Emotion Classification and Network Discovery: GNNs Without Predefined Graphs

Gabriele Dall'Aglio (377397), Zhuofu Zhou (370337), Cristiano Sartori (396315)

***Abstract*—This study explores emotion classification using fMRI data from a movie-watching paradigm. We compare Graph Neural Networks (GNNs) and traditional machine learning models (e.g., Random Forest, KNN). Despite applying Graph Structure Learning (GSL) techniques, advanced models like (Graph Attention Network) GAT and Variational Information Bottleneck model (VIB) performed at chance level, with fMRI's low temporal resolution and high inter-subject variability posing challenges. Further improvements are needed for effective emotion classification using fMRI. The baseline KNN model remains the leading model in the analysis, achieving an F1-score 9.74% using 300 neighbors and the Euclidean distance as the similarity metric.**

I. Introduction

Classifying emotions remains one of the greatest challenges in neuroscience [1], with current models often performing only slightly better than random guessing. In this study, we explore the potential of classifying emotions using fMRI data collected during a movie-watching paradigm.

Unlike EEG, which has been extensively studied for emotion detection [2] [3], fMRI remains underutilized for this purpose. While fMRI sacrifices temporal resolution, generating images approximately every 1.3 seconds, it offers unparalleled access to deep brain regions like the amygdala, which have been linked to emotion processing [4]. This trade-off makes fMRI a valuable tool for emotion classification.

A significant challenge in emotion research lies in the lack of robust paradigms to reliably evoke and measure emotional responses. Here, we adopt a movie-watching paradigm to elicit naturalistic emotional reactions [5]. This approach complicates the analysis since it blends elements of resting-state and task-based fMRI, making traditional functional connectivity measures less applicable [5]. Moreover, emotions manifest in the brain over varying timescales [5], which limits the effectiveness of using fixed-size sliding windows to capture dynamic functional connectivity.

To address these challenges, we leverage Graph Neural Networks (GNNs) [6] as a machine learning approach to classify emotions. Brain activity is often modeled as a network or graph [7], making GNNs particularly well-suited for this task. However, no definitive connectivity patterns or adjacency matrices currently exist to describe the brain's activation during emotional processing. In this study, we aim not only to classify emotions but also to uncover latent networks that may emerge during different emotional states. To achieve this, we apply Graph Structure Learning (GSL) techniques [8] alongside *a priori* knowledge to construct graph representations of brain activity (see Fig 1).

As a baseline for comparison, traditional machine learning models such as k-Nearest Neighbors (KNN), Random Forest (RF), Graph Convolutional Networks (GCN), and Feedforward Neural Networks (FNN) will be used. For GSL analysis, Graph Attention Networks (GAT) and Variational Information Bottleneck (VIB) [9] will be applied.

II. Models and Methods

A. Data Acquisition & Labels

The dataset used in this project, *Emo-FilM*, originates from the publication[10]. It comprises fMRI recordings of BOLD signals from 30 participants (18 female, age: 25.83 ± 3.60 years) while watching 14 short films. MRI scans were conducted using a 3T Siemens Magnetom TIM Trio scanner (Siemens, Erlangen, Germany) with a 32-channel head coil at the Brain and Behaviour Laboratory at the University of Geneva (BBL). Data preprocessing followed the steps described in the article [10]. The resulting dataset consisted of time-series data with 414 nodes (400 derived from the Schaefer atlas [11] and 14 from deep brain regions) for each movie and subject.

Emotion labels were assigned to each TR (time repetition) of the fMRI data by a group of 44 raters [10]. Labels were provided in two formats: one based on the Circumplex model [12] (representing emotions in a 3D space) and another using 50 discrete basic emotions scores[13].

As this study focuses on classification, the latter set of labels was used. From the initial 50 emotions, 13 were retained. Labels for each TR were extracted by selecting the emotion with the largest absolute value.

Finally, to account for the delay in hemodynamic response, all labels were shifted by 4 TRs ($\sim$6 seconds) to ensure the BOLD signal had sufficient time to reflect neural activity.

B. Train Test split

Due to the non-independent nature of movie emotion labels (where the same emotion can span multiple scenes within a movie), splitting the dataset is particularly challenging. Random splitting is unsuitable, as it may result in models being trained and tested on nearly identical data: for example, from the same participant watching the same movie, with only a few seconds separating training and test samples. Additionally, movies have varying distributions of emotions, making the splitting process even more difficult (see Fig. 2). Specifically, after rebalancing the number of emotions through subsampling, the data was split into training and test sets using 11 movies for training and 3 for testing (an 80-20% split). As mentioned earlier, due to the non-independent nature of the data, the skewed distribution of emotions across different movies, and the relatively small dataset size compared to the complexity of the task, cross-validation was not performed. This decision was made under the supervision of our lab mentor. All the results will thus

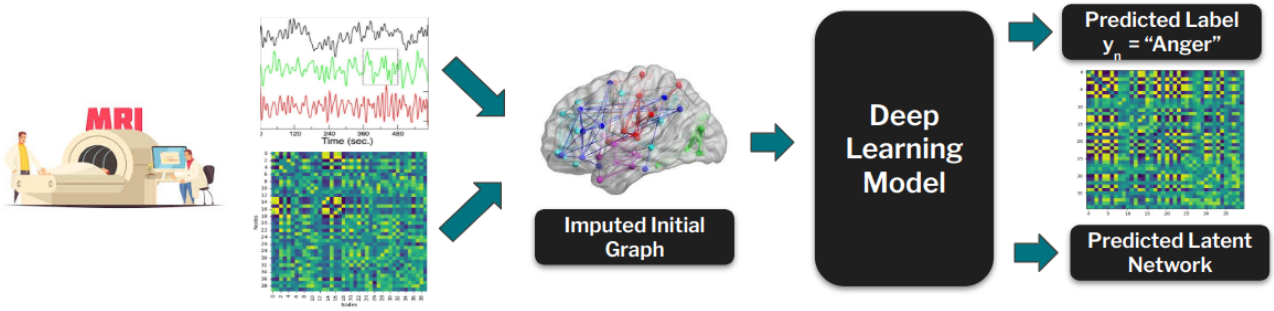

Figure 1. **Schema illustrating the main pipeline of the project:** Participants watch movies inside an fMRI machine, during which BOLD signals are recorded. The recorded signals are then preprocessed, polished, and parcellated using the Schaefer atlas. Subsequently, the raw time series data is merged with an initially imputed connectivity matrix to create an initial input graph for each timepoint, corresponding to each movie for every subject. These graphs are then passed to a machine learning model. The model predicts both the target label for the current timepoint and a refined version of the initial connectivity. This refined connectivity aims to represent the learned emotional network as inferred by the model.

report only the accuracy reached in the test set, without stardard deviation.

C. Features and Initial Imputed Connectivity

The features utilized for both baseline methods and the nodes in GNNs are divided into two categories. The first category involved using raw time series extracted from a symmetric window of variable size (tested sizes: 3, 5, 11, and 21 TRs). Additionally, kernelization of the time series using a Gaussian kernel was applied to emphasize time points closer to the target prediction point.

The second approach consisted of handcrafted features extracted from the window, including mean, standard deviation, minimum, maximum, autocorrelation with lag 1 TR, peak-to-peak range, mean absolute deviation, the first coefficient of Fourier transform, skewness, kurtosis, and covariance calculated across nodes.

As mentioned in the introduction, the initial data is not intrinsically a graph. Thus, to enable GNNs to function, adjacency attributes and connectivity matrices were imputed. This process introduced uncertainty in determining the optimal graph connectivity for training the GNN. To address this, various initial connectivity configurations were investigated.

For edge attributes, constant values across the entire movie were tested, such as using 1 or 0 (representing only self-loops with no connectivity). Alternatively, the functional connectivity (FC) calculated for the entire movie was used. For dynamic approaches, FC with a sliding window approach was tested.

Regarding initial node selection, two options were explored: using all 414 nodes or restricting the selection to nodes from a single functional network (see Fig 2). The functional networks tested corresponded to the seven networks proposed by Yeo et al. [7].

D. Models

Baseline models (i.e. Random Forest and KNN) were implemented using `sklearn`, while deep learning models using `Torch` and `torch geometric` libraries. For hyperparameters tested for each model, refer to Appendix B.

Baseline: baseline methods that do not account for GSL are:

1) **Random Forest (RF)**: An ensemble method that builds multiple decision trees, and the class with the most votes from the trees is chosen as the final prediction.
2) **K-Nearest Neighbors (KNN)**: A simple algorithm that classifies a data point based on the majority class of its nearest neighbors in the feature space.
3) **Fully Connected Neural Network (FNN)**: trained with the Adam optimizer, ReLU activation function, and 3 hidden layers with size 1500, 500, 500 respectively. The input vector for the FNN consisted of the signals from the different nodes stacked together.
4) **Graph Convolutional Network (GCN)**: a neural network designed for graph data, where nodes aggregate information from their neighbors. To avoid over-smoothing, a careful selection of the number of convolutional layers was taken into account.

Graph Attention Network (GAT): variant of GNN with attention mechanism. The model consists of a single-layer GAT with 13 attention heads (one for each emotion), followed by a global mean pooling layer and a fully connected layer to output the final graph-level prediction. The attention weights are computed using the `torch_geometric.nn.conv.GATConv` layer, based on the following formula:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{h}'_i || \mathbf{h}'_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^\top [\mathbf{h}'_i || \mathbf{h}'_k]))}$$

Where $\mathbf{a}$ is a learnable attention vector, $\mathbf{h}'_i$ and $\mathbf{h}'_j$ are the transformed features of nodes i and j, $\mathcal{N}(i)$ is the set of neighbors of node i, $[\mathbf{h}'_i || \mathbf{h}'_j]$ is the concatenation of the transformed node features of nodes i and j. The updated node embedding is created by averaging the results of the different heads.

The rationale behind choosing this architecture is its interpretability. It is hoped that, by the end of training, each head will specialize in a specific emotion, thereby capturing the hidden activation patterns associated with that emotion.

Variational Information Bottleneck (VIB): to explore a more widely recognized model, we employed a Graph Structure Learning (GSL) approach combined with Information Bottleneck (IB) principle[14], as proposed in VIB model[9]. This method aims to balance retaining task-relevant information while discarding irrelevant or noisy details in the graph. By dynamically learning the graph

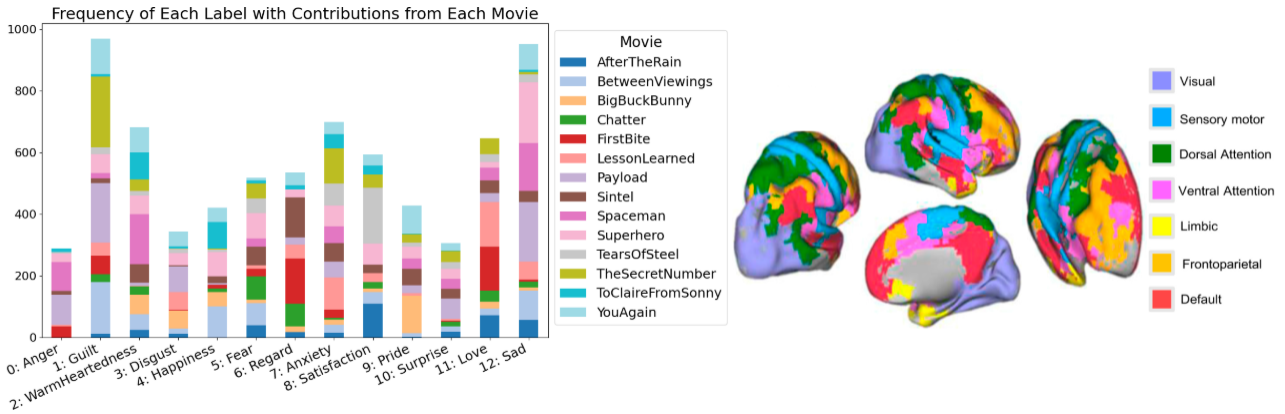

Figure 2. **On the left:** The distribution of emotions across movies. It can be observed that some emotions are underrepresented compared to others, which motivated our decision to rebalance the dataset. Additionally, different emotions are present in varying proportions across movies, complicating the identification of a clear split between the training and test sets. **On the right:** Representation of the seven functional networks proposed by Yeo. (Image source: Fan et al.)

structure during training, it is particularly well-suited for our scenario where the graph structure is uncertain.

For a more detailed explanation of the model architecture and the specific hyperparameters used during training, refer to Appendix-A.

III. Results

Table III reports the best accuracies of the different models on the test set after performing a grid search on various hyperparameters (refer to Appendix-B for hyperparameters tested for each method). Unfortunately, more complex models like GAT and VIB, despite extensive grid search, were unable to outperform the chance-level accuracy (8%) achieved by baseline models such as RF, KNN, FNN, and GCN. The best-performing model—although still at chance level—was a simple KNN model using handcrafted features, with 300 neighbors and the Euclidean distance as the similarity metric. In contrast, the worst-performing model was GCN, with an F1 score of 1.31%, primarily due to its consistent prediction of a single class.

Model	**Acc. (%)**	**F1 (%)**
K-Nearest Neighbors (KNN)	9.95	9.74
Random Forest	8.80	8.27
Feedforward Neural Network (FNN)	7.59	7.60
Graph Convolutional Network (GCN)	7.81	1.31
Graph Attention Network (GAT)	7.77	7.28
Variational Information Bottleneck (VIB)	7.40	7.35

*Refer to the appendix (C) for corresponding best hyperparameters.

As expected, the best results were obtained when prior neuroscientific knowledge was incorporated. Specifically, the best results for GCN, GAT, and VIB were achieved when a thresholded version of the functional connectivity (FC) matrix, rather than a constant matrix, was input into the models. It is important to note that the optimal results were observed with a threshold value of 0.7 for FC, which ensured the sparsity of the initial graph while maintaining its connectivity. This balance was crucial for enabling effective information propagation through the message-passing layers of the GNNs.

Taking the GAT model as an example, we observe that while the model's training loss converges by the end of training (50 epochs) the test accuracy remains at chance level (see Figure 3). Similarly, the attention weights for GAT across each head (see Fig. 4) fail to provide meaningful neuroscientific insights. The attention weights remain nearly constant across nodes, suggesting that the model did not learn to identify significant connections during training.

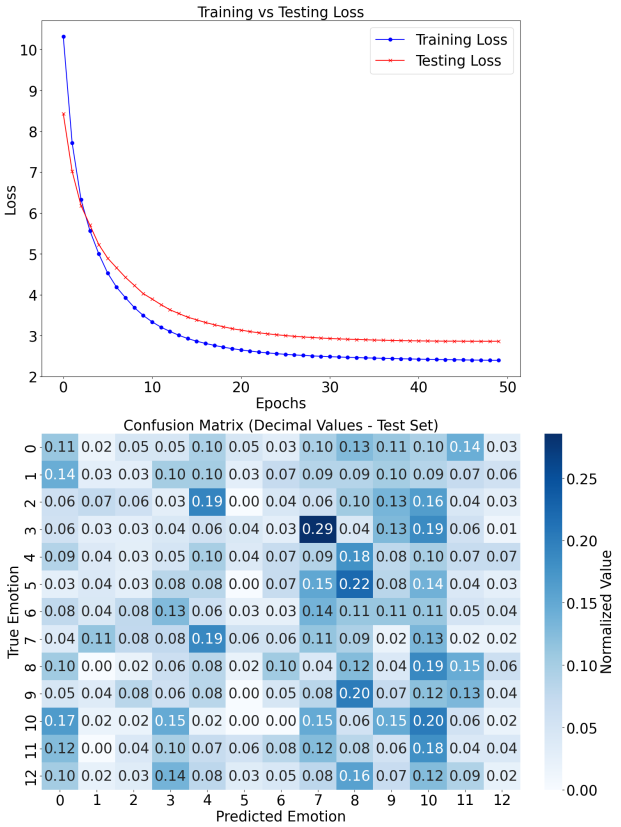

Figure 3. From top to bottom: Losses for the training and test sets of the best GAT model over 50 epochs. The model already converged after 20 epochs. The confusion matrix of the best GAT model evaluated on the test set. It is interesting to note that the highest performance is achieved in predicting *Surprise*, which may suggest an intrinsic, distinct pattern for this emotion that is easier for the model to predict.

Comparable findings are observed in the structure learned by VIB. Figure 5 illustrates the mean adjacency matrix learned by VIB for each time point labeled as "Happiness," representing the latent network for happiness inferred by VIB. While some modularity is visible, the model's poor performance and the relative homogene-

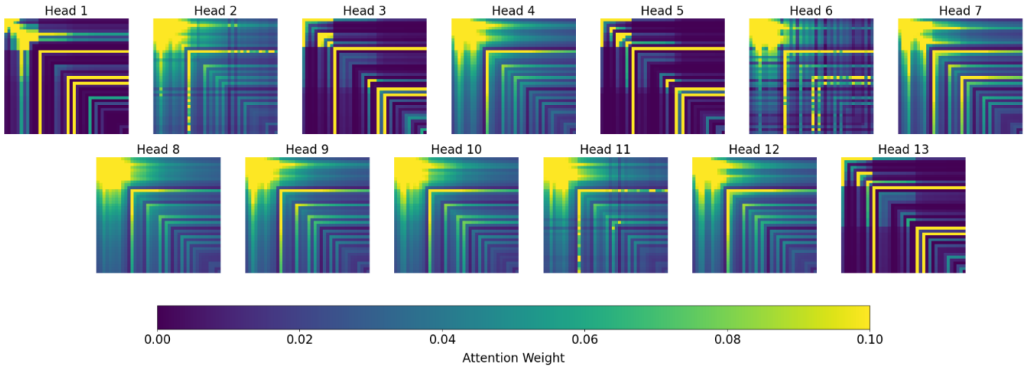

Figure 4. Attention weights extracted from the 13 heads of the GAT model, reported only for nodes relative to the Limbic Functional Network.

ity of the values suggest that it does not effectively capture the underlying activation network. Lastly, it appears that the 14 deep brain ROIs (shown as the last values at the bottom right of the adjacency matrix) have weak learned connections with the other 400 cortical ROIs from Schaefer, implying a segregation of information between cortical and core region of the brain.

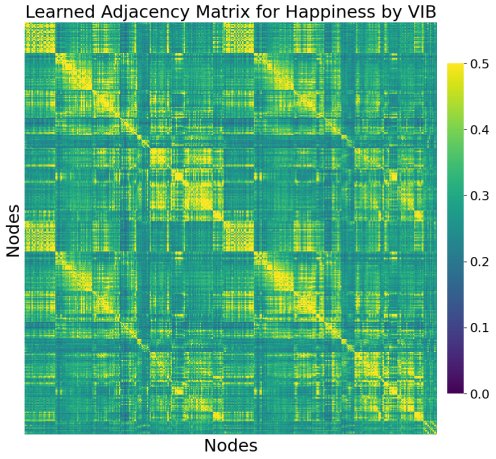

Figure 5. The learned connectivity for the 'Happiness' emotion, obtained by averaging all the relevant connectivity matrices of time-points labeled as 'Happiness' from the best VIB model, clearly shows modularity across different brain regions.

IV. Discussion

Both GSL methods GAT and VIB were unable to outperform the baseline methods. Additionally, their performances remained at random level. It is important to note that while GAT and VIB still performed at a random level when using the entire dataset, their performance improved when tested on a subsample of subjects. Most likely, this fact arises from the high inter-subject variability when coming to brain signals, which can complicate the model training.

Inter-subject variability not only complicated the task but the newly tested signal modalities also introduced additional challenges. While the presence of underlying neural networks associated with specific emotions has been well-established and validated using EEG signals, the use of fMRI data for this purpose remains largely unexplored. The coarse temporal resolution of fMRI further complicates the analysis, as emotions can emerge rapidly, one after another. As a result, the neural effects of these emotions overlap, resembling a linear time-invariant system. This overlap is difficult to disentangle due to the slow temporal resolution of fMRI.

Lastly, it can be hypothesized that the difficulty for the models in learning could be attributed to the sliding window approach. Since different emotions are elicited at different times[5], some may be penalized more than others depending on the chosen window size. A follow-up analysis using phase synchrony[15], instead of the sliding window approach, could thus be beneficial to overcome this limitation.

The research is still ongoing, and there are several directions for future work. For instance, regression models could be explored as an alternative to classification, or it may be interesting to attempt to classify multiple emotions simultaneously using a multilabel approach. Different types of initial connectivity (e.g., seed-based connectivity derived from specific brain regions, such as the amygdala) could also be investigated. To continue with the GSL methodology, unsupervised learning methods like GraphMAE [16] or diffusion models could be tested for better feature extraction or to uncover latent networks.

V. Ethics

Emotion detection and classification raise many ethical concerns [17], so much that a ban of emotion recognition systems is currently under consideration in the EU AI Act[18]. EU is exploring frameworks to regulate the use of emotion data[19], emphasizing the importance of protecting individuals from invasive or discriminatory practices.

In our study, one particularly important consideration is the privacy of brain signal data collected during fMRI scans, especially regarding its potential use in predicting the mental health status. The stakeholders most impacted by this risk include individuals undergoing fMRI scans. The privacy of these patients is at risk because fMRI data contains highly sensitive information about the brain's activity, which can be used to infer emotional states and, more critically, mental health conditions. Unauthorized access to such data could lead to discriminatory practices, such as employers using this information to make hiring decisions based on inferred emotional stability[20].

While the associated risks are severe, the likelihood of such occurrences is mitigated by the conscious and informed participation required for fMRI scans. Individuals undergoing these scans are typically informed of the risks involved and have given explicit consent for their data to be used, which significantly reduces the chance of unauthorized access. Furthermore, the long duration of

the scan and the need for patient consent at each stage provide an additional safeguard against potential misuse.

Given that the dataset for our project had already been created and preprocessed, we did not have the opportunity to directly address these risks during data collection. However, we conducted a retrospective check to ensure that all ethical approvals were correctly issued. Specifically, ethical approval was granted by the Geneva Cantonal Commission for Ethics in Research (protocol No. 2018-02006), and we verified that the study complied with the Code of Human Research Ethics (2014) [10].

Furthermore, confounding variables such as sex, age, and ethnicity were excluded from the training features to prevent any unintended biases and leakage. Finally, to ensure data protection, all data was securely stored and processed exclusively on the MIP:Lab's internal server, safeguarding it against unauthorized access or misuse.

VI. Conclusion

In this study, we explored the use of fMRI data for emotion classification during a movie-watching paradigm, applying both traditional machine learning models and advanced Graph Neural Networks (GNNs). Despite extensive experimentation, including the use of Graph Structure Learning (GSL) techniques, the models struggled to outperform baseline approaches, with results remaining at chance level, and with best model being KNN. However, interpretation methods, such as attention weight visualizations and the mean adjacency matrix, could serve as intriguing tools to study latent emotion networks in future research. The unsatisfactory performance highlights the need for further work to refine fMRI-based emotion classification methods, possibly through alternative techniques like phase synchrony or improved connectivity models.

Acknowledgments

We would like to express our sincere gratitude to Chun Hei Michael Chan, the supervisor of this project, for his invaluable guidance and support throughout the research. Our heartfelt thanks also go to MIP:Lab for providing the resources that made this work possible.

We would like to extend our appreciation to the creators of the VIB model ([9]).

Finally, we would like to thank Sevda Ögüt, our Machine Learning Teaching Assistant, for her assistance and insightful feedback during the course of this project.

References

[1] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, and W. Zhang, "A systematic review on affective computing: emotion models, databases, and recent advances," *Information Fusion*, vol. 94, p. 101810, 2023.

[2] T. Wierciński, M. Rock, R. Zwierzycki, T. Zawadzka, and M. Zawadzki, "Emotion recognition from physiological channels using graph neural network," *Sensors*, vol. 22, no. 8, p. 2980, 2022.

[3] C. Liu, X. Zhou, Y. Wu, R. Yang, Z. Wang, L. Zhai, Z. Jia, and Y. Liu, "Graph neural networks in eeg-based emotion recognition: A survey," *arXiv preprint arXiv:2402.01138*, 2024, version 3, 30 Aug 2024.

[4] R. Adolphs, S. Wang, Y. Luo, G. Pourtois, D. Sander, J. Morriss, S. Frühholz, and D. Grandjean, "The human amygdala parametrically encodes the intensity of specific emotions and their categorical ambiguity," *Nature Communications*, vol. 8, p. 14821, 2017.

[5] E. Morgenroth, D. Van De Ville, and Laur, "Probing neurodynamics of experienced emotions—a hitchhiker's guide to film fmri," *Social Cognitive and Affective Neuroscience*, vol. 18, no. 1, p. nsad063, 2023.

[6] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *arXiv preprint arXiv:1901.00596*, 2019, version 4, 4 Dec 2019.

[7] B. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011.

[8] Z. Li, L. Wang, X. Sun, Y. Luo, Y. Zhu, D. Chen, Y. Luo, X. Zhou, Q. Liu, S. Wu, L. Wang, and J. X. Yu, "Gslb: The graph structure learning benchmark," *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[9] Q. Sun, J. Li, H. Peng, J. Wu, X. Fu, C. Ji, and P. S. Yu, "Graph structure learning with variational information bottleneck," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, pp. 4165–4174, 2022.

[10] A. not specified, "Emo-film: A multimodal dataset for affective neuroscience using naturalistic stimuli," *BioRxiv*, vol. 2024.02.26.582043v1, 2024, the dataset includes detailed emotion annotations by 44 raters for 14 short films with a combined duration of over 2½ hours, as well as recordings of physiological responses in healthy participants. [Online]. Available: https://www.biorxiv.org/content/10.1101/2024.02.26.582043v1

[11] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. T. Yeo, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, 2018.

[12] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[13] P. A. Kragel and K. S. LaBar, "Multivariate neural biomarkers of emotional states are categorically distinct," *Social Cognitive and Affective Neuroscience*, vol. 10, no. 11, pp. 1437–1448, Nov 2015.

[14] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," *arXiv preprint physics/0004057*, 2000. [Online]. Available: https://arxiv.org/abs/physics/0004057

[15] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[16] Z. Hou, X. Liu, Y. Cen, Y. Dong, H. Yang, C. Wang, and J. Tang, "Graphmae: Self-supervised masked graph autoencoders," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 594–604.

[17] M. J. Farah, "Neuroethics: The ethical, legal, and societal impact of neuroscience," *Annual Review of Psychology*, vol. 63, pp. 571–591, 2012.

[18] MIT Technology Review, "Five big takeaways from europe's ai act," June 2023. [Online]. Available: https://www.technologyreview.com/2023/06/19/1075063/five-big-takeaways-from-europes-ai-act/

[19] A. Hauselmann, A. M. Sears, L. Zard, and E. Fosch-Villaronga, "Eu law and emotion data," *arXiv preprint arXiv:2309.10776*, 2023. [Online]. Available: http://arxiv.org/abs/2309.10776

[20] T. Gremsl and E. Hödl, "Emotional ai: Legal and ethical challenges," *Information Polity*, vol. 27, no. 2, pp. 163–174, 2022. [Online]. Available: https://content.iospress.com/articles/information-polity/ip211529

APPENDIX

A. VIB Model

The Variational Information Bottleneck (VIB) architecture is designed to learn a compact, task-relevant representation of input data while discarding irrelevant or noisy information. In the context of Graph Structure Learning (GSL), VIB introduces a probabilistic framework that dynamically learns an optimal graph structure during training. The architecture typically consists of three key components: an encoder, a graph learner, and a decoder.

The encoder maps input node features into a latent space where key task-related information is extracted. Simultaneously, the graph learner generates a probabilistic adjacency matrix, representing the learned graph structure. This step is guided by the information bottleneck principle, which aims to maximize the mutual information between the learned graph and task-relevant labels while minimizing redundancy or noise in the structure. The decoder then uses the latent node embeddings and the learned graph structure to make predictions for the downstream task, such as node classification or graph classification.

By integrating the graph-learning process into the architecture, VIB eliminates the need for a predefined graph and allows the model to adaptively discover the most informative graph structure. This makes it particularly effective in domains where the optimal graph topology is unknown or highly dynamic.

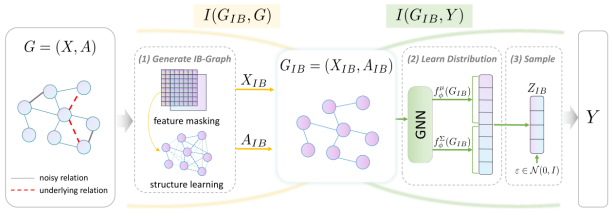

Figure 6. Overview of VIB-GSL model. Image from [9].

B. Grid Search

K-Nearest Neighbors (KNN)	
Hyperparameter	**Values Used**
K	5, 10, 100, 500
Similarity	Euclidian, Cosine
Random Forest	
Hyperparameter	**Values Used**
Number of Trees	10, 50, 100, 200, 300, 500
Max depth	5, 10, 20, full
GCN and GAT	
Hyperparameter	**Values Used**
Batch Size	16, 32, 64
Number Conv. Layers (GCN)	1, 3, 5
Dropout Rate	0.1, 0.5
Learning Rate	0.01, 0.001
Window Size	7, 11, 15, 21
Kernelize timeseries	True, False
Std. Gaussian Kernel	1, 2
Initial Edge Attributes	constant value 1, constant value 0, functional connectivity of whole movie, functional connectivity inside current time-window
Threshold FC	0.1, 0.4, 0.7
Nodes	All (414), Subset Functional Network
Functional Network (FN)	Limbic, Visual, Somato-Motor, Dorsal Attention, Ventral Attention, Default
Epochs	10, 50, 250, 500
Variational Information Bottleneck (VIB)	
Hyperparameter	**Values Used**
Intial graph connectivities	Same of GAT and GCN
Backbone	GAT, GCN, GIN
Hidden Dimension	32, 64, 256
Number of Layers	1, 3, 4
Graph Type	Bernulli, KNN, Epsilon
Top K for KNN	10, 20
Epsilon for EpsiloNN	0.3, 0.5
Graph Metric Type	multi-layer-perceptron, cosine, attention
Feature Denoise	False, True
Weigth KL divergence	0.00001, 0.0001
Bottleneck Size	16, 64, 128
Graph Skip Connection	0.0, 0.5
Graph Include Self	False, True

C. Best Hyperparameters for each model - Grid Search Results

K-Nearest Neighbors (KNN)	
Configuration	**Accuracy (%)**
$k = 300$, Distance=Euclidean	9.39
Random Forest (RF)	
Configuration	**Accuracy (%)**
Trees=100, Max-Depth=10	8.80
Feedforward Neural Network (FNN)	
Configuration	**Accuracy (%)**
Layers=3, Neurons=[1500, 500, 500], ReLu, LR=0.001	7.59
Graph Convolutional Network (GCN)	
Configuration	**Accuracy (%)**
Batch Size=16, Number convolutional layers=1, Dropout=0.5, LR=0.01, Window Size=11, Kernelize timeseries=False, Initial Edge Attributes=FC of current timeseries, Threshold FC=0.7, node=all 414, Epochs=50	7.81
Graph Attention Network (GAT)	
Configuration	**Accuracy (%)**
Batch Size=32, Number convolutional layers=1, Dropout=0.5, LR=0.001, Window Size=9, Kernelize timeseries=False, Initial Edge Attributes=FC of current timeseries, Threshold FC=0.7, node=all Visual Functional Network, Epochs=50	7.77
Variational Information Bottleneck (VIB)	
Configuration	**Accuracy (%)**
Batch Size=32, Window Size=9, Kernelize timeseries=False, Initial Edge Attributes=FC of current timeseries, Threshold FC=0.7, node=all 414, Epochs=50, Backbone=GAT, Hidden Dimension=256, Number of Layers=1, Graph Type=Bernulli, Graph Metric Type=multi-layer-perceptron, Feature Denoise=False, Weight KL Divergence=0.0001, Bottleneck Size=16, Graph Skip Connection=0.0, Graph Include Self=True	7.40