Fused Hypergraph Neural Network-Based Enhanced Emotion Recognition

Sahaya Beni Prathiba
Centre for Cyber Physical Systems
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
sahayabeni@ieee.org

Varshan A V R
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
avr.varshan20@gmail.com

Vishal Pranav
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
vishalpranav23@gmail.com

Dhanalakshmi Ranganayakulu
Centre for Cyber Physical Systems
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
dhanalakshmi.r@vit.ac.in

***Abstract*—As human-computer interaction systems rely on accurately recognizing and identifying emotional states in order to provide adaptive and personalized user experiences, traditional models for recognizing emotion have isolated the facial modality from the vocal modality, viewing them as unrelated information streams. Although ensemble models aim to integrate modalities, they usually fail to create deep connections between the modalities or utilize the connections when facial expressions and speech signals co-occur, which will often typically degrade accuracy in recognition, especially as the environments become more naturalistic and when emotional expressions incorporate multimodal cues that are intertwined. Thus, this paper proposes Fused-HGNN—a framework aimed at creating independent hypergraph representations of both facial and audio modalities. Each hypergraph is developed to represent complex emotional dependencies with the use of specialized hyperedges. The proposed method aligns and integrates these hyperedges leading to a cohesive multimodal representation that accounts for critical cross-modal dependencies in emotion perception or recognition. Fusion occurs within dedicated hypergraph convolutional layers, and rich joint embeddings are developed that represent complex, holistic affective states. Empirical evaluation of the MEAD (Multi-view Emotional Audio-visual Dataset) indicates that Fused-HGNN significantly outperforms unimodal or ensemble-based baselines by significant margin. The model achieves an accuracy of 98.5% with a loss of 2.3%, this will lead to measurable improvements of 1.4% and 0.8%, respectively, over existing ensemble methods.**

***Index Terms*—Hypergraph Neural Network (HGNN), Multimodal Emotion Recognition, Hypergraph Convolution, Hyperedge Fusion, Cross-modal Integration, Affective Computing.**

I. Introduction

Technology usage related to machine learning has transformed many sectors including healthcare diagnostics, entertainment and intelligent security systems [20]. Among this broad spectrum of application, multimodal emotion recognition has become an important area of research, allowing interactions with computers to be context-sensitive and adaptive. By analysing both facial behaviour and speech, systems can determine emotional states with greater accuracy and enhance a range of applications including assistive communication technologies; immersive entertainment; and advanced surveillance [14].

Despite continuous advancements, there are still ongoing issues to be solved which limit both deployment and scalability. Multimodal data is inherently complicated and high-dimensional, making it challenging to model complex relationships between modalities [1]. Traditional approaches (particularly ensemble techniques as well as simple feature concatenation, for example) do not have satisfactory capability of capturing nuanced relationships between auditory and visual channels [2].

In many cases these approaches apply a pairwise modeling which oversimplifies the dependence of the modalities [3]–[5], [7]. Graph based fusion approaches as well as concatenation approaches do not accurately represent higher order dependencies that are characterized by modality (intrapersonal) and between each modality. Therefore, emotion may be incomplete.

In order to mitigate these issues, we present the Fused Hypergraph Neural Network (Fused-HGNN), a model specifically designed to exploit the structural richness of hypergraphs for multimodal fusion. In distinction to existing ensemble methods, our strategy leverages hypergraph-based representations to capture not only within-modal relationships, but also inter-modal relationships, for a more thorough representation of emotional interactions across facial and audio features. Fused-HGNN enables a strong and interpretable fusing of the emotional signals by forming separate hypergraphs for each modality and merging them at the hyperedge level. The hypergraph representations are further facilitated through hypergraph convolutional processing and validated through improved recognition accuracy and system resilience.

II. Related Works

Recent studies in multimodal emotion recognition have focused on developing improved amalgamation techniques, representation learning, and interpretability. The M^{3}SA framework utilizes multi-scale feature extraction and multi-task learning with channel attention, fusing modalities by weighing and emphasizing relevant information with fusion layers. This results in superior performance compared with baseline architectures [1]. Transformer-based architectures have also seen a rise in popularity, particularly systems that model the interaction of facial expressions and physiologic signals, result in improving accuracy across multiple benchmark datasets [2].

In the area of autonomous driving, detecting driver emotions has been recognized as an important feature of adaptive safety features and coupling speech cues with facial metrics provides better monitoring of emotional state for reliability and improved user satisfaction. Most of the current methods leverage attention based facial encoders and speech emotion recognition models multimodal representations [7]. These systems perform well on these benchmark datasets RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) and SAVEE (Surrey Audio-Visual Expressed Emotion), achieving good precision and specificity [3].

Most of these methods, however, follow pairwise modeling strategies, which may fail to adequately represent complicated relational dependencies. Thus, Hypergraph-based formulations have recently emerged as a useful alternative, allowing information to be modeled beyond traditional pairwise relations. Because of hypergraph generalizations of spectral clustering methods, embedding and classification performance have improved over traditional graph structures in settings with multi-relational data [4].

Using these ideas, the Hypergraph Collaborative Network (HGCN) improves the learning process by utilizing vertex and hyperedge information, as well as incorporating reconstruction error as a regularization term. The HGCN model has been shown to be useful for semi-supervised classification problems, including evaluations on several benchmark datasets [5]. More recent advances have applied hypergraph-based methods to multimodal fusion, especially for complex diagnostic and perception tasks. For example, Hypergraph-based Multi-Modal Fusion leverages similarity matrices to encode high-order relationships among heterogeneous modalities such as imaging and genetic data. This method incorporates intra-modality and inter-modality regularization to achieve better diagnostic accuracy of neurological disorders, such as schizophrenia, but also to discover latent interactions between genetic, environmental, and neurophysiological factors [9]. In the same vein, hypergraph convolutional networks have been applied to hyperspectral image classification by incorporating multiple structural representations. By including CNN-based branches which jointly encode spectral and spatial information, hypergraph convolutional networks improved accuracy on standard benchmark data [6].

III. Proposed Work

The suggested framework proposes a new way of doing emotion recognition, referred to as Fused-HGNN (Fused Hypergraph Neural Network) that is specifically designed to overcome the issues of aggregating and modeling multimodal data. The Fused-HGNN comprises three main components that collectively facilitate efficient feature representation, multimodal aggregation, and improving the process of learning emotion-related tasks.

A. *Overview of the Framework*

The first component, Hypergraph Construction, creates its own hypergraph representations for audio and facial modalities. Each hypergraph accounts for unique structural relationships: the hypergraph for the facial data takes into consideration spatial and temporal dependencies while the hypergraph for audio facilitates the learning of temporal-semantic patterns. The second component: Hypergraph Fusion, combines the two hypergraphs for the specific modalities into a single representation maintaining the complementary features and fostering the interaction between modalities. The final part, Hypergraph Convolution, applies a convolutional layer to propagate multimodal information through the fused hypergraph structures, learning higher order representations that improve motion classification accuracy and robustness.

Figure 1 shows how the Fused-HGNN model works, illustrating the three-stage pipeline from modality-specific feature encoding to hypergraph-based fusion, and a convolutional learning stage.

B. *Hypergraph Construction*

In this stage, we formulate hypergraphs that are saliently distinctive with respect to face and audio features, to preserve the dynamic properties of each modality. Each hypergraph is denoted by sets of vertices and hyperedges capturing the features driving emotional representation.

1) Construction of Facial Hypergraph: The facial hypergraph illustrates the relationships among different facial representations. Each representation is treated as a vertex, and collections of related representations form connections known as hyperedges. These hyperedges represent facial components that tend to co-move or co-vary.

Let:

- x = a facial feature point or descriptor,
- ϵ = a hyperedge representing a group of related facial descriptors.

Formally, the incidence matrix of the facial hypergraph, denoted by $\mathcal{E}_{\text{facial}}$, defines the relationship between vertices and hyperedges as:

$$H_f(x, \epsilon) = \begin{cases} 1, & \text{if } x \in \epsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

To determine the most salient facial representations, the degree centrality of each vertex is computed as:

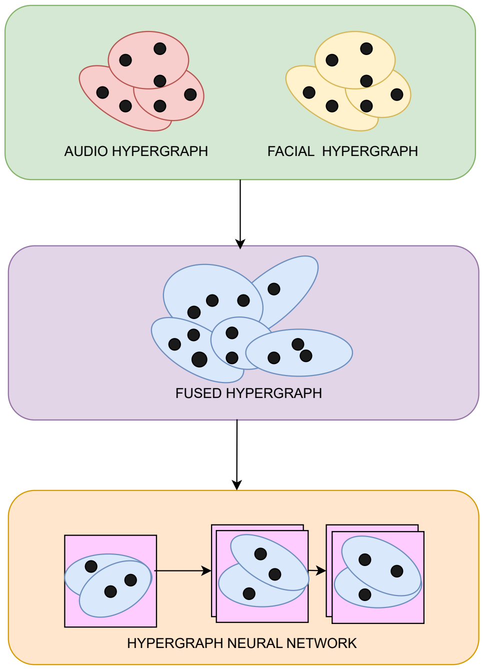

Fig. 1. Architectural overview of the proposed Fused-HGNN framework for multimodal emotion recognition. The figure illustrates the three-stage pipeline: (1) modality-specific feature encoding from facial and audio inputs through separate encoders, (2) construction of independent facial and audio hypergraphs where nodes represent feature descriptors and hyperedges capture higher-order intra-modal relationships, and (3) hyperedge-level fusion to form a unified multimodal hypergraph. Hypergraph convolutional layers subsequently propagate information across nodes and hyperedges, enabling the learning of rich joint multimodal embeddings, which are finally passed through a fully connected layer to produce the predicted emotion label.

$$C_d(x_i) = \sum_{\epsilon \in \mathcal{E}_f} H_f(x_i, \epsilon), \tag{2}$$

where $C_d(x_i)$ represents the number of hyperedges connected to vertex x_i. A higher degree centrality indicates that the vertex is more influential within the facial network since it participates in more relational connections. Vertices with high degree centrality values are grouped together into compact clusters to form dense sets of representations that signify expressive facial patterns associated with emotion.

2) Construction of Audio Hypergraph: The audio hypergraph models relationships among different acoustic features over time. Each audio feature—such as pitch, loudness, or Mel-Frequency Cepstral Coefficients (MFCCs)—is represented as a vertex, while hyperedges capture the interactions and co-variations among these features across time.

Let:

- v = an audio feature descriptor,
- e = a hyperedge connecting similar audio descriptors.

The incidence matrix of the audio hypergraph is defined as:

$$I_{\text{audio}}(v, e) = \begin{cases} 1, & \text{if } v \in e, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Hyperedges are constructed using a k-nearest neighbors (k-NN) approach, which connects each vertex to its most similar neighbors based on temporal or acoustic similarity:

$$e_i = v_i \cup \mathcal{N}_{\text{kNN}}(v_i), \tag{4}$$

where $\mathcal{N}_{\text{kNN}}(v_i)$ denotes the set of k nearest neighboring features of v_i. This captures both temporal continuity and feature-level similarity across the audio signal.

Finally, the overall audio hypergraph is represented as the combination of all hyperedges:

$$\mathcal{H}_{\text{audio}} = e_1 \parallel e_2 \parallel \ldots \parallel e_N, \tag{5}$$

where $\parallel$ denotes the concatenation operator. The resulting structure represents how groups of audio features jointly express emotion or contextual information through their temporal and spectral relationships.

C. Hypergraph Fusion

The fusion stage merges two hypergraphs, each specific to a modality, into one coherent multimodal hypergraph, $\mathcal{H}_{\text{fused}}$, which consolidates modality-specific information while incorporating explicit intermodal relationships through hyperedge concatenation and adjacency alignment with weights.

$$\mathcal{V}_{\text{fused}} = \mathcal{V}_{\text{facial}} \cup \mathcal{V}_{\text{audio}}, \tag{6}$$

$$\mathcal{E}_{\text{fused}} = \mathcal{E}_{\text{facial}} \cup \mathcal{E}_{\text{audio}}. \tag{7}$$

Algorithm 1 Hypergraph Fusion via Concatenation

Input: $\mathcal{H}_{\text{facial}}, \mathcal{H}_{\text{audio}}$

Output: $\mathcal{H}_{\text{fused}}$

1) Initialize empty sets $\mathcal{V}_{\text{fused}}$ and $\mathcal{E}_{\text{fused}}$.
2) Merge nodes and hyperedges from both modalities:
   - $\mathcal{V}_{\text{fused}} \leftarrow \mathcal{V}_{\text{facial}} \cup \mathcal{V}_{\text{audio}}$
   - $\mathcal{E}_{\text{fused}} \leftarrow \mathcal{E}_{\text{facial}} \cup \mathcal{E}_{\text{audio}}$
3) Return $\mathcal{H}_{\text{fused}} = (\mathcal{V}_{\text{fused}}, \mathcal{E}_{\text{fused}})$.

This fusion mechanism ensures the preservation of both intra-modal coherence and inter-modal complementarity, establishing a foundation for joint feature learning.

D. Hypergraph Convolution

In the last step, hypergraph convolution is used on the fused hypergraph to obtain compact, discriminative embeddings using information from neighboring nodes and hyperedges to capture high-order correlations.

Let:

- $F^{(l)}$ = node feature matrix at layer l,
- I_{fused} = fused incidence matrix,
- $W^{(l)}$ = learnable weight matrix at layer l,

- σ = nonlinear activation function,
- D_v, D_e = vertex and hyperedge degree matrices.

The convolution operation is defined as:

$$F^{(l+1)} = \sigma\left(D_v^{-\frac{1}{2}} I_{\text{fused}} W^{(l)} I_{\text{fused}}^{\top} D_e^{-\frac{1}{2}} F^{(l)}\right). \tag{8}$$

To incorporate modality-specific contributions:

$$\begin{aligned}\Phi = D_v^{-\frac{1}{2}} I_{\text{facial}} W_{\text{facial}} I_{\text{facial}}^{\top} D_e^{-\frac{1}{2}} + \\ D_v^{-\frac{1}{2}} I_{\text{audio}} W_{\text{audio}} I_{\text{audio}}^{\top} D_e^{-\frac{1}{2}},\end{aligned} \tag{9}$$

where Φ captures the combined effect of both modalities, I_{facial} and I_{audio} are the incidence matrices for facial and audio hypergraphs respectively, and W_{facial} and W_{audio} are modality-specific learnable weight matrices.

The feature update rule becomes:

$$F^{(l+1)} = \sigma\left(\Phi F^{(l)}\right), \tag{10}$$

and the final representation after two convolutional layers is:

$$Z_{\text{HGNN}} = \sigma\left(\Phi F^{(1)} W^{(2)}\right). \tag{11}$$

Finally, a fully connected layer produces the classification output:

$$\widetilde{Z}_{\text{HGNN}} = \sigma\left(W_{\text{fc}} Z_{\text{HGNN}} + b_{\text{fc}}\right), \tag{12}$$

where W_{fc} and b_{fc} are the weight matrix and bias vector of the fully connected layer.

This comprehensive architecture facilitates effective learning of multimodal relationships while guaranteeing stability, flexibility and improved accuracy of emotion recognition.

IV. Results

We carried out a substantial series of experiments to validate the performance of the proposed Fused-HGNN model using the MEAD (Multi-view Emotional Audio-visual Dataset) benchmark, which is widely used in the field of multimodal emotion recognition. The dataset contained a variety of emotional expressions, both visual and auditory. For the evaluations we trained the model on 10% of labeled samples per class and allocated an additional 10% for validation (hyperparameter tuning) and the other 80% for testing. The hyperparameters were optimised carefully (learning rate was set to 0.001 and training was done through 100 epochs).

To ensure a fair performance comparison, we applied an ensemble baseline which included three models trained independently: the facial feature extraction was completed by a CNN model, a LSTM model was used to model temporal dependencies in audio, and a SVM model was used to classify over the fused features. The predictions from the ensemble were then averaged by specific modality weights that we used based upon validation so that all modalities contribute equally. While our ensemble can promote modularity, it is unable to discover higher order dependencies that the hypergraph-based Fused-HGNN can explicitly learn.

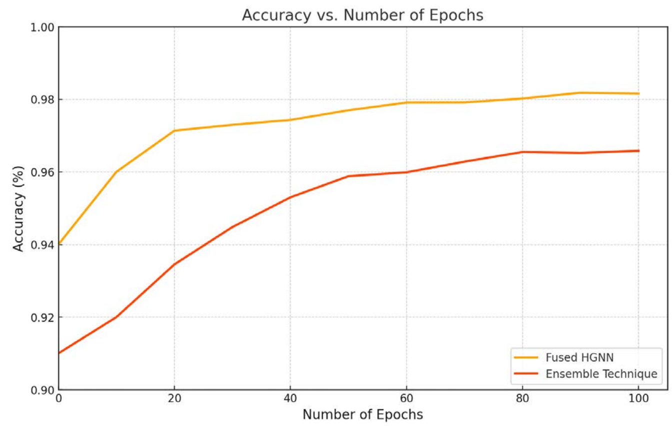

Fig. 2. Accuracy vs. Number of Epochs.

All experiments were developed and ran in PyTorch and executed on a high-performance computing cluster. Model optimization and grid search were utilized and the model was initialized with a batch size of 32 for 100 epochs. Evaluation metrics included accuracy, loss, precision, and F1-score, in order to give a holistic comparison of performance.

Figure 2 shows the results of comparison between the proposed Fused-HGNN and the reference ensemble model when it comes to accuracy over 100 epochs. The proposed model yields greater accuracy across the entire training period compared to the ensemble, achieving an overall accuracy of 98.5% compared to 97.1%. This significant improvement represents the proposed model's stronger ability to model high-order multimodal relations through hypergraph fusion.

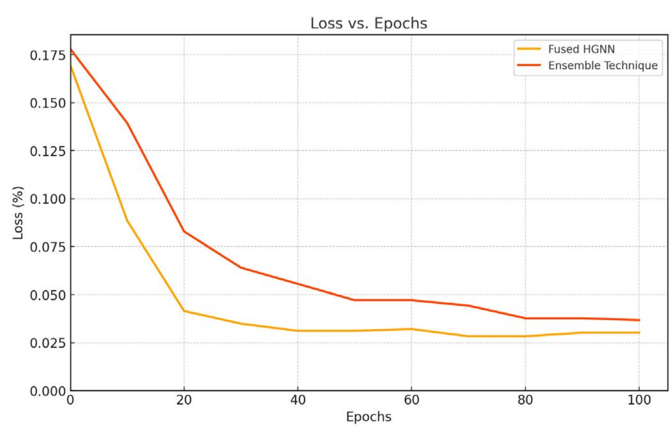

Fig. 3. Loss vs. Number of Epochs.

Figure 3 shows that both models experience a quick drop in loss during the first few epochs followed by a gradual stabilization. The Fused-HGNN maintains a lower loss curve throughout training, ending with a loss of 2.3%, compared to the ensemble model at 3.1%. This suggests that the hypergraph-based learning framework not only converges quicker, but exhibits optimized stability as well.

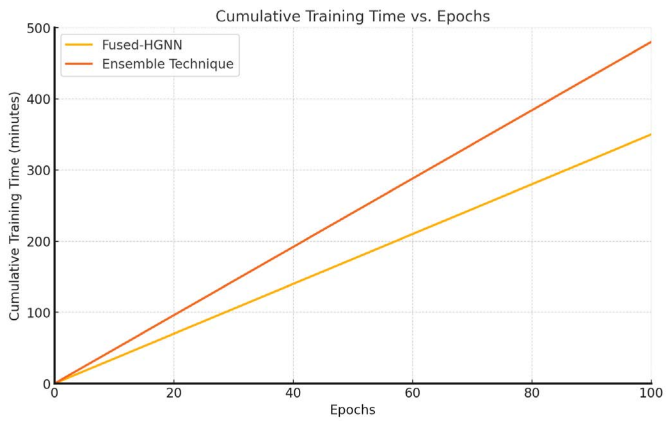

Fig. 4. Cumulative Training Time vs. Number of Epochs.

Figure 4 compares total training time between the two models and shows the proposed approach to take less computation time to complete the training. The Fused-HGNN requires less time during each epoch and was able to achieve convergence in 350 minutes as opposed to the ensemble model which took 480 minutes. This fact that the Fused-HGNN took less time to converge compared to the ensemble is because of the model's ability to use end-to-end hypergraph convolutional architecture to learn compact representations across modalities simultaneously.

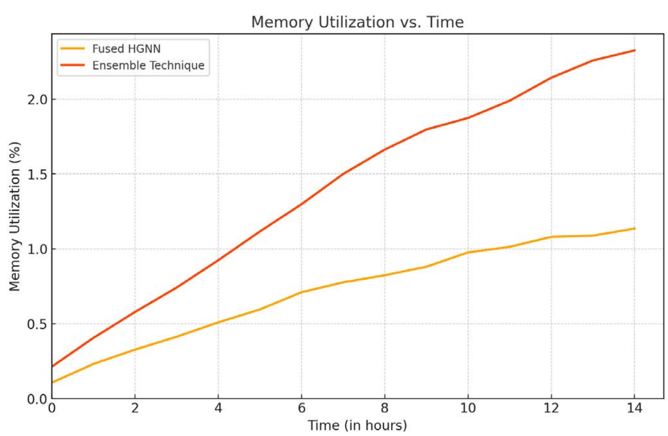

Fig. 5. Memory Utilization vs. Time.

Figure 5 compares the memory usage of the proposed Fused-HGNN with the ensemble model, and the total memory consumption of the proposed model is significantly lower, reaching a maximum capacity of 1.5%, while the ensemble model reached 2.2%. This is an important finding, as it points to the scalability and suitability for deploying hypergraph models without excessive memory constraints.

The proposed model is able to run with less computational demand and train effectively faster, because of their unified architecture, whereby both modalities are fused together from the beginning through hypergraph convolutional layers. The ensemble model, in contrast to the proposed model, trained separate independent models from the beginning, which naturally takes longer for training time and demand greater memory resources.

Overall, the experimental results suggest that the Fused-HGNN outperformed all ensemble-based systems in three dimensions, including accuracy, convergence speed, and efficiency. The hypergraph representation successfully exploited, characterized, and represented complex high-order relationships between facial and auditory features, which improved generalization and robustness for the emotion recognition tasks using the MEAD dataset.

V. Conclusion

This research presents Fused-HGNN, a hypergraph-based framework designed to improve multi-modal emotion recognition by capturing higher-order dependencies between the facial and audio modalities. This approach overcomes the shortcomings of unimodal and ensemble strategies through a single representation learned using hypergraph fusion and convolution. Results on the MEAD benchmark dataset demonstrate the model's superiority, with a classification accuracy of 98.5% (loss metric 2.3%), surpassing ensemble baselines by 1.4% accuracy and 0.8% loss. The framework is extremely efficient in computational resource utilization as training was completed after spending just 350 minutes in contrast to 480 minutes in the ensemble. The results of resource utilization also shows feasibility, where the peak memory is 1.5% and baselined system 2.2%. The usage of hypergraphs support efficient encoding of intra-modal and cross-modal dependencies using dedicated convolutional operations, in support of the Fused-HGNN for richer and integrated joint representation. In addition, Fused-HGNN uses modality specific weight matrices that's embedded in the convolutional layers, so there can be evenly balanced contribution of facial and audio features improving robustness for classification.

Future directions should be to test the model in real-time, affective computing contexts; broaden the adaptability in rate-latencies in contexts such as emotion aware virtual agents, intelligent surveillance systems, and human-computer interaction contexts. Lastly, there are broader possibilities to expand lightweight hypergraph formulations, or adaptive fusion. These two possibilities could be another approach to improve scale and responsiveness in context, and proposal.

References

[1] C. Lin, H. Cheng, Q. Rao, and Y. Yang, "M^{3}SA: Multimodal Sentiment Analysis Based on Multi-Scale Feature Extraction and Multi-Task Learning," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 1416–1429, 2024.

[2] J. Li and J. Peng, "End-to-End Multimodal Emotion Recognition Based on Facial Expressions and Remote Photoplethysmography Signals," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 10, pp. 6054–6063, 2024.

[3] N. Ying, Y. Jiang, C. Guo, D. Zhou, and J. Zhao, "A Multimodal Driver Emotion Recognition Algorithm Based on the Audio and Video Signals in Internet of Vehicles Platform," *IEEE Internet of Things Journal*, vol. 11, no. 22, pp. 35812–35824, 2024.

[4] B. Schölkopf, J. Platt, and T. Hofmann, "Learning with Hypergraphs: Clustering, Classification, and Embedding," in *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*, pp. 1601–1608, 2007.
[5] H. Wu, Y. Yan, and M. K.-P. Ng, "Hypergraph Collaborative Network on Vertices and Hyperedges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3245–3258, 2023.
[6] Y. Wang, Z. Xue, M. Jia, Z. Liu, and H. Su, "Hypergraph Convolutional Network With Multiple Hyperedges Fusion for Hyperspectral Image Classification Under Limited Samples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–18, 2024.
[7] M. M. Garasuie, M. Shabankhah, and A. Kamandi, "Improving Hypergraph Attention and Hypergraph Convolution Networks," in *Proc. 11th International Conference on Information and Knowledge Technology (IKT)*, pp. 67–72, IEEE, 2020.
[8] Y. Gao, Y. Feng, S. Ji, and R. Ji, "HGNN+: General Hypergraph Neural Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3181–3199, 2023.
[9] Y. Zhang, H. Zhang, L. Xiao, Y. Bai, V. D. Calhoun, and Y.-P. Wang, "Multi-Modal Imaging Genetics Data Fusion via a Hypergraph-Based Manifold Regularization: Application to Schizophrenia Study," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2263–2272, 2022.
[10] Y. Wang, Z. Xue, M. Jia, Z. Liu, and H. Su, "Hypergraph Convolutional Network With Multiple Hyperedges Fusion for Hyperspectral Image Classification Under Limited Samples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–18, 2024.
[11] X. Yan, Z. Shi, Z. Sun, and C.-A. Zhang, "Multisensor Fusion on Hypergraph for Fault Diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 8, pp. 10008–10018, 2024.
[12] F. Yao, X. Sun, N. Liu, C. Tian, L. Xu, L. Hu, and C. Ding, "Hypergraph-Enhanced Textual-Visual Matching Network for Cross-Modal Remote Sensing Image Retrieval via Dynamic Hypergraph Learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 688–701, 2023.
[13] Q. Wang, J. Huang, T. Shen, and Y. Gu, "EHGNN: Enhanced Hypergraph Neural Network for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024.
[14] W. Zhang, Y. Qin, and X. Tao, "MMPHGCN: A Hypergraph Convolutional Network for Detection of Driver Intention on Multimodal Physiological Signals," *IEEE Signal Processing Letters*, vol. 31, pp. 1730–1734, 2024.
[15] D. A. Nguyen, C. H. Nguyen, and H. Mamitsuka, "Central-Smoothing Hypergraph Neural Networks for Predicting Drug–Drug Interactions," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 8, pp. 11620–11625, 2024.
[16] Q. Xu, S. Xu, J. Liu, and L. Huang, "Dynamic Hypergraph Convolution and Recursive Gated Convolution Fusion Network for Hyperspectral Image Classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, pp. 1–5, 2023.
[17] J. Zhang et al., "Learnable Hypergraph Laplacian for Hypergraph Learning," Proc. IEEE ICASSP, pp. 4503–4507, 2022.
[18] Z. Liu et al., "A Multi-Modal Hypergraph Neural Network via Parametric Filtering," *IEEE Transactions on Big Data*, vol. 9, pp. 1365–1379, 2023.
[19] X. Hao et al., "Hypergraph Neural Network for Skeleton-Based Action Recognition," *IEEE TIP*, vol. 30, pp. 2263–2275, 2021.
[20] S. Akinpelu and S. Viriri, "Deep Learning Framework for Speech Emotion Classification: A Survey," *IEEE Access*, vol. 12, pp. 152152–152182, 2024.