Content

Appendix I: Scikit- learn cheat sheet

To determine appropriate machine learning algorithms for predicting Freezing of Gait (FoG), model selection was guided by the decision flow provided in the Scikit-Learn cheatsheet. The dataset used satisfies the following conditions:

- More than 50 samples
- Categorical target variable (presence or absence of FoG)
- Labeled data
- Fewer than 100,000 samples
- No textual data

Following this pathway, the following classification models were identified as suitable candidates:

Linear Support Vector Classifier (Linear SVC)

This model is effective for high-dimensional datasets and is particularly suitable when the data is linearly separable. Given the absence of text data and the binary nature of the target variable (FoG vs. no FoG), Linear SVC is a logical baseline model.

K-Nearest Neighbors (KNN) Classifier

KNN is a simple, instance-based learning algorithm that makes predictions based on the majority class among the k closest training samples in the feature space. It is non-parametric and can capture local patterns in the data that may indicate the onset of FoG episodes.

Decision Tree Classifier (as part of ensemble methods)

Decision Trees offer a transparent and interpretable structure for classification. They are capable of capturing non-linear relationships between features and the target class, which can be valuable when modeling complex motor symptoms such as FoG.

Random Forest Classifier (as part of ensemble methods)

Random Forest builds multiple decision trees and aggregates their predictions to improve generalization and reduce overfitting. This ensemble method is particularly robust in handling noisy or imbalanced data, which is often the case in medical datasets.

These models were selected to explore a range of classification strategies, from simple instance-based approaches to more complex ensemble methods. Their performance will be evaluated to determine the most effective model for accurately predicting FoG events.

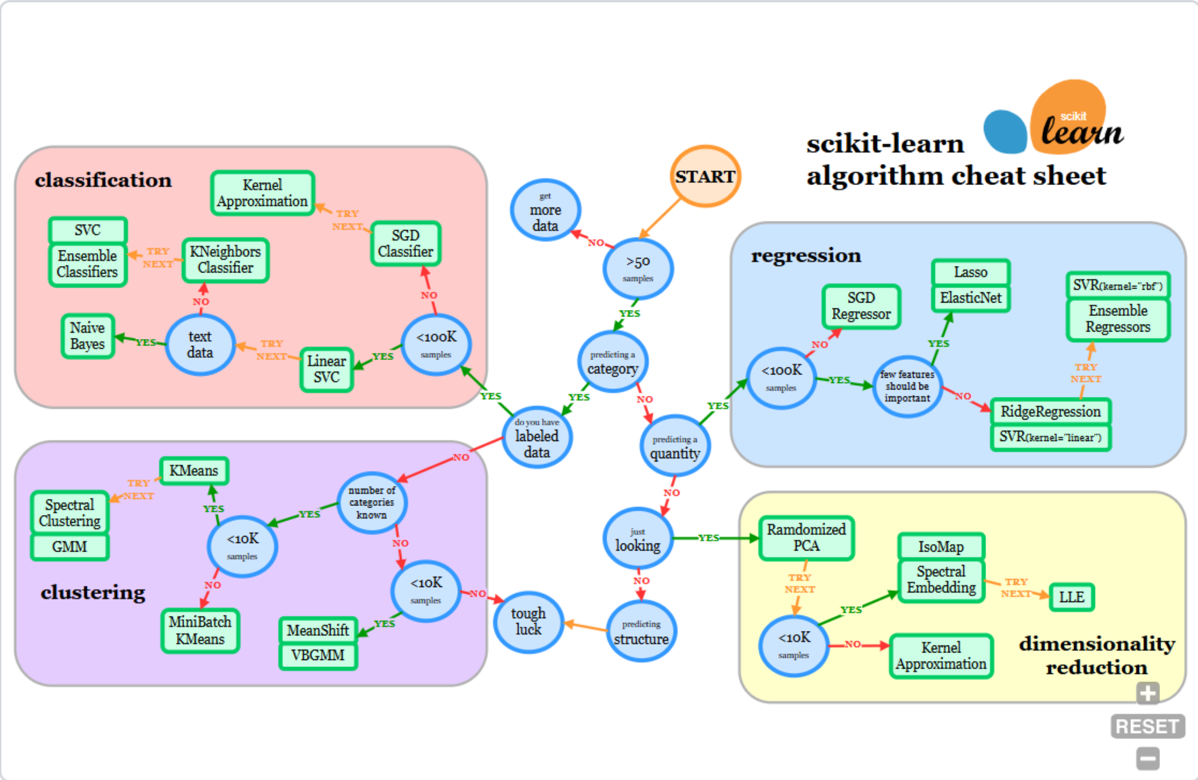


Figure 1: scikit-learn, algorithm cheat sheet, 2025

Appendix II: Model performance

Evaluation Metrics

To evaluate the performance of the models used for detecting Freezing of Gait (FoG), several commonly used machine learning metrics were applied. These metrics help assess both overall accuracy and the model's ability to identify freezing episodes in real-world scenarios.

Precision, Recall, F1-Score, and Support

These values are derived from the classification report and provide detailed insights into how the model handles each class:

- Precision:
  Indicates how many of the predicted positive cases (e.g., predicted FoG) were actually correct. High precision means few false positives.
- Recall (Sensitivity):
  Measures how many of the actual positive cases were correctly identified. High recall reduces the number of missed FoG events (false negatives), which is important in this context.
- F1-Score:
  The harmonic mean of precision and recall. It provides a balanced metric that is especially useful when the class distribution is uneven or when both types of errors matter.
- Support:
  The number of true instances for each class in the dataset. This gives context to the above metrics and shows how balanced the dataset is.

ROC Curve and AUC (Area Under the Curve)

The ROC curve visualizes the model's ability to separate the two classes across different classification thresholds:

- It plots the true positive rate against the false positive rate.
- A curve that bends towards the top-left corner represents better performance.

The AUC score summarizes the curve into a single value:

- 1.0 means perfect classification,
- 0.5 means random guessing,
- Scores above 0.8 generally indicate strong performance.

This is particularly relevant for medical applications, where both sensitivity and specificity are important.

Confusion Matrix

The confusion matrix provides a detailed breakdown of correct and incorrect predictions:

- True Positives: Correctly predicted FoG events.
- True Negatives: Correctly predicted non-FoG events.
- False Positives: Non-FoG incorrectly predicted as FoG.
- False Negatives: Missed FoG events (predicted as non-FoG).

This matrix allows a quick visual check of model performance and helps identify common types of misclassification. It was especially valuable during live testing, where real-time feedback was compared to actual simulated freezing behavior.

Machine Learning Models Performance

Linear Support Vector Classifier (Linear SVC)

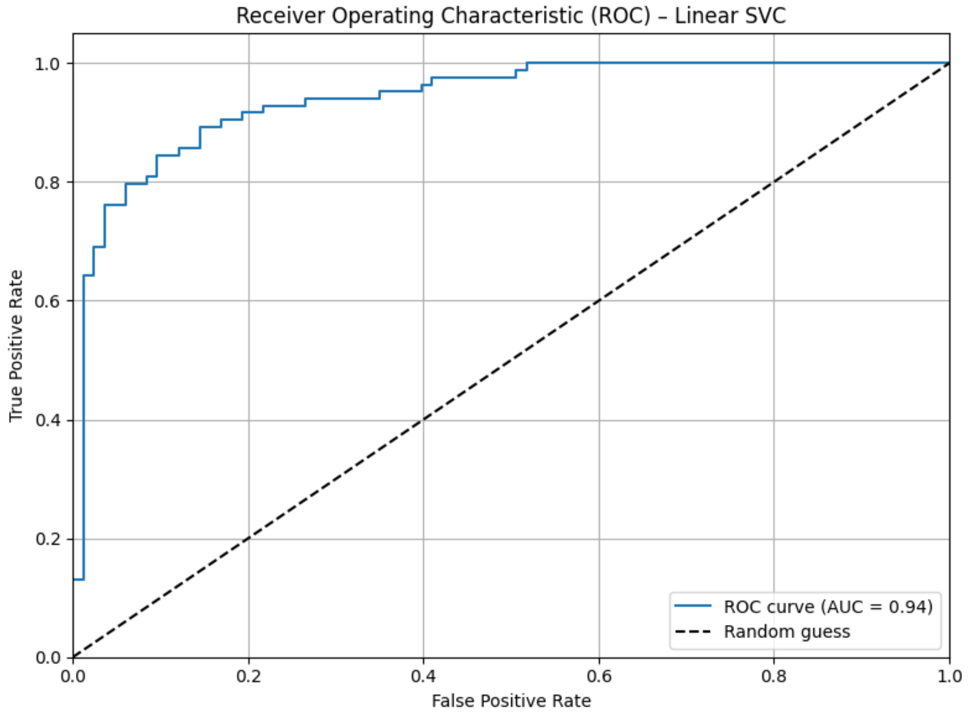
Model Performance – Linear Support Vector Classifier (Linear SVC)

```
Training Performance:
              precision    recall  f1-score   support

       False      0.842     0.807     0.824       389
        True      0.815     0.848     0.831       389

    accuracy                          0.828       778
   macro avg      0.828     0.828     0.828       778
weighted avg      0.828     0.828     0.828       778

[[314  75]
 [ 59 330]]
Test Performance:
              precision    recall  f1-score   support

       False      0.866     0.855     0.861        83
        True      0.859     0.869     0.864        84

    accuracy                          0.862       167
   macro avg      0.862     0.862     0.862       167
weighted avg      0.862     0.862     0.862       167

[[71 12]
 [11 73]]
```

The Linear SVC model demonstrates strong and balanced performance in detecting Freezing of Gait (FoG) events. On the training set, the model achieves an overall accuracy of 82.8%. It shows slightly higher precision for the non-freezing class (82.8%) as well as the recall. This suggests the model is able to identify freezing episodes effectively while still limiting false positives. On the test set, the model maintains a high accuracy of 86.2%, indicating good generalization to unseen data. Both classes are detected with similar effectiveness:

- FoG detection (True): precision of 85.9%, recall of 86.9%
- Non-FoG detection (False): precision of 86.6%, recall of 85.5%

ROC Curve – Linear Support Vector Classifier (Linear SVC)

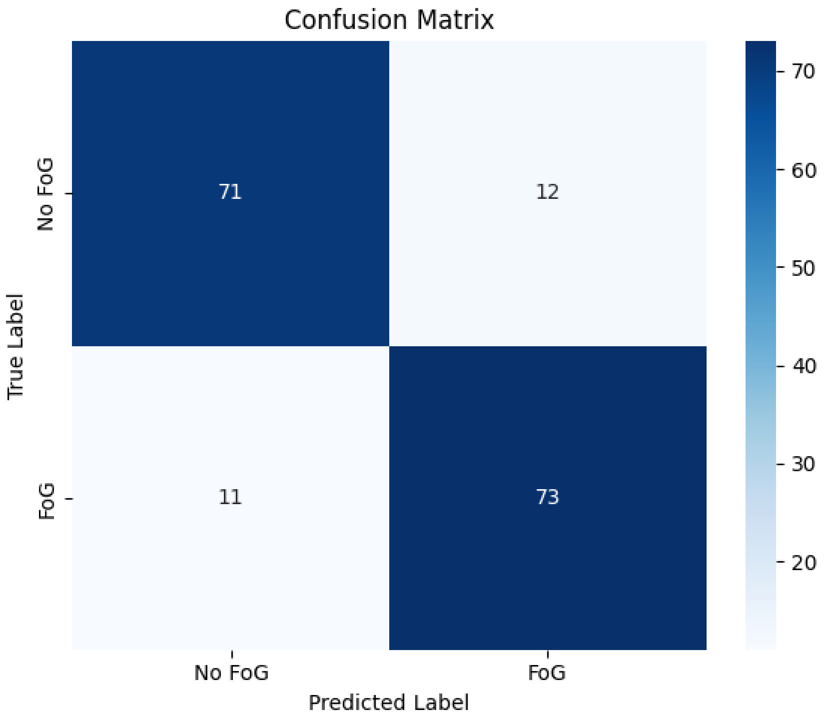


The curve rises sharply toward the top-left corner, which indicates that the model achieves a high true positive rate (sensitivity) while maintaining a low false positive rate, an ideal characteristic for a reliable classifier.

The AUC (Area Under the Curve) is 0.94, which is considered very good. A perfect classifier would have an AUC of 1.0, while a value of 0.5 corresponds to random guessing. An AUC of 0.94 indicates that the model can correctly differentiate between FoG and non-FoG events with a high degree of confidence.

This performance confirms that the Linear SVC model is highly effective in capturing the relevant patterns in the sensor data and has strong potential for use in real-time, wearable FoG detection systems.

Confusion Matrix – Linear SVC (Test Set)



The confusion matrix for the Linear SVC model illustrates its ability to classify freezing (FoG) and non-freezing events with a good balance between sensitivity and specificity:

- 71 non-freezing instances were correctly predicted as No FoG (True Negatives).
- 73 freezing events were correctly predicted as FoG (True Positives).
- 12 non-freezing instances were incorrectly predicted as FoG (False Positives).
- 11 freezing events were missed and predicted as No FoG (False Negatives).

The model performs well and has a low number of missed FoG classifications.

K-Nearest Neighbors (KNN) Classifier

Model Performance – K-Nearest Neighbors (KNN) Classifier

```
Training Performance:
              precision    recall  f1-score   support

       False      0.852     0.874     0.863       389
        True      0.871     0.848     0.859       389

    accuracy                          0.861       778
   macro avg      0.861     0.861     0.861       778
weighted avg      0.861     0.861     0.861       778

[[340  49]
 [ 59 330]]
Test Performance:
              precision    recall  f1-score   support

       False      0.818     0.867     0.842        83
        True      0.861     0.810     0.834        84

    accuracy                          0.838       167
   macro avg      0.839     0.838     0.838       167
weighted avg      0.840     0.838     0.838       167

[[72 11]
 [16 68]]
```

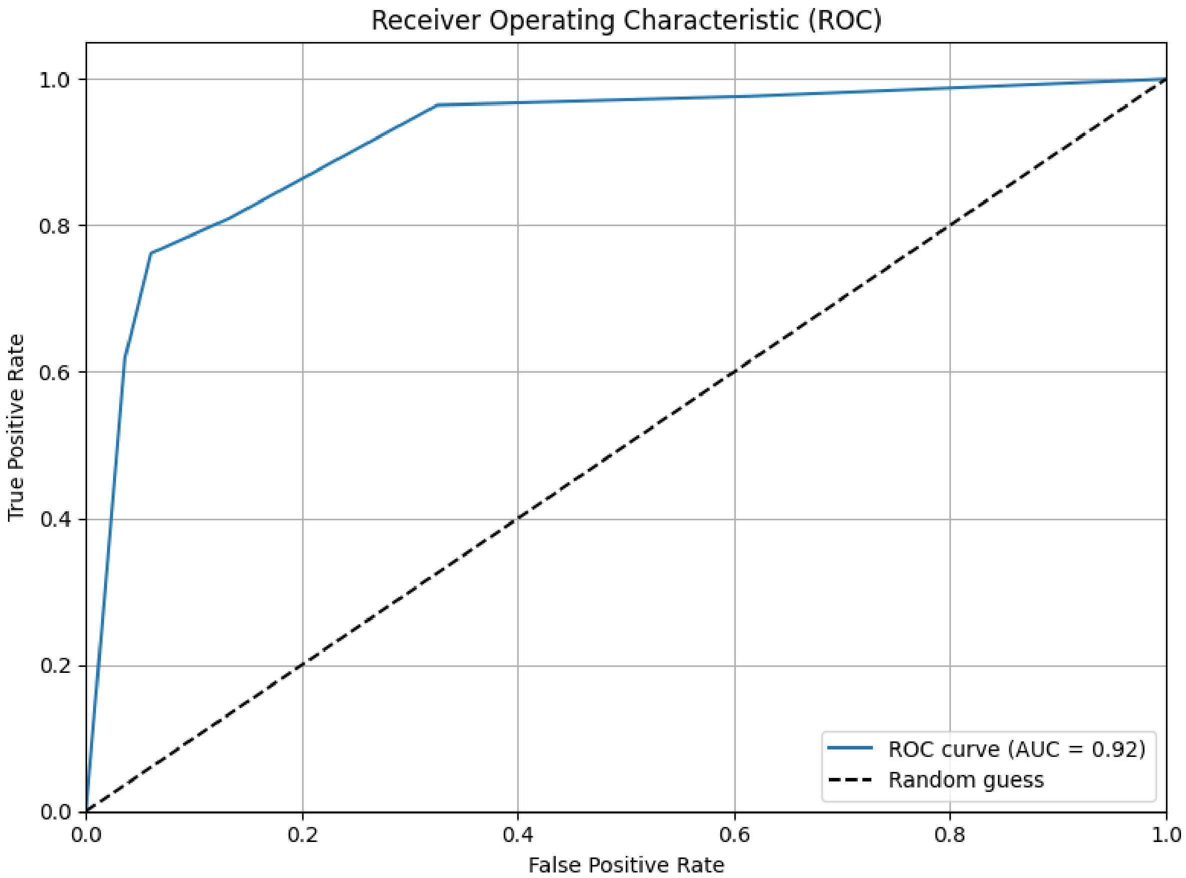
The KNN model demonstrates consistently strong performance in predicting Freezing of Gait (FoG) events.

On the train set, the model has a high accuracy of 86.1%, showing good ability to identify classes. The precision and recall for both classes are mostly balanced (around 85.2–86.1%), confirming the model's robustness. The confusion matrix shows only 108 misclassifications out of 778 samples (49 false positives and 59 false negatives), which is low.

On the test set, the classifier achieves an overall accuracy of 83.4%, with a preference for FoG over no FoG classes. With the accuracy of the lowest class (no FoG) of 81.8% and the highest class (FOG) of 86.1%, the model can correctly identify both freezing and non-freezing events. The confusion matrix confirms this balance, with 68 true negatives and 72 true positives.

Overall the KNN classifier has a high accuracy with 83.8%. The model is slightly biased towards FOG events, with an accuracy of 86.1% for 81.8% FoG events and for non FoG events. This model is a candidate with its high precision, but the bias should be improved before the model is used.

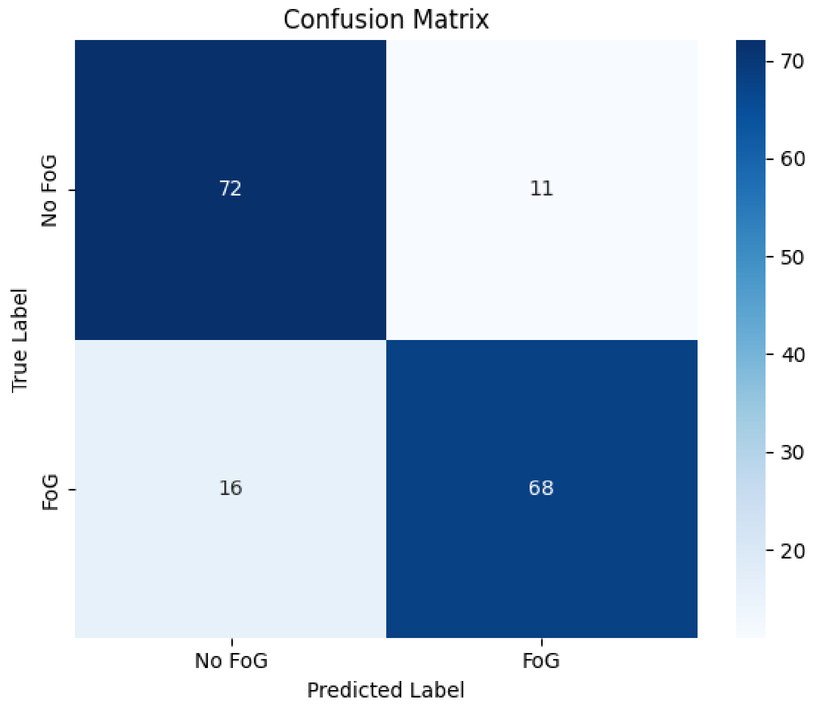
ROC Curve – K-Nearest Neighbors (KNN) Classifier



The ROC curve shows that the KNN model performs very well in distinguishing between freezing (FoG) and non-freezing events, with a small bias toward true positives. The curve quickly rises toward the top-left corner, which indicates a high true positive rate with a low false positive rate — exactly what is desired for a classification model.

The Area Under the Curve (AUC) is 0.92, which is considered excellent. AUC values above 0.9 suggest that the model is highly capable of separating the two classes across all classification thresholds.
This confirms that the KNN classifier not only performs strongly in terms of accuracy, but also maintains a balance between sensitivity and specificity, making it a reliable tool for detecting FoG episodes based on sensor data.

Confusion Matrix – K-Nearest Neighbors (KNN) Classifier (Test Set)



The confusion matrix shows that the KNN classifier makes very few mistakes in classifying both FoG and non-FoG instances:

- 72 non-freezing events were correctly predicted as No FoG (True Negatives).
- 68 freezing events were correctly classified as FoG (True Positives).
- 11 non-freezing instances were incorrectly predicted as FoG (False Positives).
- 16 freezing events were missed and predicted as No FoG (False Negatives).

The model has both high sensitivity (detecting most true FoG events) and specificity (correctly recognizing when no FoG is present), with low rates of false positives and false negatives. The confusion matrix supports the high accuracy, precision, recall, and AUC score previously observed, making KNN a strong and consistent performer in this classification task.

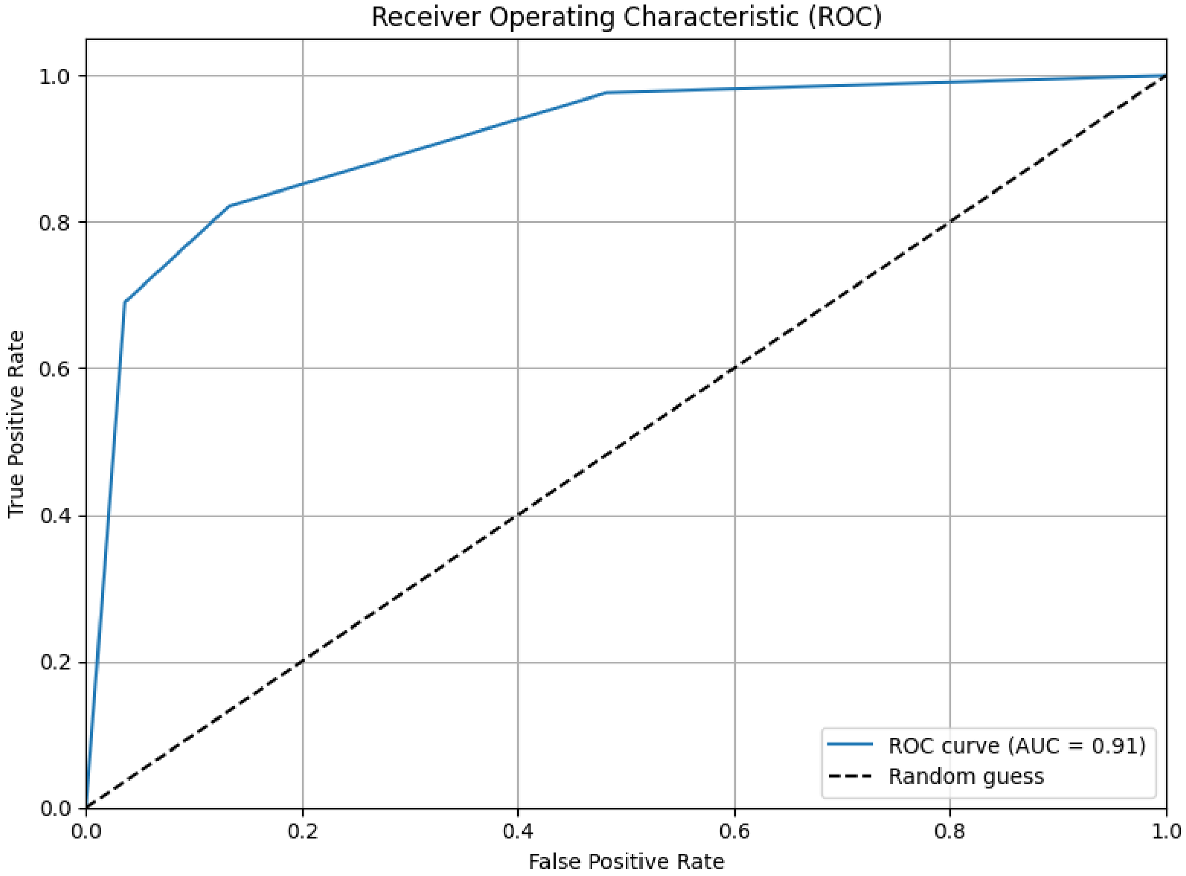
Decision Tree Classifier (as part of ensemble methods)

Model Performance – Decision Tree Classifier

```
Training Performance:
              precision    recall  f1-score   support

       False      0.785     0.843     0.813       389
        True      0.831     0.769     0.798       389

    accuracy                          0.806       778
   macro avg      0.808     0.806     0.806       778
weighted avg      0.808     0.806     0.806       778

[[328  61]
 [ 90 299]]
Test Performance:
              precision    recall  f1-score   support

       False      0.828     0.867     0.847        83
        True      0.863     0.821     0.841        84

    accuracy                          0.844       167
   macro avg      0.845     0.844     0.844       167
weighted avg      0.845     0.844     0.844       167

[[72 11]
 [15 69]]
```

The Decision Tree classifier shows strong and balanced performance in predicting Freezing of Gait (FoG) events. On the training set, the model achieves an accuracy of 80.6%. It performs well in detecting both classes, with a precision of 78.5% for FoG events and 83.1% for non-FoG. The recall for FoG is 76.9% indicating that the model misses some freezing episodes and for non-FoG 84.3%. On the test set, the accuracy improves slightly to 84.4%, suggesting good generalization to unseen data. The model achieves a precision of 84.5% for predicting FoG events and a recall of 84.4%.

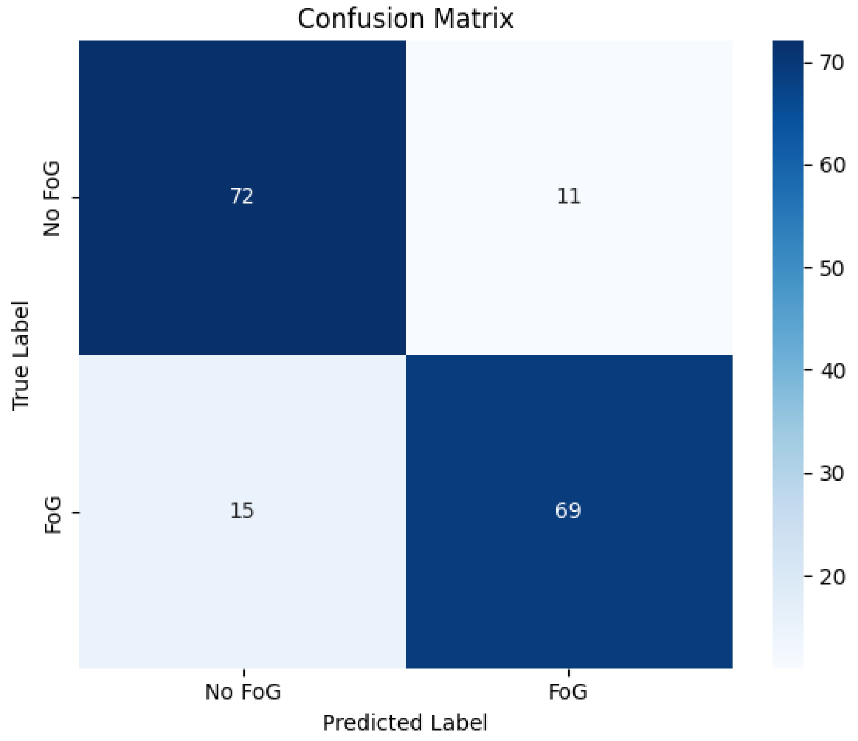
ROC Curve – Decision Tree Classifier



The ROC curve provides a visual representation of how well the Decision Tree classifier can distinguish between freezing (FoG) and non-freezing events. The curve shown rises steeply toward the top-left corner, which indicates that the model achieves a high true positive rate while maintaining a low false positive rate.

The AUC (Area Under the Curve) is 0.91, which reflects strong overall performance. Since an AUC of 1.0 represents perfect classification and 0.5 corresponds to random guessing, a value of 0.91 means the model is able to effectively separate FoG from non-FoG instances. This confirms that the Decision Tree classifier performs well not only in terms of accuracy but also in distinguishing between the two classes across all classification thresholds.

Confusion Matrix – Decision Tree Classifier (Test Set)



The confusion matrix shows how the Decision Tree model performs in classifying freezing (FoG) and non-freezing events:

- 72 non-freezing events were correctly identified as No FoG (True Negatives).
- 69 freezing events were correctly detected as FoG (True Positives).
- 11 non-freezing instances were wrongly predicted as FoG (False Positives).
- 15 actual freezing events were missed and classified as No FoG (False Negatives).

This performance indicates that the model is strong at recognizing non-freezing events. However, while it still detects the majority of freezing events, it misses some of them, as shown by the 15 false negatives.
The matrix reflects the model's conservative tendency favoring fewer false positives, but at the cost of missing some true FoG instances. In applications where it is critical not to overlook freezing episodes, this trade-off should be carefully considered.

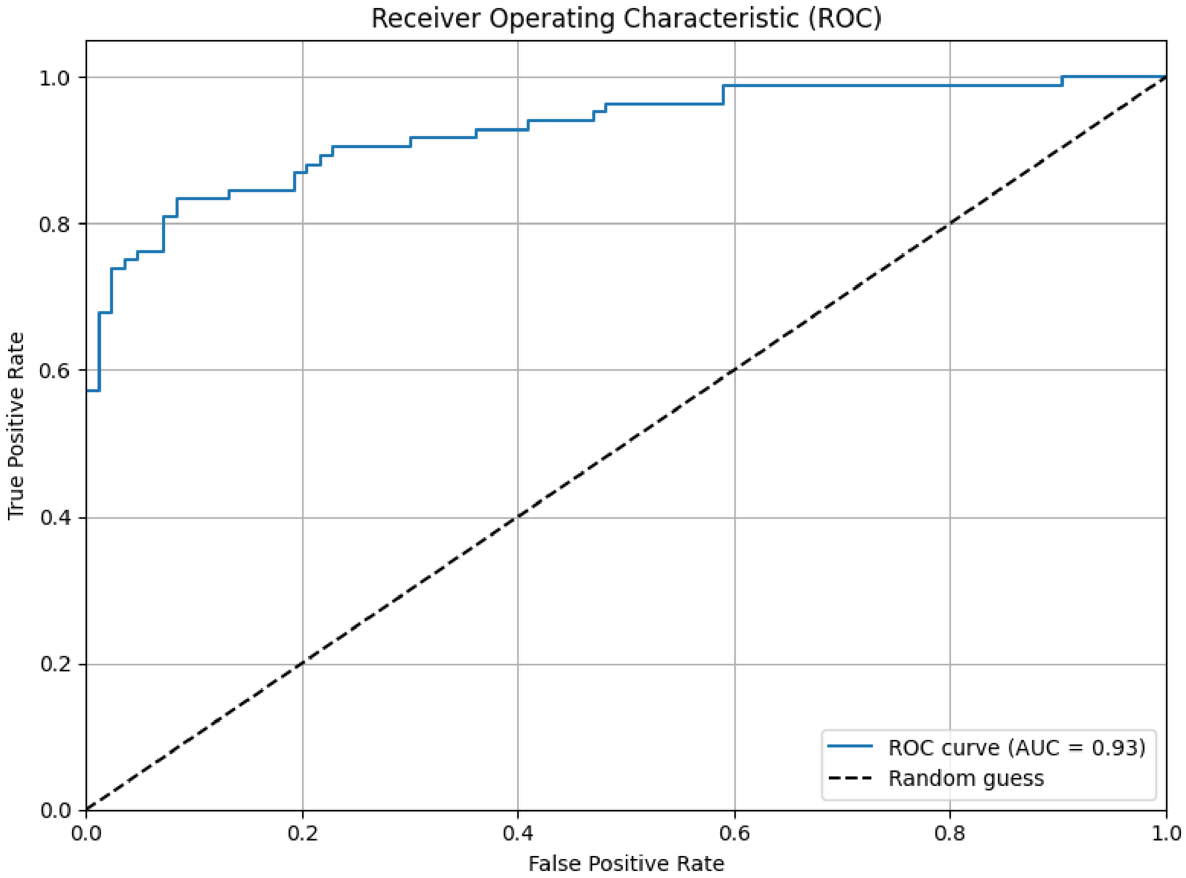
Random Forest Classifier (as part of ensemble methods)

Model Performance – Random Forest Classifier

```
Training Performance:
              precision    recall  f1-score   support

       False      0.773     0.882     0.824       389
        True      0.862     0.740     0.797       389

    accuracy                          0.811       778
   macro avg      0.817     0.811     0.810       778
weighted avg      0.817     0.811     0.810       778

[[343  46]
 [101 288]]
Test Performance:
              precision    recall  f1-score   support

       False      0.835     0.916     0.874        83
        True      0.908     0.821     0.863        84

    accuracy                          0.868       167
   macro avg      0.872     0.869     0.868       167
weighted avg      0.872     0.868     0.868       167

[[76  7]
 [15 69]]
```

The Random Forest classifier demonstrates solid performance in predicting FoG events. On the training set, the model achieved an accuracy of 86.8%. It obtained a precision of 86.2% and a recall of 74.1% for detecting FoG instances and a precision of 77.3% and a recall of 88.2% for non-FoG instances. On the test set, the accuracy remained consistent at 86.8%, indicating good generalization. The model showed a higher precision (87.2%) for identifying FoG events, also the recall improved (86.9%).Overall, the model performs well with relatively balanced sensitivity and specificity, making it a promising candidate for detecting FoG based on sensor data.

ROC Curve – Random Forest Classifier (Test Set)

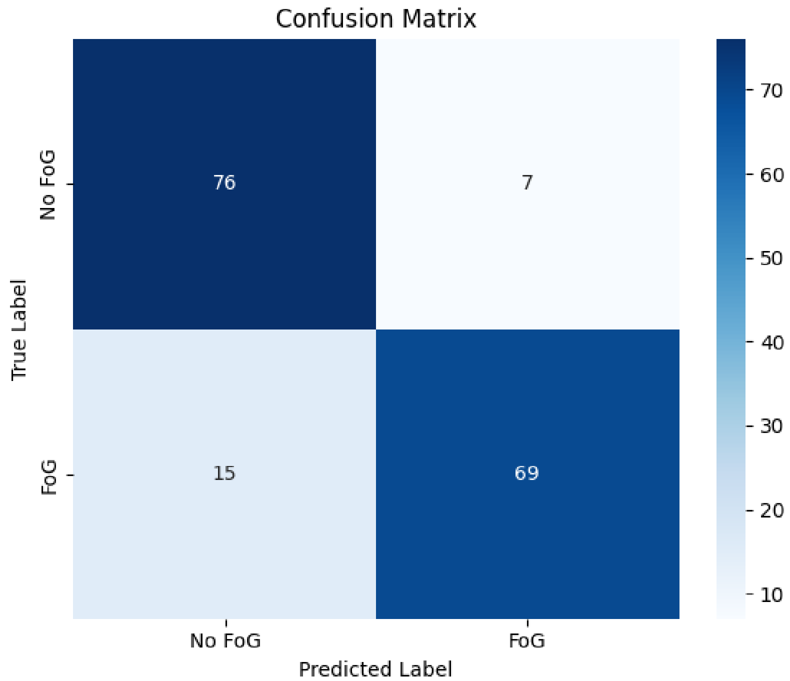


The ROC curve displayed above illustrates the diagnostic ability of the Random Forest classifier to distinguish between freezing and non-freezing events. The curve shows a strong upward bend toward the top-left corner, indicating a good balance between the true positive rate (sensitivity) and the false positive rate.

The AUC (Area Under the Curve) is 0.93, which is considered good. An AUC of 0.5 represents random guessing, while an AUC of 1.0 indicates perfect classification. Therefore, a score of 0. suggests that the model is highly effective at discriminating between FoG and non-FoG instances.

In summary, the ROC curve confirms that the model performs well, with a strong ability to correctly classify freezing episodes while minimizing false alarms.

Confusion Matrix – Random Forest Classifier (Test Set)



The confusion matrix provides insight into how well the model distinguishes between freezing (FoG) and non-freezing events on the test set:

- 76 non-freezing instances were correctly classified as No FoG (True Negatives).
- 69 freezing events were correctly identified as FoG (True Positives).
- 7 non-freezing cases were incorrectly classified as FoG (False Positives).
- 15 freezing events were missed and labeled as No FoG (False Negatives).

This shows that the model performs well overall, especially in recognizing non-freezing moments. While it correctly detects a large portion of actual FoG events, there is still a tendency to miss some (false negatives), which is important to consider in clinical or real-time applications where missing a freezing episode can be critical.
The confusion matrix confirms the model's balanced performance and supports the findings from the classification report and ROC analysis.

Appendix III: Real-World Testing of the Freezing of Gait Detection Model

To better understand how Freezing of Gait (FoG) presents itself in real-life situations, reference videos were first studied:

The following YouTube video provided a clear visual example of typical Parkinsonian gait and freezing episodes:
https://www.youtube.com/watch?v=EQ0HG16EC3g

In addition, this dataset on Figshare was reviewed, which includes synchronized video and sensor data of Parkinson’s patients during turning-in-place tasks:
A public dataset of video, acceleration, angular velocity, and clinical scales in individuals with Parkinson's disease during the turning-in-place task

These videos clearly demonstrate characteristic movement patterns, including small shuffling steps, reduced arm swing, and the sudden inability to initiate or continue movement typical signs of FoG. Freezing often appears during turns, narrow passages, or when approaching an obstacle, where the feet seem "glued" to the ground despite visible effort from the upper body to move forward.

Live Model Testing with Sensor Setup

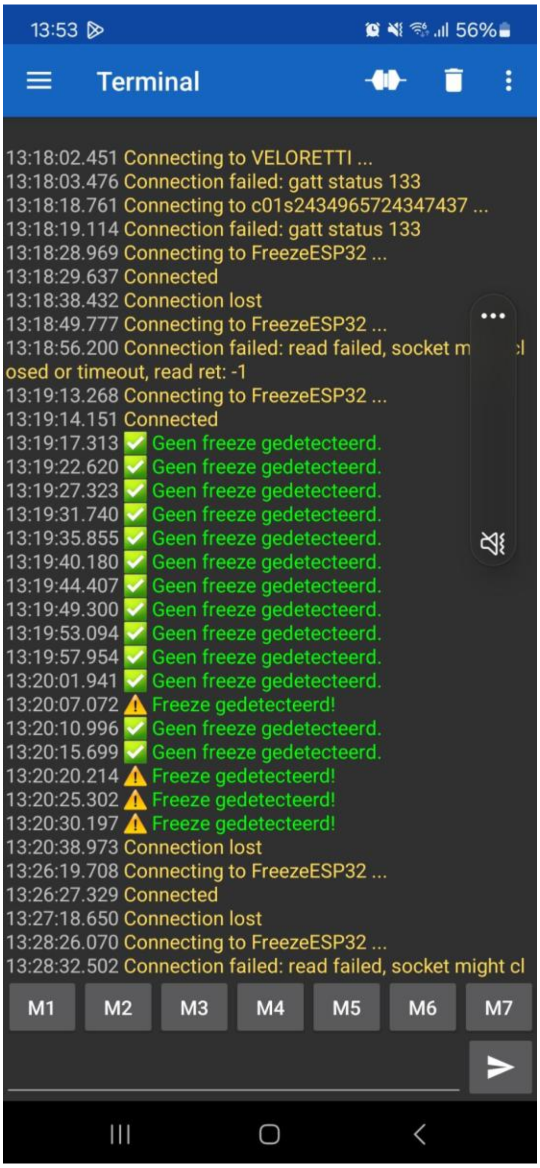
After studying these examples, the model was tested in a real-world scenario. The hardware setup consisted of a LOLIN D32 development board connected to a motion sensor (MPU 6050), all enclosed in a custom-built casing. This sensor unit was strapped to the upper and lower leg (ankle area) to simulate placement in a practical use case.

In the initial phase of testing, the system was still physically connected to a laptop via USB. This allowed live monitoring of the model's output while a test subject imitated both regular Parkinsonian walking and freezing episodes, based on the patterns observed in the reference videos. During this stage, the system successfully provided real-time detection of simulated freezing moments, which could be directly observed on the laptop.

Wireless Testing via Bluetooth Connection

In a more advanced phase of testing, the sensor system was made fully portable. A battery pack was connected to power the LOLIN board, and the model output was transmitted via Bluetooth. Using the Serial Bluetooth Terminal app, the system could now be monitored in real-time on a smartphone, eliminating the need for a constant connection to the laptop.

The device remained strapped to the leg, and further walking trials were performed. Again, Parkinsonian gait and freezing were simulated. During these tests, freezing events were accurately detected by the model and clearly displayed in the app interface, validating the system's ability to function wirelessly and in motion.


13:53
Terminal
13:18:02.451 Connecting to VELORETTI ...
13:18:03.476 Connection failed: gatt status 133
13:18:18.761 Connecting to c01s2434965724347437 ...
13:18:19.114 Connection failed: gatt status 133
13:18:28.969 Connecting to FreezeESP32 ...
13:18:29.637 Connected
13:18:38.432 Connection lost
13:18:49.777 Connecting to FreezeESP32 ...
13:18:56.200 Connection failed: read failed, socket m
osed or timeout, read ret: -1
13:19:13.268 Connecting to FreezeESP32 ...
13:19:14.151 Connected
13:19:17.313 ✅ Geen freeze gedetecteerd.
13:19:22.620 ✅ Geen freeze gedetecteerd.
13:19:27.323 ✅ Geen freeze gedetecteerd.
13:19:31.740 ✅ Geen freeze gedetecteerd.
13:19:35.855 ✅ Geen freeze gedetecteerd.
13:19:40.180 ✅ Geen freeze gedetecteerd.
13:19:44.407 ✅ Geen freeze gedetecteerd.
13:19:49.300 ✅ Geen freeze gedetecteerd.
13:19:53.094 ✅ Geen freeze gedetecteerd.
13:19:57.954 ✅ Geen freeze gedetecteerd.
13:20:01.941 ✅ Geen freeze gedetecteerd.
13:20:07.072 ⚠️ Freeze gedetecteerd!
13:20:10.996 ✅ Geen freeze gedetecteerd.
13:20:15.699 ✅ Geen freeze gedetecteerd.
13:20:20.214 ⚠️ Freeze gedetecteerd!
13:20:25.302 ⚠️ Freeze gedetecteerd!
13:20:30.197 ⚠️ Freeze gedetecteerd!
13:20:38.973 Connection lost
13:26:19.708 Connecting to FreezeESP32 ...
13:26:27.329 Connected
13:27:18.650 Connection lost
13:28:26.070 Connecting to FreezeESP32 ...
13:28:32.502 Connection failed: read failed, socket might cl
M1
M2
M3
M4
M5
M6
M7

Appendix IV: 3D-Printed Enclosure for ESP32 and IMU

To house the Lolin32 development board (which includes the ESP32), the MPU6050 IMU sensor, and the required jumper wires, a custom enclosure was designed and 3D printed. The design was created in Autodesk Fusion 360, allowing for precise placement of components and openings for connectivity.

The enclosure is printed using PETG filament (Polyethylene Terephthalate Glycol), which was chosen for its durability, temperature resistance, and slight flexibility making it ideal for protecting electronics while remaining easy to modify. Compared to PLA, PETG offers greater impact resistance and can better withstand the heat that might build up during continuous device operation.

Printing was done using PrusaSlicer version 2.9.0 and a Prusa Original Mini & Mini+ printer. The print quality and layer adhesion were excellent, producing a sturdy structure.
The form factor of the housing is tailored to fit the setup perfectly:

- The enclosure's length precisely accommodates the Lolin32 board and the IMU sensor.
- The width provides sufficient space for jumper wires, allowing them to be tucked inside neatly without stress or bending.
- On both ends of the enclosure, two cutouts are integrated on the bottom: one for connecting a battery, and another for Micro-USB access, enabling charging or data transfer while the device remains in its case.
- The top lid is currently not permanently fixed. It can be placed or removed as needed, ensuring that the reset button on the Lolin32 remains accessible, which is essential during testing and reprogramming phases.

This custom housing not only improves the portability and protection of the system, but also contributes to its overall usability as a wearable or standalone prototype for real-time Freezing of Gait detection.