KU LEUVEN

AI & HEALTH

18/05/2024

Iñigo Aduna, KU Leuven (Brugge)
Ahmad Beigrezaei, KU Leuven (Brugge)

Inigo.adunaalonso@student.kuleuven.be

ahmad.beigrezaei@student.kuleuven.be

Master of Artificial Intelligence in Business and Industry

Applied AI – Academic Perspectives: AI & Health

KU LEUVEN

Contents

KU LEUVEN

1. Introduction

1.1. Freezing of Gait (FOG)

Freezing of Gait (FOG) is a common and debilitating symptom of Parkinson's disease, defined as "a brief, episodic absence or marked reduction of forward progression of the feet despite the intention to walk" (Nutt et al., 2011). FOG severely impacts the quality of life, as it can lead to falls and a loss of independence. The manifestation of FOG varies among patients but typically includes shuffling with small steps, trembling in place, or complete akinesia, where movement ceases entirely.

1.2. State of the Art in FOG Monitoring and Analysis

Current approaches to monitoring FOG involve the use of **inertial measurement units** (IMUs), which are placed on various parts of the patient's body. Typically, six IMUs are utilized, located on the chest, lumbar region, left ankle, right ankle, left foot, and right foot. Each IMU records data from three gyroscopes (θ, ϕ, φ) and three accelerometers (a_x, a_y, a_z) capturing detailed movement patterns during episodes of FOG.

The collected data are primarily used by physiotherapists to analyse the patient's gait and to adjust therapeutic interventions accordingly. Furthermore, neurologists employ this data to assess the effectiveness of ongoing treatments in alleviating the symptoms of Parkinson's disease. Through precise monitoring, clinicians can observe the progression of the disease and make informed decisions about treatment modifications.

1.3. Economic and Practical Challenges of Current FOG Monitoring Techniques

Despite the sophistication of current monitoring systems, they present several challenges that limit their widespread adoption. Firstly, the setup involving multiple sensors is complex and time-consuming, requiring professional assistance. Secondly, the cost associated with acquiring and maintaining this equipment is significant, making it inaccessible for many healthcare settings, particularly in low-resource environments. Lastly, the analysis of the vast amounts of data generated by these sensors demands considerable expertise and time from specialized personnel, further escalating the operational costs.

1.4. The Role of AI in Optimizing FOG Management

The integration of Artificial Intelligence (AI) into the management of FOG presents a promising alternative to traditional methods. AI can automate the processing and analysis of sensor data, significantly reducing the time and expertise required to interpret complex datasets. By employing machine learning algorithms, AI systems can predict FOG episodes in real-time, providing instant feedback to patients and clinicians. This capability not only enhances the efficiency of interventions but also allows for personalized treatment adjustments based on the patient's immediate needs.

Moreover, AI-driven tools can be developed to be more cost-effective, with algorithms optimized to run on less expensive hardware or even integrated directly into consumer-grade smartphones and wearable devices. This democratization of technology could significantly

broaden access to advanced FOG management tools, making it feasible for routine use in diverse healthcare settings.

In conclusion, while current methods provide valuable insights into the mechanics and treatment of FOG, they are hampered by high costs and operational complexities. Artificial Intelligence offers a transformative potential to overcome these barriers, enhancing both the efficacy and accessibility of FOG management in Parkinson's disease.

2. Data Availability

For this project, data were collected from seven patients performing various types of activities, each equipped with six sensors located on the chest, lumbar region, left and right ankles, and left and right feet. The dataset comprises 2,554 files, each containing 256 rows and 36 columns. Each file represents a two-second window and is named in the format "pt{1,2,3,4,5,6,7} *visit_24_tbc_walklr_0_trial_2.0*{PD/FOG}__input.csv". The file names encode critical information: the patient number and the event type, where 'FOG' denotes a Freezing of Gait event and 'PD' indicates a normal event associated with Parkinson's Disease.

Each file corresponds to a 2-second interval, and with 256-time steps per file, the data sampling rate is determined to be 128Hz. At each time step, there are $36\ (0-35)$ values corresponding to the outputs from three gyroscopes and three accelerometers across the six sensors. These sensors record data from the specified body locations during the activity.

Columns	IMU Location
0-5	Chest
6-11	Lumbar
12-17	Left Ankle
18-23	Right Ankle
24-29	Left Foot
30-35	Right Foot

Each file is labelled either 'FOG=1' for a Freezing of Gait event or 'PD=0' for a non-FOG event. Models will be trained on this data, taking an input shape of $(256, N)$, where N is the number of sensors considered, to predict a binary output indicating either a FOG event (1) or a non-FOG event (0).

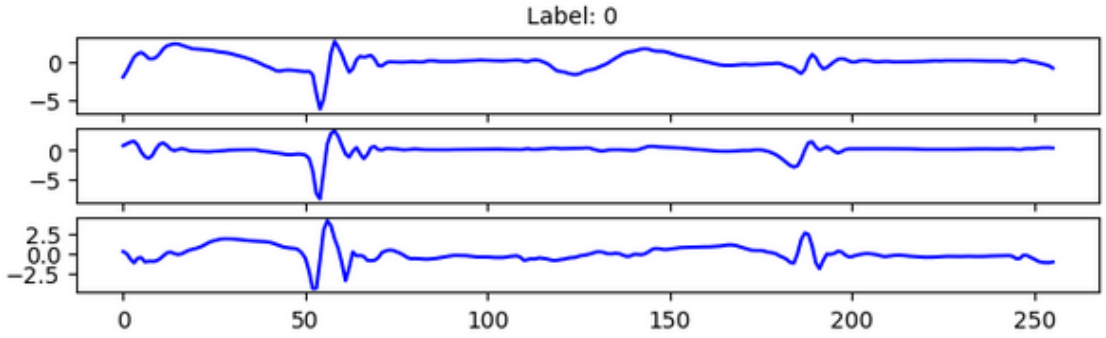
In conclusion, when we use the signals from all the sensors, X takes a shape of (2554, 256, 36) and Y would be in a shape of (2554, 1).

3. First Approach

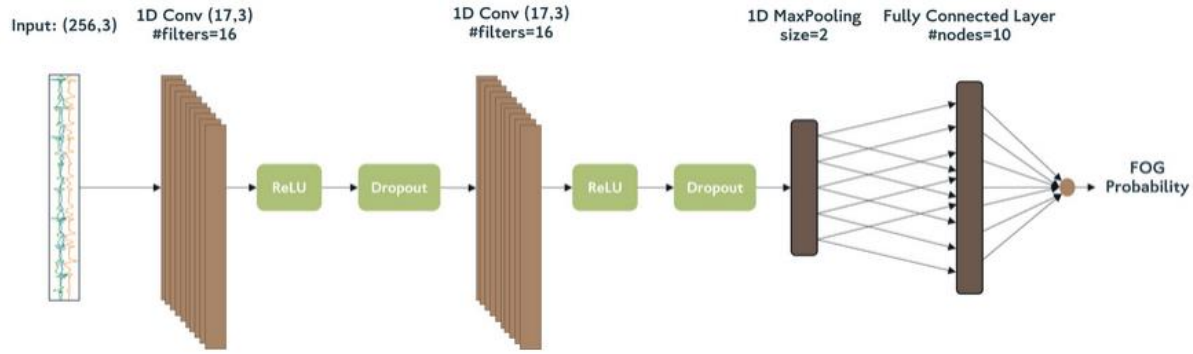
3.1. Train-Test 80/20 Split

In this initial study, we only consider data from three accelerometers on the IMU located at the lumbar level, specifically columns $(6-7-8)$ from all CSV documents. The data are loaded into variables X and Y, resulting in a dataset shape of $(2554, 256, 3)$ and label shape of $(2554, 1)$. We observe an imbalance in the dataset, with only 645 out of 2554 events, or 25.25%, experiencing Freezing of Gait. Next, we normalize the data to achieve zero mean and unit

variance using the formula $x' = \frac{x-\mu}{\sigma}$, where μ is the mean across all documents for a sensor, and σ is the standard deviation. The following figure displays the values of the accelerometer signals in the lumbar IMU from the first document.

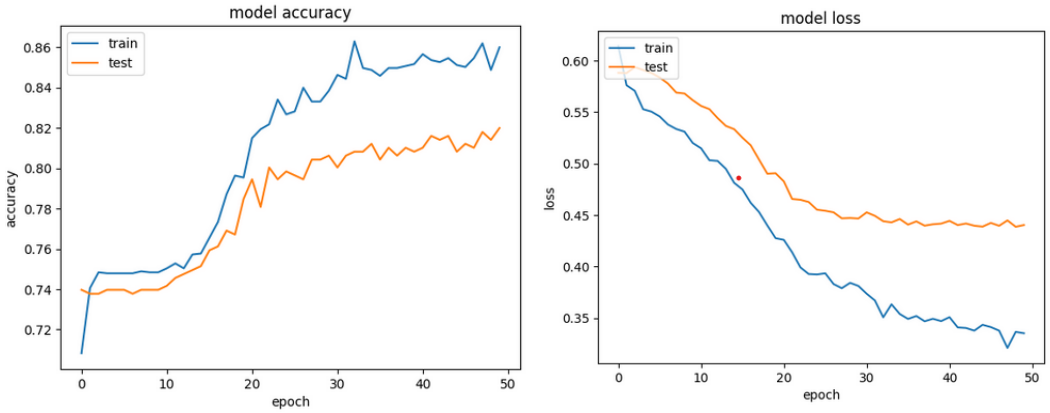


Normalized signals from the x, y, z accelerometers of the lumbar IMU sensor in the first patient during the first 2 seconds of activity are shown. We will train a CNN model as proposed in the following paper (O'Day et all.) with the following architecture:



The image reveals an error in the second convolution where the filters are $(17{,}16)$ as it is a 1D convolution. The input X represents accelerometer values over each 2-second window, consisting of 256-time steps with 3 samples each. This is followed by a 1D convolution layer with 16 filters of size (17x3), a stride of one, and zero padding, resulting in activations to the next layer of $((256-17)/1+1=240)x16$. A nonlinear activation unit and a dropout with a probability of 0.5 are then applied to prevent overfitting. Another 1D convolution layer follows, again with 16 filters of size $17x16$, yielding an output of $((240-17)/1+1=224)x16$. Another nonlinear activation and a dropout are applied to avoid overfitting. A Maxpooling layer with size 2 and stride 2 reduces the dimension by half, resulting in a linearized output of $(112x16=1792,1)$. We add a fully connected layer with 10 nodes, which connect to a single neuron followed by a sigmoid activation function to yield a result between 0 and 1. This model has $23{,}141$ parameters, with 832 in the first convolution $(16x3x17\ +\ 16)$, 4368 in the second convolution $(16x16x17\ +\ 16)$, and 17,930 in the fully connected layer $(1792x10\ +\ 10)$.

For this initial approach, we separate the training and validation data in an $80-20$ ratio, using 2043 files for training and 511 for validation. The model training uses an Adam optimizer with a learning rate of 0.0001, binary cross entropy for the error function, and accuracy metric to evaluate model performance. The model is trained over 50 epochs with a batch size of 32.



In this figure, we observe the training and testing graphs of the model. Both graphs show that the model is learning during training; as the epochs progress, that is, as the model iterates all the data points more times, we see an increase in accuracy and a decrease in loss.

We can observe that the model is not overfitting, but we would have expected the gap between the training and validation sets to be lower. In the final stretch, we see that despite the training dataset having a steep slope indicating that its accuracy is increasing rapidly, the test dataset has a much smaller slope, close to zero. This may indicate that if we continue increasing the epochs, overfitting could occur, and the model's generalization performance may decrease.

However, given the data imbalance, we consider accuracy not to be a suitable metric in this case. As previously mentioned, only 33.8% of the cases are FOG, while 66.2% are NON-FOG. A model that predicts everything as NON-FOG would achieve an accuracy of 66.2%, yet it would not be useful. Therefore, we will use Recall and Precision instead.

Precision, also known as Specificity or Positive Predictive Value, is defined as $P = TP/(TP + FP)$. It indicates how many of the windows labelled as FOG events are FOG events. Recall, or Sensitivity, is defined as $R = TP/(TP + FN)$, representing how many FOG events are correctly classified as FOG by our model. Combining both, we define the F1 score as $F1 = 2 \times (P \times R)/(P + R)$, a much more suitable metric in the case of unbalanced data.

Classification Report				
	Precision	**Recall**	**F1 score**	**Support**
0.0	0.84	0.94	0.88	378
1.0	0.73	0.49	0.59	133
Accuracy			0.82	511
Micro avg.	0.78	0.71	0.74	511
Weighted avg.	0.81	0.82	0.81	511

The classification report includes several key metrics: micro average, weighted average, and support. The micro average aggregates the contributions of all classes to compute the average metric, offering a global view of performance. The weighted average takes the average of the metrics for each class, weighted by the number of instances in each class, thus reflecting the model's performance in relation to class distribution. Support indicates the number of actual occurrences of each class in the validation data, providing context for the other metrics.

In the given table, the model demonstrates a precision of 0.84 and a recall of 0.94 for class 0 (NON-FOG), indicating high accuracy in identifying non-FOG cases with few false positives. The F1-score for this class is 0.88, reflecting a strong balance between precision and recall. For class 1 (FOG), the precision is lower at 0.73, and the recall drops to 0.49, indicating that while the model correctly identifies some FOG instances, it misses a significant portion, leading to an F1-score of 0.59. The overall accuracy of the model is 0.82, but this is heavily influenced by the class imbalance. The weighted averages for precision, recall, and F1-score are 0.81, 0.82, and 0.81, respectively, showing reasonable performance across classes. However, the relatively low recall for class 1 suggests a need for further improvement in detecting FOG cases, despite the high accuracy for NON-FOG instances.

3.2. Leave-One-Subject-Out (LOSO) Cross Validation

Often when dealing with medical data and given the limited data, conducting a single train-test split is not sufficient as it does not provide an accurate estimate of the model's performance. To address this situation, we will apply the Leave-One-Subject-Out Cross-Validation method. This is particularly useful in cases where we want to evaluate the model's performance on patients never seen. In this approach, we leave one patient out for testing and use the rest for training. We repeat this for each patient, and the results obtained are as follows:

Subject Out	N of Samples	FOG Samples	Train/Test Split
1	324	67	20,78%
2	273	51	18,68%
3	418	117	27,99%
4	392	243	61,99%
5	369	43	11,65%
6	401	22	5,49%
7	377	102	27,06%

The results are presented in the following table:

Subject Out	Accuracy	F1 score of FOG
1	0.84	0.65
2	0.86	0.48
3	0.72	0.47
4	0.74	0.76
5	0.74	0.37
6	0.96	0.43
7	0.73	0.22
Average	0.80	0.48

The table displays the results of Leave-One-Subject-Out cross-validation, highlighting variability in model performance across different subjects. The accuracy values range from 0.72 to 0.96, indicating that the model's performance is subject-dependent. Notably, subject 6 achieves the highest accuracy (0.96) but has a relatively low F1 score (0.43), suggesting that high accuracy does not necessarily translate to balanced precision and recall.

Subject 7 shows a significantly lower F1 score (0.22) compared to other subjects, despite having an average accuracy of 0.73. This indicates a substantial imbalance between precision and recall, possibly due to the model's difficulty in correctly identifying FOG events for this subject. In contrast, subject 4 has a higher F1 score (0.76) with a similar accuracy (0.74), pointing to a better balance in detecting both classes.

The average accuracy across all subjects is 0.80, with an average F1 score of 0.48, highlighting a general trend of moderate performance. The notable discrepancy in F1 scores, especially the significantly lower score for subject 7, warrants further investigation to understand the underlying causes, which could include subject-specific variations in data or model sensitivity to particular features.

3.3. Leave-One-Subject-Out (LOSO) Cross Validation + Class Weighting

To improve the results, we consider mitigating the bias towards the minority class by introducing class weighting. This involves assigning greater weights to the minority class, in this case to windows with FOG=1, to emphasize their importance. This approach ensures that errors in mislabelling a FOG event as NON-FOG are penalized more heavily than errors in mislabelling a NON-FOG event as FOG. This strategy is particularly relevant in this application were identifying as many FOG events as possible is crucial.

We retrained the model with a weight of 1 for NON-FOG examples and 3 for FOG examples. The results obtained are shown in the following table:

Subject Out	Accuracy	F1 score of FOG
1	0.70	0.56
2	0.86	0.65
3	0.58	0.50
4	0.76	0.79
5	0.47	0.27
6	0.96	0.56
7	0.71	0.44
Average	0.72	0.54

****Green: Increased with respect to the last table**, **Red: Decreased with respect to the last table**

Results:

This table presents the results after applying class weighting, giving three times more importance to the FOG (Freezing of Gait) positive class. This approach aimed to address class imbalance and improve FOG detection.

Comparing the two tables, the average accuracy decreased from 0.80 to 0.72, while the average F1 score of FOG increased from 0.48 to 0.54, indicating better balance between precision and recall for FOG.

Subject-specific performance shows mixed results. Subjects 2, 4, and 7 showed significant improvements in their F1 scores, with subject 2's F1 score rising from 0.48 to 0.65, subject 4's from 0.76 to 0.79, and subject 7's from 0.22 to 0.44. Despite this, subject 1 saw a decrease in

accuracy from 0.84 to 0.70, offset by a slight drop in F1 score from 0.65 to 0.56. Subject 5 experienced decreases in both accuracy (0.74 to 0.47) and F1 score (0.37 to 0.27).

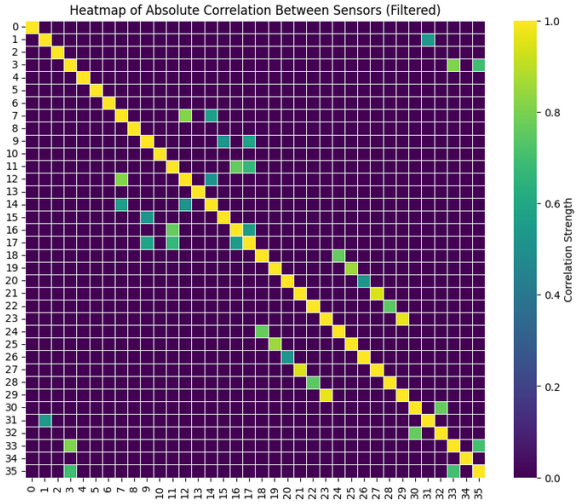
Overall, while accuracy declined, the increased F1 scores for several subjects indicate that class weighting improved the model's ability to detect FOG events, suggesting that this method enhances detection balance at the cost of some accuracy. Further refinements are needed to achieve consistent improvements across all subjects.

4. Second Approach

Incorporating signal data from additional IMU sensors is an appealing prospect, yet this comes with increased model complexity, extended training times, and longer prediction periods. As a preliminary step, we propose calculating the correlation among sensor signals for each patient to discern any discernible patterns. Signals that exhibit high correlation are likely redundant, offering minimal new information due to their similarity. In contrast, signals with low correlation could significantly enrich the model's data pool by providing diverse perspectives crucial for detecting when FOG occurs.

We suggest analysing the correlation of sensors located on various body parts. A high correlation would indicate that adding such sensors would not substantially enhance the model's information capacity. On the other hand, a low correlation suggests that the sensor's data could considerably boost model performance.

Upon analysing these correlations, we can strategically select sensors that contribute the most valuable information, incorporating these features into our model. Consequently, the input X would be adjusted to (2554, 256, N), where N represents the number of features incorporated, followed by application of a 1D convolution to assess how model performance varies with the addition of more sensors.



The heatmap analysis reveals a strong correlation (1-1) between the accelerometers and gyroscopes of sensors (18-23) and (24-29), indicating a link between the Right Ankle IMU and the Left Foot IMU. This finding appears counterintuitive given their distinct locations and warrants further investigation. It is equally noteworthy to observe correlations such as that

between the z-axis accelerometer of the chest IMU and the gyroscopes on the Right Foot, or between sensors (12-17) and (6-11). The absence of more extensive correlations between accelerometers and gyroscopes within the same sensor is surprising.

Based on these findings and to prevent excessive model size, it seems prudent to select sensors from both feet, the chest, and the lumbar region, as they exhibit minimal correlation with each other. Considering the limited angular mobility at the lumbar and chest, we could consider incorporating features from the chest and lumbar accelerometers, along with the accelerometers and gyroscopes of both ankles, resulting in $X.shape\ of\ (2554, 256, 18)$. We would then apply a 1D convolution with number of features = 18.

Note: In this approach we also used the weight class giving 3 times more importance to the FOG datapoints since we saw an improvement in the last study.

Subject Out	Accuracy	F1 score of FOG
1	0.83	0.65
2	0.82	0.60
3	0.63	0.53
4	0.78	0.83
5	0.58	0.31
6	0.85	**0.29**
7	0.72	0.41
Average	0.75	0.52

****Green: Increased with respect to the last table**, **Red: Decreased with respect to the last table**

Results:

Comparing the new results with the previous ones, we observe mixed outcomes. The average accuracy increased from 0.72 to 0.75, and the average F1 score of FOG slightly decreased from 0.54 to 0.52, indicating a marginal decline in balancing precision and recall for FOG detection.

For individual subjects, the results varied. Subject 1's accuracy improved from 0.70 to 0.83, and the F1 score increased from 0.56 to 0.65, indicating better overall performance. Subject 2 saw a slight decrease in accuracy from 0.86 to 0.82, with the F1 score dropping from 0.65 to 0.60. Subject 3 experienced improvements in both metrics, with accuracy increasing from 0.58 to 0.63 and the F1 score from 0.50 to 0.53. Subject 4 also improved, with accuracy rising from 0.76 to 0.78 and the F1 score from 0.79 to 0.83, reflecting better detection of FOG events. Subject 5's accuracy increased from 0.47 to 0.58, while the F1 score showed a slight improvement from 0.27 to 0.31.

Subject 6, however, showed a significant drop in F1 score from 0.56 to 0.29 despite a decrease in accuracy from 0.96 to 0.85. This subject has highly imbalanced classes, with only 22 FOG cases out of 401 examples, which likely explains the poor F1 score despite the high accuracy. Subject 7 saw slight improvements, with accuracy increasing from 0.71 to 0.72 and the decrease in F1 score from 0.44 to 0.41.

The decrease in performance for Subject 6 suggests that incorporating more features might have led to overfitting to the dominant class (NON-FOG) due to severe class imbalance. To address this, techniques to handle class imbalance effectively are necessary.

Possible solutions include using synthetic data generation methods such as SMOTE (Synthetic Minority Over-sampling Technique) to balance the classes, applying more robust regularization techniques, or utilizing ensemble methods to improve the model's generalization capabilities. Additionally, further fine-tuning of the selected features and model parameters may be necessary to enhance the model's performance across all subjects.

5. Third Approach

In this approach, we focus on the frequency components of signals to aid in the diagnosis and study of Parkinson's disease. The dataset comprises sensor readings with dimensions (2554, 256, 36), representing 2554 windows, each consisting of 256-time steps and in each time step we get the information of the 36 sensors. Each window covers a span of 2 seconds, reflecting a sensor sampling rate of 128 Hz. For this approach we take into account the information of the 4 sensors mentioned in the previous section, so in this case the X.Shape is (2554, 256,18).

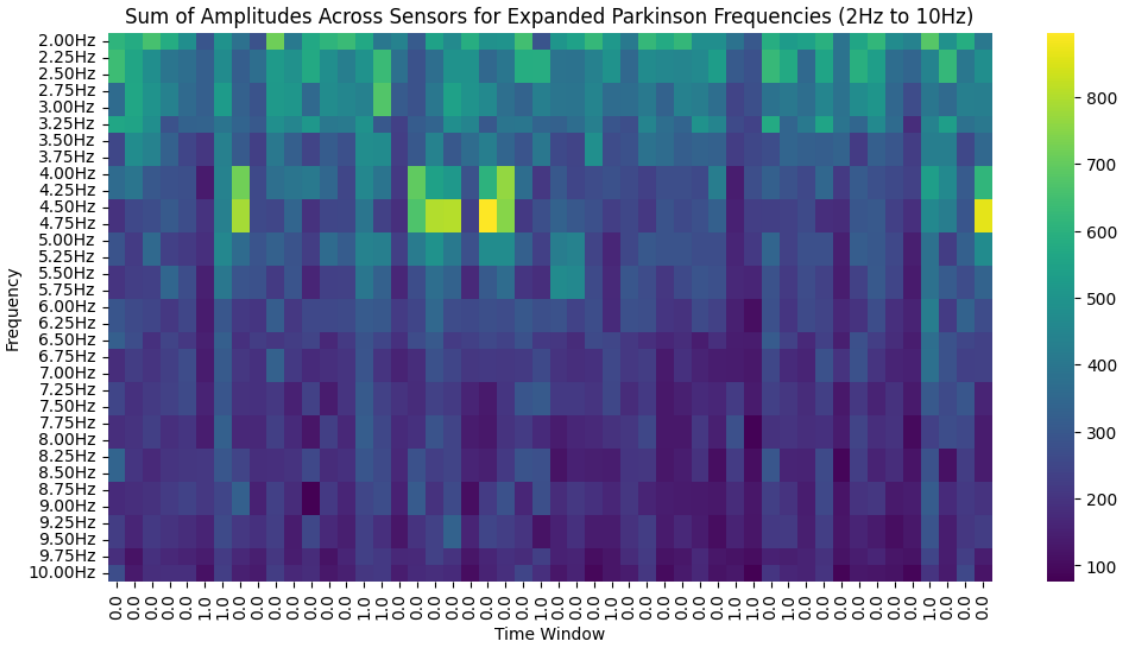
We processed the data as follows:

1. **DC Component Removal:** Initially, we eliminate the DC component from each sensor reading by subtracting the mean value. Removing this constant offset is essential to concentrate on the signal's fluctuating properties rather than its stationary value.
2. **Windowing:** To mitigate spectral leakage and preserve signal integrity at the window edges, we apply a Hamming window to segment the data into overlapping windows. The window size is set to 276 samples, with 10 samples of zero padding at both the beginning and end of each window, and a stride of 256 samples to maintain continuity between consecutive windows.
3. **Discrete Fourier Transform (DFT):** Each window undergoes a Discrete Fourier Transform (DFT), transitioning the data from the time domain to the frequency domain. This transformation is pivotal for analysing signal frequency components critical in identifying patterns related to Parkinson's disease.
4. **Frequency Filtering:** After transforming the signals, we focus on the frequency range from 3 to 7 Hz, pertinent to symptoms such as tremors commonly observed in Parkinson's disease. We apply a frequency resolution of 0.25 Hz to precisely examine the spectrum and highlight the most indicative frequencies. The indices corresponding to these frequencies are calculated and extracted from the Fourier-transformed data.
5. **Amplitude Summation:** We then compute the amplitude of these selected frequencies for each frame, providing a detailed view of the signal strength at each frequency.
6. **Feature Extraction:** The amplitude data of the frequencies of interest are organized into a feature matrix, where each row corresponds to a time window and each column corresponds to a specific frequency. This feature matrix can then be used for further analysis or fed into machine learning models for classification or diagnostic purposes.

This methodology is designed to highlight the most relevant frequency components associated with Parkinson's disease, facilitating effective analysis and interpretation of sensor data.

The results for the first 50-time samples are presented in the following figure. In the left corner we see the frequency, the colour represents the amplitude of that specific frequency in that time step, summing the amplitudes of all the sensors. Finally, in the x axis, we observe the label

per time frame. Additionally, we can observe that we have included the frequency range from 2 to 10 to cover a slightly wider range and capture more information.



The dataset now has the shape of X.shape = (2554, 33) and Y.shape = (2554, 1), as we have obtained data from 33 frequency features for each window of 256 time steps. Another approach could have been to consider separately the frequency components of the 18 sensors, which would result in an X.shape of (2553, 33, 18). This is left for future projects. It is also possible to concatenate the temporal features with the frequency features and interpret the results. However, this falls outside the scope of this project.

To apply the aforementioned model, we reshape the dataset format to (2554, 33, 1), which would be equivalent to having 33 time steps and one feature from the previous examples. Since in this case, the number of time steps is much smaller than in previous cases, it was considered to reduce the filter size to 5. That is, in the first 1D convolution, we learn 15 filters of size (5, 1).

Given this, the following results are obtained:

Subject Out	Accuracy	F1 score of FOG
1	0.81	0.54
2	0.74	0.57
3	0.57	0.45
4	0.57	0.52
5	0.60	0.28
6	0.86	0.33
7	0.68	0.35
Average	0.69	0.44

We observe that the results are slightly worse than the other approaches but given the limited number of frequency values (33) being considered, we believe it is a promising alternative and warrants a more detailed study.

6. Conclusion

In this study, we explored various approaches to enhance the detection and management of Freezing of Gait (FOG) in Parkinson's disease through the application of Artificial Intelligence (AI) and machine learning techniques. FOG, a debilitating symptom of Parkinson's, significantly impacts the quality of life, leading to falls and loss of independence. Although current monitoring techniques are sophisticated, they are often complex and costly.

Our first approach employed a convolutional neural network (CNN) to analyse data from a single inertial measurement unit (IMU) located at the lumbar region. This method achieved reasonable accuracy (82%), but it revealed a significant class imbalance problem, with a higher rate of false negatives for FOG events. To address this, we applied Leave-One-Subject-Out (LOSO) cross-validation, which highlighted variability in model performance across different subjects. Introducing class weighting to this approach improved the F1 score from 0.48 to 0.54, indicating a better balance between precision and recall for FOG detection.

The second approach incorporated additional sensors to provide a richer dataset. We conducted correlation analysis to identify and eliminate redundant signals, thereby balancing model complexity with the potential for enhanced performance. This approach resulted in an increase in average accuracy to 75%, though the F1 score slightly decreased to 0.52, suggesting issues with overfitting and the need for more robust handling of class imbalance.

In the third approach, we focused on the frequency components of sensor signals by utilizing the Discrete Fourier Transform (DFT) to analyse the relevant frequency ranges associated with Parkinson's symptoms. While this method provided a novel perspective, the results were less favourable compared to time-domain analysis, with an average accuracy of 69% and an F1 score of 0.44. These outcomes indicate that while frequency-based analysis is promising, it requires further refinement and a more detailed study to optimize its effectiveness.

Throughout our study, class imbalance emerged as a critical challenge. Both the first and second approaches demonstrated that the class imbalance significantly affected model performance. The introduction of class weighting improved the detection of the minority class (FOG events), underscoring the importance of addressing this issue. Additionally, our sensor data correlation analysis revealed unexpected correlations among certain sensors, suggesting the necessity for careful sensor selection to avoid redundant data while maximizing the information value.

Moreover, the frequency-based analysis in our third approach, although initially less effective, indicated potential when combined with time-domain features. This highlights the need for integrated approaches that leverage the strengths of both domains.

Future work should focus on several key areas. Firstly, implementing synthetic data generation techniques, such as the Synthetic Minority Over-sampling Technique (SMOTE), could create a more balanced dataset, further mitigating class imbalance. Secondly, employing ensemble learning techniques to combine multiple models could enhance generalization and robustness in FOG detection. Additionally, exploring the integration of temporal and frequency features

could leverage the strengths of both domains, potentially improving model performance. Applying advanced regularization methods will also be essential to prevent overfitting, particularly when increasing model complexity with additional features or sensors.

Another promising direction is the development of adaptive models tailored to individual patient profiles. These personalized models could offer more accurate and responsive FOG detection, catering to the unique characteristics of each patient.

In conclusion, while significant progress has been made in applying AI to manage FOG in Parkinson's disease, further research and development are necessary to enhance model accuracy, robustness, and generalization across diverse patient populations. Addressing class imbalance, optimizing sensor selection, and integrating multi-domain features will be crucial steps towards improving the detection and management of FOG.

7. Bibliography

[1] Nutt JG, Bloem BR, Giladi N, Hallett M, Horak FB, Nieuwboer A. Freezing of gait: moving forward on a mysterious clinical phenomenon. Lancet Neurol. 2011 Aug;10(8):734-44. doi: 10.1016/S1474-4422(11)70143-0. PMID: 21777828; PMCID: PMC7293393.

[2] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. University of Toronto.

[3] O'Day J, Lee M, Seagers K, Hoffman S, Jih-Schiff A, Kidziński Ł, Delp S, Bronte-Stewart H. Assessing inertial measurement unit locations for freezing of gait detection and patient preference. J Neuroeng Rehabil. 2022 Feb 13;19(1):20. doi: 10.1186/s12984-022-00992-x. PMID: 35152881; PMCID: PMC8842967.

[4] O'Day, J., Lee, M., Seagers, K. et al. Assessing inertial measurement unit locations for freezing of gait detection and patient preference. J NeuroEngineering Rehabil 19, 20 (2022). https://doi.org/10.1186/s12984-022-00992-x