M.Sc. in Mathematical Engineering - Biomedical Engineering A.Y. 2022/2023
Project Report of Statistical Learning for Healthcare Data (056867, 5 CFU)
Professors: Anna Maria Paganoni, Emanuela Ferrario

SLHD Project Report: Challenge 1

Filippo Cipriani* Andrea Puricelli† Sofia Sonnacchi‡

June 1th, 2023

Abstract

This report presents a statistical analysis of the UniZgFall dataset and explores the development of a wearable fall detection system utilizing machine learning algorithms. The system employs accelerometer data and advanced algorithms to identify patterns and distinguish between everyday activities and fall events. To enhance model performance and reduce complexity, feature selection and resampling techniques were applied. This study evaluates the performance of six machine learning models (k-Nearest Neighbour, Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest, Ensemble of LR and RF) using precision, recall, F-score, and accuracy metrics. The results show the effectiveness of the Random Forest model in accurately detecting falls while minimizing false alarms. These findings emphasize the potential of machine learning-based wearable fall detection systems to improve the well-being of the elderly, presenting a promising solution to address the growing risks of falls in an ageing society.

The source code of the entire project is available at
`SmearyTundra/SLHD-challenge`

*`filippo.cipriani@mail.polimi.it`, 10596877
†`andrea3.puricelli@mail.polimi.it`, 10632135
‡`sofia.sonnacchi@mail.polimi.it`, 10679538

Contents

1 Introduction

The rise in life expectancy observed in numerous countries has led to a growing elderly population, presenting a range of prospects and challenges [1]. While healthcare advancements have contributed to extended lifespans, they have also brought forth distinct concerns, including an increased risk of falls among the elderly [2]. According to the World Health Organization, falls rank as the second leading cause of unintentional injury-related fatalities worldwide and hospitalization rates are projected to increase by an average of 2% annually until 2030 [3].

Fall-related injuries incur significant financial costs that have a substantial impact on healthcare systems (hospitalizations, surgeries, and rehabilitative), individuals (productivity, reduced quality of life, and long-term care requirements), and society as a whole [4].

Wearable fall detection systems utilizing accelerometers and machine learning algorithms have emerged as promising solutions [5]. These systems analyze real-time data to detect anomalous changes indicative of a fall event [6]. By providing timely alerts, they can mitigate the severity of fall-related injuries.

In this report, we present the development of a classification system able to distinguish several distinct activities of daily living based on accelerometer measurements.

2 Material and Methods

2.1 Dataset presentation

The data we analyzed were gathered from 16 healthy individuals (age 21.9 ± 2.2 years, height 178.1 ± 7.8 cm, and weight 70.0 ± 12.1 kg) performing falls and various non-fall activities. Each subject was required to stage 12 types of activities of daily living (ADL) and 3 types of simulated falls while wearing an inertial sensor unit attached sideways to their waist at a belt high and in a controlled environment.

This study implements a three-class classification approach, where (as shown in Table 1) all variations of falls were considered into a single "Falls" category while the activities of walking, running, jumping, walking down the stairs, and walking up the stairs were grouped under the "Moving" category. The remaining activities of ADLs were labeled as "Other classes".

The dataset comprises 468 events, each representing an activity performed by a single participant. Each event includes the activity label and three time series measurements corresponding to the accelerometer readings on the X, Y, and Z axis. The length of these measurements differs for each event, as it depends on the duration of the simulated activity being performed.

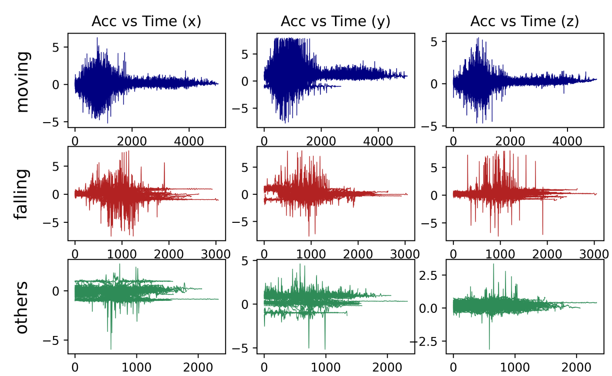

Figure 1: Time series by axis and class

Table 1: Division of the 10 original classes in 3 main categories.

Class division		
Categories	**Types**	**Total Instances**
Moving	Walking Running Jumping Walking down the stairs Walking up the stairs	230
Falls	Forward fall Sideways fall Backward fall	102
Other Classes	Lying down Inactive	136

2.2 Preprocessing

First, we examined the dataset for any missing values, and none were found. Subsequently, a resampling technique was utilized to ensure a consistent time interval for the accelerometer data.

In the original data, time was saved with a 5 digits precision *regardless of starting time and decimal precision*. If a signal was collected with a 3-digits starting time e.g. 104.47 seconds, then the recorded times would have a $5 - 3 = 2$ decimal points precision e.g. $t_1 = 104.47\text{s}, t_2 = 104.47\text{s}, t_3 = 104.48\text{s}$: not enough to accurately represent the sampling frequency of the device of 200Hz(0.005s). Thus, the signals were resampled at a frequency of 200 Hz and adjusted to begin at 0s.

Due to the major challenges posed by the differing lengths of the signals, we decided to discard a functional data analysis approach in favour of a more traditional multivariate analysis, performing feature extraction on the time series. The feature set to be extracted was the same for each axis and carefully constructed using statistical measures inspired by different areas, ranging from spectral analysis to linear fitting.

The extracted features are 20 for each axis of the time series of the observation (X,Y,Z), yielding a total of 60 features divided into three macro areas.

- Basic Statistical Measures (Max, Mean, etc.);
- Time Series Analysis inspired measures (Slope of linear fit, Zero Cross count, etc.);
- Signal Processing inspired measures (Area Under Curve, Fundamental Frequency, etc.).

The entire list of extracted features can be found in Table 5 of Appendix A.

The dataset was divided into two subsets, with 70% allocated for training purposes and the remaining 30% designated for testing, while maintaining the proportion among classes.

To ensure fair confrontation and prevent one feature from overshadowing others due to different magnitudes, we performed data scaling. By scaling the data, variables with distinct ranges were transformed to a common scale, allowing for an equal contribution of all features in the analysis. This was necessary due to developing models based on distances (e.g. KNN), to ease numerical convergence (e.g. Logistic Regression) and to allow models that performed well with features of different magnitude (e.g. Random Forest) to still work in Ensembles with the first type of models, even if scaling was not necessary for the single model.

After careful consideration, it was determined that removing outliers was not necessary for our analysis. This decision was based on a deep understanding of the dataset and the feature generation process. We are confident that these outliers are genuine observations and not the result of errors or measurement issues. Furthermore, the algorithms employed in our analysis have shown inherent robustness to outliers and better generalization. Therefore outlier removal was deemed unnecessary for our analysis.

2.3 Feature selection

To improve model performance and enhance interpretability, we performed feature selection.

We employed a feature selection process for each model. This involved removing highly correlated features (with absolute correlation value $\geq$ 0.7), leaving 33 features among all axes. After running a model, we selected a subset of the most relevant ones by using either feature importance or permutational feature importance. This reduction in the number of features allows for a reduction in the complexity of the models leading to a better generalization to unseen data and enhanced interpretability. The number of features to be selected using feature importance was determined using an elbow criterion based on accuracy and false positive rate (FPR) of the **Falling** class.

2.4 Models and Performance metrics

We developed the following classification models:

- Multinomial Logistic Regression (LR)
- Random Forests (RF)
- K-Nearest Neighbour (k-NN)
- Support Vector Machine (SVM)
- Gaussian Naive Bayes Classifier (NB)

using the Scikit-learn library. Lastly, we performed an ensemble model of RF and LR by using both soft and hard voting methods.

LR, RF, and SVM had their hyper-parameters optimized using Grid Search in order to maximize their performance.

As the number of fall data is small compared to ADLs, and given the importance of correctly classifying this class in the context of this project, the accuracy alone cannot be used to measure the classifier's performance. Thus, precision, recall, and F-score were used to assess the models' accuracy in detecting falls and minimizing false negatives and false positives.

The performance of the predictive models can be easily visualised using the confusion matrix, with a particular focus on accurately classifying falls. In the context of fall detection systems, the primary objective is to minimize false negatives, ensuring that instances of falls are correctly identified. This sensitivity is especially critical in situations where the consequences of missed fall detection can be severe, such as causing serious injuries with the need for immediate medical intervention.

However, attention should also be given to false positives, as they can trigger unnecessary alarms within fall detection systems. Such false alarms not only inconvenience users but can also undermine user acceptance and compliance. Thus, minimizing false positives is also important for maintaining user trust and upholding the reliability of the system.

3 Results

3.1 Results with 3 classes

The results of the best performing models (Logistic Regression, Random Forest, Ensemble of LR and RF) after feature selection are shown in Table 2. A detailed overview of the performance metrics and results obtained from all the models, including those that were discarded, can be found in Table 6 of Appendix B.

Overall, Random Forest emerged as the final and top-performing model among the evaluated options, taking into account factors such as interpretability and the number of selected features.

All the models exhibited an exceptional overall accuracy. The comparison of the recall metric for class Falling, essential for capturing relevant instances by correctly identifying all positive instances out of the total actual positives in the dataset, showed that Random Forest outperformed Logistic Regression. Ensemble model performed equally well: however, due to its enhanced interpretability and simplicity, Random Forest was selected as the preferred model.

By opting for such model, notable advantages are visible in terms of simplicity. This is attributed to Random Forest's feature selection process, which results in a small number of selected features.

The Confusion Matrix and ROC of the RF model are shown in Figure 2.

Table 2: Results of the three best models for the three classes.

Model performances						
Model	**Class**	**Metrics**				**Features**
		Precision	Recall	F-score	Test Acc	
LR	Moving	0.99	1.00	0.99	95.74%	13
	Falling	0.88	0.94	0.91		
	Others	0.97	0.90	0.94		
RF	Moving	1.00	1.00	1.00	95.74%	4
	Falling	0.86	0.97	0.91		
	Others	0.97	0.88	0.92		
Ensemble	Moving	0.99	1.00	0.99	96.45%	16
	Falling	0.94	0.97	0.95		
	Others	1.00	0.95	0.97		

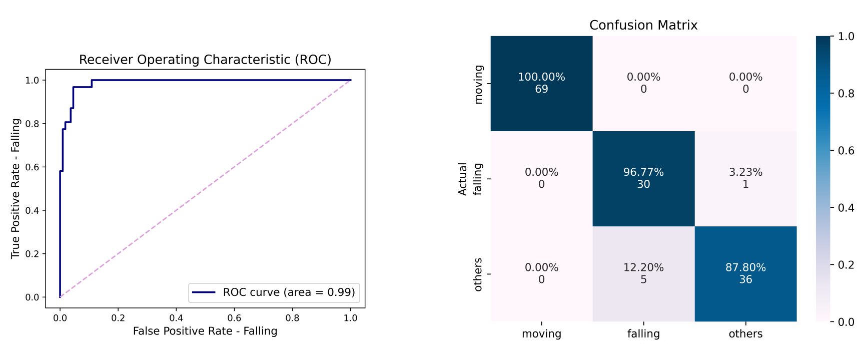

Figure 2: ROC and Confusion Matrix relative to the Random Forest model.

3.2 Final Model Interpretation

The features selected in the RF model, in decreasing order of feature importance, are:

- Fundamental Frequency (Y-axis)
- Mean (Y-axis)
- Kurtosis (X-axis)
- Min (Y-axis)

By inspecting the selected model using Partial Dependence Plots (PDP), we evaluated the effects of the four features on class prediction. The PDPs for classes Moving, Falling, and Others are shown in Figure 3.

Fundamental Frequency, the most important feature, has an enormous effect on the prediction of all classes. Low values are associated with Falling, middle values with Others, and high values with Moving. The Mean (Y-axis) values follow this pattern as well.

This statement could be reasonably explained: during falls, the Y-axis acceleration tends to be pointing downward. Moreover, the oscillation associated with Falling, and its relative fundamental frequency, could be fit in one larger, sudden movement compared to Moving, where the oscillations would be associated to a shorter, more repetitive movement.

Kurtosis (X-axis) separates well Moving from Others, and Min (Y-axis) separates well Falling from Others, but both have problems with separating a third class.

3.3 Results with 10 classes

Additionally, as part of our research, we conducted a classification task involving the ten categories outlined in the original challenge. Employing the same data pipeline and techniques utilized for the three-class classification, we proceeded with the analysis. The performance results for the LR and RF were satisfactory, indicating their effectiveness in tackling the task. However, the Ensemble model of LR and RF employing soft voting exhibited the overall best performance. The detailed results for this model are shown in Table 3. The outcome reaffirms the effectiveness and robustness of our analysis methods, showing their ability to handle a broader range of categories within the dataset.

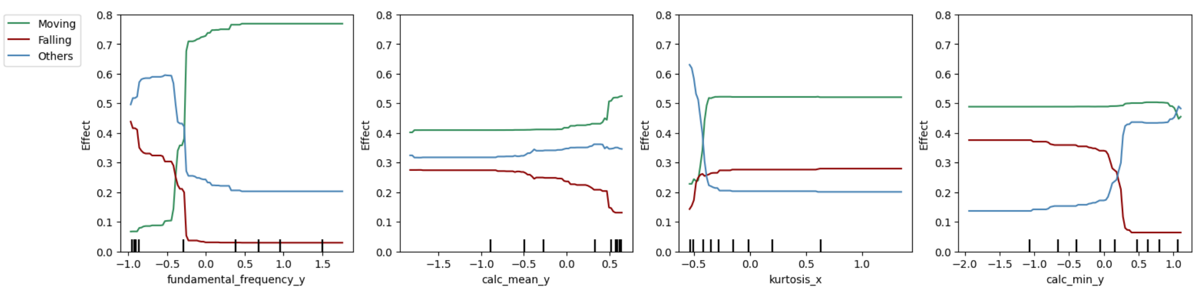

Figure 3: Partial dependence plots for the Random Forest

Table 3: Results of the Ensemble model for all 10 classes.

Model performances						
Model	**Class**		**Metrics**			**Features**
		Precision	Recall	F-score	CV-Acc	
	Walking	0.95	0.95	0.95		
	Running	1.00	1.00	1.00		
	Jumping	1.00	0.95	0.97		
	Walking down	1.00	0.75	0.86		
Ensemble	Walking up	1.00	1.00	1.00	92.74%	14
	Forward fall	0.89	0.80	0.84		
	Sideways fall	1.00	1.00	1.00		
	Backward fall	0.82	0.90	0.86		
	Lying down	0.90	0.90	0.90		
	Inactive	0.91	0.97	0.94		

Table 4: Features selected for the Ensemble model for 10 classes.

Features selected		
Axis	**Features**	
X	Area Under Curve	Min
	Std	Mean Abs Difference
	Median Abs Deviation	Zero Cross Count
Y	Max	Min
	Fundamental frequency	Zero Cross Count
	Mean Abs Deviation	Mean Difference
Z	Area Under Curve	Zero Cross Count

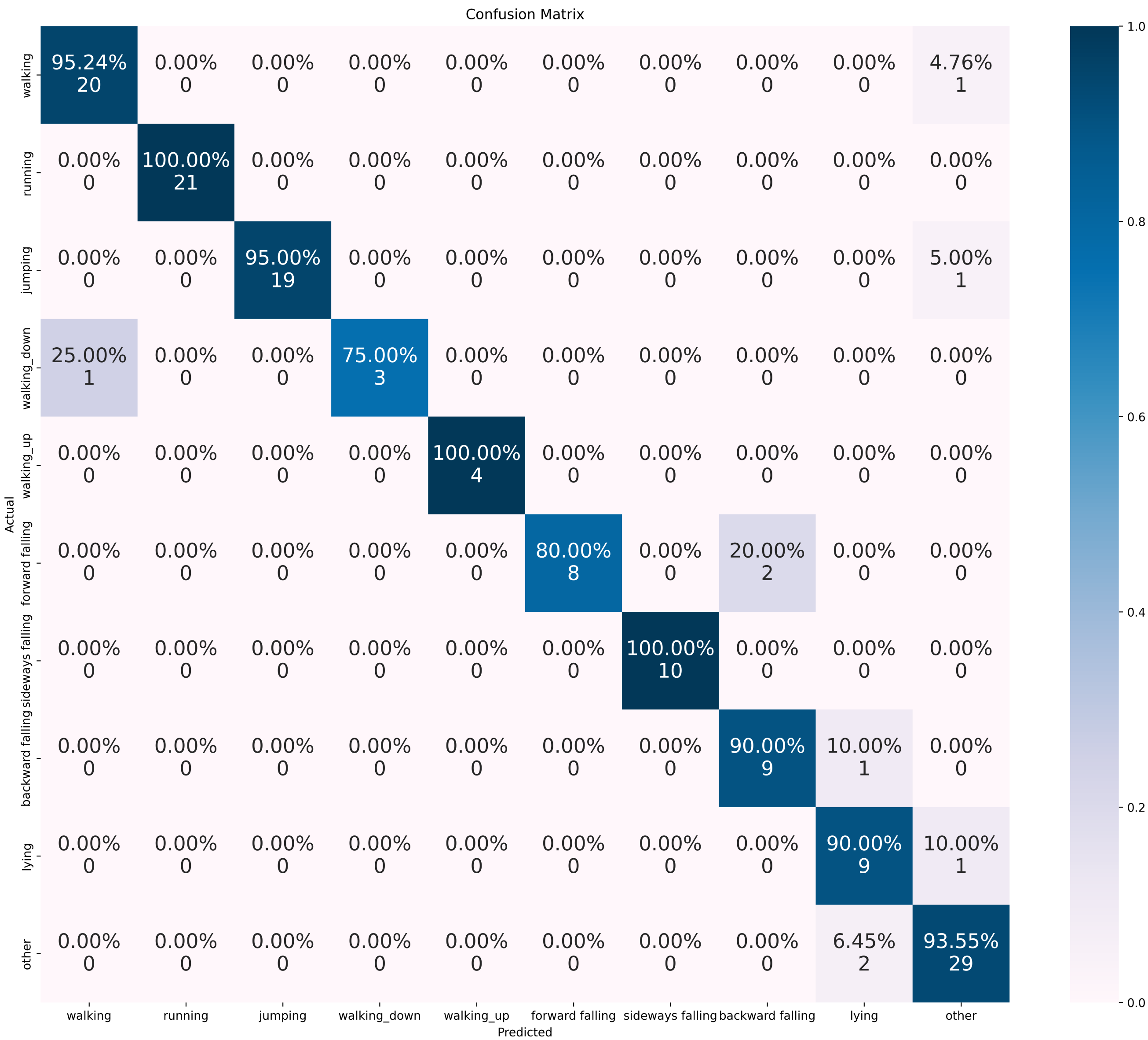

Figure 4: Confusion Matrix of the Ensemble Model with 10 classes.

4 Conclusions

In conclusion, this study has focused on the UniZgFall dataset and the development of a wearable fall detection system using machine learning algorithms. Through the evaluation of various techniques such as Logistic Regression, Random Forests, K-Nearest Neighbour, Support Vector Machine, and Gaussian Naive Bayes Classifier, we have demonstrated the potential of machine learning in accurately detecting falls while minimizing false negatives and false positives. Particularly, Random Forests emerged as the top-performing model, exhibiting remarkable precision, recall, and F-score for fall classification, resulting in an impressive overall accuracy rate of 95.74%.

Lastly, it is crucial to stress the fundamental role played by the feature extraction and selection steps in the development of our model. The extraction of meaningful features allowed us to obtain a notable separation of the classes (as shown in Figure 5) with minimal effort, achieving good results using relatively simple and interpretable models.

Particularly, the Y-axis features, associated with vertical movement, appeared to be the most important. Among them, the **fundamental frequency** feature consistently scored as the most important feature throughout all models, managing to encode well the spectral content of the signals.

Overall, the developed classification system, based on accelerometer measurements, shows great promise in accurately identifying and distinguishing different activities of daily living, especially fall events. This advancement in technology holds significant potential for enhancing the safety and quality of life for the elderly population. Further research and development in the field of wearable fall detection systems are crucial for mitigating the adverse consequences of falls and ensuring the well-being of older adults in an ageing society. By continuously improving and refining these systems, we can contribute to a safer and more secure environment for older adults.

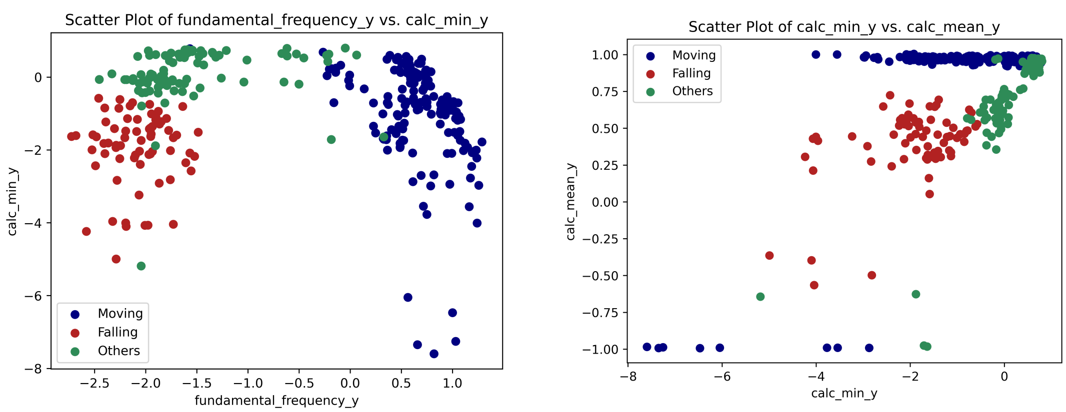

Figure 5: Scatter plot using RF model features.

References

[1] URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health#:~:text=By%5C%202030%5C%2C%5C%201%5C%20in%5C%206,will%5C%20double%5C%20(2.1%5C%20billion.

[2] Vianda S Stel et al. 'Consequences of falling in older men and women and risk factors for health service use and functional decline'. en. In: *Age Ageing* 33.1 (Jan. 2004), pp. 58–65.

[3] URL: https://www.who.int/news-room/fact-sheets/detail/falls.

[4] Curtis S Florence et al. 'Medical costs of fatal and nonfatal falls in older adults'. en. In: *J. Am. Geriatr. Soc.* 66.4 (Apr. 2018), pp. 693–698.

[5] Raul Igual, Carlos Medrano and Inmaculada Plaza. 'Challenges, issues and trends in fall detection systems'. en. In: *Biomed. Eng. Online* 12.1 (July 2013), p. 66.

[6] Eduardo Casilari, Jose A. Santoyo-Ramón and Jose M. Cano-García. 'UMAFall: A Multisensor Dataset for the Research on Automatic Fall Detection'. In: *Procedia Computer Science* 110 (2017). 14th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2017) / 12th International Conference on Future Networks and Communications (FNC 2017) / Affiliated Workshops, pp. 32–39. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2017.06.110. URL: https://www.sciencedirect.com/science/article/pii/S1877050917312899.

A Complete list of extracted features

Table 5: Features extracted for each axis

Feature table		
Area	**Feature Name**	**Description**
Basic statistical measures	Max	Minimum value of the acceleration
	Min	Maximum value of the acceleration
	Mean	Mean value of the acceleration
	Median	Median value of the acceleration
	Std	Standard deviation of the acceleration
	Variance	Variance deviation of the acceleration
Time Series Analysis	Kurtosis	"Tailedness" of the distribution
	Mean (Abs) Deviation	Mean (absolute) deviation of the signal
	Median Abs Deviation	Value of the median absolute deviation of the signal
	Mean (Abs) Difference	Mean (absolute) difference of the signal
	Median Abs Difference	Median absolute difference of the signal
	Slope	Slope obtained by fitting a linear equation to the observed data
	Sum Abs Difference	Sum of absolute differences of the signal
	Zero Cross Count	Zero-crossing rate of the signal
Signal Processing	Area Under Curve	Area under the curve of the signal computed with trapezoid rule
	Fundamental Frequency	Fundamental frequency of the signal.
	Peak to Peak distance	Peak to peak distance of the signal
	Total Energy	Total energy of the signal

B Complete list of model performances

Table 6: Results of all the models for the three classes.

Model performances						
Model	**Class**	**Metrics**				**Features**
		Precision	Recall	F-score	Test Acc	
	Moving	0.99	1.00	0.99		
LR	Falling	0.88	0.94	0.91	95.74%	13
	Others	0.97	0.90	0.94		
	Moving	1.00	1.00	1.00		
RF	Falling	0.86	0.97	0.91	95.74%	4
	Others	0.97	0.88	0.92		
	Moving	0.99	1.00	0.99		
KNN	Falling	0.85	0.94	0.89	93.26%	4
	Others	0.95	0.85	0.90		
	Moving	1.00	0.99	0.99		
SVM	Falling	0.93	0.90	0.92	96.00%	6
	Others	0.91	0.95	0.93		
	Moving	1.00	1.00	1.00		
NB	Falling	0.88	0.90	0.89	95.71%	7
	Others	0.93	0.90	0.91		
	Moving	0.99	1.00	0.99		
Ensemble	Falling	0.94	0.97	0.95	96.45%	16
	Others	1.00	0.95	0.97		