emotion recognition technique was the brain-computer interface. Their best-performing model Spectogram features 2DCNN + XGBoost with about 97.7% accuracy for arousal, valance, and dominance. They present a CNN-XGBoost fusion approach based on signal spectrogram images for recognizing three aspects of emotion. Arousal, calm or agitation, valence, positive or negative mood, and dominance are the three aspects, without control or empowerment. They employed the DREAMER benchmark dataset, which captured EEG signals from different stimuli as well as self-evaluation ratings. The suggested approach has an arousal accuracy of 99.712% and a valence accuracy of 99.770%. [8].

The author uses self-supervised deep multi-task learning to propose an ECG-based emotion identification solution in this study [10]. Perhaps this is the first instance of self-supervised learning being used for ECG-based emotion identification. This work uses the four publicly available datasets AMIGOS, DREAMER, WESAD, and SWELL to conduct emotion recognition. They establish new benchmarks for the categorization of arousal, valence, affective states, and stress for the four datasets, demonstrating that the proposed method significantly outperforms a fully-supervised technique. Analyzing the effects of various self-supervised transformation recognition tasks for learning the ECG representations. The analysis demonstrates that the network learns improved ECG representations, which may be used for emotion categorization, for an ideal level of difficulty in the pretext tasks. The tests also demonstrate that a multi-task CNN outperforms a single-task network at learning ECG representations. Last but not least, its consideration of the utilization of different datasets for training the self-supervised network demonstrates that doing so improves the classification performed downstream.

With the help of electrocardiogram (ECG) signals, Dissanayake, T., Rajapaksha, Y., Ragel, and Nawinne [11] propose an ensemble learning strategy for creating a machine learning model that can identify the four main human emotions of anger, sadness, joy, and pleasure. This analysis integrates the four ECG signal-based feature extraction approaches of empirical mode decomposition, with-in-beat analysis, frequency spectrum analysis, and heart rate variability. The first three approaches of feature extraction are well-known ECG-based methods that have been cited in the literature, while the fourth way is a brand-new one that was suggested in this work. Using feature selection as a first step to ensemble model training, the machine learning process of this inquiry assesses the performance of a group of well-known ensemble learners for emotion classification and further enhances the classification outcomes. The created ensemble learner has an accuracy improvement of 10.77% when compared to the literature's top-performing single biosensor-based model. The created model also performs better than the majority of the multi-biosensor-based emotion detection models with a markedly higher classification accuracy gain.

III. Methodology

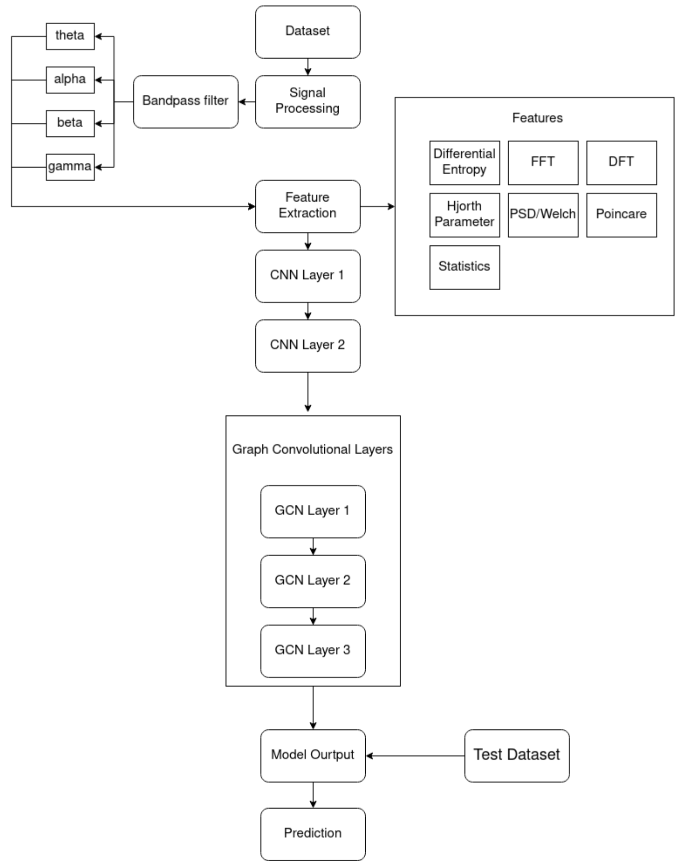

Fig. 1. Top-level overview of the Proposed System

Fig 1 shows the methodology of our proposed model. Here in the first step, the diagram shows the Dataset collection step, I have collected our Dataset from the website and which has EEG data for numerous subjects. After data collection, the next step is the signal processing step, where the raw signal from the dataset is passed between a bandpass filter to extract signals between a specific range, and then theta, alpha, and beta signals are filtered. These signals are then used in feature extraction methods like signal statistics, Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT), Hjorth, Poincare, Power Spectral Density (PSD), etc. Following are two 1D convolution layers to further preprocessing the feature extracted data of the signal. This is the final feature value which is split into 80% training data and 20% testing data. All the data is passed to the GCN layers and training is performed in the training dataset and finally, I have tested my Model with the test data. These steps performed a classification with an average accuracy of 90% on both datasets combined. Further explanations of each layer are provided on the following pages.

A. Data collection

EEG/brain signals were collected from the Dreamer dataset which consists of 14 channel EEG signal and ECG signal data collected at a 128Hz sampling rate, from 23 subjects. Signals were taken on and off stimuli and the stimuli consist of 18 different emotional video clips. A similar case was done with another dataset used called gameemo where stimuli were playing games of 4 types of emotions. In both datasets, written assessments were taken after each stimulus for measuring valence, arousal, and dominance. In this research, I have used both the dataset to increase the number of data to fit in the GCN model and reduce the dataset limitation, because GCN is a deep learning model and the deep learning model works when the number of data is larger.

B. Signal Processing

EEG signals contain a lot of noise that needs to be filtered hence, for emotion recognition, the frequency range is considered to be within a collection of frequency bands within 4-50 Hz. theta band (4-8 Hz), the alpha band (8-13 Hz), the beta band (13-30 Hz), and the gamma band (30-50 Hz). For that, I have used the band-pass filtering technique to remove the redundant frequencies under the low pass and above the high pass and keep the frequencies remaining frequencies. Each signal from the datasets was filtered into the theta, alpha, beta, and gamma bands to remove redundant signals, which are then passed to the next step of pre-processing and feature extraction.

C. Feature Extraction

The first feature I used is the Fast Fourier Transform (FFT), which is one of the most popular methods of signal processing. FFT is an optimized version of Discrete Fourier Transform (DCT), which uses a signal in the time domain and divides it into multiple sub-signals of equal time domain, this causes the signal to split into different spectral signals in the time domain. Then, each signal in the time domain is converted to a frequency and then all the frequencies in the frequency domain are put into a sequence. Hence it can be said that FFT converts the time domain values of the signal to the frequency domain and I have used this FFT and DCT transform to find out the frequency coefficients of the corresponding time domain signals and took the maximum and mean values as features.

And differential entropy is also suitable for signal processing as it measures the randomness in continuous random variables and converts the variables to a continuous probability distribution. Following is the equation of differential entropy.

$$\mathrm{h(X)} = E[-\log f(X)]\mathrm{h(X)} = -\int_x f(x)\log f(X)dx \quad (1)$$

The next feature I have used is the Hjorth parameter, this method is also very useful in providing statistical details about the time domain signal. Hjorth's method gives three parameters activity, mobility, and complexity. Activity parameter equation 2 define the variance of the time function and shows the power spectrum value of the frequency domain, by providing the value of high-frequency components. Mobility parameter equation 2 gives the proportion of standard deviation of the power spectrum. Complexity parameter equation 2 shows the change in frequency of the signal relative to a pure sine wave and converges to 1 if it is similar to a sine wave.

$$\begin{aligned} \text{activity} &= var(y(t)), \\ \text{mobility} &= \sqrt{\frac{var(y'(t))}{var(y(t))}}, \\ \text{complexity} &= \frac{mobility(y'(t))}{mobility(y(t))} \end{aligned} \quad (2)$$

Furthermore, I have used Welch's method to form a power spectrum to find power spectral density (PSD), which computes the power/amplitude of a signal relative to the frequency domain. Power spectrum can be used to identify frequency patterns in signals and this may vary depending on the emotion type, in the case of brain signal, and can give useful information about the signal. Welch's method is an improved method of periodograms equation of 3, which estimates the spectral density and can be used to create spectrograms. Welch's method divides the time domain signal into discrete time intervals and makes a spectrogram of each segment then all the spectrograms are averaged as shown in the equation 4. This method is smoother than the full FFT approach and thus I extracted the maximum power from the signals.

$$\hat{\mathrm{P}}^{\mathrm{i}}_{\mathrm{M}}(f) = \frac{1}{L\left|\sum_{n=1}^{L-1} w[n]x_i[n]e^{j2fn}\right|^2} \quad (3)$$

$$\hat{\mathrm{P}}_{\mathrm{B}}(f) = \frac{1}{K\sum_{i=1}^{K} \hat{P}^i_M(f)} \quad (4)$$

Finally, in feature extraction various statistical data including mean, median, maximum, variance, and skewness has been taken as features from the signals. After all this feature extraction the features are ready to be trained.

1) Data distribution: From observation and datasets analysis, which shows sample valence and arousal distribution from the dreamer dataset, average people with excitement and horror have valence and arousal greater than 3 and vice versa for sad and calm. So from intuition valance and arousal above 3 are counted as high and low for vice versa. Each valence and arousal were mapped to high or low as "1" or "0" respectively and then as described in the previous section the valence and arousal were used to map the emotion to labels as "0" (boring), "1" (calm), "2" (horror), "3" (excited), for classification. In the case of the Gameemo dataset, the emotions were already set for each experiment to 4 different emotions, which was also incorporated with the intuition of the dreamer, and thus I have used the same method for mapping this dataset. Figure 2 shows the final label distribution of both of the datasets which will be used for classification.

D. Proposed Model

I have collected our features and labels to be trained in our proposed model which is a hybrid system of two types of convolutional models, 1D convolutional neural network (CNN) and graph convolutional network (GCN). CNN and GCN are

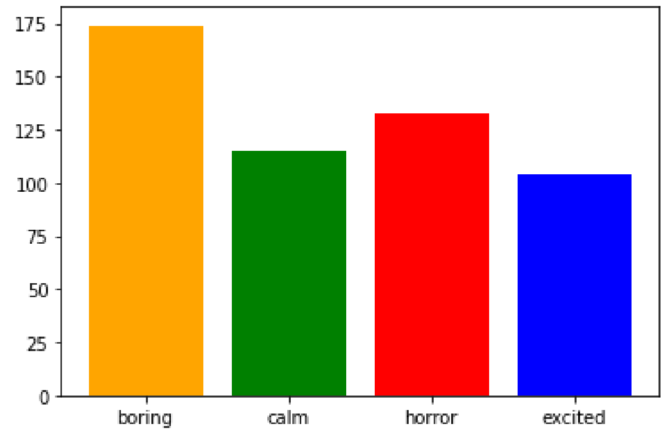

Fig. 2. Overall data distribution of happy and sad category on DREAMER dataset

very well-known convolutional models in the field of machine learning. 1D CNN was used so that the collected features can go through a further feature extraction process, which is the specialty of CNN as it collects features from the input and 1D CNN takes values from rows rather than matrix which is computationally faster than 2D or 3D CNN. GCN uses a graph-based convolutional layer that has the ability to find intrinsic patterns and behaviors which can distinguish between data of different labels and it also uses a message-passing technique to train the neighboring nodes of the same attributes and using this a classification model can be obtained.

1) CNN Layer: Convolutional neural networks (CNN) are very famous for image processing and image feature extractions since they can extract different features from image metrics using filter/kernel metrics to create convoluted metrics by sliding the filter and taking product throughout the image metrics. I have used CNN in our research for extracting distinct features specified to each EEG feature vector, however, 2D-CNN increases computational complexity due to a high number of matrix multiplication, so a more optimized way is 1D-CNN. 1D-CNNs are computationally optimized and work best with signals as signals are 1d vectors and they can extract unique attributes within the signal vectors [5]. As it works on vectorized data rather than the matrix and the kernels used in this research are 64 3-dimensional vectors sliding over each feature vector, this model brings out the features of each signal feature vector with less computation and better signal attributes for each type of signal [5]. Also, I have used 2 of such CNN layers for more intrinsic properties within each type of signal feature vector.

2) Graph Convolutional Network: Graph convolutional network (GCN) is also famous for the graph-based approach to node classification. For that, it requires an adjacency matrix for the graph (edges), the feature vectors (nodes), and the edge weight vector (degree matrix). GCN uses a sparse matrix that maps the graph's nodes and edges and the graph relations, in this research I have created the sparse matrix using the labels that are connected to the nodes of the same type of label and as I am working on 4 different emotion nodes are connected to each of the same types of labels. illustrates our stacked architecture of the GCN model, working with node embeddings/messages classified in nodes and then passed on to the neighboring nodes through multiple layers combined to classify nodes from the resulting graph. As shown, the initial graph has convoluted attributes from the CNN layers and each node randomizes each feature vector, and through message passing technique using neural networks and aggregations, described below, the neighboring nodes of similar features connected form clusters of the same type of label/emotion as showing in the figure, colors of connected nodes are same at the end of each layer. Similarly, throughout the 3 GCN layers used in this research, the final graph shows nodes with the most common attributes and intrinsic properties are connected together closely forming different colored clusters depending on the type of emotion [13].

$$H^{(l)} = \hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} \tag{5}$$

$$f\left(H^{(l)}, A\right) = \sigma\left(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}\right) \tag{6}$$

After the GCN model is initialized with the required parameters mentioned above, the training data are passed to the dataset. At the start, first, the training data is passed on to a pre-convolution process, through two dense layers of 32 neurons, each, embedded inside the nodes, acting as a feed-forward network to create a message in equation 5 [12], then there is a batch normalization step which prevents any kind of overfitting in the message vectors [6].

Next is the convolution layer, where the message vectors are passed on to the neighboring nodes and are aggregated using an aggregation function. I have tried various aggregation functions like sum, mean, and concatenated functions but I obtained the best result with mean aggregation and row concatenation which is the generalized equation of GCN. After that, the node embeddings are created in the convolution layer, from the neural network inside the nodes and the messages are updated and passed to the next convolution layer [12]. A total of three convolution layers were used and after each iteration, the weight and the messages were updated in the nodes continuously until the final node embeddings, after each convolution layer a batch normalization has been used to reduce overfitting. These node embeddings/weights inside the graph convolution network are used to predict the emotion. For the optimization of calculations, the "adam" optimizer was used along with "sparse categorical cross-entropy" as a loss function for each training set.

IV. Result and Analysis

A. Result

Figure 3 illustrates the training graph of the model with the highest accuracy I have collected from three training ses-

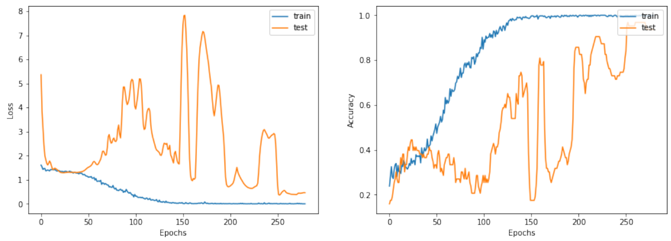

Fig. 3. Proposed model loss accuracy graph

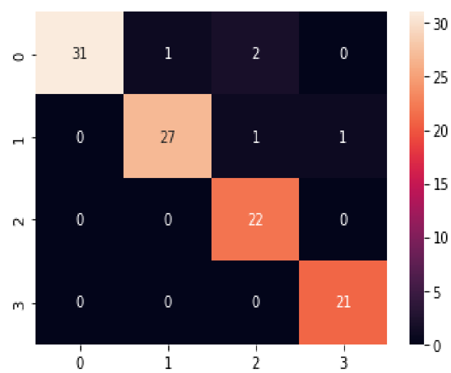

Fig. 4. Proposed model confusion matrix

sions and every training graph shows similar results. Training epochs were set to 300 for each training session and the accuracy graph on the right side of figure 3 demonstrates an upward trend for both graphs, even though the validation graph fluctuates it still reaches a validation score above 95%. Figure 4 is showing the accuracy score on the test dataset that has been used to predict untrained data, and it can be interpreted that among a total of 106 untrained data of 4 different emotions (boring, calm, horror, excited), 102 of them were classified correctly to the belonged class and 4 were incorrect. Conclusively, it can be said that our proposed model can classify human emotions with very high accuracy.

V. Conclusion

Detecting human emotion has been a great contribution to the research of brain-computer interface and much research has been done on multiple emotion detection, this research has proposed a way to detect four types of emotions boring, calm, horror, and sad by using a fusion of graph-based approach using semi-supervised graph model graph convolutional model and convolutional neural network model for better classification than present research. In the future, this research can simply do multi-class classification and properly evaluate emotional intelligence in addition to gathering datasets utilizing an EEG machine and novel stimuli. Conclusively, through various signal processing, feature extraction, and model training I have successfully created a classification model with an average classification score of 97.6% accuracy, which supports our research proposal and is worth investigating for more emotion analysis in the further future.

References

[1] Khan, C. M. T., Ab Aziz, N. A., Raja, J. E., Nawawi, S. W. B., Rani, P. (2022). Evaluation of Machine Learning Algorithms for Emotions Recognition using Electrocardiogram. Emerging Science Journal, 7(1), 147-161.

[2] Sepúlveda A, Castillo F, Palma C, Rodriguez-Fernandez M. Emotion Recognition from ECG Signals Using Wavelet Scattering and Machine Learning. Applied Sciences. 2021; 11(11):4945. https://doi.org/10.3390/app11114945

[3] Xefteris, V. R., Tsanousa, A., Georgakopoulou, N., Diplaris, S., Vrochidis, S., Kompatsiaris, I. (2022). Graph Theoretical Analysis of EEG Functional Connectivity Patterns and Fusion with Physiological Signals for Emotion Recognition. Sensors, 22(21), 8198.

[4] Duan, R.-N., Zhu, J.-Y., Lu, B.-L. (2013). Differential entropy feature for EEG-based emotion classification. 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). doi:10.1109/ner.2013.6695876

[5] Kiranyaz, S., Ince, T., Abdeljaber, O., Avci, O., Gabbouj, M. (2019). 1-D Convolutional Neural Networks for Signal Processing Applications. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp.2019.8682194

[6] Zhang, Tong & Wang, Xuehan & Xu, Xiangmin & Chen, C.. (2019). GCB-Net: Graph Convolutional Broad Network and Its Application in Emotion Recognition. IEEE Transactions on Affective Computing. PP. 1-1. 10.1109/TAFFC.2019.2937768.

[7] Song, Tengfei & Zheng, Wenming & Song, Peng & Cui, Zhen. (2018). EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. IEEE Transactions on Affective Computing. PP. 1-1. 10.1109/TAFFC.2018.2817622.

[8] Khan, Md. Sakib, et al. "CNN-XGBoost Fusion-based Affective State Recognition Using EEG Spectrogram Image Analysis - Scientific Reports." Nature, 19 Aug. 2022, www.nature.com/articles/s41598-022-18257-x.

[9] Canal, F. Z., Muller, T. R., Matias, J. C., Scotton, G. G., Sa Junior, A. R. de, Pozzebon, E., & Sobieranski, A. C. (2021, October 7). A survey on facial emotion recognition techniques: A state-of-the-art literature review - ScienceDirect. A Survey on Facial Emotion Recognition Techniques: A State-of-the-Art Literature Review - ScienceDirect. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0020025521010136

[10] P. Sarkar and A. Etemad, "Self-Supervised ECG Representation Learning for Emotion Recognition," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1541-1554, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.3014842.

[11] Dissanayake, T.; Rajapaksha, Y.; Ragel, R.; Nawinne, I. An Ensemble Learning Approach for Electrocardiogram Sensor Based Human Emotion Recognition. Sensors 2019, 19, 4495. https://doi.org/10.3390/s19204495

[12] Jiang, B., Zhang, Z., Lin, D., Tang, J., Luo, B. (2019). Semi-Supervised Learning With Graph Learning-Convolutional Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2019.01157

[13] Kipf, Thomas Welling, Max. (2016). Semi-Supervised Classification with Graph Convolutional Networks.