Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

June 2024, Special Issue on AI 4 All- 1

Evaluating Parkinson's Disease Severity Through Attention-Based STGCN and S2AGCN Models Utilizing Kinect Skeleton Images

Fatemeh Fadaie Ardestani✉, Code Orcid: 0009-0005-5865-4407
Dept.of Computer Engineering, AI and Robotics, University of Isfahan, Isfahan, Iran, fatemeh.fe76it@gmail.com
Nima Asadi, Code Orcid: 0000-0002-5102-6927
Doctor of Philosophy - PhD, Computer Science, University of Maryland, College Park, Maryland, United States, Nima228@gmail.com

Abstract— Parkinson's Disease (PD) is a prevalent neurological disorder marked by motor symptoms such as rigidity and tremors. Accurate and timely assessment of disease severity is essential for judging the efficacy of various treatment interventions. This study presents an innovative approach that employs computer vision technology—specifically Kinect cameras—paired with advanced deep learning techniques to enable precise evaluations of Parkinson's severity.

Leveraging the high accuracy of Kinect cameras in capturing essential movement patterns, our proposed system employs advanced convolutional neural networks, specifically incorporating mechanisms from the Spatial-Temporal Graph Convolutional Network (STGCN) and the Two-Stream Adaptive Graph Convolutional Network (2SAGCN). These architectures are adept at detecting movement anomalies and generating precise quantitative severity measures. To further enhance the performance of the 2SAGCN, we introduce distinct temporal and spatial attention modules, resulting in improved classification outcomes. The model achieves outstanding metrics, with accuracy, precision, recall, and F1 score recorded at 94.14 ± 0.26, 98.1 ± 0.12, 98.6 ± 0.05, and 98.2 ± 0.02, respectively, when incorporating spatial attention into the 2SAGCN model and utilizing 3D augmented data.

The severity classification framework distinguishes between 11 specific classes of Parkinson's symptoms, which are derived from 9 distinct motion categories. Within this framework, class 0 represents healthy individuals, while classes 0 to 1 correspond to varying degrees of severity in Parkinson's symptoms, resulting in a comprehensive classification system encompassing 99 distinct outcomes.

To further enhance the model's accuracy, we have implemented strategies such as transfer learning and data 3D augmentation. This research marks a significant step forward in the realm of non-invasive, quantitative assessments of Parkinson's Disease, showcasing the potential of cutting-edge technology and state-of-the-art neural network architectures. (Abstract)

Keywords—3D Motion tracking, Computational neurology, Deep learning diagnostics, Motor function analysis, Parkinson's assessment.

I. Introduction

Parkinson's disease is a prevalent neurodegenerative disorder characterized by motor, speech, and writing impairments, as well as other symptoms[1]. Accurately assessing disease severity is essential for monitoring progression and evaluating treatment efficacy. Traditional clinical evaluations often depend on subjective assessments, leading to potential biases. Consequently, there has been a growing interest in modern technology to develop objective, non-invasive assessment methods.

Recent advancements in computer vision, particularly with the Kinect sensor, provide an innovative avenue for automated analysis of motor functions in PD patients. Originally designed for gaming, the Kinect captures 3D skeletal data and analyzes movement patterns, making it an ideal tool for assessing motor impairments associated with Parkinson's disease. By employing deep learning techniques, this data can be processed to identify key motor dysfunctions and quantitatively measure disease severity.

This research aims to create a system that leverages Kinect's skeletal data and advanced deep learning models, specifically spatial-temporal graph convolutional networks (STGCN) and two stream adaptive graph convolutional networks (2SAGCN). The system will analyze movement patterns such as gait, tremors, and bradykinesia, providing a robust and accurate means of quantifying motor symptoms. The goal is to enhance the precision of PD diagnosis and monitoring while reducing dependence on subjective clinical evaluations.

Parkinson's is ranked second among neurological disorders after Alzheimer's and is associated with muscle stiffness, slow movements, and balance issues, ultimately leading to significant physical and cognitive decline without treatment[2]. Gait

Submit Date: 2024-10-15
Accept Date: 2025-04-21
✉ Corresponding author

alterations and walking challenges are prevalent concerns for affected individuals, highlighting the urgent need for improved diagnostic methods[3]. Analyzing human movement serves as a valuable resource for assessing motor disorders like Parkinson's. Advances in computer vision, particularly in video-based human movement detection, have shown promise for identifying and quantifying disease severity[5]. Modern methodologies utilize deep learning models to enhance detection accuracy compared to earlier techniques.

Human motion analysis spans two-dimensional and three-dimensional frameworks, with depth-sensing cameras like Kinect proving advantageous due to their non-invasiveness and cost-effectiveness. Kinect sensors minimize human error in diverse environments by enabling precise modeling of walking patterns and speeds[6]. Graph neural networks (GNNs) are recognized for their effectiveness in processing graph-represented data, excelling in tasks such as classification and robust learning[7],[8]. For detecting human states and movements, GNNs can provide superior accuracy by emphasizing joint connectivity and continuity. Models like STGCN[9] and 2SAGCN[10] have been introduced for detecting human joint movements, and leveraging these models can significantly enhance Parkinson's disease diagnosis.

This study investigates the impact of utilizing STGCN and 2SAGCN models on skeletal data to assess Parkinson's disease severity. The need for accurate diagnostic modalities for Parkinson's is critical due to its notable motor impairments. While advanced computer vision techniques offer potential solutions, their implementation and evaluation remain limited. Incorporating deep learning, artificial neural networks, and transfer learning may improve research outcomes. Key objectives of this study include:

- Evaluating Parkinson's disease severity on the IRDS dataset through expert medical labeling.
- Utilizing STGCN and 2SAGCN models for classifying severity based on skeletal data.
- Investigating the effects of Temporal Attention and Spatial Attention, within these models.
- Applying 3D augmentation techniques to enhance dataset accuracy.

II. RELATED WORK

Various methods have been proposed for assessing the severity of Parkinson's disease, including frequency component analysis[11], linear accelerometer sensors[12], fingertip magnetic sensors[13], and notably Kinect-based approaches[14], which have demonstrated superior performance.

In 2013, Liu et al.[15] utilized the Microsoft Kinect system for human motion detection, combining Kmeans clustering and Hidden Markov Models (HMMs), achieving an accuracy of approximately 91.4%. A 2015 study by Ondřej Tupa et al.[16] focused on motion tracking and gait feature estimation using Kinect, also reaching over 90% accuracy in detecting Parkinson's symptoms. In 2016, Ioannis Pachoulakis and colleagues reported an accuracy of 78% with a Kinect-based platform for mild to moderate Parkinson's patients[17]. Further enhancing this field, Dranca and colleagues achieved a classification accuracy of 93.40% for Parkinson's disease stages using Kinect technology in 2018[18]. Ilaria Bortone's team evaluated various classification algorithms, finding that Artificial Neural Networks (ANN) reached accuracies of 89.40% and 95.02% with fewer features for PD diagnosis and severity ranking, respectively[19].

Recent advancements include the CST-GCN network proposed by Tian Haoyu et al. in 2024, which leverages skeletal data captured by Kinect to estimate severity based on the MDS-UPDRS scale, yielding a Root Mean Square (RMS) error of 1.762 and an accuracy of 70%[20]. The spatio-temporal graph convolutional network (STGCN) has also shown promise for assessing Parkinson's severity[21]. This model captures both spatial and temporal features simultaneously by processing joint coordinate vectors organized into a spatial-temporal graph. Each joint acts as a node connected to others in a manner that reflects the human body's natural structure, enabling the model to track dynamics across frames.

STGCN operates with a structured input tensor of dimensions (C, V, T), representing channels, joints, and time steps. The system captures spatial dependencies using graph convolutional layers, while temporal dependencies are modeled through temporal convolutional layers with residual connections that address the vanishing gradient problem. The resulting network predicts outcomes for regression or classification through optimization techniques like backpropagation.

The 2SAGCN, elaborated in recent research, is designed for action recognition using skeleton data. This model integrates spatial and temporal information, processing skeleton sequences represented as 3D tensors to classify human actions effectively. Through an adaptive graph formulation, it adjusts its structure based on input sequences, enhancing performance in action prediction tasks.

Together, these innovative techniques leveraging Kinect technology and advanced neural networks offer significant advancements in accurately assessing and monitoring Parkinson's disease.

III. MATERIALS AND METHODS

I. Dataset Description

In this research study, the dataset is crucial for training and validating deep learning models aimed at evaluating the severity of Parkinson's disease (PD) using skeletal motion data collected through the Kinect sensor. This dataset features 3D skeletal representations of joint movements from PD patients engaged in standard motor tasks, with each participant generating thousands of frames containing 3D coordinates for 25 distinct joints.

The study utilizes the IntelliRehabDS (IRDS) dataset[22], curated by Myron et al., which comprises skeletal data collected via Kinect during rehabilitation exercises. The dataset includes 2,589 3D images from 29 participants—15 with Parkinson's and 14 healthy controls—providing a diverse representation of movement patterns. The primary goal is to classify Parkinson's severity into 11 levels based on nine essential movements. Important 3D augmentation techniques are employed to enhance the dataset, crucial for training neural networks for accurate classification.

Participants performed nine key rehabilitation gestures in their preferred positions (seated or standing), reflecting realism and variability in the data. Both correct and incorrect gestures were recorded to capture the natural variances of real-life performance. The average age of participants with Parkinson's was 43 years, while healthy controls averaged 26, allowing for nuanced comparisons across life stages. Multiple repetitions of each gesture were recorded for comprehensive performance analysis, and

the dataset maintains balanced gender distribution, enhancing generalizability.

Initial classifications of participants as healthy or PD-affected were refined with severity labels assigned by a physician and validated by a neurologist. Data was collected at the Pusat Rehabilitasi Perkeso Melaka, Malaysia, with participants in comfortable attire suitable for movement and exercises performed according to personal preference (standing, sitting, or using a wheelchair). A Microsoft Kinect One sensor recorded joint data at 30 frames per second, ensuring high-quality data for analysis. The dataset was meticulously annotated by a medical expert, with labels categorized into 11 distinct classes (ranging from 0 to 1). This labeling approach represents a unique methodological contribution, as previous studies have not employed this specific technique.

To enhance model generalization, the NTU RGB+D dataset[23] was also utilized, containing 56,880 samples across 60 action classes from 40 subjects. This multimodal dataset includes RGB videos, depth maps, 3D skeletal data, and infrared sequences, all captured under varying conditions by three cameras. Each sample comprises RGB videos (1920×1080), depth maps, IR videos (512×424), and skeletal data detailing 25 body joints, facilitating cross-subject and cross-view evaluations.

In the data collection process using the Kinect depth sensor, patients diagnosed with Parkinson's were recorded performing various common motor tasks. Key assessments included walking a predefined path to analyze gait patterns, maintaining posture and balance, and monitoring tremors during rest or simple hand exercises. The Kinect sensor tracked movements by detecting 25 key joints, including the head, neck, spine, shoulders, elbows, wrists, hips, knees, and ankles. Each frame illustrates the 3D position of these joints in real-time, allowing for a detailed analysis of both spatial and temporal features of motion.

II. Image Preprocessing

Data augmentation is a vital technique to enhance model accuracy by expanding datasets. When faced with limited or imbalanced data, artificial expansion through data augmentation significantly improves a model's generalization capabilities, particularly in medical imaging, where tailored augmentation is essential to avoid unrealistic scenarios[24].

For analyzing human skeletal positions, especially using Kinect data, 3D augmentation[25] is effective. This approach involves rotating skeletal data in 3D space to create new training data. It allows for the extraction of valuable insights, such as movement stability and bone relationships, reflecting structural changes during various activities. The process involves randomly selecting angles for rotation along the x, y, and z axes, updating joint position values accordingly.

The transformation of joint positions after rotation follows a specific equation where "old joint values" indicate positions before rotation, "new joint values" represent positions after rotation, and "angle values" denote the randomly selected rotation angles. Figures 1 and 2 illustrate data imbalances and the impact of 3D augmentation in balancing datasets. In cases of significant imbalance, uniform data augmentation is applied to address sample shortages, particularly when data instances are scarce.

$$\boldsymbol{R} = \begin{bmatrix} cos\alpha & -sin\alpha & 0 \\ sin\alpha & cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} cos\beta & 0 & sin\beta \\ 0 & 1 & 0 \\ -sin\beta & 0 & cos\beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\gamma & -sin\gamma \\ 0 & sin\gamma & cos\gamma \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

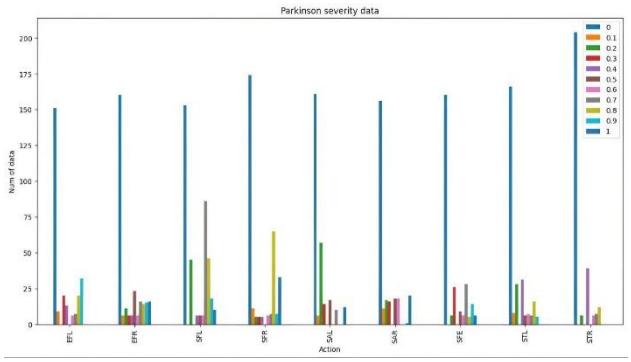

Fig 1. Comparing Data Distribution Before Balancing and 3D Augmentation [22]

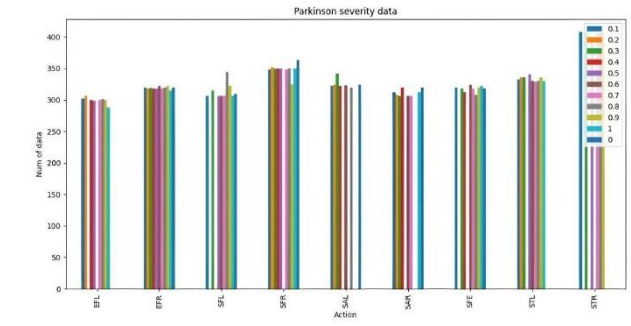

Fig 2. Comparing Data Distribution After Balancing and 3D Augmentation [22]

III. Preposed Architecture

In this study, we investigate the effectiveness of combining Spatio-Temporal Graph Convolutional Networks (STGCN) and 2-Stream Additive Graph Convolutional Networks (2SAGCN) with various attention mechanisms—including Temporal Attention, Spatial Attention, and Spatial-Temporal Attention—to accurately assess the severity of Parkinson's disease. The integration of the Temporal Attention module allows the model to dynamically adjust its focus, significantly enhancing performance in tasks characterized by complex spatial and channel-wise features.

The proposed model architecture is depicted in Figure 3, which combines convolutional layers for feature extraction with a temporal attention mechanism to capture sequential dependencies, making it well-suited for processing spatiotemporal data. Following the convolutional layers, batch normalization (BN) and ReLU activations are employed, which are designed to effectively extract hierarchical features from the spatiotemporal input data. The convolutional layers capture local and spatial patterns within the frames, gradually increasing the complexity of the feature representation. Batch normalization aids in stabilizing training by normalizing the input distributions of each layer, reducing the internal covariate shift, and allowing for faster convergence. ReLU activations introduce non-linearity, enabling the model to learn complex relationships while mitigating the vanishing gradient problem often encountered in deep networks. This combination ensures that feature extraction remains efficient, stable, and interpretable, providing a solid foundation for downstream processing, including the temporal attention block that focuses on sequential dependencies for enhanced predictive performance.

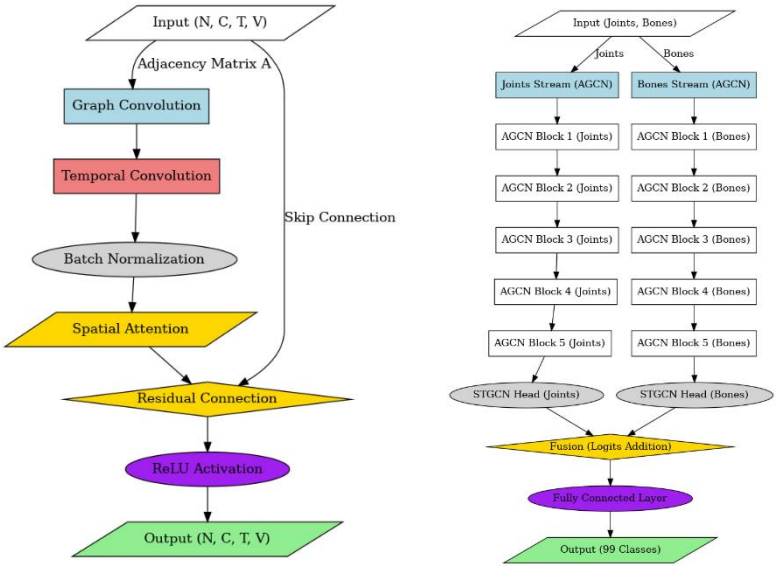

Fig 3. Right: the architecture of 2AGCN. Left: the architecture of each

AGCN block with the proposed attention mechanism within the 2AGCN model. A similar architecture is proposed with the temporal attention as well.

The temporal attention is a crucial component for focusing on the most relevant time steps in the sequence, ensuring the network emphasizes important frames or segments of data. The attention module dynamically computes weights for different time steps, allowing the model to adaptively attend to critical moments in the sequence, which is especially useful for tasks like action recognition, where not all frames are equally important.

A variation of this architecture utilizes a spatial attention block instead of the temporal attention depicted in Fig 1. Similarly, the spatial attention mechanism plays a vital role in highlighting the most critical joints of the body when analyzing movement patterns over Kinect video data. By dynamically computing weights for different nodes (joints), the spatial attention module enables the network to focus on areas that provide the most informative cues about the subject's stage of Parkinson's disease.

Finally, the features extracted by the convolutional layers and refined by the attention mechanism are passed through a fully connected layer, which combines them into a final representation suitable for the prediction task. For classification, this would involve mapping the learned features to action categories, while in other tasks, the fully connected layer could output continuous values. This overall architecture is designed to effectively capture both spatial and temporal dependencies in sequential data, making it well-suited for tasks like human action recognition

In the following sections we explain each block of the architecture separately.

1.**Graph Construction:** The AGCN Block incorporates a ConvTemporalGraphical module that dynamically builds graphs to capture the intricate relationships among different joints in the skeletal structure. This is achieved through a series of convolutional operations that adaptively modify the weights assigned to each joint, enabling the network to evaluate the significance of various connections.

2.**Temporal Convolution:** Within the ConvTemporalGraphical module, temporal convolutional operations are conducted on the constructed graphs. This process involves multiple convolutional layers (conva, convb) working along the temporal dimension, which enhances the model's ability to recognize temporal patterns and dependencies.

3.**Spatial Convolution:** Concurrently, spatial convolutional operations (convd) are employed to extract spatial features from the skeletal data. Utilizing 1x1 convolutions allows the network to capture meaningful spatial information effectively.

4. **Downsampling and Normalization:** To manage the complexity of the graph and enhance computational efficiency, a downsampling operation is implemented. This includes a convolutional layer followed by batch normalization, which effectively reduces the dimensionality of the data while retaining critical information.
5. **Activation and Normalization:** A rectified linear unit (ReLU) activation function is introduced to add non-linearity, facilitating the learning of complex patterns. Additionally, batch normalization is utilized to stabilize and accelerate the training process by normalizing inputs at each layer.
6. **Softmax Activation:** The AGCN Block integrates a softmax activation function along the second-to-last dimension of the constructed graph. This mechanism enables the network to assign appropriate attention scores to different joints, facilitating adaptive adjustments of graph connections based on their importance.
7. **AGCN Block:** The AGCN Block (a, b, c) is central to our proposed methodology for evaluating the severity of Parkinson's Disease. It leverages STGCN and S2AGCN networks to analyze Kinect skeleton images effectively.

Temporal Attention: This mechanism is a vital component frequently used in deep learning architectures, particularly within Convolutional Neural Networks (CNNs). It is specifically designed to enhance the model's focus on salient temporal features while reducing the influence of less informative elements. By effectively emphasizing critical temporal signals, Temporal Attention bolsters the model's learning and predictive abilities, leading to superior performance across various applications.

Spatial Attention: Similarly, the Spatial Attention mechanism is crucial in Graph Convolutional Networks (GCNs). Its primary aim is to improve the model's focus on significant nodes or regions within a graph while minimizing the impact of less relevant areas. By dynamically assigning varying levels of importance to different graph components, Spatial Attention enables GCNs to capture essential structural and contextual information more effectively. This mechanism significantly enhances model performance in tasks such as node classification, link prediction, and graph-based recommendations, ensuring that attention is directed toward the most informative parts of the graph, thereby facilitating more efficient learning and representation.

Through the strategic integration of the fundamental components of STGCN, 2SAGCN, and attention mechanisms, this study aims to enhance diagnostic capabilities for evaluating the severity of Parkinson's disease, contributing to the advancement of medical diagnostics in this domain.

IV. Evaluation Metrics

The model's performance is evaluated by comparing its predictions to actual labels through several key classification metrics: precision, recall, F1 score, accuracy, and area under the curve (AUC), alongside mean square error (MSE). These initial five metrics are derived from a confusion matrix, which includes True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Precision measures the proportion of correctly identified positive cases among all instances predicted as positive, while recall (sensitivity) evaluates the proportion of

correctly identified positive cases out of the total actual positive cases. The F1 score provides a single metric that harmonizes precision and recall, delivering a holistic view of the model's effectiveness.

To rigorously assess the performance of the deep learning models, the study employs k-fold cross-validation (CV). This method divides the dataset into K folds, allowing the model to be trained K times, each time using a different fold as the test set. This process not only enhances the robustness of performance estimates but is especially valuable for smaller datasets that may be susceptible to overfitting.

For reporting the results of baseline and ablation studies, a 10-fold CV approach is utilized, alongside comparisons to state-of-the-art techniques using 3-fold and 5-fold CV methods. This comprehensive strategy ensures a thorough and unbiased assessment of the model's generalization capabilities on unseen data.

$$\text{Percision}=TP/(TP+FP) \quad (1)$$

$$\text{Recall}=TP/(TP+FN) \quad (2)$$

$$\text{F1Score}=2(\text{Percision}.\text{Recall})/(\text{Percision}+\text{Recall}) \quad (3)$$

V. Validation Methods

The study utilizes an unbiased k-fold cross-validation (CV) method to thoroughly assess the performance of deep learning (DL) models. This approach partitions the dataset into K distinct folds, allowing the model to be trained K times, each time using a different fold as the test set. The final performance metric is derived from the average of the test results across all iterations, providing a robust estimate of model performance. This methodology is particularly advantageous for small datasets, which are often susceptible to overfitting.

For reporting the baseline and ablation study results, the study adopts the 10-fold CV method, while also conducting comparisons with state-of-the-art techniques that employ 3-fold and 5-fold CV methods. This comprehensive approach ensures an unbiased evaluation of the DL model's ability to generalize to unseen data.

VI. Study Design And Ablation Study

A comprehensive ablation study was conducted to assess the contribution of various model components and architectural enhancements. This involved comparing performance metrics—accuracy, precision, recall, F1-score, and AUC—with and without critical modules like Temporal and Spatial Attention, as well as evaluating different graph configurations. The impact of the 2SAGCN model versus standard STGCN models was also measured. The results underscore the importance of attention mechanisms and graph-based enhancements in capturing movement patterns associated with Parkinson's Disease severity. Additionally, the study demonstrates that the proposed architectural choices lead to significant improvements in classification performance. Figure 3 presents the confusion matrix for the 2SAGCN model with spatial attention.

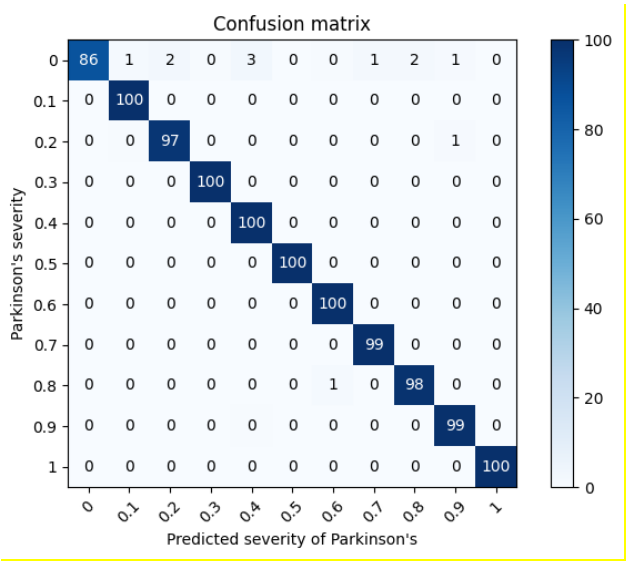

Fig 3. Confusion Matrix for the 2SAGCN Model with Spatial Attention (Each intensity corresponds to 9 different movement)

Tables I and II present a comparative analysis of the performance between STGCN and 2SAGCN models in classifying the severity of Parkinson's disease, utilizing various attention mechanisms.

I. PERFORMANCE COMPARISON OF STGCN MODEL IN PARKINSON'S DISEASE SEVERITY CLASSIFICATION UNDER VARIOUS ATTENTION MECHANISMS

Model	3D Aug	Attention	Acc	Precision	Recall	F1-Score
STGCN	No	0	85.99± 0.33	87.5 ± 0.05	92.4 ± 0.05	89.9 ± 0.02
		Temporal Attention	89.13± 0.54	90.7 ± 0.08	93.0 ± 0.04	91.8 ± 0.05
		Spatial Attention	89.68± 0.48	91.2 ± 0.22	93.8 ± 0.05	92.2 ± 0.12
	Yes	0	88.23± 0.11	90.9 ± 0.05	91.7 ± 0.03	91.3 ± 0.03
		Temporal Attention	92.63± 0.32	98.5 ± 0.07	96.8 ± 0.01	97.6 ± 0.05
		Spatial Attention	92.88± 0.13	96.7 ± 0.22	95.6 ± 0.4	96.2 ± 0.04

II. PERFORMANCE COMPARISON OF 2SAGCN MODEL IN PARKINSON'S DISEASE SEVERITY CLASSIFICATION UNDER VARIOUS ATTENTION MECHANISMS

Model	3D Aug	Attention	Acc	Precision	Recall	F1-Score
2SAGCN	No	0	88.11± 0.24	90.1± 0.02	92.5 ± 0.03	91.3± 0.05
		Temporal Attention	92.65± 0.32	93.6 ± 0.05	92.2 ± 0.08	92.9 ± 0.05
		Spatial Attention	93.15± 0.08	93.9 ± 0.21	94.1 ± 0.24	93.8 ± 0.06
	Yes	0	91.40± 0.33	93.2 ± 0.02	91.5 ± 0.05	92.2 ± 0.02
		Temporal Attention	93.71± 0.40	97.9 ± 0.05	98.2 ± 0.03	98.1 ± 0.08
		Spatial Attention	**94.14± 0.26**	**98.1 ± 0.12**	**98.6 ± 0.05**	**98.2 ± 0.02**

TRAINING THE NETWORK

The neural network was trained using the back propagation algorithm on a skeletal dataset comprising individuals classified as either healthy or diagnosed with Parkinson's disease. The initial weights of the network were set randomly to introduce variability. To ensure a diverse and comprehensive training experience, images were randomly selected from each class and organized into batches. The loss function employed to assess the training process was categorical cross-entropy loss, which effectively measures the performance of the model across the multiple classes.

IV. DISCUSSION

The presented results in Table 1 highlight the performance of different model architectures for classifying Parkinson's Disease severity using skeletal motion data captured by Kinect sensors. The comparison indicates that models incorporating attention mechanisms, particularly spatial attention, achieve superior accuracy and performance compared to baseline models like STGCN

and 2SAGCN without attention modules. This suggests that attention mechanisms significantly enhance the network's ability to focus on the most critical aspects of the input data, such as relevant joints in the case of spatial attention.

The improvement observed with spatial attention over temporal attention further underscores the importance of capturing spatial dependencies in the context of Parkinson's Disease, where specific joint movements and body postures provide crucial diagnostic cues. The marginal difference between the two attention mechanisms suggests that while both contribute positively, spatial relationships might carry more relevant discriminative power in this application.

Moreover, the results reveal that attention-enhanced models not only improve classification accuracy but also boost other metrics such as precision, recall, and F1 scores. This demonstrates their robustness and effectiveness in identifying varying levels of Parkinson's symptoms, ensuring reliable assessments. The use of both attention modules in conjunction with graph-based convolutional networks effectively captures complex spatio-temporal patterns, emphasizing the utility of combining structural graph features with dynamic focus through attention.

Overall, these findings support the conclusion that incorporating attention mechanisms into graph convolutional networks is beneficial for complex tasks like severity classification of Parkinson's Disease. This approach offers a pathway to more accurate, non-invasive, and automated assessments, contributing to advancements in medical diagnostics and treatment monitoring. Future work could explore further refinements of these models, including more sophisticated attention mechanisms, to capture even more subtle and context-dependent features in movement data.

V. CONCLUSION

In this study, we proposed an innovative framework for evaluating the severity of Parkinson's Disease using skeletal motion data captured via Kinect sensors. By integrating Spatio-Temporal Graph Convolutional Networks (STGCN) and Two-Stream Adaptive Graph Convolutional Networks (2SAGCN) with specialized Temporal and Spatial Attention modules, we aimed to improve classification performance and enhance the model's ability to capture critical movement patterns. Our findings demonstrate that the inclusion of attention mechanisms significantly boosts model accuracy. Specifically, the spatial attention mechanism outperformed the temporal attention mechanism, highlighting its superior ability to focus on critical joints and capture movement nuances that are most relevant for Parkinson's severity classification. Furthermore, both attention-enhanced models provided better results compared to the baseline STGCN and 2SAGCN architectures without attention modules. These results validate the effectiveness of our approach in improving the precision and robustness of Parkinson's Disease assessments using non-invasive motion data, showcasing the potential for broader applications in clinical diagnostics and rehabilitation monitoring.

References

[1] U. Walter, L. Niehaus, T. Probst, R. Benecke, B. Meyer, D. Dressler, Brain parenchyma sonography discriminates pakinsons disease and atypical parkinsonian syndromes, Neurology 60 (2003) 74–7. doi:10.1212/WNL. 61.6.871.

[2] S. Annesley, S. Lay, S. De Piazza, O. Sanislav, E. Hammersley, C. Al440 lan, L. Francione, M. Bui, Z.-P. Chen, K. Ngoei, F. Tassone, B. Kemp, E. Storey, A. Evans, D. Loesch, P. Fisher, Immortalized parkinson's disease lymphocytes have enhanced mitochondrial respiratory activity, Disease models mechanisms (2016). doi:10.1242/dmm.025684.

[3] G. Pahuja, N. T N, A comparative study of existing machine learning 445 approaches for parkinson's disease detection, IETE Journal of Research 67 (2018) 1–11. doi:10.1080/03772063.2018.1531730.

[4] T. L. Munea, Z. Yalew, H. Tekle, L. Chen, C. Huang, C. Yang, The progress of human pose estimation: A survey and taxonomy of models applied in 2d human pose estimation, IEEE Access PP (2020) 1–1. doi:10.1109/ 450 ACCESS.2020.3010248.

[5] C. Zheng, W. Wu, T. Yang, S. Zhu, C. Chen, R. Liu, J. Shen, N. Kehtarnavaz, M. Shah, Deep learning-based human pose estimation: A survey (12 2020).

[6] R. Das, Das, r.: A comparison of multiple classification methods for di455 agnosis of parkinson disease. expert systems with applications 37, 1568- 1572, Expert Systems with Applications 37 (2010) 1568–1572. doi: 10.1016/j.eswa.2009.06.040.

[7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P. Yu, A comprehensive survey on graph neural networks (01 2019).

[8] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, M. Sun, Graph neural networks: A review of methods and applications (12 2018).

[9] S. Yan, Y. Xiong, D. Lin, Spatial temporal graph convolutional networks for skeleton-based action recognition, Proceedings of the AAAI Conference on Artificial Intelligence 32. (2018) doi:10.1609/aaai.v32i1.12328.

[10] L. Shi, Y. Zhang, J. Cheng, H. Lu, Two-stream adaptive graph convolutional networks for skeleton-based action recognition, 2019, pp. 12018– 12027. doi:10.1109/CVPR.2019.01230.

[11] G. Ye, Y. Liu, N. Hasler, X. Ji, Q. Dai, C. Theobalt, Performance capture of interacting characters with handheld kinects, 2012, pp. 828–841.doi:10.1007/978-3-642-33709-3_59.

[12] J. Fujikawa, R. Morigaki, N. Yamamoto, H. Nakanishi, T. Oda, Y. Izumi, Y. Takagi, Diagnosis and treatment of tremor in parkinson's disease using mechanical devices, Life 13 (2022) 78. doi:10.3390/life13010078.

[13] J. Li, H. Zhu, Y. Pan, H. Wang, Z. Cen, D. Yang, W. Luo, Three dimensional pattern features in finger tapping test for patients with parkinson's disease, Vol. 2020, 2020, pp. 3676–3679. doi:10.1109/EMBC44109. 2020.9176652.

[14] Z. Wu, H. Gu, R. Hong, Z. Xing, Z. Zhang, K. Peng, Y. He, L. Xie, J. Zhang, Y. Gao, Y. Jin, X. Su, H. Zhi, Q. Guan, L. Pan, L. Jin, Kinect based objective evaluation of bradykinesia in patients with parkinson's disease, DIGITAL HEALTH 9 (2023) 205520762311766. doi:10.1177/ 20552076231176653.

[15] T. Liu, Y. Song, Y. Gu, A. Li, Human action recognition based on depth images from microsoft kinect, 2013, pp. 200–204. doi:10.1109/GCIS.2013.38.

[16] O. Tupa, A. Proch´azka, O. Vy˘sata, M. Sch¨atz, J. Mares, M. Valis, V. Ma˘r´ık, ˘ Motion tracking and gait feature estimation for recognising parkinson's disease using ms kinect, BioMedical Engineering OnLine 14 (2015). doi:10.1186/ s12938-015-0092-7.

[17] I. Pachoulakis, N. Xilourgos, N. Papadopoulos, A. Analyti, A kinect-based physiotherapy and assessment platform for parkinson's disease patients, Journal of Medical Engineering 2016 (2016) 1–8. doi:10.1155/2016/ 9413642

[18] L. Dranca, L. Mendarozketa, A. Goni, A. Illarramendi, I. Navalpotro, M. Delgado-Alvarado, M. Rodr´ıguez-Oroz, Using kinect to classify parkinson's disease stages related to severity of gait impairment, BMC Bioinformatics 19 (2018) 471. doi:10.1186/s12859-018-2488-4.

[19] I. Bortone, M. Quercia, N. Ieva, G. Cascarano, G. Trotta, S. Tat`o, V. Bevilacqua, Recognition and Severity Rating of Parkinson's Disease from Postural and Kinematic Features During Gait Analysis with Microsoft Kinect, 2018, pp. 613–618. doi:10.1007/978-3-319-95933-7_70.

[20] H. Tian, H. Li, W. Jiang, X. Ma, X. Li, H. Wu, Y. Li, Cross-spatiotemporal graph convolution networks for skeleton-based parkinsonian gait mds-updrs score estimation, IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society PP (2024). doi:10.1109/TNSRE.2024.3352004.

[21] A. Sabo, S. Mehdizadeh, A. Iaboni, B. Taati, Estimating parkinsonism severity in natural gait videos of older adults with dementia (05 2021).

[22] A. D. Miron, N. Sadawi, W. Ismail, H. Hussain, C. Grosan, Intellirehabds (IRDS) - A dataset of physical rehabilitation movements, Data 6 (5) (2021) 46. doi:10.3390/DATA6050046. URL https://doi.org/10.3390/data6050046.

[23] Y. Xiao, J. Chen, Y. Wang, Z.-G. Cao, J. Zhou, X. Bai, Action recognition for depth video using multi-view dynamic images, Information Sciences 480. doi:10.1016/j.ins.2018.12.050.

[24] A. Miko lajczyk-Bare la, M. Grochowski, Data augmentation for improving deep learning in image classification problem, 2018, pp. 117–122. doi: 10.1109/IIPHDW.2018.8388338.

[25] A. Morozov, D. Zgyatti, P. Popov, Equidistant and uniform data augmentation for 3d objects, IEEE Access PP (2021) 1–1. doi:10.1109/ACCESS. 2021.3138162.

[26] M.-H. Guo, T. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. Martin, M.-M. Cheng, S.-M. Hu, Attention mechanisms in computer vision: A survey, Computational Visual Media 8 (2022). doi:10.1007/s41095-022-0271-y.

[27] X. Li, F. Xu, F. Liu, X. Lyu, Y. Tong, Z. Xu, J. Zhou, A synergistical attention model for semantic segmentation of remote sensing images, IEEE Transactions on Geoscience and Remote Sensing PP (2023) 1–1. doi: 10.1109/TGRS.2023.3243954.

[28] K. Ting, Confusion Matrix, 2017, pp. 260–260. doi:10.1007/ 978-1-4899-7687-1_50.

Fatemeh Fadaie Ardestani
Dept.of Computer Engineering, AI and Robotics
University of Isfahan
Isfahan, Iran
Orcid: 0009-0005-5865-4407

Nima Asadi
Doctor of Philosophy - PhD, Computer Science
University of Maryland
College Park, Maryland, United States
Orcid: 0000-0002-5102-6927