InceptoFormer: A Multi-Signal Neural Framework for Parkinson's Disease Severity Evaluation from Gait

Safwen Naimi[†,*], Arij Said[†], Wassim Bouachir[†], Guillaume-Alexandre Bilodeau[‡]
[†] University of Quebec (TÉLUQ), Montreal, QC, Canada
[‡] Polytechnique Montréal, Montreal, QC, Canada

Abstract

We present *InceptoFormer*, a multi-signal neural framework designed for Parkinson's Disease (PD) severity evaluation via gait dynamics analysis. Our architecture introduces a 1D adaptation of the Inception model, which we refer to as Inception1D, along with a Transformer-based framework to stage PD severity according to the Hoehn and Yahr (H&Y) scale. The Inception1D component captures multi-scale temporal features by employing parallel 1D convolutional filters with varying kernel sizes, thereby extracting features across multiple temporal scales. The transformer component efficiently models long-range dependencies within gait sequences, providing a comprehensive understanding of both local and global patterns. To address the issue of class imbalance in PD severity staging, we propose a data structuring and preprocessing strategy based on oversampling to enhance the representation of underrepresented severity levels. The overall design enables to capture fine-grained temporal variations and global dynamics in gait signal, significantly improving classification performance for PD severity evaluation. Through extensive experimentation, *InceptoFormer* achieves an accuracy of 96.6%, outperforming existing state-of-the-art methods in PD severity assessment. The source code for our implementation is publicly available at https://github.com/SafwenNaimi/InceptoFormer.

Keywords: Inception1D, Transformers, Parkinson's disease staging, H&Y scale, VGRF Signals

1. Introduction

The prevalence of Parkinson's disease is rising globally with an estimated 10 million people currently living with the condition [1]. This neurodegenerative disorder primarily affects the brain regions responsible for coordinating movement, leading to symptoms such as tremors and difficulties in walking. While there is no cure for Parkinson's disease, various treatments are available to control its symptoms. To better analyze the condition, gait analysis is commonly used to detect any irregularities in movement. This analysis is reliant on assessing a range of clinical symptoms and signs. These factors can pose challenges since they may overlap with other neurological disorders. For that, the extraction of gait information using foot sensors is essential to understand our movement patterns and identify abnormalities. To classify Parkinson's disease into five distinct stages based on symptom severity, the Hoehn and Yahr (H&Y) scale [2] is commonly employed. Another frequently utilized method for assessing symptom severity is the Unified Parkinson's Disease Rating Scale (UPDRS) [3]. Both of these scales play crucial roles in clinical settings and research for tracking disease progression and evaluating the effectiveness of treatments.

Despite advancements in gait-based diagnostic techniques, assessing the severity of the disease remains a significant challenge. Several deep learning-based techniques have been proposed for detecting Parkinson's disease (PD) from gait data, and have produced encouraging results [4–6]. However, these methods are used for binary classification to detect PD based on gait patterns. The classification of PD severity into specific stages is less explored by the research community. Most of the methods use the H&Y criteria derived from the

[*] safwen.naimi@teluq.ca

publicly available Physionet gait dataset [7]. Despite the good results obtained from the studies focusing on PD staging [8, 9], they often face the problem of data imbalance, which causes a performance drop. This imbalance arises because certain severity stages, particularly the more advanced ones, are generally underrepresented, leading to biased learning and reduced classification performance. Models trained on such imbalanced data tend to overfit to the majority classes while struggling to correctly classify the minority classes, ultimately affecting their generalization capability.

In this work, we introduce *InceptoFormer*, a Multi-Signal neural framework specifically designed for PD staging based on gait dynamics. Our approach not only improves classification accuracy but also ensures robust performance across imbalanced classes of gait data. By integrating multi-signal temporal feature extraction with attention mechanisms, *InceptoFormer* effectively captures the complex temporal and spatial patterns in gait data, while employing an oversampling strategy to enhance the representation of minority classes.

2. Related Work

Various deep-learning approaches have been applied to assess the severity stages of Parkinson's disease by analyzing gait data. In particular, the Vertical Ground Reaction Force (VGRF) signal is widely used, as it has been proven to be a crucial and distinguishing kinematic feature for detecting and evaluating Parkinson's disease stages [10]. Ertugrul et al. [11] introduced an algorithm utilizing shifted 1D local binary patterns (1D-LBP) in conjunction with machine learning classifiers. Their approach involved applying a shifted 1D-LBP to construct 18 histograms of the corresponding patterns. Zhao et al. [12] developed an algorithm featuring two parallel networks, a 2D Convolutional Neural Network (2D-ConvNet) to analyze the spatial distribution of forces, and a recurrent neural network (RNN) to examine temporal distributions. The final classification was determined by averaging the outputs from both networks. Aşuroğlu et al. [13] explored a Perceiver-based multimodal model for predicting UPDRS scores from VGRF gait data. The Perceiver architecture leverages self-attention mechanisms to process sequences of varying lengths and complexities, significantly enhancing performance in severity assessment for Parkinson's disease. Balaji et al. [14] proposed a correlation-based feature extraction method. Biomarkers were extracted from spatiotemporal VGRF gait data using correlation. They employed four supervised machine learning algorithms K-nearest neighbors (KNN), Naive Bayes (NB), Ensemble classifier (EC), and Support Vector Machine (SVM) to classify the severity of PD based on the Hoehn and Yahr (H&Y) scale. Naimi et al. [15] introduced a hybrid ConvNet-Transformer model capable of capturing both spatial and temporal features. Their model is designed for both detection and severity evaluation of PD. Veeraragavan et al.[16] introduced an approach that uses feedforward neural networks (FNN) to classify severity levels. Mirelman et al. [17] employed random forest classifiers and support vector machines (SVM) to differentiate between stages of Parkinson's disease, identifying key features correlated with disease severity.

While existing methods have shown good performances in terms of classification accuracy for detecting and staging PD, they face two main challenges that hinder their overall performance: class imbalance and limitations in capturing the complex patterns of physiological data, mainly due to architectural constraints. These issues contribute to a noticeable drop in model performance. Our approach addresses these issues by integrating multi-signal feature extraction through Inception1D blocks to capture fine-grained temporal variations and by employing Transformer-based encoders to model both long-range dependencies and spatial correlations in gait dynamics. The proposed design mitigates class imbalance through a tailored data structuring and preprocessing strategy. It also enhances the ability to learn complex physiological patterns, ultimately leading to improved performance in PD severity evaluation.

3. Method

Our approach focuses on identifying the severity stage of Parkinson's disease (PD) based on the Hoehn and Yahr (H&Y) staging scale, which classifies symptoms into five stages. However, the Physionet gait dataset [7], used in most existing studies including ours, includes only four main stages: Stage 0 (healthy), Stage 1 (mild, severity 2), Stage 2 (moderate, severity 2.5), and Stage 3 (severe, severity 3). These stages are derived from gait signals collected through foot sensors attached to patients. Given the inherent complexity and variability of gait data, accurately distinguishing between these stages remains challenging, particularly in the presence of class imbalance due to the scarcity of data on advanced stages of the disease. The following sections provide a detailed presentation of our approach, including the data structuring and preprocessing strategy to deal with data imbalance and *InceptoFormer* architecture for improving PD severity evaluation.

3.1. Data Structuring and Preprocessing

The Physionet dataset [7] used in our study presents multiple 1D VGRF signals captured from patients' walks and measured using 18-foot sensors. We started by dividing the walks into small segments of 100-time steps with 50% overlap composed of groups of elements. This segmentation not only increases the amount of training data but also preserves the temporal continuity of gait patterns, enhances feature extraction by capturing finer motion variations, and ensures a more robust representation of the gait dynamics across different severity levels. For each segment, we assign a specific label or category that identifies it. This is essential for time series data, as it helps capturing temporal patterns. A common issue in the dataset used in our study is the imbalance between the four classes of the H&Y scale as depicted in Figure 1. For that, we employed an oversampling strategy of the minority class using the Synthetic Minority Over-sampling Technique [18] to balance the classes. It is a preprocessing technique used to address the class imbalance by creating synthetic samples. We started by selecting the K-nearest neighbors of every sample that belongs to a minority class, then we generate samples along the line segments joining the minority class sample

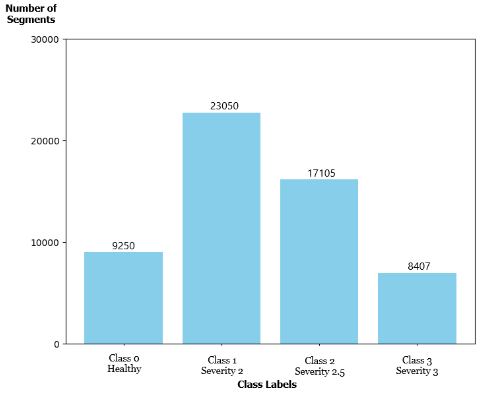

Figure 1. Initial Class distribution: Before applying our data structuring and preprocessing strategy

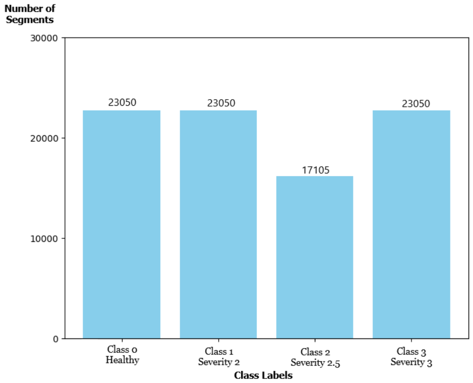

Figure 2. Final Class distribution: After applying our data structuring and preprocessing strategy

to its nearest neighbors. The minority classes, class 0 and class 3 have been oversampled while majority classes 1 and 2 maintain a stable number of samples, as they are sufficiently represented and not considered minority classes. Given the two minority class samples class 0 and class 3, respectively x_i and x_j, the synthetic sample x_{new} for each one is generated as follows:

$$x_{new} = x_i + \lambda \cdot (x_j - x_i) \tag{3.1}$$

where x_i and x_j are two randomly chosen samples from the minority class, and λ is a random number sampled from a uniform distribution: $\lambda \sim U(0,1)$.

For the minority classes 0 and 3, N_{c_i} the number of samples for class c_i and $N_{new,\, c_i}$ the number of synthetic samples generated for class c_i, the synthetic samples are generated as follows:

$$N_{new,\, c_i} = N_{majority} - N_{c_i}, \quad c_i \in \{0,3\} \tag{3.2}$$

where $N_{majority}$ is the number of samples in the majority class class 1. The oversampling ensures that the number of samples in classes 0 and 3 equals that of the majority class:

$$N_{c_i} + N_{new,\, c_i} = N_{majority}, \quad c_i \in \{0,3\} \tag{3.3}$$

For classes 1 and 2, the sample count remains unchanged:

$$N_{new,\, c_i} = 0, \quad c_i \in \{1,2\} \tag{3.4}$$

Figures 1 and 2 illustrate the class distribution before and after applying our data structuring and preprocessing strategy to the Physionet gait dataset.

3.2. **InceptoFormer Architecture**

The detailed architecture of *InceptoFormer* is provided in Figure 3. This model is designed to predict the severity of Parkinson's disease based on gait analysis. Since we are

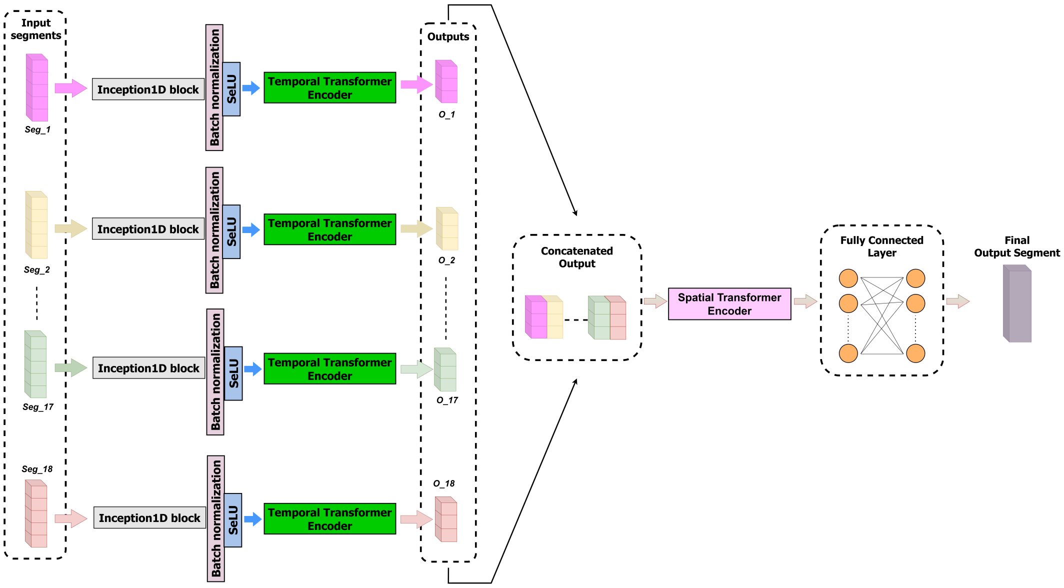

Figure 3. The proposed InceptoFormer architecture. We have 18 signals captured from the sensors. To fully leverage the information in each signal, we process them independently using 18 parallel Inception1D blocks. The outputs are then concatenated and passed to a temporal transformer, and then we applied a spatial transformer encoder. In the end, a classifier block is used to generate the final classification.

processing data from 18 sensors, we have 18 distinct signals, each capturing different aspects of the patient's foot movements during a walk. To fully explore the information in each signal, we process them independently using 18 parallel Inception1D blocks to capture dependencies in different scales and extract unique features which are subsequently concatenated together to give a better description of the data. Following this stage, each feature stream is processed by a temporal transformer encoder, which models long-range dependencies within each gait sequence. This component enables the model to capture the progression of movement patterns over time, ensuring a more comprehensive understanding of gait dynamics. Before further processing, the outputs of these temporal transformers undergo dimensionality reduction to optimize computational efficiency. The reduced feature representations from all 18 signals are subsequently concatenated to form a unified feature vector, which serves as the input to the spatial transformer encoder. It is responsible for capturing spatial dependencies between the sensors, learning how different regions of the foot interact during movement. By leveraging self-attention mechanisms, it effectively identifies correlations between sensor signals, refining the extracted feature representation. The final stage consists of a classifier block used to predict the PD stage and generate the final classification. The following sections provide a more detailed breakdown of each *InceptoFormer* component, highlighting their individual contributions to the overall architecture.

3.2.1. **Inception1D Block**

We adapted the inception architecture which was initially designed for image processing for 1D data (see Figure 4). It introduces a new version of the inception module that can be easily integrated with our 1D VGRF signals. It consists of three convolutional streams in parallel in which the previous layer is common to all of them. The first stream applies a 1D convolution with 32 filters and a kernel size of 1 (K=1). The second stream uses a

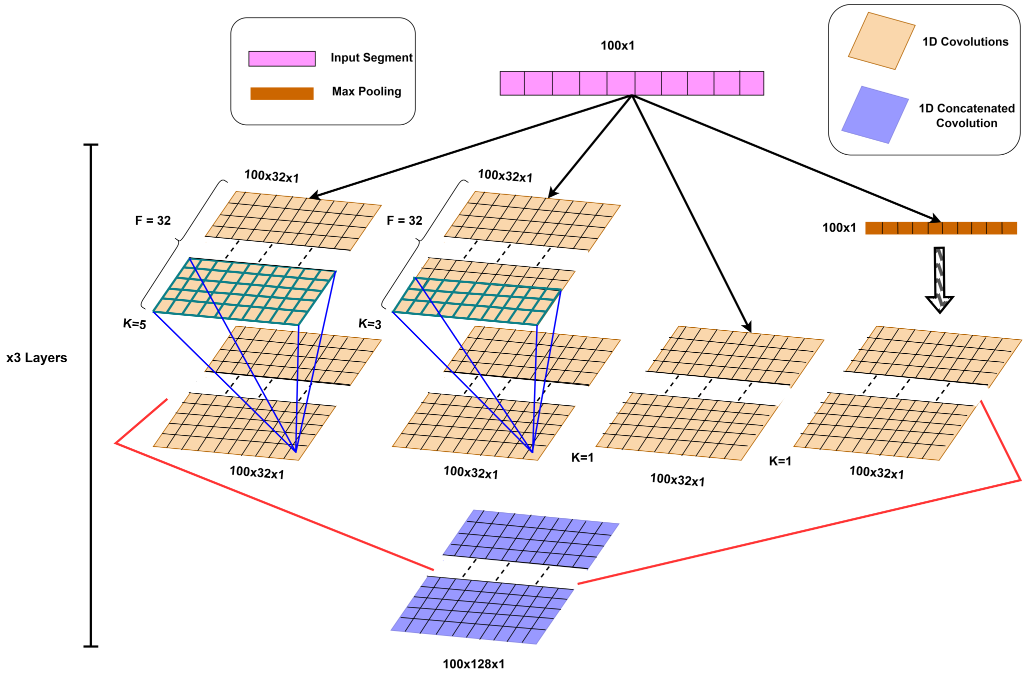

Figure 4. Architecture of the Inception1D block

convolutional layer with kernel size K=3, while the third stream employs a larger kernel size K=5, both having 32 filters as well. These streams allow for capturing features at various scales, with the smaller kernel sizes focusing on more localized patterns and the larger kernels extracting regional patterns. The convolutional layers in the Inception1D module can be summarized as follows:

$$y_k(t) = f\left(\sum_{i=0}^{k-1} W_k(i) \cdot x + b_k\right) \tag{3.5}$$

with x as the input segment, i indexes the positions in the kernel, $W_k(i)$ represents the weights of the kernel of size k, b_k is the bias term for the convolution with kernel size and $f(\cdot)$ is the SeLU activation function applied element-wise after convolution.

The 3 streams later converge to form a single output by concatenating their outputs. This output is presented as:

$$y_{inception} = [y_1, y_3, y_5] \tag{3.6}$$

We follow the output with Batch Normalization to normalize the output from the module, improving the stability and speed of training, and an activation function is applied to introduce non-linearity. In this way, both local and regional information were extracted on the same features map. We repeated the inception architecture 3 times in cascade to achieve a deeper network, which made it possible to extract information from increasingly larger regions.

3.2.2. **Temporal Transformer Block**

In *InceptoFormer*, we implemented a temporal transformer encoder to minimize intraclass variance and capture long-range dependencies in the data. This encoder plays a key role in our architecture by encoding the input sequence and producing representations that reflect the underlying dependencies within the data [19]. This temporal transformer block is implemented after each Inception1D Block. It includes a multi-head attention layer with two heads, followed by a feed-forward network, mirroring the architecture used in BERT [20] for natural language processing.

We applied a fixed positional encoding with a constant step size corresponding to the segment length since the gait data is segmented into fixed-length and constant intervals. The use of a fixed positional encoder enabled us to effectively capture the temporal patterns within these segments by maintaining the temporal ordering of the data. Furthermore, we applied normalization to the positional encoding to ensure the original vector information is not entirely masked.

3.2.3. **Spatial Transformer Block**

The outputs from the 18 parallel temporal transformer encoders are concatenated. This step aids in creating a more compact representation of the data and eliminates redundant information. The resulting concatenated vector is then fed into the spatial transformer encoder block. A fixed positional encoding is also applied here, providing the spatial transformer encoder with details about the relative positions of elements within the concatenated vector, maintaining the spatial information of the sensors with respect to each other. Same as the temporal transformer encoder block, the spatial transformer includes a multi-head attention layer with two heads and a feed-forward network.

The primary function of the spatial transformer is to identify dependencies between the sensors by considering their spatial arrangement on the foot and potentially uncovering correlations among them.

3.2.4. Classifier Block

To predict the stage of Parkinson's Disease, we employed a classifier consisting of two fully connected layers and an output layer, which serve as the final part of our *Incepto-Former*. This block receives the output from the spatial transformer encoder and generates a probability distribution across the PD severity levels. The classifier was trained using the categorical cross-entropy loss function to adjust the weights and biases of the model. The resulting probabilities were then used to determine the predicted PD stage for each input signal.

4. Experiments

This section outlines the experiments conducted to evaluate the performance of our approach. We describe the dataset, the evaluation metrics, and the training details employed. Additionally, we discuss the results and illustrate the ablation study of our method.

4.1. Dataset Description

We used the Physionet gait dataset [7] which contains multiple 1D VGRF signals recorded from patients' walks using 18 foot-mounted sensors. The dataset includes gait measurements from 93 patients with Parkinson's disease (PD) and 73 healthy individuals. It was assembled by three research teams. The first data was collected by Yogev et al. [21], it contains a gait cycle for normal walking on a leveled surface. The second data was reported by Hausdorff et al. [22], it contains the gait cycle for walking at a casual speed with RAS, and the final data was collected by Toledo et al. [23], it contains time series data of a subject for walking on a treadmill. This database includes demographic information, measures of severity rating scale such as the Hoehn & Yahr scale, UPDRS scale, and other related measures.

4.2. Evaluation Metrics

We evaluated our method using 10-fold cross-validation. We divided the Parkinsonian (Pd) and control (Co) groups into 10 folds, ensuring that each fold maintained the same dataset balance (70% for Pd and 30% for Co). For the evaluation metrics, we used the following notations: TP for the true positives, TN for the true negatives, FP for the false positives, and FN for the false negatives. Our method was assessed using precision, recall, F1-score, and accuracy. The utilized metrics equations are given below.

$$\textbf{Precision: } Pr = \frac{TP}{TP + FP} \tag{4.1}$$

$$\textbf{Recall: } Re = \frac{TP}{TP + FN} \tag{4.2}$$

$$\textbf{F1-Score } = 2 \times \frac{Pr \times Re}{Pr + Re} \tag{4.3}$$

$$\textbf{Accuracy (\%): } Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{4.4}$$

4.3. Training Details

The proposed approach was trained and tested utilizing a batch size of 64 samples for each iteration. All fully connected layers, except for the output layer, employed the SeLU activation function. The model is trained using the Nadam stochastic optimization method [24] which adjusts learning rates for each parameter, making it a robust optimizer for our approach. The learning rate is set to $\eta = 10^{-4}$. We combine the Nadam optimizer with the

Adam optimizer and Nesterov momentum. The update rule for the weights w in Nadam is given by:

$$w_{t+1} = w_t - \eta \cdot \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \right) \tag{4.5}$$

where η is the learning rate, $\hat{m}_t$ is the moving average of the gradients, $\hat{v}_t$ is the moving average of the squared gradients, and ϵ is a small constant to avoid division by zero. To enhance the performance of the *InceptoFormer* and mitigate overfitting, we implemented a dropout rate of 0.2 and employed an early stopping based on the validation loss.

4.4. **Results**

Table 1 presents a comparative analysis between our proposed approach and several existing methods. Through an examination of the average metric variations across these methods, we show that *InceptoFormer* outperforms the others in terms of precision, recall, and F1-score, achieving a superior final accuracy of 96.6%. This performance is largely attributed to the integration of Inception1D, which enables the model to capture critical multi-scale temporal features, and the Transformer-based design, which effectively models long-term dependencies and intricate gait dynamics. Furthermore, a key factor contributing to this performance gain is our data structuring and preprocessing strategy. By addressing the severe class imbalance inherent in the Physionet dataset, we ensure a better representation of all severity levels, preventing the model from being biased toward the majority classes. Table 1 demonstrates the effectiveness of our data structuring and preprocessing strategy. We retrained the method marked with * using this strategy under the same settings as reported in its corresponding paper. Notably, the application of our strategy led to a measurable performance improvement as seen in the enhanced accuracy, precision, recall, and F1-score. This further highlights the necessity of addressing the class imbalance in PD severity evaluation and underscores the robustness of our proposed approach in PD staging compared to other methods on the same Physionet dataset.

Figure 5 illustrates the average confusion matrix across all 10 folds for our proposed approach. We achieved a 97% accuracy in classifying instances of Healthy patients. For severity 2 cases, 97% were accurately predicted, with minor misclassifications into severity 2.5 and severity 3. The model demonstrated good performance in classifying severity 2.5 cases, correctly predicting 98%, with a slight tendency to misclassify them as severity 2. In the case of severity 3 patients, the model accurately classified 94% of instances. However, this category experienced a notable degree of misclassification into severity 2.

Table 1. Comparison of our proposed architecture with the state-of-the-art methods for PD severity evaluation. Method marked with * indicates that we retrained it using the SMOTE technique and the same settings as reported in the corresponding paper. The best results are in bold and the second best results are in italic red.

Method	Accuracy	Precision	Recall	F1-Score
OF-DDNet [25]	84.00%	*90.25%*	86.00%	—
1D-ConvNet [9]	85.30%	89,48%	82.68%	85.04%
Feed Forward Network [16]	86.50%	87.73%	87.55%	87.67%
1D-Convolutional Transformer [15]	88.00%	87.25%	85.25%	85.50%
1D-Convolutional Transformer * [15]	*89.12%*	89.52%	*89.25%*	*88.11%*
InceptoFormer (Ours)	**96.60%**	**93.97%**	**94.15%**	**93.60%**

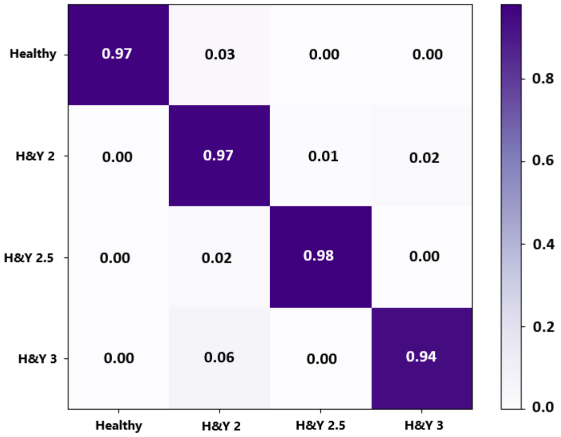

Figure 5. Confusion matrix of InceptoFormer

4.5. **Ablation Study**

To evaluate the contribution of each key component in *InceptoFormer*, we conducted ablation experiments on the Physionet gait dataset to assess their impact on the final performance in terms of accuracy. A detailed analysis is provided in Table 2. In Model 1, we removed both the Temporal and Spatial Transformer blocks, retaining only the Inception1D module. This configuration resulted in a significant drop in performance, with accuracy decreasing by 18.26%, precision by 13.82%, recall by 13.33%, and the F1-Score by 17.83%. These results underscore the critical role played by the Temporal and Spatial Transformer blocks in capturing the temporal and spatial dependencies that are essential for recognizing gait patterns. The absence of these attention mechanisms leaves the model reliant solely on multi-scale feature extraction from the Inception1D component, which, while useful, is insufficient to capture complex temporal relationships, leading to reduced overall performance. Conversely, in Model 2, we excluded the Inception1D module and retained the Temporal and Spatial Transformer blocks. The model performance also degraded but to a lesser extent than in Model 1, with accuracy decreasing by 7.43%, precision by 6.88%, recall by 1.27%, and the F1-Score by 5.98%. These results highlight the importance of the Inception1D component, which enhances the model capacity to extract multi-scale features from the input data, thereby improving the model ability to differentiate between subtle variations in gait patterns. However, the Transformer blocks were able to compensate for the absence of multi-scale features to some degree, emphasizing their strong contribution to the overall effectiveness of our model. Finally, Model 3, representing our *InceptoFormer* that integrates both the Temporal and Spatial Transformer blocks alongside the Inception1D module, achieves the best performance across all metrics. The combination of these components leads to the highest accuracy (96.6%), precision (93.97%), recall (94.15%), and F1-Score (93.6%), demonstrating that the synergy between multi-scale feature extraction and attention-based temporal-spatial modeling is essential for superior performance in gait analysis tasks. This comprehensive integration allows our *InceptoFormer* to capture both local and global dependencies, offering a more nuanced and complete understanding of gait dynamics.

Table 2. Ablation study results on the effect of each component of our InceptoFormer.

Variations	Inception1D	Temporal and Spatial Transformers	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Model 1	✓	✗	78,34 (↓18.26)	80.15 (↓13.82)	80.82 (↓13.33)	75.77 (↓17.83)
Model 2	✗	✓	89.17 (↓7.43)	87.09 (↓6.88)	92.88 (↓1.27)	87.62 (↓5.98)
Model 3	✓	✓	**96.6**	**93.97**	**94.15**	**93.6**

5. Conclusions

In this work, we introduced a multi-signal neural framework for Parkinson's disease severity evaluation using gait dynamics, integrating Inception1D for multi-scale feature extraction and transformer encoders for temporal and spatial modeling. We proposed a data structuring and preprocessing strategy to address class imbalance and improve classification robustness. Our model achieved 96.6% accuracy on the Physionet dataset, surpassing existing methods. Our findings highlight the potential of our framework for fine-grained Parkinson's disease staging, which could be expanded to other clinical applications.

References

[1] E. R. Dorsey, R. Constantinescu, R. Constantinescu, J. P. Thompson, K. M. Biglan, R. G. Holloway, K. K. Kieburtz, F. J. Marshall, B. M. Ravina, G. Schifitto, A. Siderowf, and C. M. Tanner. "Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030". In: *Neurology* 68 (2007), pp. 384 –386.

[2] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. E. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning, M. D. Yahr, and L. Seidl. "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations The Movement Disorder Society Task Force on rating scales for Parkinson's disease". In: *Movement Disorders* 19 (2004).

[3] L. M. Shulman, A. L. Gruber-Baldini, K. E. Anderson, P. Fishman, S. G. Reich, and W. J. Weiner. "The clinically important difference on the unified Parkinson's disease rating scale." In: *Archives of neurology* 67 1 (2010), pp. 64–70.

[4] N. P. Narendra, B. W. Schuller, and P. Alku. "The Detection of Parkinson's Disease From Speech Using Voice Source Information". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), pp. 1925–1936.

[5] C. Quan, K. Ren, and Z. Luo. "A Deep Learning Based Method for Parkinson's Disease Detection Using Dynamic Features of Speech". In: *IEEE Access* 9 (2021), pp. 10239–10252.

[6] Ö. F. Ertuğrul, Y. Kaya, R. Tekin, and M. N. Almalı. "Detection of Parkinson's disease by Shifted One Dimensional Local Binary Patterns from gait". In: *Expert Syst. Appl.* 56 (2016), pp. 156–163.

[7] *Physionet Dataset.* `https://www.physionet.org/content/gaitpdb/1.0.0/`.

[8] S. Naimi, W. Bouachir, and G.-A. Bilodeau. "HCT: Hybrid Convnet-Transformer for Parkinson's Disease Detection and Severity Prediction from Gait". In: *2023 International Conference on Machine Learning and Applications (ICMLA)* (2023), pp. 814–819.

[9] I. E. Maachi, G.-A. Bilodeau, and W. Bouachir. "Deep 1D-ConvNet for accurate Parkinson disease detection and severity prediction from gait". In: *Expert Syst.App* 143 (2019).

[10] Y. Guo, J. Yang, Y. Liu, X. Chen, and G.-Z. Yang. "Detection and assessment of Parkinson's disease based on gait analysis: A survey". In: *Frontiers in Aging Neuroscience* 14 (2022).

[11] F. Ertugrul, Y. Kaya, R. Tekin, and M. N. Almali. "Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait". In: *Expert Syst.Appl* 56 (2016), pp. 156–163.

[12] A. Zhao, L. Qi, J. Li, J. Dong, and H. Yu. "A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data". In: *Neurocomputing* 315 (2018), pp. 1–8.

[13] T. Aşuroğlu and H. Oğul. “A deep learning approach for Parkinson’s disease severity assessment”. In: *Østfold University College, Norway* (2022).

[14] E. Balaji, D. Brindha, V. K. Elumalai, and K. Umesh. “Data-driven gait analysis for diagnosis and severity rating of Parkinson’s disease”. In: *Department of Biomedical Engineering, PSG College of Technology and School of Electrical Engineering, VIT* (2020).

[15] S. Naimi, W. Bouachir, and G.-A. Bilodeau. “1D-Convolutional Transformer for Parkinson disease diagnosis from gait”. In: *Neural Computing and Applications* (2023).

[16] S. Veeraragavan, A. A. Gopalai, D. Gouwanda, and S. A. Ahmad. “Parkinson’s disease diagnosis and severity assessment using ground reaction forces and neural networks”. In: *Frontiers in Physiology* 11 (2020).

[17] A. Mirelman et al. “Detecting sensitive mobility features for Parkinson’s disease stages via machine learning”. In: *Movement Disorders* 36 (2021).

[18] N. Chawla, K. Bowyer, L. O. Hall, and W. P. Kegelmeyer. “SMOTE: Synthetic Minority Over-sampling Technique”. In: *ArXiv* abs/1106.1813 (2002).

[19] S. Lohit, Q. Wang, and P. K. Turaga. “Temporal Transformer Networks: Joint Learning of Invariant and Discriminative Time Warping”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 12418–12427.

[20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”. In: *North American Chapter of the Association for Computational Linguistics.* 2019.

[21] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon, and J. M. Hausdorff. “Dual tasking, gait rhythmicity, and Parkinson’s disease: Which aspects of gait are attention demanding?” In: *European Journal of Neuroscience* 22 (2005).

[22] J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi. “Rhythmic auditory stimulation modulates gait variability in Parkinson’s disease”. In: *European Journal of Neuroscience* 26 (2007).

[23] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff. “Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson’s disease”. In: *Movement Disorders* 20 (2005).

[24] T. Dozat. “Incorporating Nesterov Momentum into Adam”. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016).* 2016.

[25] M. Lu, Q. Wang, P. Duh, and J. Ponce. “Vision-based estimation of MDS-UPDRS gait scores for assessing Parkinson’s disease motor severity”. In: *arXiv preprint arXiv:2007.08920* (2020).