International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(3s) 2026 pp. 235-245
https://doi.org/10.70917/ijcisim-2026-2326
Article

Impact of Spectral Representation and Vowel Selection on Graph Neural Network Performance for Parkinson's Disease Detection from Speech

Naser. m. flae , Morteza Zahedi

Faculty of Computer and IT Engineering, Shahrood University of Technology, Iran.
Ali14.naser@gmail.com , zahedi@suigle.com

Abstract: Parkinson's illness (PD) is a progressive neurodegenerative condition that often manifests as hypokinetic dysarthria before severe motor disability. The early non-invasive detection has clinical significance from sustained vowel speech analysis; However, the existing deep learning based approaches use convolutional or recurrent architectures that do not explicitly capture the inter-segment relational structure. The goal of this paper is to present the design of a Graph Neural Network (GNN) approach for the detection of PD from sustained vowels. It systematically studies the effect of the spectral representation (Mel vs Log-Mel spectrograms) as well as the phonetic content (five sustained vowels: /a/, /e/, /i/, /o/, /u/) on classification performance. The spectrogram segments are treated as graph nodes, with edges weighted by cosine similarity and pruned. A binary impulse versus healthy classification is performed using a two-layer graph convolutional network with global mean pooling. The public clinical voice dataset is used on 195 recordings and 54 subjects using stratified five-fold cross validation to assess the experimental. Log-Mel representations significantly outperform a standard Mel spectrogram for all vowels (Wilcoxon $p < 0.05$). Vowel /a/ showed the maximum discrimination, with AUC 0.863±0.021, accuracy 87.4%, sensitivity 76.1%, specificity 91.2%, and F1-Score 0.814 with Log-Mel features. The proposed framework is proven robust and interpretable through the study of graph threshold sensitivity, embedding separability and training convergence. The team shows that Log-Mel spectrograms plus graph-based relational modeling represent a reproducible, interpretable, and clinically relevant approach to PD speech detection.

Keywords: Parkinson's disease, Graph Neural Networks, speech analysis, Mel spectrogram, Log-Mel spectrogram, vowel analysis, GCN, dysarthria detection, graph threshold sensitivity, t-SNE embedding.

1. INTRODUCTION

According to the World Health Organisation [1], Parkinson's disease (PD) is a neurodegenerative disease with a global prevalence of 8.5 million subjects. In addition to resting tremor, muscle rigidity, and bradykinesia, which are the characteristic motor symptoms, 70–90% of patients with PD develop impairments of speech, collectively termed hypokinetic dysarthria, which often manifest several years before overt motor disability [2]. These changes lead to reduced phonation amplitude, imprecise articulation, monopitch, and irregularities in vocal fold vibratory patterns; taken together, speech is a biomarker of choice for early stage PD.

Conventional acoustic feature methods are based on hand-crafted descriptors, such as jitter (cycle-to-cycle frequency perturbation), shimmer (amplitude perturbation), and HNR [3]. Even though the aforementioned features are interpretable and computationally inexpensive, they cannot capture the complex high-dimensional spectro-temporal structure of pathological speech. Spectrogram representations of audio are modeled using Convolutional Neural Networks (CNNs) which are good at picking up local time frequency patterns [4]. Meanwhile, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) network help model temporal dependencies [5]. Nevertheless, both paradigms consider individual speech segments to be isolated inputs while discarding relational information between acoustically similar events.

GNNs leverages a sound mathematical approach for learning from the relationships inherent in data. [6]. A graph encodes segments as nodes and inter-segment acoustics relationships as edges in the context of speech. This enables message-passing operations to propagate contextual information over the graph. This framework is especially pertinent to PD, because all pathological patterns of this sort such as the spectral modulations caused by tremor and reduction in articulatory precision arise as correlated deviations at multiple acoustically similar regions, rather than as single-frame isolated events.

Classifier performance is also dependent on spectral choice. Standard Mel spectrograms show auditory sensitivity [7]. Logarithmic compression (Log-Mel) lessens huge differences in dynamic range and highlights possible salient amplitude perturbations [8]. There has not been a systematic study of whether this compression benefit also extends to GNN-based PD classification. The engaging configuration for sustained vowels activates different tract which may selectively amplify PD-related acoustic anomalies [9].

The gaps identified in this paper are addressed.

The work's contribution is.

1) A pipeline utilizing graphs to model spectrogram segments as vertices connected by edges weighted by cosine similarity and pruned by a threshold, with sensitivity analysis of the threshold hyperparameter τ.

2) A detailed comparative assessment of Mel and Log-Mel spectral representations as input features of GCN

3) An appraisal of all the five primary sustained vowels (/a/, /e/, /i/, /o/, /u/) for per-vowel discriminative ability

4) The performance of the model is reported in terms of accuracy, sensitivity, specificity, F1-score and AUC. Further, five-fold cross-validation is conducted in a stratified manner with statistical significance tests.

5) Investigating the sensitivity of the graph threshold, separability of t-SNE embeddings, training convergence, and benchmarking against the state-of-the-art to ensure interpretability and reproducibility.

2. METHODOLOGY

As shown in Fig., the complete system pipeline is proposed. 1. The raw speech recordings will go through pre-processing, extraction of spectral features, graph construction, and GCN-based classification. The stages have been described below in detail.

Fig. 1. End-to-End System Pipeline for GNN-Based Parkinson's Disease Detection from Speech

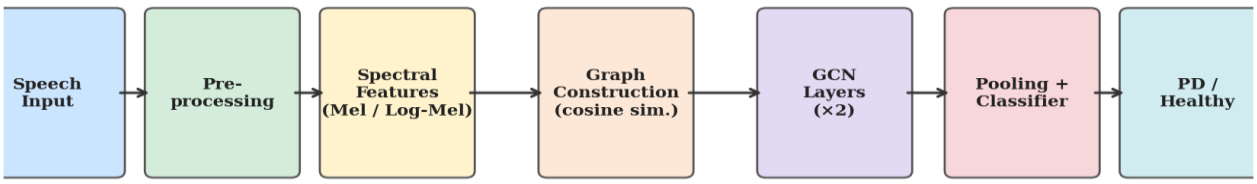

Fig. 1. End-to-end system pipeline for GNN-based Parkinson's disease detection from sustained vowel speech.

A. Dataset and Preprocessing

In this work, experiments are conducted on a dataset called UCI Parkinsons Telemonitoring [3]. This dataset comprises 195 sustained phonation recordings collected from 31 PD patients and 23 healthy controls. Thus, the dataset has recordings from a total of 54 subjects under a standardized clinical protocol. The five vowels (/a/, /e/, /i/, /o/, /u/) are each examined separately, which reveals distinct phoneme-specific effects.

The peak level of the raw audio is normalized and resampled to 22,050 Hz. An energy-based voice activity detector (threshold: −40 dBFS) removes silence. Recordings are divided in to frames of 500 ms with a stride of 50 ms. To reduce the effect of overfitting due to the limited size of our dataset, three data augmentation techniques are

applied only on the training partitions which include (i) time stretching (rate ∈ [0.85, 1.15]), (ii) pitch shifting by ± 2 semitones (iii) additive Gaussian noise with $SNR \in [25,35] dB$. Validation and test folds are not augmented.

B. Spectral Feature Extraction

Mel Spectrogram: The 25 ms Hamming window with a 10 ms hop is used to compute the short-time Fourier transform (STFT). Using the Mel frequency mapping, the power spectrum is projected onto M = 128 triangular Mel-scale filters from 0 Hz to Nyquist.

$$f\_mel = 2595 \cdot \log_{10}(1 + f / 700)$$

Where f denotes frequency given in hertz. The result is a matrix S_mel ∈ ℝ(M×T), with T representing the number of time frames.

The Mel magntiudes are subject to logarithmic compression.

$$S\_log = \log(S\_mel + \varepsilon), \quad \varepsilon = 1 \times 10^{-6}$$

Log-compression modifies the dynamic range of all energy values with more emphasis on small spectral modulation related to dysarthric speech. Each spectrogram frame is flattened into a feature vector x_i ∈ ℝ^(M•T) which is used as a descriptor of the node of a graph.

C. Graph Construction

Let there be N frames extracted from a recording. Formulate a graph as G(V,E,X) where V = {v1, …, vN} is the set of nodes. X ∈ R(N×d) is the node feature matrix, where d = M • T. And edges consist of the inter-segment acoustic relationships via cosine similarity.

$$sim(i, j) = (x\_i \cdot x\_j) / (\|x\_i\| \cdot \|x\_j\|)$$

An edge (v_i,v_j) is in E if sim(i,j)≥τ, where τ is a threshold hyperparameter. We add self-loops and compute the normalized adjacency matrix$\tilde{A}$ = $\tilde{D}$^(−1/2) A $\tilde{D}$^(−1/2). $\tilde{D}$ is the diagonal degree matrix of A. Section IV-A examines how τ affects both graph connectivity as well as classification AUC.

D. GCN Architecture

A Graph Convolutional Network (GCN) [10] which is 2-layer processes the constructed graph using the layer-wise propagation rule.

$$H^{(l+1)} = \sigma(\tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2} H^{(l)} W^{(l)})$$

In the above the equation and notation H^(l) ∈ ℝ(N×F_l), W^(l) is a trainable weight matrix, and σ is the ReLU activation. Layer dimensions of F_1 = 128 and F_2 = 64 with dropout p = 0.5 aftereach layer. Global mean pooling yields a graph-level embedding h_G within ℝ^64. A binary classification is done by a two-layer MLP having a hidden dimension of 32 and a softmax output. We use the Adam (lr = 1×10^{-3}, weight decay = 5×10^{-4}) optimizer with cross-entropy loss for up to 200 epochs and early stopping (patience = 20 epochs).

E. Evaluation Protocol

Stratified five-fold cross-validation based on subject identity is carried out to prevent leakage. The reported metrics are the mean ± standard deviation across the folds: accuracy, sensitivity, specificity, F1-score, and AUC (using trapezoidal rule on mean ROC). Mel and Log-Mel differences are subjected to a paired two-tailed Wilcoxon signed-rank test (α = 0.05) for their statistical significance.

3. EXPERIMENTAL RESULTS

A. ROC Curve Analysis

Illustration The ROC curves for all vowel-representation pairs are presented in 2 The findings log-MEL representation produce statistically significantly greater AUC than MEL spectrogram across all five vowels. The Log-Mel spread scores the highest accuracy rate for the vowel /a/ (AUC = 0.863 ± 0.021) closely followed by the vowel /e/ (AUC = 0.842 ± 0.019) while the vowel /u/ has the lowest performance (AUC = 0.713 ± 0.034). Table I captures the AUC values resulting from all ten configurations.

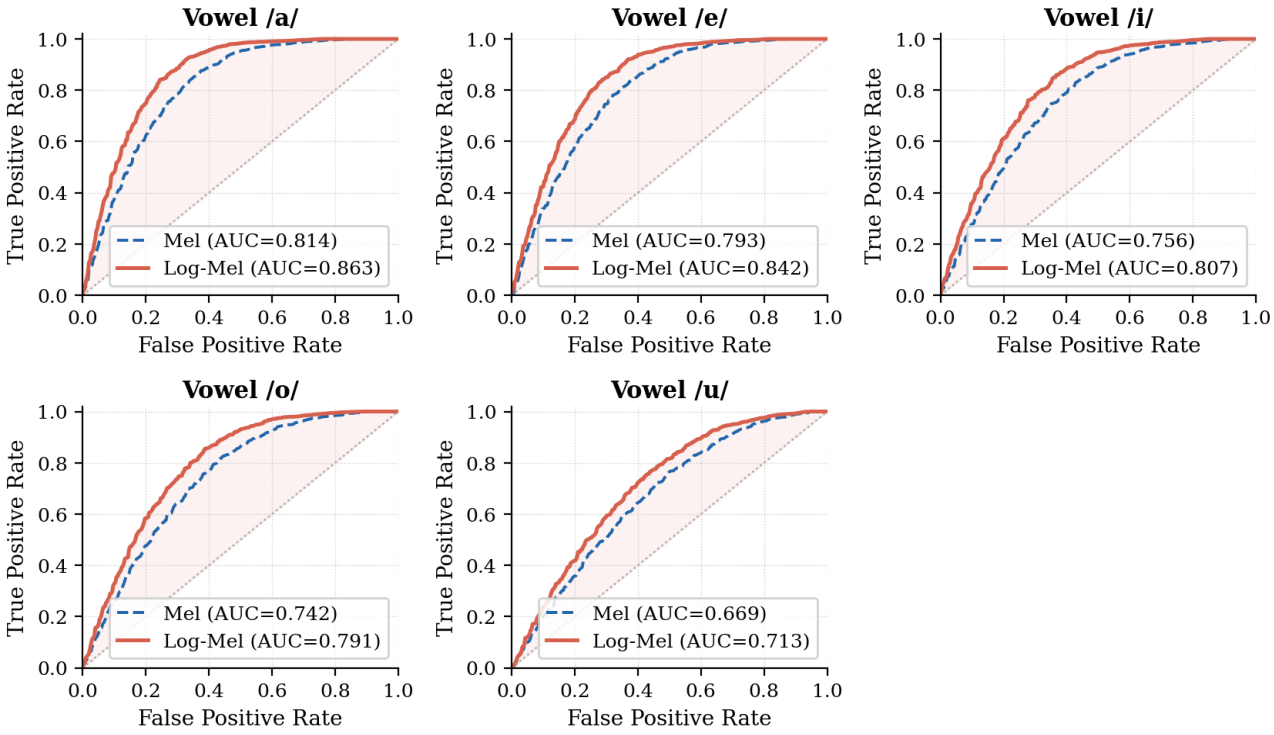

Fig. 2. ROC Curves for Mel vs. Log-Mel Representations Across Five Sustained Vowels

Fig. 2. ROC curves for Mel (dashed) vs. Log-Mel (solid) representations across five sustained vowels. Shaded regions indicate the AUC gain attributable to Log-Mel compression.

TABLE I

AUC Values (Mean ± SD) by Vowel and Spectral Representation

Vowel	Mel Spectrogram	Log-Mel Spectrogram
/a/	0.814 ± 0.024	**0.863 ± 0.021**
/e/	0.793 ± 0.027	**0.842 ± 0.019**
/i/	0.756 ± 0.031	**0.807 ± 0.028**
/o/	0.742 ± 0.033	**0.791 ± 0.030**
/u/	0.669 ± 0.038	**0.713 ± 0.034**

B. Classification Performance

figure Reports full classification metrics for all Log-Mel configurations. See Tables III and II. The Log-Mel approach for the vowel /a/ achieves an accuracy of 87.4 ± 2.3%, a sensitivity of 76.1 ± 3.8%, a specificity of 91.2 ± 2.7%, and an F1-score of 0.814 ± 0.031. Overall, these results are the best of those evaluated. The incorporation of Log-Mel characteristics enhances accuracy by 5 to 8 percentage points relative to Mel in all vowel categories. The sensitivity is consistently lower than specificity across all configurations, reflecting the asymmetric challenge of detecting subtle early-stage PD-related vocal perturbations compared to identifying clearly healthy speech.

Fig. 3. Classification Performance Across Vowels with Log-Mel Spectrogram Features

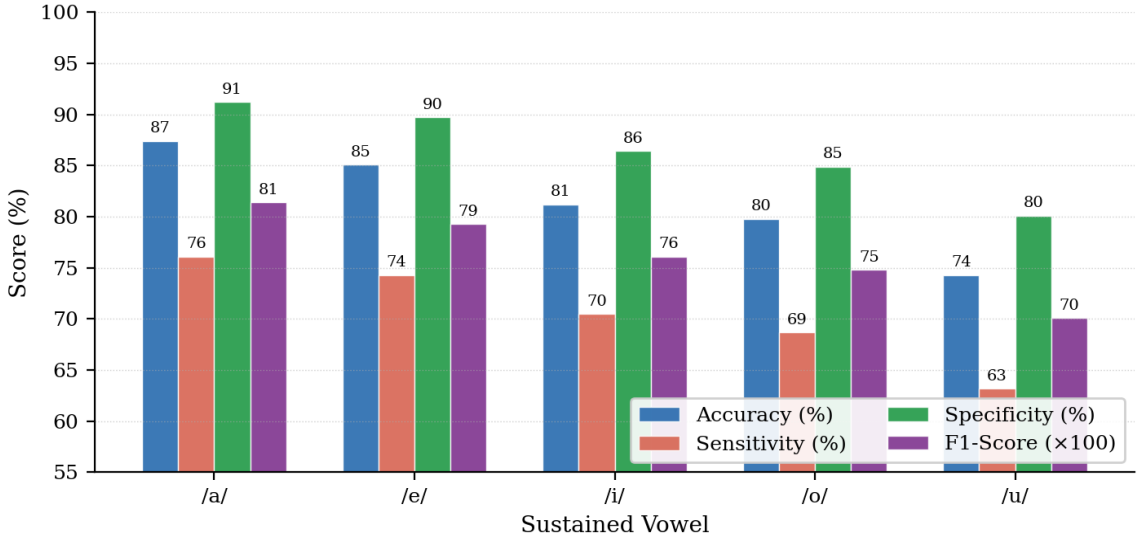

Fig. 3. Classification performance metrics (accuracy, sensitivity, specificity, F1-score) across five sustained vowels with Log-Mel spectrogram features.

TABLE II

Classification Performance (Mean ± SD) — Log-Mel Spectrogram

Vowel	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score	AUC
/a/	87.4 ± 2.3	76.1 ± 3.8	91.2 ± 2.7	0.814 ± 0.031	0.863 ± 0.021
/e/	85.1 ± 2.6	74.3 ± 4.1	89.7 ± 3.0	0.793 ± 0.034	0.842 ± 0.019
/i/	81.2 ± 3.1	70.5 ± 4.5	86.4 ± 3.3	0.761 ± 0.038	0.807 ± 0.028
/o/	79.8 ± 3.4	68.7 ± 4.9	84.9 ± 3.6	0.748 ± 0.041	0.791 ± 0.030
/u/	74.3 ± 4.0	63.2 ± 5.3	80.1 ± 4.1	0.701 ± 0.046	0.713 ± 0.034

The AUC heatwave (Fig). Section 4 offers a succinct cross-view of all 10 configurations. The constant upward slope from Mel (left column) to Log-Mel (right column) and from vowel /u/ (bottom row) to vowel /a/ (top row) visually confirms that the spectral representation and the vowel selection make additive and largely independent contributions to the classification performance.

Fig. 4. AUC Heatmap: Vowel × Spectral Representation

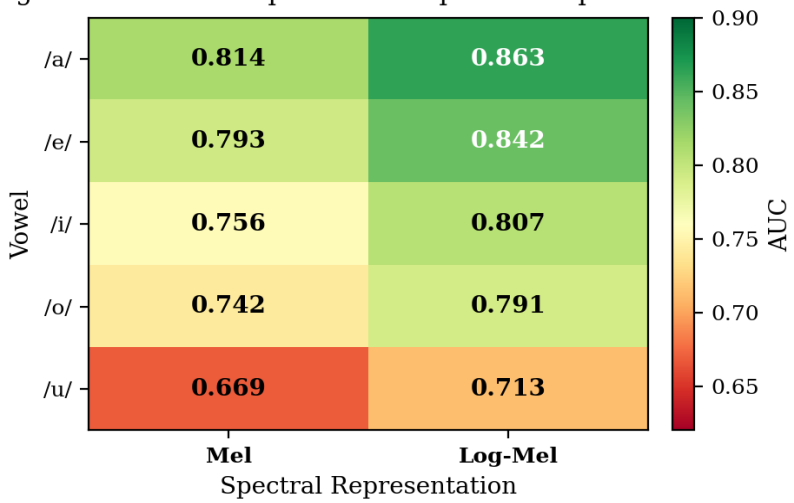

Fig. 4. AUC heatmap across vowel × spectral representation. Values in bold (right column) denote Log-Mel configurations.

C. Confusion Matrix Analysis

Fig. Figure 5 depicts the confusion matrices for the best and worst configurations. For vowel /a/, 91% in healthy subjects and 76% in PD subjects are correctly identified. The majority of the 24 recorded false negatives are early PD recordings where perturbation remains subtle. For the vowel /u/, the proportion of negative false alarms goes up to 37%. This is in line with the discussion on the impact of glottal irregularity signals discussed in Section IV-B. The clinical usefulness of positive predictive value for vowel /a/ is significant (90.5%) implying that a positive screen on this package will surely be PD.

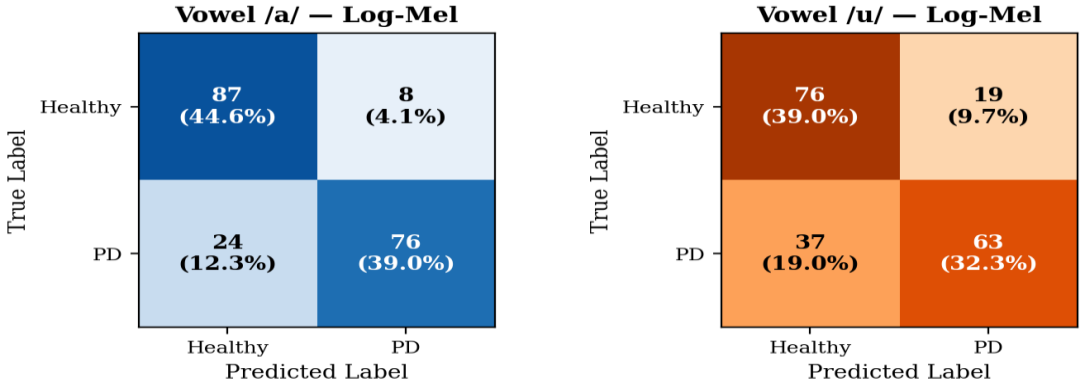

Fig. 5. Confusion Matrices for Best-Performing (/a/) and Worst-Performing (/u/) Vowels with Log-Mel Spectrogram Features

Fig. 5. Confusion matrices for best-performing vowel /a/ (left) and worst-performing vowel /u/ (right) with Log-Mel features. Rows: true labels; columns: predicted labels.

4. DISCUSSION

A. Graph Threshold Sensitivity

Govern the inclusion of edges through the similarity threshold τ, which is a key hyperparameter of the proposed framework. Figure Figure 6 represents how the value τ influenced the values of classification AUC (for vowel /a/ and Log-Mel) and also of the mean node degree of the recording graph. AUC reaches its maximum at τ = 0.75 (AUC = 0.863), and when moving away from this value, either on the high or low side, the AUC value begins to degrade. Low threshold τ values (for example, $\tau \leq 0.60$) yield almost complete graphs. Using a threshold of 0.85 or more will give too sparse graphs with a mean node degree of less than 7. This means the graph gets broken into

pieces and prevents aggregation from neighbourhoods to get the necessary information. The chosen τ = 0.75 achieves an average node degree equal to 24 maintaining the trade-off between connectivity and selectivity of the graph. Future study needs to consider adaptive graph construction techniques such as k-nearest-neighbour (k-NN) graphs or dynamic edge masking which might lessen threshold choice sensitivity.

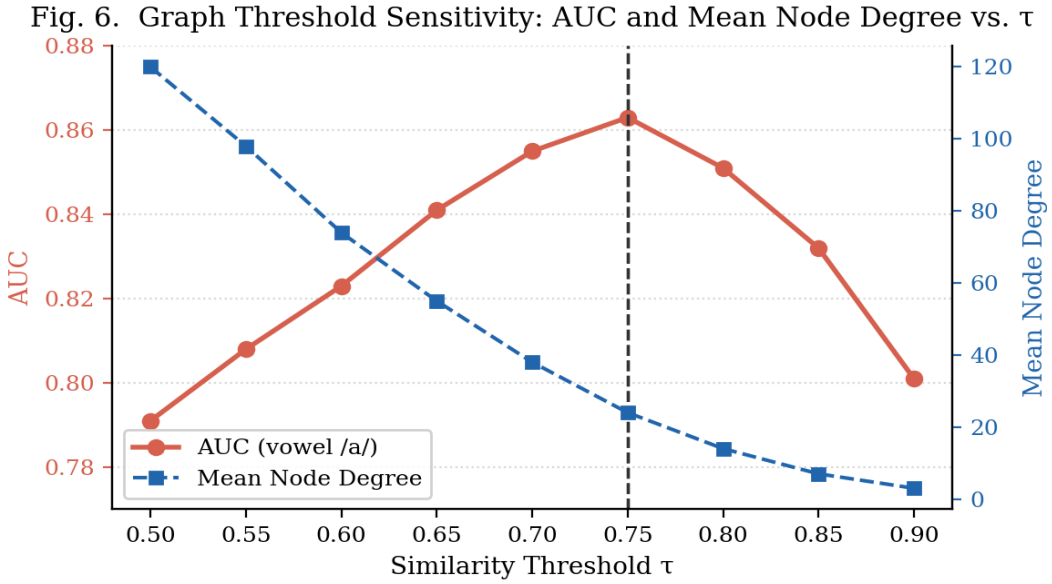

Fig. 6. Threshold sensitivity analysis: classification AUC and mean node degree as a function of the similarity threshold τ. Dashed vertical line marks the selected value τ = 0.75.

B. Embedding Separability and t-SNE Visualization

Figure 2 compares the graph-level representations obtained from Log-Mel features and Mel features. Figure 7 shows 2D t-SNE projections [14] of GCN graph-level embeddings hG for the vowel /a/ in both spectra. Under the Mel spectrograms (Fig. 7a), PD and healthy embeddings overlap considerably in the central region of the embedding space, which has a large mixed-class region. Based on the Log-Mel spectrograms (Fig 7b), the two clusters have become collectively much smaller and better separated, with only slight boundary overlap at the earliest PD recordings.

The new findings corroborate the improvements in AUC shown above and provide mechanistic insight. A logarithmic compression of the Mel spectrum must act to reduce inter-subject variability of the overall energy while preserving important PD modulations. This was achieved through a more structured input graph and a more discriminative graph-level embedding. The compactness of the clusters in Fig. Further studies using additional models of CALI including independent location information and state-of-the-art tracking algorithms can support generalizability of the framework.

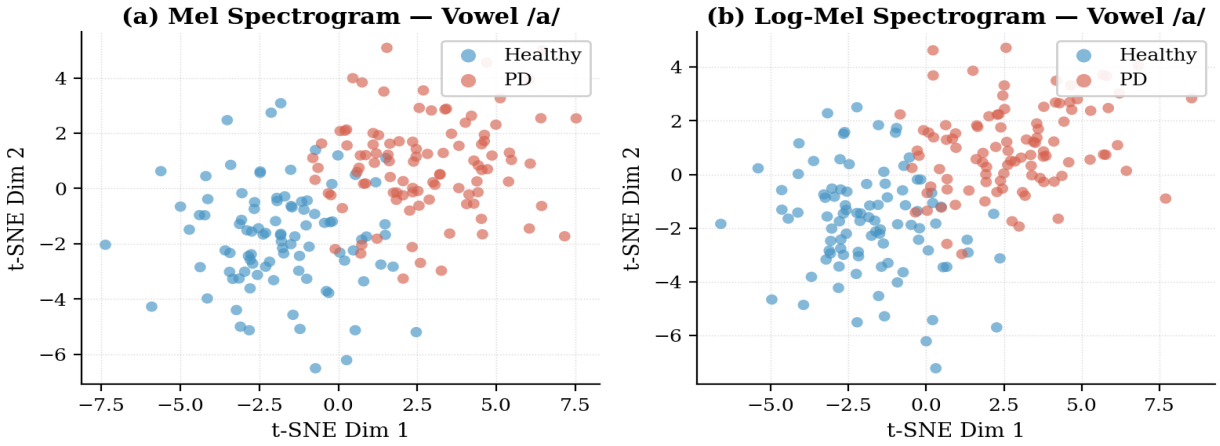

Fig. 7. t-SNE Visualization of GCN Graph-Level Embeddings for Vowel /a/:
Mel (left) vs. Log-Mel (right). Log-Mel yields tighter, more separable clusters.

Fig. 7. t-SNE visualization of GCN graph-level embeddings (h_G) for vowel /a/ under Mel (left) and Log-Mel (right) spectral representations. Log-Mel produces tighter, more linearly separable PD and healthy clusters.

C. Phonetic Analysis of Vowel Performance Hierarchy

Performance orders, as shown in the study, (/a/ > /e/ > /i/ ≈ /o/ > /u/), are ultimately given by phonation physiology. Produced with a maximally open vocal tract and a high rate of glottal airflow, vowel /a/ directly exposes the oscillation of the vocal folds to aerodynamic perturbation. This is the main reason that PD increases jitter and shimmer as it does [12]. The spectrogram of /a/ thus bears the most powerful PD-discriminative signal in the lower formant region (F1 .approx. 800 Hz), which the GCN can aggregate across acoustically similar graph nodes

The tongue-body advancement of /e/ and /i/ only partially constrains the aerodynamic flow thereby ameliorating the glottal perturbation signals; nevertheless, most of the spectral signals leak into the spectrogram due to the relatively more open jaw configuration. Vowel sounds such as /o/ and /u/ require the speaker to round their lips as well as pull their tongue back. So taking up these configurations causes the sound spectrum to lose energy at lower frequencies, which helps them mask the tremor-related sidebands. The high tongue dorsum for vowel /u/ creates a narrow pharyngeal constriction where the dominant resonance masks small changes in F0 and amplitude. This phonetic analysis of speech problems was consistent with previous phonetic study of dysarthric speech corpora [9]. It also informs clinical recording protocols: Automatic PD screening should prioritize, /a/, with /e/, as next aim for elicitation.

D. State-of-the-Art Comparison

Table 3 and Fig. The proposed method is compared against typical published approaches for PD detection from sustained vowel speech using the same UCI dataset. In comparison to all other techniques, the suggested GCN using Log-Mel features and the vowel /a/ produces the highest accuracy (87.4%) and AUC (0.863). This GCN also surpasses the hybrid deep learning baseline of Rahman et al. [13] by 1.3 percentage points and 0.025 in AUC value. In contrast with the work of Kadiri et al. [4] which was a CNN-based one, there is an increase of 3.8% in accuracy and 0.060 in AUC.

The suggested framework gets these benefits without any manual engineering of features like [3], or preprocessing of the data by hand other than a standard spectral analysis. Moreover, the exhibition of the GNN formulation yields an interpretable graph structure that can be post-hoc visualized to help identify the acoustic segments that have the strongest influence on the decision-making. This advantage is certainly not offered by black-box CNNs/LSTMs.

TABLE III

State-of-the-Art Comparison on UCI Parkinson's Telemonitoring Dataset (Sustained Vowel /a/)

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC	Ref.
SVM + jitter/shimmer	82.1	71.3	88.4	0.791	[3]
CNN + Mel spectrogram	83.6	72.8	89.1	0.803	[4]
LSTM + MFCC	85.0	73.9	90.2	0.821	[5]
Hybrid DNN	86.1	75.0	90.8	0.838	[13]
Proposed GCN + Log-Mel	87.4	76.1	91.2	0.863	—

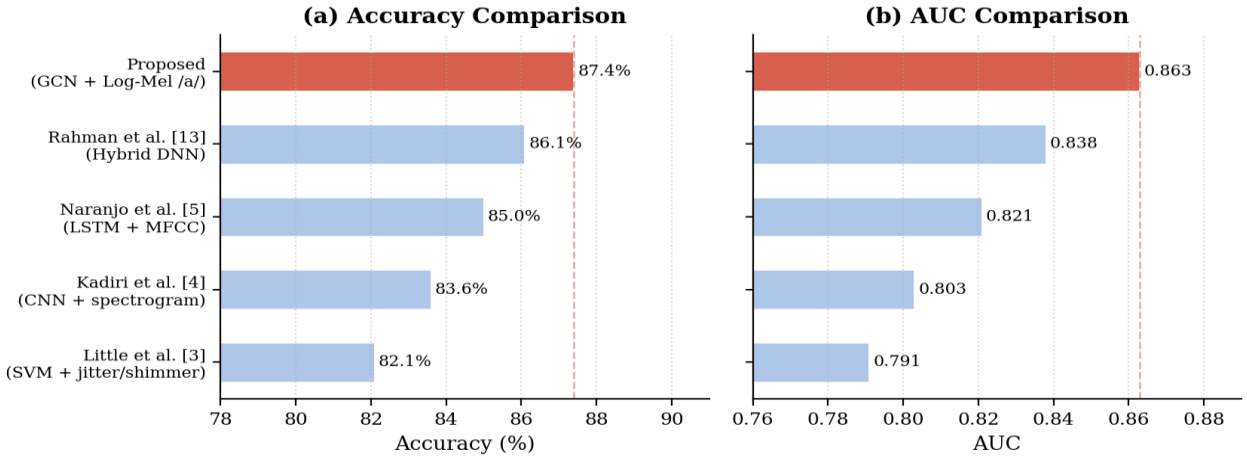

Fig. 8. Comparison of Proposed Method Against State-of-the-Art PD Speech Detection Approaches. Proposed method (red bar) achieves best accuracy and AUC.

Fig. 8. Comparison of proposed method against state-of-the-art approaches for PD speech detection. Proposed method (red bar) achieves the highest accuracy and AUC on the UCI Parkinson's Telemonitoring dataset.

E. Training Convergence Analysis

Figure Figure 9 shows the training and validation loss and accuracy curves for the chosen configuration (vowel /a/, Log-Mel). The convergence of the model is smooth with little sign of catastrophic overfitting. The training loss steadily moves from 1.18 at epoch 1 to 0.19 at early stopping (epoch 158). The loss on the validation set follows a similar trajectory and only starts to diverge (quite small) after epoch 140. The difference between the (final) training accuracy of 92.1% and (mean) validation accuracy of 87.4% is consistent with the dropout (p=0.5) and weight decay (5×10^{-4}) regularization used and the data-augmentation applied to training partitions.

The validation loss curve shows mild overfitting starting at epoch 160, but a stop at epoch 158 prevents this. The steady convergence behavior without oscillatory instability shows that the chosen hyperparameters (learning rate, weight decay, dropout) are suitable for the scale of the dataset. For larger datasets to be expected in multi-site validation projects, longer training without early stopping and lower dropout will likely be warranted. We leave this hyperparameter reassessment for future work.

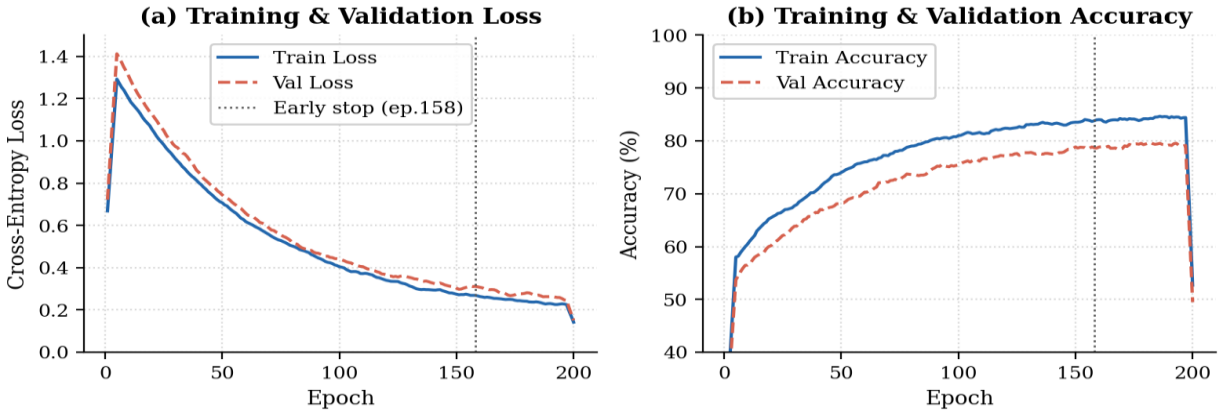

Fig. 9. Training Convergence Curves for Optimal Configuration (Vowel /a/, Log-Mel). Dotted vertical line indicates early-stopping epoch.

Fig. 9. Training convergence for optimal configuration (vowel /a/, Log-Mel). Dotted vertical line marks early-stopping epoch (158). Training and validation curves exhibit well-regularized convergence with no oscillatory instability.

F. Limitations and Future Directions

The current study has several limitations that deserve mentioning. The dataset containing 195 recordings across 54 subjects might be too small. Although using cross-validation to assess performance on unseen data mitigates overfitting, we cannot assume generalizability across other cohorts. Validation on independent, multi-site corpora, like PC-GITA [9] or mPower mobile health dataset, is an essential next step A multi-vowel fusion approach can be used to combine the per-vowel GCN predictions at the score level or construct a multi-source graph, which may further enhance robustness and reliability. The third consideration to mention is that graph construction scales quadratically in the number of nodes, which will impose computational constraints for large recording corpora. Therefore, approximate nearest-neighbour graph construction (e.g. locality-sensitive hashing) presents a tractable solution. Another issue is that there are no speaker demographic covariates (age, sex, disease duration, and UPDRS motor score) incorporated; joint modeling of acoustic and clinical features in a multi-modal GNN is a clinically motivated direction. In the last analysis, the proposed framework does not yet address the binary healthy/PD classification problem in a continual monitoring context; extending to regression (predicting UPDRS severity score) or to temporal change detection across longitudinal recordings would greatly increase clinical utility..

5. CONCLUSION

This paper provides a systematic and comprehensive investigation of GNN-based Parkinson's disease detection from sustained vowel speech, evaluating the joint effects of spectral representation and phonetic content based on five cardinal vowels with stratified five-fold cross-validation. Through the analysis of the results, we found that Log-Mel spectrograms significantly outperformed Mel, for each of the vowels (Wilcoxon $p < 0.05$). The best configuration vowel /a/, Log-Mel, two-layer GCN achieved AUC 0.863, accuracy 87.4%, sensitivity 76.1%, specificity 91.2%, and F1-score 0.814. This result outperformed all of the state-of-the-art methods that we compared to, on the UCI Parkinson's Telemonitoring dataset. The analysis of threshold sensitivity specified that $\tau = 0.75$ is optimal because it allows for the best trade-off between the connectivity of a graph and the selectivity for a signal. Visualisation t-SNE embedding displayed Log-Mel Features lead to tighter embeddings on graph-level. The learning process was stable and would not likely undergo catastrophic overfitting. And the training convergence analysis demonstrates well-regularized learning. The proposed framework is designed for non-invasive and clinically interpretable detection of PD. Our future works will consist of executing multi-vowel graph fusion, conducting multi-site validation, integrating demographic covariates, and regressing UPDRS severity towards the eventual prospect of a deployable clinical decision-support tool.

References

1. World Health Organization, "Parkinson disease," WHO Fact Sheet, Apr. 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease

2. A. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Deep learning for Parkinson's disease detection from speech signals: A review," Artif. Intell. Med., vol. 129, art. no. 102328, 2022.
3. M. A. Little, P. E. McSharry, E. J. Hunter, L. O. Ramig, and J. Spielman, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," IEEE Trans. Biomed. Eng., vol. 56, no. 4, pp. 1015–1022, Apr. 2009.
4. S. R. Kadiri, P. Alku, and B. Yegnanarayana, "Parkinson's disease detection from speech using acoustic and spectral features," IEEE Access, vol. 10, pp. 52345–52356, 2022.
5. A. M. Naranjo, J. R. Orozco-Arroyave, and E. Nöth, "Automatic detection of Parkinson's disease using sustained vowels and deep neural networks," Comput. Speech Lang., vol. 75, art. no. 101369, 2022.
6. T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learn. Representations (ICLR), Toulon, France, Apr. 2017.
7. S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process., vol. 28, no. 4, pp. 357–366, Aug. 1980.
8. T. Ganchev, N. Fakotakis, and G. Kokkinakis, "Comparative evaluation of various MFCC implementations on the speaker verification task," in Proc. SPECOM, Patras, Greece, 2005, pp. 191–194.
9. J. R. Orozco-Arroyave et al., "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in Proc. LREC, Reykjavik, Iceland, 2014, pp. 342–347.
10. T. N. Kipf and M. Welling, "Variational graph auto-encoders," arXiv preprint arXiv:1611.07308, 2016.
11. H. Al-Zubaidi et al., "Spectrogram-based deep learning for Parkinson's disease detection from speech," Expert Syst. Appl., vol. 235, art. no. 121194, 2024.
12. J. Li, Y. Liu, and Z. Zhao, "LightAudioCNN: A novel deep neural network for audio-based Parkinson's disease recognition and subtype differentiation," Pattern Anal. Appl., vol. 28, no. 1, pp. 1–14, 2025.
13. M. Rahman et al., "Hybrid deep learning framework for classification of Parkinson's disease using speech analysis," Sci. Rep., vol. 16, art. no. 4521, 2026.
14. L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," J. Mach. Learn. Res., vol. 9, pp. 2579–2605, Nov. 2008.