

Deep Learning for Frailty Classification Using Raw Inertial Sensor Gait Data

Arslan Amjad¹^a, Agnieszka Szczęśna¹^b and Monika Błaszczyszyn²^c

¹Department of Computer Graphics, Vision and Digital Systems, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

²Department of Physical Education and Sport, Faculty of Physical Education and Physiotherapy, Opole University of Technology, Prószkowska 76, 45-758 Opole, Poland

Keywords: Deep Learning, Frailty, Gait, Healthcare, IMU Sensor.


Abstract: The Frailty is a significant health issue in older adults that increases the risk of disability, decline in physiologic reserve and function, hospitalization, and even death. The social and economic impact of frailty increased due to the higher healthcare costs and the medical resources. The intervention of early frailty detection can prevent its progression and delay the disability, ultimately improving the quality of life in the elderly population. This study aims to propose a frailty classification system based on gait data collected from an Inertial Measurement Unit (IMU) sensor with the utilization of the Deep Learning (DL) approach. The individual's frailty status is classified as robust, pre-frail, or frail. A publicly available dataset of 163 participants was utilized to analyze the raw gait signals and find the most effective DL for extracting gait patterns for frailty classification. DeepConvLSTM model has shown effective performance on raw IMU gait data with a balanced accuracy, precision, recall, and F1-score of 91%. The results show that the proposed methodology successfully classifies the pre-frail individuals, which demonstrate its potential to enhance frailty detection and intervention in clinical settings. This ultimately provides an improved healthcare system and a quality of life in elderly populations.


1 INTRODUCTION


The number of elderly individuals is increasing dramatically as the world's population grows (Hernigou et al., 2024). World Health Organization data show this demographic trend: among the 8.1 billion population, people aged 60 years and older will become 1.4 billion by 2030 and 2.1 billion by 2050 (Sun et al., 2024; United Nation, 2024; World Health, 2024). Frailty is one of the most common and fatal disorders in the elderly population (Hakeem et al., 2023). Physical frailty is a multidimensional condition that is defined as a decline in physiological reserves. This makes older persons more vulnerable to stresses and increases the possibility of negative health effects (Kojima et al., 2018). Considering its consequences link to increased illness, disability, and death, this raises a significant public health concern

(Pasieczna et al., 2023). To reduce the burden of frailty on the healthcare system and enhance the quality of life for the aging population, it is essential to address it through early identification, precise assessment, and effective management.

In order to reduce the risk of frailty among older adults, it is essential to develop an objective healthcare solution. Traditional clinical frailty assessment methods are time-consuming and need specialized equipment and experienced healthcare personnel (Obbia et al., 2020). To solve this issue, wearable technology and advanced Machine Learning (ML) algorithms have emerged as a potential solution (Fan et al., 2023; Minici et al., 2022). These technologies provide continuous, real-time remote monitoring, allowing for early identification and classification of frailty stages.

^a <https://orcid.org/0000-0002-6711-4382>

^b <https://orcid.org/0000-0002-4354-8258>

^c <https://orcid.org/0000-0002-1723-4001>

This study offers a smart stack of an Inertial Measurement Unit (IMU) sensor and Deep Learning (DL) technologies as a promising solution for frailty classification. An IMU sensor was worn by each participant to collect their raw IMU gait signals. These signals were then pre-processed and converted to the frequency domain in order to capture the underlying patterns. Following this, a Deep Learning (DL) algorithm was used to extract intrinsic gait parameters for frailty classification into frail, pre-frail, or robust stages.

This study has two main objectives: 1) analyzing the raw IMU gait signals for frailty classification and 2) finding the most effective DL algorithm for frailty classification using raw IMU gait data. The ultimate goal of this research is to develop an early frailty detection system that will detect the frailty stage timely and prevent the frailty from progressing in older adults. Early detection of frailty allows individuals to seek medical advice and take appropriate measures, which lowers the total cost of healthcare for society. This study proposed an intelligent frailty assessment system that will be expanded into a real-time application, increasing its use and impact in clinical settings.

The paper is organized as follows: Section 2 outlines the relevant literature work; Section 3 explains the research methodology, including the dataset and the application of the DL algorithm for frailty classification; Section 4 provides the results with discussion; and finally, Section 5 concludes the paper and suggests directions for future work.

2 LITERATURE REVIEW

For clinical gait analysis, the most commonly used DL algorithms in the previous studies are: Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Auto-Encoders (AE). These algorithms became popular due to their ability to analyze complex time-series data and automatically extract features from raw IMU sensor data, making them appropriate for applications such as gait analysis.

The studies (García et al., 2022; Kou et al., 2024), and (Li et al., 2024) classify the fall risks using CNN-LSTM and CNN-BiLSTM algorithms. The study (García et al., 2022) used a 3-D IMU device placed on a wrist and leveraged the CNN-LSTM model to achieve an accuracy of 93.60%. Whereas the study (Kou et al., 2024) achieved an F1-score of 95.18% and the study (Li et al., 2024) obtained an accuracy of 98.40%.

The study (Kamran et al., 2021) explored the utility of 1-D CNN for automatically assessing balance using data from a single IMU worn on the lower back. They also compared the results with handcrafted features. DL provided significant results with an AUROC of 0.81. Another study (Hauth et al., 2021) utilized three IMU sensors while performing daily activities. The BiLSTM model outperformed with an AUROC score of 0.87.

Another approach used in the previous studies (Butt et al., 2020; San-Segundo et al., 2019; Sánchez-DelaCruz et al., 2019) is the transformation of raw IMU signals into images. This structured format of input leverages the DL algorithms to extract more enhanced features. The overview of previous studies that utilized raw IMU gait signals with DL algorithms is shown in Table 1.

Table 1: Overview of relevant studies that utilized raw IMU gait data for frailty analysis.

Ref.	Algorithms	Task	Outcomes
(García et al., 2022)	CNN-LSTM	Falls risks	Accuracy = 93.60%
(Kou et al., 2024)	CNN-LSTM		F1-score = 95.18%
(Li et al., 2024)	CNN-BiLSTM		Accuracy = 98.40%
(Kamran et al., 2021)	1-D CNN	Assess balance	AUROC = 0.81
(Hauth et al., 2021)	BiLSTM		AUROC = 0.87

3 METHOD

The research methodology consists of key steps, which include: 1) the analysis of raw IMU gait data and assigning the frailty status label to each participant; 2) pre-processing of the raw IMU data and data formatting using sliding window technique and wavelet transformation; and 3) implementation of the DL algorithm to classify the frailty into frail, pre-frail, or robust stages. The research methodology is illustrated in Figure 1.

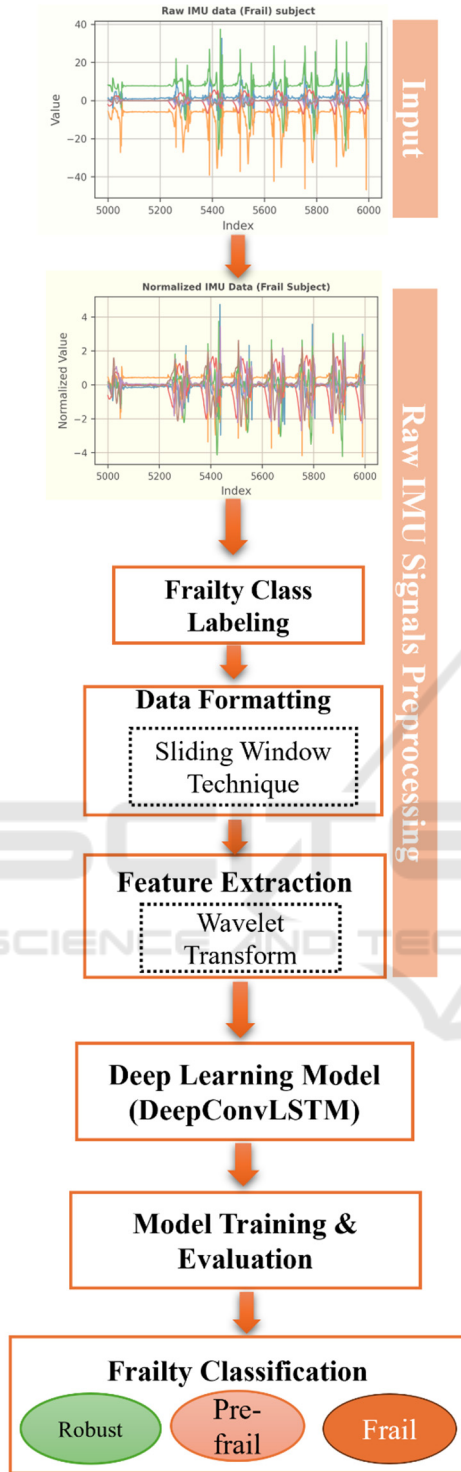


Figure 1: Research methodology.

3.1 Dataset

The GSTRIDE (García-de-Villa et al., 2023) dataset was utilized in this study. It's a publicly available

dataset that consists of 163 (45 men and 118 women) older adults. Their ages range from 70 to 98 years and an average weight of 64.2 to 77.3 kg. The list of parameters available in the GSTRIDE database is shown in Table 2.

Table 2: List of Parameters available in GSTRIDE database.

Category	Parameter	Description
Socio Demographic	Age	Average age of the subjects (years)
	Gender	Male/Female
	Living Environment	Type of living environment (e.g., Home, Assisted Living)
Anatomical	Weight	Average weight of the subjects (kg)
	Height	Average height of the subjects (cm)
	Body Mass Index (BMI)	Average BMI of the subjects (kg/m ²)
Cognitive	Global Deterioration Scale (GDS) Index	Average GDS index of the subjects (scale 1-7)
Functional	4-metre Gait Speed Test	Average time taken (seconds)
	Hand Grip Strength	Average hand grip strength (kg)
	Timed Up and Go (TUG)	Average time taken for TUG test (seconds)
	Short Physical Performance Battery (SPPB)	Average SPPB score
	Short Falls Efficacy Scale International (FES-I)	Average FES-I score

For raw IMU signal acquisition, two IMU sensors (CSIC and Gaitup) were used, with only one sensor worn on the foot of each participant during 15 minutes of gait (García-de-Villa et al., 2023). The reason for using two different sensors with varying frequencies was to assess the effect of varying configurations of sensors on the spatio-temporal estimation. The authors reported a minimal effect of these varying configurations of sensors on spatio-temporal estimation, although there was a slight

variation in the accuracy of estimation (García-Villamil et al., 2021).

3.2 Class Labelling of Participants

The Standardized Fried's phenotype (Fried et al., 2001) test was adopted to label the frailty status of each participant. In this test, the Frailty Index (FI) score is calculated using the five parameters. Each parameter's score is assigned a score of 0 or 1. The final FI score is calculated by summing the score of all parameters (ranges from 0 to 5) (García-de-Villa et al., 2023). The class label is assigned to each participant based on FI score. If the FI score is 0, then the frailty class label is "Robust". If the FI score is 1 or 2, then the frailty class label is "Pre-frail", otherwise the frailty label is "Frail", as shown in Figure 2.

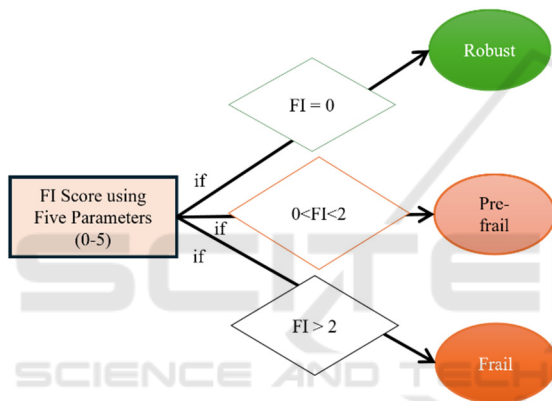


Figure 2: Criteria for assigning the frailty label to each participant.

3.3 Data Pre-Processing

After raw IMU signal data acquisition and labelling, pre-processing became a critical step for further analysis. In this stage, outliers were removed from raw IMU signals, and the signals were normalized using the "StandardScaler" function. The data was then segmented into smaller chunks using the "Sliding Window" technique (Jaén-Vargas et al., 2022), which allows the extraction of spatio-temporal features from the time-series IMU signals. The window size set in this study was 200, with a stride of 50. Next, a wavelet transformation is applied to each segment to capture both time and frequency domain features. The "pywt.wavedec" function was used for Daubechies wavelet of order 1 ("db1"). This frequency transformation is suitable for raw IMU signals to capture the sharp changes in the signals

(Chakraborty et al., 2020; Kuduz et al., 2023; Michau et al., 2022).

At the end, the segmented windows were divided into 75% training set. The remaining segments were divided into equal sets for validation and testing. The code is implemented in Python, version 3.5, using Spyder as the development environment.

3.4 Deep Learning (DL) Algorithm Architecture

The input data is ready after pre-processing steps. It can be input to a DL algorithm for frailty classification. The DeepConvLSTM (Ordóñez et al., 2016) algorithm was utilized for this purpose. The model consists of a convolutional layer with Long Short-Term Memory (LSTM) layers to capture both the spatial and temporal (spatio-temporal) features in raw IMU signals, which makes it an effective algorithm for frailty classification.

Two different DeepConvLSTM models were created and trained on the training dataset. The best model was selected based on high accuracy and minimum losses on both training and validation datasets. After finalizing the training process, the best model's hyperparameters were saved and tested on the test dataset. The models were created using an open-source Python's library, McFly (van Kuppevelt et al., 2020).

In this study, the architecture of the best DeepConvLSTM model was initialized with a "BatchNormalization" layer followed by a reshape operation. Following this, a 2D convolutional layer with 54 filters was applied, followed by normalized and activated layers. After convolution, the resulting tensor is reshaped to prepare it for recurrent processing. Mathematically, the convolutional process is defined as: for an input $\mathbf{x} \in \mathbb{R}^{T \times W \times H \times C}$ (where T is time, W is width, H is height, and C is channels), the convolutional process is depicted in (1).

$$\mathbf{x}_{Conv} = Conv^{2D}(\mathbf{x}) \quad (1)$$

$Conv^{2D}$ represents the 2D convolutional operation in the model; the overall equation of the convolutional process with filters \mathbf{F} is:

$$\mathbf{x}_{conv}[t, w, h, c] = \sum_{i=0}^{F_W-1} \sum_{j=0}^{F_H-1} \sum_{k=0}^{F_C-1} \mathbf{x}[t, w+i, h+j, k] \cdot \mathbf{F}[i, j, k, c] + b[c] \quad (2)$$

Table 3: DeepConvLSTM model performance on test dataset.

	DeepConvLSTM performance (Test Set)				
	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>	<i>Accuracy</i>
Robust	0.94	0.93	0.93	39063	0.91
Pre-frail	0.92	0.90	0.91	38269	
Frail	0.67	0.85	0.75	6541	
Weighted Avg.	0.91	0.91	0.91	83873	

In (2), the F_w and F_H are filter weight and height, respectively. Whereas the F_c represents input channels and bias is represented with $b[c]$.

The convolutional operation results were input into the stack of LSTM layers. Four LSTM layers were used in the model with 29, 95, 94, and 46 units, respectively. After that, the dropout layer was added to prevent the model overfitting. The model concluded with the “TimeDistributed” layer with 3 units followed by “softmax” activation for the classification task. The final description of a model can be represented as: Conv(54)–BN–LSTM(29)–LSTM(95)–LSTM(94)–LSTM(46)–D–TD(3)–S.

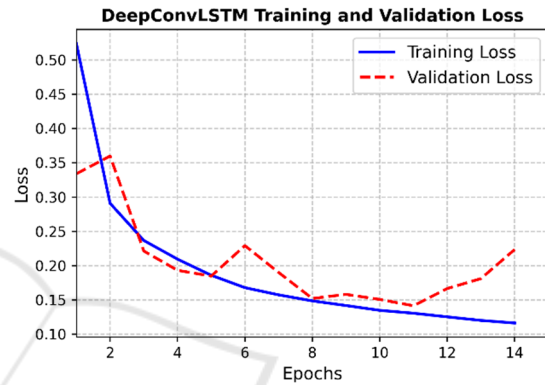


Figure 3: Training and validation losses of DeepConvLSTM in training phase.

4 RESULTS AND DISCUSSION

The evaluation criteria in this study consist of two phases. In the first phase, the two created DeepConvLSTM models were evaluated in the training process based on their best training and validation accuracy and minimum losses, respectively. The second phase evaluated the best selected DeepConvLSTM model on the testing dataset using metrics such as accuracy, precision, recall, and F1-score (Wasikowski et al., 2010).

In the training phase, the hyperparameters of DeepConvLSTM models were fine-tuned using 25 epochs with a batch size of 64 and a stopping patience of 3. The hyperparameters of the best model were reported as a learning rate of 0.0268, a regularization rate of 0.0004, and convolutional filters and LSTM dimensions as 54 and (29, 95, 94, 46), respectively. The model achieved training and validation accuracy of 95.18% and 94.14% with corresponding losses of 0.1163 and 0.1415, respectively, as shown in Figure 3.

The DeepConvLSTM model’s performance on test data is reported in Table 3. It shows that the DeepConvLSTM achieved an accuracy of 91% on test data. Whereas the overall frailty stage-wise confusion matrix is depicted in Figure 4.

The results in Table 3 suggested that the model effectively classified the pre-frail and robust individuals but reported low precision in the case of the frail class. This is due to the highly class-imbalanced, as the frail class has fewer instances, which may overlap the features with other classes. This problem can be overcome by adding more frail instances utilizing data augmentation techniques to synthetically increase the number of frail samples or applying class-weighted loss functions and oversampling methods like Synthetic Minority Over-Sampling Technique (SMOTE) (Hosseini et al., 2024) during model training. However, this study used raw IMU sensor signals as input, keeping the original data with its spatio-temporal properties. This ensures the effectiveness of the DL model for frailty classification.

robust	36202	2228	633
pre-frail	1970	34258	2041
frail	434	571	5536
	robust	pre-frail	frail

Figure 4: Confusion matrix represents the performance of DeepConvLSTM model.

Accurate classification is a major concern in clinical settings, as it directly influences patient care and intervention strategies. Overall, the DL model performed effectively; better performance on frail individuals will enhance clinical decision-making and personalized care.

The proposed approach may have some challenges when applied in the clinical healthcare system. These challenges include maintaining the privacy of data, facilitating real-time processing with wearable IMU sensors, and smoothly integrating into clinical workflows. Furthermore, it is crucial to validate the system in real-world settings and achieve generalizability across a variety of demographics. The method's potential for the early frailty detection task is highlighted by its adaptability to diverse operational circumstances and scalability to multiple sensor configurations.

5 CONCLUSION

In the world of a growing elderly population, frailty is an important factor in the adverse health outcomes among elders. Early and accurate detection of frailty can significantly enhance clinical decision-making, leading to better patient care and management.

This study proposed a sensor-based approach with a DL algorithm to classify the frailty into robust, pre-frail, or frail stages. The DeepConvLSTM model demonstrated its effectiveness in frailty classification using raw IMU sensor data, with an overall accuracy of 91%. The performance of the DL model has shown its potential to develop a frailty classification system that depicts the real-world clinical scenario.

The limitations of this research work are: 1) The small size of the dataset limited the performance of the DL model; and 2) A diverse dataset and the

selection of features may also affect the DL performance. Future studies should focus on the diverse types of sensors for the data collection. There is also a need to develop a real-time application to monitor the frailty status in a real-world clinical environment.

REFERENCES

- Butt, A. H., Cavallo, F., Maremmani, C., & Rovini, E. (2020). Biomechanical parameters assessment for the classification of Parkinson Disease using Bidirectional Long Short-Term Memory. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 5761–5764. doi: 10.1109/EMBC44109.2020.9176051
- Chakraborty, J., & Nandy, A. (2020). Discrete wavelet transform based data representation in deep neural network for gait abnormality detection. *Biomedical Signal Processing and Control*, 62, 102076. doi: 10.1016/j.bspc.2020.102076
- Fan, S., Ye, J., Xu, Q., Peng, R., Hu, B., Pei, Z., Yang, Z., & Xu, F. (2023). Digital health technology combining wearable gait sensors and machine learning improve the accuracy in prediction of frailty. *Frontiers in Public Health*, 11. doi: 10.3389/fpubh.2023.1169083
- Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., Seeman, T., Tracy, R., Kop, W. J., Burke, G., & McBurnie, M. A. (2001). Frailty in Older Adults: Evidence for a Phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 56(3), M146–M157. doi: 10.1093/gerona/56.3.M146
- García-de-Villa, S., Neira, G. G.-V., Álvarez, M. N., Huertas-Hoyas, E., Ruiz, L. R., Del-Ama, A. J., Sánchez, M. C. R., & Jiménez, A. R. (2023). A database with frailty, functional and inertial gait metrics for the research of fall causes in older adults. *Scientific Data*, 10(1), 566. doi: 10.1038/s41597-023-02428-0
- García-Villamil, G., Ruiz, L., Jiménez, A. R., Granja, F. S., & Rodríguez-Sánchez, M. C. (2021). Influence of IMU's Measurement Noise on the Accuracy of Stride-Length Estimation for Gait Analysis. *IPIN-WiP*. Retrieved from <https://api.semanticscholar.org/CorpusID:247314771>
- García, E., Villar, M., Fañez, M., Villar, J. R., de la Cal, E., & Cho, S.-B. (2022). Towards effective detection of elderly falls with CNN-LSTM neural networks. *Neurocomputing*, 500, 231–240. doi: 10.1016/j.neucom.2021.06.102
- Hakeem, F. F., Maharani, A., Todd, C., & O'Neill, T. W. (2023). Development, validation and performance of laboratory frailty indices: A scoping review. *Archives of Gerontology and Geriatrics*, 111, 104995. doi: 10.1016/j.archger.2023.104995
- Hauth, J., Jabri, S., Kamran, F., Feleke, E. W., Nigusie, K., Ojeda, L. V., Handelzalts, S., Nyquist, L., Alexander, N. B., Huan, X., Wiens, J., & Sienko, K. H. (2021).

- Automated Loss-of-Balance Event Identification in Older Adults at Risk of Falls during Real-World Walking Using Wearable Inertial Measurement Units. *Sensors*, 21(14), 4661. doi: 10.3390/s21144661
- Hernigou, P., Bumbasirevic, M., Pecina, M., & Scarlat, M. M. (2024). Eight billion people, sixteen billion hip joints today: are future orthopedists prepared to treat a world of ultra-old patients and centenarians in 2050? *International Orthopaedics*, 48(8), 1939–1944. doi: 10.1007/s00264-024-06245-x
- Hosseini, I., Rojas, R. F., & Ghahramani, M. (2024). Fall Risk Assessment Using Single IMU. *2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 1–6. doi: 10.1109/MeMeA60663.2024.10596880
- Jaén-Vargas, M., Reyes Leiva, K. M., Fernandes, F., Barroso Gonçalves, S., Tavares Silva, M., Lopes, D. S., & Serrano Olmedo, J. J. (2022). Effects of sliding window variation in the performance of acceleration-based human activity recognition using deep learning models. *PeerJ Computer Science*, 8, e1052. doi: 10.7717/peerj-cs.1052
- Kamran, F., Harrold, K., Zwier, J., Carender, W., Bao, T., Sienko, K. H., & Wiens, J. (2021). Automatically evaluating balance using machine learning and data from a single inertial measurement unit. *Journal of NeuroEngineering and Rehabilitation*, 18(1), 114. doi: 10.1186/s12984-021-00894-4
- Kojima, G., Iliffe, S., & Walters, K. (2018). Frailty index as a predictor of mortality: a systematic review and meta-analysis. *Age and Ageing*, 47(2), 193–200. doi: 10.1093/ageing/afx162
- Kou, J., Xu, X., Ni, X., Ma, S., & Guo, L. (2024). Fall-risk assessment of aged workers using wearable inertial measurement units based on machine learning. *Safety Science*, 176, 106551. doi: 10.1016/j.ssci.2024.106551
- Kuduz, H., & Kaçar, F. (2023). A deep learning approach for human gait recognition from time-frequency analysis images of inertial measurement unit signal. *International Journal of Applied Methods in Electronics and Computers*. doi: 10.58190/ijamec.2023.44
- Li, C., Cai, Y., Li, Y., & Zhang, P. (2024). Fusion of Dual Sensor Features for Fall Risk Assessment with Improved Attention Mechanism. *Traitement Du Signal*, 41(1), 73–83. doi: 10.18280/ts.410106
- Michau, G., Frusque, G., & Fink, O. (2022). Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series. *Proceedings of the National Academy of Sciences*, 119(8). doi: 10.1073/pnas.2106598119
- Minici, D., Cola, G., Giordano, A., Antoci, S., Girardi, E., Bari, M. Di, & Avvenuti, M. (2022). Towards Automated Assessment of Frailty Status Using a Wrist-Worn Device. *IEEE Journal of Biomedical and Health Informatics*, 26(3), 1013–1022. doi: 10.1109/JBHI.2021.3100979
- Obbia, P., Graham, C., Duffy, F. J. R., & Gobbens, R. J. J. (2020). Preventing frailty in older people: An exploration of primary care professionals' experiences. *International Journal of Older People Nursing*, 15(2). doi: 10.1111/opn.12297
- Ordóñez, F., & Roggen, D. (2016). Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors*, 16(1), 115. doi: 10.3390/s16010115
- Pasieczna, A. H., Szczepanowski, R., Sobiecki, J., Katarzyniak, R., Uchmanowicz, I., Gobbens, R. J. J., Kahsin, A., & Dixit, A. (2023). Importance analysis of psychosociological variables in frailty syndrome in heart failure patients using machine learning approach. *Scientific Reports*, 13(1), 7782. doi: 10.1038/s41598-023-35037-3
- San-Segundo, R., Navarro-Hellín, H., Torres-Sánchez, R., Hodgins, J., & De la Torre, F. (2019). Increasing Robustness in the Detection of Freezing of Gait in Parkinson's Disease. *Electronics*, 8(2), 119. doi: 10.3390/electronics8020119
- Sánchez-DelaCruz, E., Weber, R., Biswal, R. R., Mejía, J., Hernández-Chan, G., & Gómez-Pozos, H. (2019). Gait Biomarkers Classification by Combining Assembled Algorithms and Deep Learning: Results of a Local Study. *Computational and Mathematical Methods in Medicine*, 2019, 1–14. doi: 10.1155/2019/3515268
- Sun, Q., Xia, X., & He, F. (2024). Longitudinal association between Body mass index (BMI), BMI trajectories and the risk of frailty among older adults: A systematic review and meta-analysis of prospective cohort studies. *Archives of Gerontology and Geriatrics*, 124, 105467. doi: 10.1016/j.archger.2024.105467
- United Nation. (2024). *World Population Prospects 2024*. Retrieved from <https://population.un.org/wpp/>
- van Kuppevelt, D., Meijer, C., Huber, F., van der Ploeg, A., Georgievska, S., & van Hees, V. T. (2020). Mcfly: Automated deep learning on time series. *SoftwareX*, 12, 100548. doi: 10.1016/j.softx.2020.100548
- Wasikowski, M., & Chen, X. (2010). Combating the Small Sample Class Imbalance Problem Using Feature Selection. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1388–1400. doi: 10.1109/TKDE.2009.187
- World Health, O. (2024). *Ageing*. Retrieved from https://www.who.int/health-topics/ageing#tab=tab_1.