Older Adult Fall Risk Prediction with Deep Learning and Timed Up and Go (TUG) Test Data

Josu Maiora *, Chloe Rezola-Pardo, Guillermo García, Begoña Sanz, Manuel Graña

Posted Date: 14 September 2024

doi: 10.20944/preprints202409.1021.v1

Keywords: Inertial sensors; fall prediction; fall risk assessment; deep learning; machine learning

Article

Older Adult Fall Risk Prediction with Deep Learning and Timed Up and Go (TUG) Test Data

Josu Maiora [1,2], Chloe Rezola-Pardo [3], Guillermo Garcia [4], Begoña Sanz [3,5] and Manuel Graña [2]

[1] Electronic Technology Department Faculty of Engineering of Gipuzkoa University of the Basque Country San Sebastian, Spain
[2] Computational Intelligence Group Department of CCIA University of the Basque Country San Sebastian, Spain
[3] Department of Physiology University of the Basque Country Leioa (Bizkaia) Spain; chloe.rezola@ehu.eus
[4] Systems and Automation Department, Faculty of Engineering of Gipuzkoa University of the Basque Country San Sebastian, Spain
[5] Biobizkaia Health Research Institute, Barakaldo (Bizkaia) Spain
* Correspondence: j.maiora@ehu.eus; Tel.: +34943018617

Abstract: Falls are a major health hazard for older adults, therefore, in the context of an aging population, predicting the risk of a patient suffering falls in the near future is of great impact for health care systems. Currently, the standard prospective fall risk assessment instrument is the relies on a set of clinical and functional mobility assessment tools, one of them is the Timed Up and Go (TUG) test. Recently, wearable inertial measurement units (IMUs) have been proposed to capture motion data that would allow to build estimates of fall risk. The hypothesis of this study is that the data gathered from IMU readings while the patient is performing the TUG test can be used to build a predictive model that would provide an estimate of the probability of suffering a fall in the near future, i.e., assessing prospective fall risk. This study applies deep learning convolutional neural networks (CNN) and recurrent neural networks (RNN) to build such predictive models based on features extracted from IMU data acquired during TUG test realizations. Data were obtained from a cohort of 106 older adults wearing a wireless IMU sensor with a sampling frequency of 100 Hz while performing the TUG test. The dependent variable is a binary variable that is true if the patient suffered a fall in a six-month follow-up period. This variable was used as the output variable for the supervised training and validations of the deep learning architectures and competing machine learning approaches. A hold out validation process using 75 subjects for training and 31 subjects for testing was repeated one hundred times to obtain robust estimations of model performances At each repetition, 5-fold cross-validation was carried out to select the best model over the training subset. Best results were achieved by a bidirectional long short-term memory (BLSTM), obtaining an accuracy of 0.83 and AUC of 0.73 with good sensitivity and specificity values.

Keywords: Inertial sensors; fall prediction; fall risk assessment; deep learning; machine learning

1. Introduction

Older people suffering falls often require medical attention [1,2], hence falls are becoming a major public health problem due to the increasing aging of the population. The rising incidence of accidental falls has a great economic impact on healthcare systems and for society: 20-30% of falls lead to mild to severe injuries, and falls are the underlying cause of 10-15% of all emergency department visits of older people in the United Kingdom in 1999 [3], and these figures are growing with population aging since then. Moreover, falls often cause mobility impairments that lead to dependency for activities of daily living, along with psychological consequences such as anxiety and fear for future falls [4,5]. According to the World Health Organization, approximately, worldwide yearly incidence of falls for people over 65 years old is 28-35%, increasing to 32-42% for people aged over 70 years [6]. In particular, older adults living in nursing homes are especially prone to falling. In fact, the fall incidence in this population is three times that of older people living in the community [7]. The financial toll from older adult falls in the United States was estimated in $67.7 billion in 2016 [8]. Therefore, fall prevention in older adults is of upmost socioeconomic importance.

 doi:10.20944/preprints202409.1021.v1

To this end, clinical questionnaires and clinical assessment-based fall risk prediction tools have been proposed reporting a wide range in performance scores (sensitivity in the range 14-94%, specificity in the range 38-100%) [9]. Additionally, fall risk assessment protocols like the STEADI (stopping elderly accidents, deaths & injuries) proposed by the Centers for Disease Control (CDC) rely on functional mobility assessment tools in the form of questionnaires, physical tests, gait analysis, and physical activity measurements [10]. Some of the most widely used assessment tools are the Timed Up and Go (TUG) test [11], the Tinetti Assessment Tool [12], the STRATIFY score [13], and the Five-Times-Sit-to-Stand (FTSS) test [14]. Specifically, the TUG test has proven valuable in early assessment of balance and mobility [15–17]. However, all of these tools are in fact used qualitatively by the clinician trying to assess prospective fall risk.

The main hypothesis of this study is that the information extracted from IMU readings during the realization of the TUG test can be used to build predictive models that provide an estimate of the probability of the patient suffering a fall in the near future. In other words, this information may be used for quantitative and predictive fall risk assessment. This information would be of great importance to guide fall prevention for older adults, and especially for those living in nursing homes due to their greater fall incidence.

The paper reports two computational experiments. The first corresponds to application of supervised machine learning algorithms to some descriptive variables of the TUG test phases. The second corresponds to the application of deep learning architectures over the raw data of the IMU wearable.

The contributions of this paper are the following ones: (a) the collection of a dataset of IMU readings while a large number of subjects are realizing a TUG test whose {F,NF} labels are generated in a follow-up period of 6 months; (b) the proposal of deep learning architectures to deal with this prediction problem; (c) the proposal of feature extraction processes and conventional machine learning for comparison with the deep learning approaches.

2. Materials and Methods

Recent surveys on the application of machine learning methods for prospective and retrospective discrimination between patients who experience falls, i.e., fallers, (F) from non-fallers (NF) using IMU information report widely different predictive performance results (accuracy: 62-100%, sensitivity: 55-99%, specificity: 35-100%) in populations over 65 years old [18–21]. These surveys also report a large heterogeneity of sensor placement, tasks assessed, and sensor features. Specifically, some authors found that data from wearable IMU sensors add meaningful information to the TUG test [22].

Deep learning architectures have been applied successfully in many areas of computer vision [23], medical image analysis [24], assisted/autonomous driving [25], and machine anomaly monitoring [26], to name a few applications. Deep Learning has already been applied to the classification of IMU sensor data [27–30] for human activity recognition. However, multiple data sources and adequate assessment tests are necessary to generalize fall risk predictions. Nait Aicha et al. [31] compared deep learning approaches to traditional machine learning methods to model fall risk on the basis of daily-life body trunk accelerometer data. They acquired data of participants wearing a triaxial accelerometer for 1 week. They evaluated convolutional neural network (CNN), the long short-term memory (LSTM) model, and a combination of both which they refer to as the "ConvLSTM", reporting good results in modelling the training data, but it generalized poorly over new subjects and the relatively long period during which subjects must wear the inertial sensor is a barrier to its implementation.

Due to the multidimensional nature of the risk of falls in older adults, there is no single ideal tool that performs a perfect risk assessment in any context. For this reason, the simultaneous application of multiple tools is recommended [32].

The present article presents a secondary analysis of two single-blinded and multicenter randomized controlled trials that were registered with codes [ACTRN12618000536268; NCT03996083] whose primary outcomes have previously been published [33,34]. This study includes

 doi:10.20944/preprints202409.1021.v1

106 subjects (68 women and 38 men) from 9 long-term nursing homes (LTNHs) (Gipuzkoa, Basque Country, Spain). Subject's ages ranged from 70 and 104 years old and their physical and cognitive characteristics were described previously [35]. After providing written consent, participants performed the TUG test twice wearing a wireless inertial sensor (G-Walk, BTS Bioengineering Corp.) and the best (fastest) trial was selected. This sensor was placed on the lower back area in order to quantify the center of mass movement. This study was approved by the Committee on Ethics in Research at the University of the Basque Country (Humans Committee Code M10/2016/105). All feature extraction and classification cross-validation was carried out in Matlab using wavelet, statistics and machine learning, and deep learning toolboxes. For performance evaluation we split the data in 5 groups and in each iteration, we hold out one group/fold and train the algorithms in the remaining 4 groups. We perform this method to get a less biased model than other methods, such as a simple train/test split. This process is carried out for all evaluated classifiers and feature extraction techniques. A fall was defined as an unintentional event in which the person comes to rest on the ground, not as a result of an epilectic seizure or an acute stroke [36]. Falls suffered by the residents are systematically detected and immediately recorded in the database by the staff of each nursing home. Information regarding residents who experienced any fall during 6-month follow-up period was extracted from the participant's medical record as provided by the medical staff. Participants were labeled as faller (F) or non-faller (NF). The number of falls was not taken into consideration in the present study.

2.2. Data and Feature Extraction

2.2.1. TUG Test Realization for Data Capture

The TUG test process is decomposed into six phases, as shown in Figure 1, which are described as follows:

1. The time elapsed from the beginning of standing-up motion up to the instant when the subject stands up;
2. The time elapsed walking from the initial standing up position to the position where s/he starts turning down;
3. The time elapsed while turning down;
4. The time elapsed walking back to the chair from the end of the first turn to the beginning of the second turn;
5. The time elapsed turning to prepare to sit down, and;
6. The time elapsed sitting down in the chair, completing the TUG test.

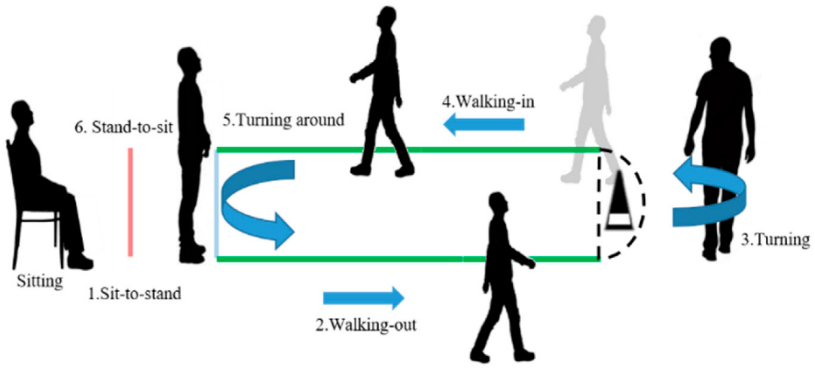

Figure 1. The process of the realization of the TUG test decomposed into six phases.

 doi:10.20944/preprints202409.1021.v1

2.2.1. Raw IMU Data and Labels

The G-Walk IMU sensor acquires acceleration, angular velocity and magnetic field data. Its components are: a triaxial accelerometer (x, y, z), a triaxial gyroscope (x, y, z) and a triaxial magnetometer (roll, pitch, yaw). Sampling frequency was adjusted to 100 Hz. The accelerometer has a resolution of 16 bits *per* axe and its sensitivity was adjusted to 2g. The gyroscope also has a resolution of 16 bits *per* axe and its sensitivity was adjusted to 2000 ◦/s. The magnetometer has a resolution of 13 bits with a sensitivity of 1200 µT. Figure 2 shows example readings from the sensor.

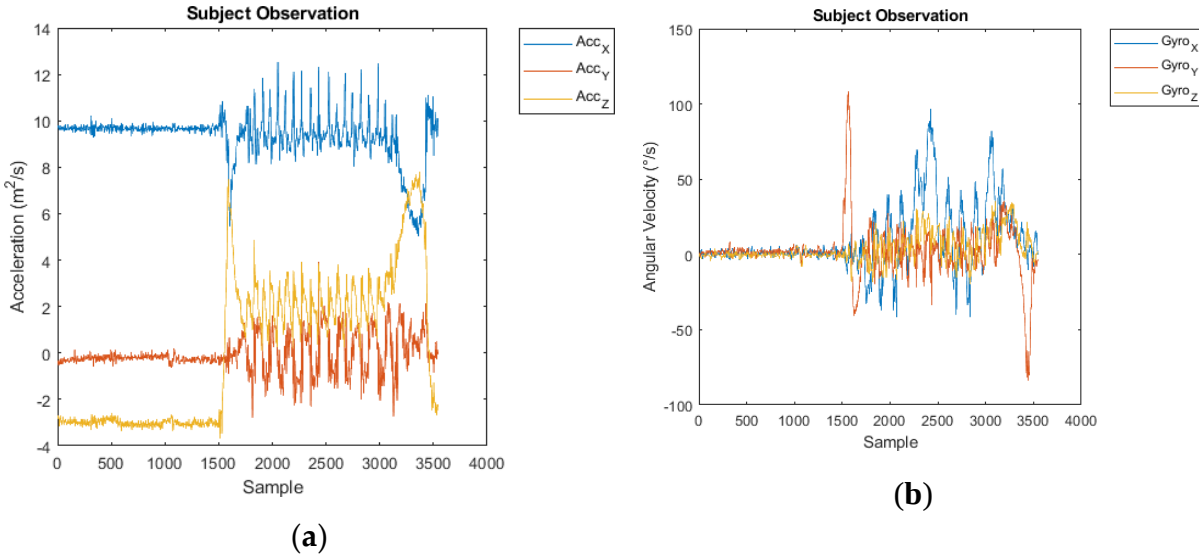

Figure 2. An instance of the readings of the G-walk during a TUG test realization shown as raw data plots: (a) triaxial accelerometer, and (b) triaxial gyroscope.

We collected raw IMU data for each TUG test realization by a subject. Due to variability in the time taken to perform the TUG test, the number of samples per subject varies from 1364 to 9975 as shown in Figure 3. Additionally, data of patients suffering fall occurrences during a 6-month follow-up period were collected and provided to the researchers by the staff of the LTNHs. In this period, 21 subjects (19%) were labeled as fallers (F). This label data is used the dependent variable in the training and validation of the classification algorithms, both deep learning networks and conventional machine learning approaches.

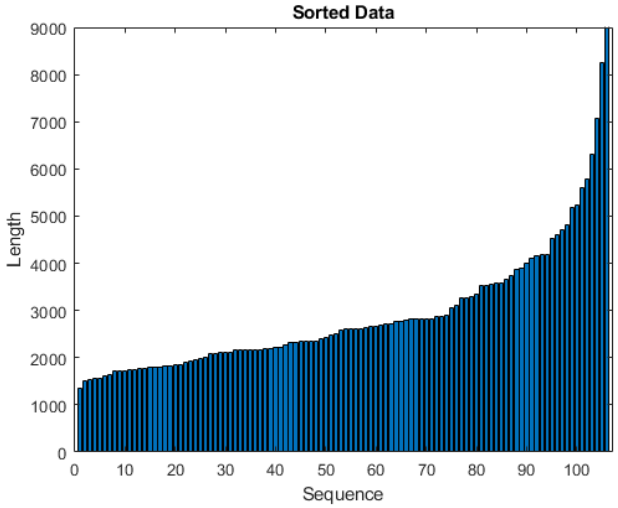

Figure 3. Number of samples per recorded IMU sequence during the realization of TUG tests sorted in ascending order.

 doi:10.20944/preprints202409.1021.v1

In the pre-processing steps we remove the subjects with missing values of IMU sensors or without label information (faller/non-faller). Then, we sort the subjects regarding their number of samples, observing that the great majority of them have less than 5000 IMU data samples, and that those above this number could be considered outliers. However, these subjects are precisely the ones that have a higher fall risk. Consequently, we train our model with data from all the subjects.

The class imbalance in the dataset is moderate (ratio 1:5), however conventional machine learning approaches are usually biased towards the majority class, which in this study is the non-fallers (NF) class, suffering of low sensitivity even when reporting high accuracy [37].

The norm of the 3D acceleration vectors is computed at each instant in order to obtain a scalar time series. In this way, significant changes in acceleration magnitude, which occur in events such as walking, turning or getting up / sitting in the chair, are easily detected regardless of the orientation of the device.

2.2.3. TUG Test Variables *per* Phase

We recorded spatiotemporal measurements of the IMU wearable sensor during TUG test realizations decomposed into standing phase, sitting phase, walking phases, and body trunk rotations (flexion and/or extension angle). These measurements are used as input variables by the conventional machine learning classification algorithms. Table 1 shows the maximum, minimum and average values of each of these parameters across subjects. The first group of variables are the duration of the different phases. During “Sit to Stand" and "Stand to Sit” phases we recorded the vertical, media-lateral, and anterior-posterior accelerations, as well as extension and bending angles. In “Turning” phases we recorded the angular accelerations. We recorded the duration of each activity phase for all subjects computing the mean and variance of each of them.

Table 1. Descriptive statistics of the spatiotemporal measurements of the TUG test realizations corresponding to standing phase, sitting phase and rotations body trunk kinematics (flexion and/or extension angle). Accelerations (acc) are measured in m/s². Body trunk rotations are measured in degrees. Anterior-posterior (AP), Medio-Lateral (ML), and Vertical (Vert) axis accelerometer data are shown.

Variable	Max	Min	Average
PHASE DURATION			
Sit_to_Stand (s)	4.7	0.33	1.73
Walking_out (s)	24.99	0.78	4.66
Turning (s)	17.19	1.5	4.81
Walking_in (s)	10.1	0.57	3.15
Turning_around (s)	11.43	1.5	3.73
Stand_to_Sit (s)	4	0.5	1.95
SIT TO STAND			
Sit_to_Stand_Vert_Min_acc (m/s²)	4.53	0.26	1.83
Sit_to_Stand_Vert_Min_acc (m/s²)	-0.72	-4.12	-1.93
Sit_to_Stand_ML_Max_acc (m/s²)	2.26	0.1	0.86
Sit_to_Stand_ML_Min_acc (m/s²)	-0.34	-2.34	-0.94
Sit_to_Stand_AP_Max_acc (m/s²)	4.27	0.29	1.60

 doi:10.20944/preprints202409.1021.v1

Sit_to_Stand_AP_Min_acc (m/s²)	-0.42	-2.69	-1.12
Sit_to_Stand_Extension_Peak (°)	60.2	4.9	33.36
Sit_to_Stand_Extension_Range (°)	39	0.1	14.21
Sit_to_Stand_Bending_Peak (°)	70	19.9	47.42
Sit_to_Stand_Bending_Range (°)	69.9	10.4	44.85
TURNING			
First_Turn_Avg__Angular_Acc (m/s²)	88.4	11.4	43.76
First_turn_Peak_Angular_Acc (m/s²)	181.4	28.5	88.63
TURNING AROUND			
Second_Turn_Avg_Angular_Acc (m/s²)	109.2	14.1	50.71
Second_turn_Peak_Angular_Acc (m/s²)	194.3	40.3	100.20
STAND TO SIT			
Stand_to_Sit_Vert_Max_acc (m/s²)	9.84	0.39	4.88
Stand_to_Sit_Vert_Min_acc (m/s²)	-0.59	-4.84	-2.39
Stand_to_Sit_ML_Max_acc (m/s²)	3.69	0.67	1.78
Stand_to_Sit_ML_Min_acc (m/s²)	-0.53	-6.59	-1.87
Stand_to_Sit_AP_Max_acc (m/s²)	6.19	0.43	3.02
Stand_to_Sit_AP_Min_acc (m/s²)	0.11	-2.5	-0.98
Stand_to_Sit_Extension_Peak (°)	55.8	1	11.82
Stand_to_Sit_Extension_Range (°)	66.5	0.6	42.28
Stand_to_Sit_Bending_Peak (°)	75.5	19.5	53.27
Stand_to_Sit_Bending_Range (°)	62.6	0	24.94

Figure 4 show a box plot of the duration of each phase. The turning phase has the longest average duration followed by the walking out, turning around and walking-in phases. The sitting and standing activities have the shortest average durations. We compute the univariate Chi-Square Test [38] of each feature relative to the {F, NF} class label, obtaining the feature importance ranking shown in Figure 5.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2024 doi:10.20944/preprints202409.1021.v1

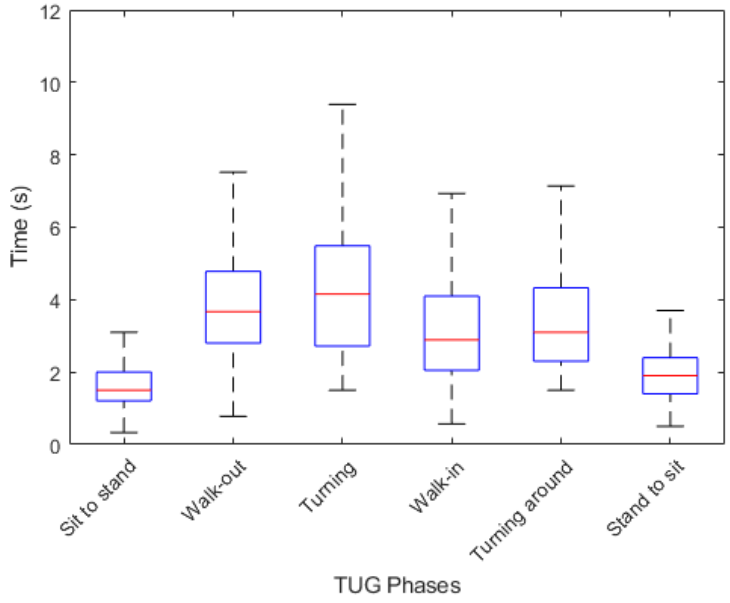

Figure 4. Box-plot of each phase duration in TUG test. The median, upper- lower quartiles and maximum-minimum values are shown.

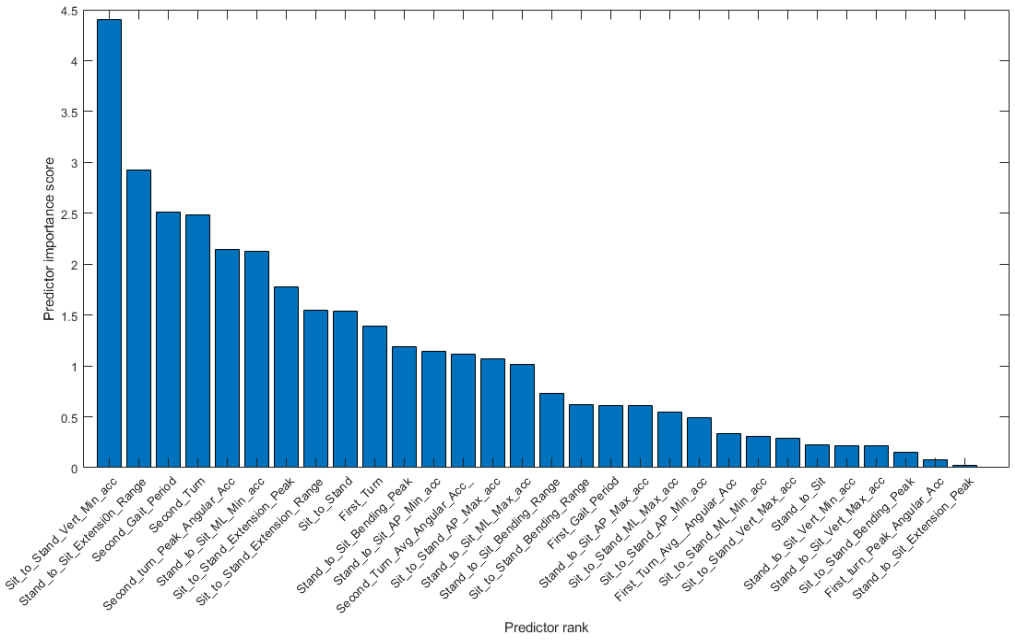

Figure 5. Univariate Chi-Square Test importance ranking of TUG test phase input variables used by conventional machine learning classifiers.

2.2.4. Wavelet Features

Wavelet Transforms (WT) are used to represent a signal in terms of localized basis functions called wavelets. WT use a wavelet function and a lowpass scaling function to generate low-variance representations of real-valued time series data at different time scales. The general formulation of the wavelet is like the following equation:

$$\gamma(s,\tau) \quad \frac{1}{\sqrt{s}}\int x(t)\cdot\psi\left(\frac{\tau \quad \tau}{s}\right)$$

Traditional frequency analysis methods such as the Fourier Transform yield only frequency-domain information without any indication of the temporal location/extent of a given frequency

 doi:10.20944/preprints202409.1021.v1

component. Wavelet transforms on the other hand provide both temporal and frequency information, as the basis functions it relies upon are localized in both time and frequency.

The IMU readings are transformed by the wavelet time scattering decomposition using the Gabor wavelet [39] that yields representations insensitive to translations in the input signal without sacrificing class discriminability and separate the data from different classes as far clear as possible. These wavelet features are obtained after applying the filter banks of the wavelet transform to our signals The scattering sequences are 38-by-1250 where 1250 is the number of time steps and 38 is the number of scattering scales. This matrix constitutes the input features for our 1-D CNN approach to fall risk prediction. Additionally, for we consider each element of the matrix as an independent feature. As a result, we get 47500 independent wavelet features with this decomposition. Due to the large number of features, we need to carry out a feature selection process to enhance the efficiency of the model. The importance of each wavelet feature to discriminate faller vs. non faller is evaluated by individual Chi-square tests [40]. Finally, we choose the 20 most significant wavelet features as the optimal ones. Increased number of wavelet features did not improve the classification performance.

2.3. Machine Learning

The fall risk assessment is stated as a binary classification problem, where the classes are {F, NF} labels assigned in the follow-up period after the IMU measurements (hence, we deal with a prospective problem).

2.3.1. Conventional Machine Learning Algorithms

We have applied the following 5 conventional Machine Learning (ML) algorithms to classify the subjects according to their fall risk assessment: Random Forest (RF), Support Vector Machines (SVM), K nearest neighbors (KNN), Naive Bayes (NB). The hyper-parameters of the machine learning algorithms are set as follows: RF: #splits=105, #learners= 30 SVM: quadratic kernel; KNN: K=10; NB: Gaussian kernel. The implementations used are the standard ones provided in MATLAB. Conventional ML algorithms are applied over TUG test phase variables described in Table 1, because the raw IMU signals have an extremely large dimensionality to be used as inputs for the selected ML models.

2.3.2. Deep Learning Neural Network Models

One of the most distinctive characteristics of deep learning approaches is that they learn a hierarchy of abstract representations from the raw data [41] overcoming the need to define and tune specific features for the problem at hand. In fact, most deep learning approaches are artificial neural networks, so that the term “deep” refers to the number of layers in the network—the more layers, the deeper the network. Two of the most popular deep learning networks are the convolutional neural network (CNN) [42] and the long short-term memory (LSTM) [43]. CNNs built up a hierarchy of convolution filters trained from the data. We use a specific brand of CNNs whose input data is extracted by means of Scattering Wavelet Transforms [43,44] in its 1D version.

An LSTM is good for classifying sequential and time-series data, when the prediction or output of the network must be based on a remembered sequence of data points. An LSTM is a type of recurrent neural network (RNN) [45] that can learn long-term dependencies between time steps of sequence data. Unlike a CNN, a LSTM can remember the state of the network between predictions [23]. The core components of a LSTM network are a sequence input layer and a LSTM layer. A sequence input layer incorporates time-series data into the network. A LSTM layer learns long-term dependencies between time steps of sequence data over time. The LSTM is trained over the raw IMU readings, after computing the norms of the 3D vectors of each measure.

3. Results

We have performed 4 different computational experiments evaluating the different fall risk predictors performance in terms of accuracy, sensitivity, specificity. In the case of raw data, we have also computed the area under the receiving operator curve (AUC). In all cases, we have carried out

 doi:10.20944/preprints202409.1021.v1

100 repetitions of the holdout cross validation with 75 subjects for training and 31 for testing using stratified sampling in the sample extraction, and 5-fold cross-validation over the training set to select the best model for testing at each holdout repetition.

3.1. Conventional Machine Learning Classifiers

We have carried out two different computational experiments with conventional ML classifiers that will serve as benchmarks for the deep learning approaches. In the first experiment, we use as features the aggregated spatiotemporal measurements of the realizations of TUG test corresponding to standing phase, sitting phase and rotations body trunk kinematics from Table 1. The results are shown in Table 2. We have carried out the classifier validation experiments over three distinct subsets of features: (a) the most important TUG phase descriptive variables selected by independent Chi-square tests, (b) the duration of each phase of the TUG test, and (c) the entire set of TUG phase descriptive variables. Results are rather poor for all models and features, with accuracy below 0.7, and sensitivity below 0.33.

Table 2. Average test performance results after 100 repetitions of hold-out cross-validation of different classifiers for sets of features extracted from the TUG test phases enumerated in Table 1.

Feature Set	Classifier	Accuracy	Sensitivity	Specificity
6 Most Important Feature Set	RF	0.62	0.08	0.84
	SVM	0.65	0.25	0.80
	KNN	0.69	0.04	0.95
	NB	0.65	0.25	0.80
	LR	0.66	0.04	0.90
	LD	0.67	0.04	0.92
Phase Duration Features	RF	0.60	0.17	0.77
	SVM	0.65	0.21	0.82
	KNN	0.65	0.46	0.72
	NB	0.66	0.21	0.84
	LR	0.66	0.13	0.87
	LD	0.66	0.13	0.87
All Feature Set	RF	0.68	0.17	0.89
	SVM	0.54	0.21	0.67
	KNN	0.58	0.33	0.67
	NB	0.59	0.25	0.72
	LR	0.48	0.21	0.59
	LD	0.57	0.36	0.62

In the second experiment, we apply the ML classifiers to the selection of the 20 most significant wavelet scattering features extracted from the magnitude of the acceleration signal. Results presented in Table 3 show significant improvement over results reported in Table 2. The increase in specificity may be due to the class imbalance induced bias, while the naive Bayes approach achieves an average sensitivity of 0.52, which is the best result found.

Table 3. Average test performance results after 100 repetitions of hold-out cross-validation of different classifiers using the 20 most significant wavelet scattering features extracted from the acceleration magnitude signal recorded along the TUG test.

Classifier.	Accuracy	Sensitivity	Specificity	AUC
RF	0.81	0.10	0.99	0.75
SVM	0.69	0.29	0.79	0.64
KNN	0.77	0.10	0.94	0.61
NB	0.79	0.52	0.86	0.77

 doi:10.20944/preprints202409.1021.v1

LR	0.71	0.19	0.84	0.61
LD	0.73	0.19	0.86	0.70

3.2. Deep Learning Results

3.2.1. CNN

We evaluate 1-D CNN using as inputs the wavelet scattering matrices computed over the acceleration magnitude. The scattering sequences are 38-by-1250 where 1250 is the number of time steps and 38 is the number of scattering paths. Results are shown in Table 4 for various selections of gradient descent optimization methods (RMSProp, SGDM, and Adam). Results improve over the ML conventional classifiers in terms of accuracy; however, they are not above of RF in terms of AUC, which for many authors is a more appropriate performance measure for class imbalanced datasets.

Table 4. Average test performance results after 100 repetitions of hold-out cross-validation for the 1D CNN architectures.

	1D CNN		
	RMSProp	SGDM	Adam
Accuracy	0.84	0.81	0.81
Sensitivity	0.33	0.33	0
Specificity	0.96	0.92	1
Precision	0.66	0.50	0
AUC	0.63	0.65	0.5

3.2.2. LSTM

We evaluate LSTM deep learning algorithms over raw inertial sensor data (triaxial accelerometer, gyroscope and magnetometer). Both standard LSTM and bidirectional LSTM (BLSTM) were used as we have access to the entire sequence data. We evaluated mini-batch sizes from 5 to 25 with number of hidden units set to 40 and a learning rate of 0.005. The best accuracy results were obtained for mini-batch sizes of 10, 11 and 15. To find the optimal number of hidden units, we set the mini-batch size to 11 and we evaluated the accuracy beginning from 10 until 100 units with increments of 10. The best values are obtained for 40 hidden units. We chose a mini- batch size of 11. Subjects were ordered according to their number of samples and shuffle was disabled to reduce the "padding effect".

Table 5 shows the average test performance results after 100 repetitions of hold-out cross validation of various LSTM architectures. We found that BLSTM performance measures are significantly better than standard LSTM results for every mini-batch size and the best size for BLSTM is ten. The BLSTM trained with SGDM outperforms significantly all other approaches in terms of sensitivity and AUC. Figure 6 shows the corresponding ROC curve with point-wise confidence bounds.

Table 5. Average test performance results after 100 repetitions of hold-out cross validation for the LSTM architectures.

	LSTM	**LSTM**	**LSTM**	**BLSTM**	**BLSTM**	**BLSTM**
	RMSProp	SGDM	Adam	RMSProp	SGDM	Adam
Accuracy	0.87	0.80	0.80	0.83	0.83	0.87
Sensitivity	0.33	0.16	0.16	0.33	0.50	0.33
Specificity	1	0.96	0.96	0.96	0.92	1
Precision	1	0.50	0.50	0.66	0.85	1
AUC	0.60	0.64	0.62	0.66	0.73	0.78

 doi:10.20944/preprints202409.1021.v1

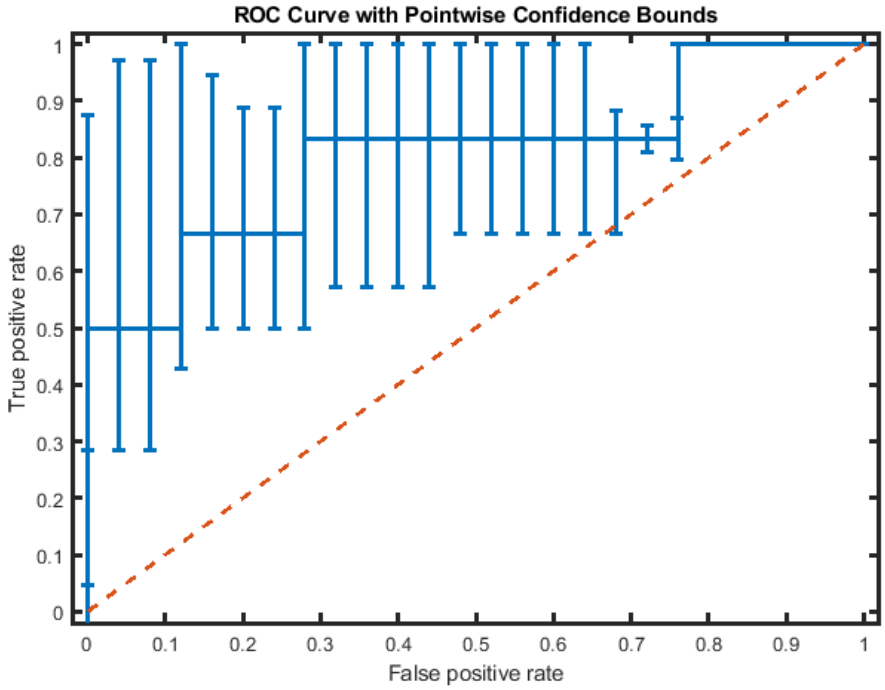

Figure 6. ROC curve with Point-wise Confidence Bounds of an instance of the 5-fold cross-validation of the BILSTM architecture. The dashed lines represent the chance ROC.

4. Discussion

In the present study, conventional machine learning classifiers and deep learning networks have been applied to prospective fall risk prediction over IMU sensor data captured during the realization of the TUG test for a cohort of older adults (N=106, of which 21 are fallers). The hypothesis of this work is that processing this data with machine learning and deep learning approaches would allow prospective fall risk prediction. We have explored several signal features including the raw signal and several machine learning and deep learning approaches. Best results in terms of sensitivity (i.e., accurate prediction of fallers) have been obtained by the naive Bayes approach on wavelet scattering features (sensitivity=0.52), and by the BLSTM trained with SGDM on the raw IMU signal data (0.50). We got high specificity in many instances, however the cost of misclassification of a faller is higher than misclassification of a non-faller, hence sensitivity is a more relevant performance measure. It was argued that the ability of the TUG test to assess prospective fall risk was limited [14], however our results show that processing the IMU sensor data, that implicitly takes into account postural stability, gait, stride length, and sway, a fair prediction of fall risk can be achieved. In the future, we will be testing our approach in larger cohorts. Additionally, we will be exploring the application of Generative Adversarial Networks (GAN) for the enrichment of the faller class in order to obtain more balanced datasets for training and synthetic data generation techniques like SMOTE (Synthetic Minority Over-sampling Technique). We believe our results are promising and could contribute to fall prevention enhancement. This is important and would directly benefit older adults themselves, as those at risk of falling would be identified beforehand and it would enable the relevant entities to consider proper measures and to implement strategies to prevent falling, ultimately preserving their independence and reducing medical care costs.

5. Conclusion

Falls are among the most significant challenges faced by older adults, making their assessment and prevention critically important, particularly in the current demographic context. Although several tools exist for assessing fall risk, these are typically based on time, distance, or visual observation metrics. In fact, these tools are of qualitative nature helping to guide the medical staff

assessment. Our approach, by contrast, leverages the large amount of information that can be collected by wearable IMU sensors on individuals being studied while performing the Timed Up and Go (TUG) test , specifically we can use the raw data from the accelerometer, gyroscope, and magnetometer. Given the relatively high sampling frequency (100 samples per second), the duration of the test, and the three-dimensional data produced by each of the three sensors, a substantial volume of data is generated. The most effective way to analyze such data, with current technological capabilities, is through the application of artificial intelligence. The study includes 106 subjects (68 women and 38 men) from 9 long-term nursing homes (LTNHs). Upon comparing traditional machine learning methods with deep learning approaches, it was found that the latter yielded the most accurate results, specifically the BLSTM algorithm. We believe that our method complements traditional fall risk screening methods and adds valuable information to improve the assessment of subjects with frailty.

Author Contributions: Conceptualization, J.M., G.G., C.R. and M.G.; methodology, J.M.; software, J.M.; validation, G.G., C.R. and B.S.; investigation, G.G., B.S.; resources, C.R. and B.S.; data curation, J.M.; writing—original draft preparation, J.M.; writing—review and editing, M.G., C.R. and B.S.; visualization, J.M.; supervision, M.G.; funding acquisition, M.G. All authors have read and agreed to the published version of the manuscript.

Funding: The Grupo de Inteligencia Computacional, Universidad del Pais Vasco, UPV/EHU, received research funds from the Basque Government from 2007 until 2025. The current code for the grant is IT1689-22. The Spanish MCIN has also granted the authors a research project under code PID2020-116346GB-I00.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Committee on Ethics in Research at the University of the Basque Country (Humans Committee Code M10/2016/105). for studies involving humans.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from Matia Fundazioa and are available from the authors with the permission of Matia Fundazioa.

Acknowledgments: We would like to thank the staff of the centers that participated in our study for their support: Bermingham, Lamourous, Julián Rezola (Matia Fundazioa), Anaka, Betharram (Fundación Caser), Villa Sacramento, Berra (DomusVi), Zorroaga, and San Markosene. We especially thank the study participants and their families for their participation and cooperation.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. P. Kannus et al., "Fall-induced injuries and deaths among older adults," J Am Med Assoc, vol. 281, no. 20, pp. 1895–1899, 1999, doi: 10.1001/jama.281.20.1895.
2. D. A. Sterling, J. A. O'Connor, and J. Bonadies, "Geriatric falls: Injury severity is high and disproportionate to mechanism," *Journal of Trauma - Injury, Infection and Critical Care*, vol. 50, no. 1, pp. 116–119, 2001, doi: 10.1097/00005373-200101000-00021.
3. P. Scuffham and S. Chaplin, "Incidence and costs of unintentional falls in older people in the United Kingdom," *J Epidemiol Community Health*, vol. 57, pp. 740–744, 2003, doi: 10.1136/jech.57.9.740.
4. P. Eggenberger, N. Theill, S. Holenstein, V. Schumacher, and E. de Bruin, "Multicomponent physical exercise with simultaneous cognitive training to enhance dual-task walking of older adults: a secondary analysis of a 6-month randomized controlled trial with 1-year follow-up," *Clin Interv Aging*, p. 1711, Oct. 2015, doi: 10.2147/CIA.S91997.
5. D. J. Hallford, G. Nicholson, K. Sanders, and M. P. McCabe, "The Association Between Anxiety and Falls: A Meta-Analysis," *J Gerontol B Psychol Sci Soc Sci*, p. gbv160, Jan. 2016, doi: 10.1093/geronb/gbv160.
6. United Nations Department Of Economic and Social Affairs, "WHO Global Report on Falls Prevention in Older Age.," *Community Health*, p. 53, 2007, doi: 978 92 4 156353 6.
7. L. Z. Rubenstein, "Falls in older people: epidemiology, risk factors and strategies for prevention," *Age Ageing*, vol. 35, no. suppl_2, pp. ii37–ii41, Sep. 2006, doi: 10.1093/ageing/afl084.
8. National Council on Aging, "Falls Prevention Facts," 2016.

9. V. Scott, K. Votova, A. Scanlan, and J. Close, "Multifactorial and functional mobility assessment tools for fall risk among older adults in community, home-support, long-term and acute care settings," *Age Ageing*, vol. 36, no. 2, pp. 130–139, Jan. 2007, doi: 10.1093/ageing/afl165.
10. K. L. Perell, A. Nelson, R. L. Goldman, S. L. Luther, N. Prieto-Lewis, and L. Z. Rubenstein, "Fall Risk Assessment Measures: An Analytic Review," 2001. Accessed: Jul. 10, 2019. [Online]. Available: http://biomedgerontology.oxfordjournals.org/
11. S. Mathias, U. S. Nayak, and B. Isaacs, "Balance in elderly patients: the "get-up and go" test.," *Arch Phys Med Rehabil*, vol. 67, no. 6, pp. 387–9, Jun. 1986, Accessed: Jul. 10, 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/3487300
12. M. E. Tinetti, T. Franklin Williams, and R. Mayewski, "Fall risk index for elderly patients based on number of chronic disabilities," *Am J Med*, vol. 80, no. 3, pp. 429–434, Mar. 1986, doi: 10.1016/0002-9343(86)90717-5.
13. D. Oliver, M. Britton, P. Seed, F. C. Martin, and A. H. Hopper, "Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: case-control and cohort studies," *BMJ*, vol. 315, no. 7115, pp. 1049–1053, Oct. 1997, doi: 10.1136/bmj.315.7115.1049.
14. M. Csuka and D. J. McCarty, "Simple method for measurement of lower extremity muscle strength," *Am J Med*, vol. 78, no. 1, pp. 77–81, Jan. 1985, doi: 10.1016/0002-9343(85)90465-6.
15. O. Bruyere *et al.*, "Controlled whole body vibration to decrease fall risk and improve health-related quality of life of nursing home residents," *Arch Phys Med Rehabil*, vol. 86, no. 2, pp. 303–307, 2005, doi: 10.1016/j.apmr.2004.05.019.
16. T. Herman, N. Giladi, and J. M. Hausdorff, "Properties of the 'Timed Up and Go' test: More than meets the eye," *Gerontology*, vol. 57, no. 3, pp. 203–210, 2011, doi: 10.1159/000314963.
17. D. Schoene *et al.*, "Discriminative Ability and Predictive Validity of the Timed Up and Go Test in Identifying Older People Who Fall: Systematic Review and Meta-Analysis," *J Am Geriatr Soc*, vol. 61, no. 2, pp. 202–208, Feb. 2013, doi: 10.1111/jgs.12106.
18. Y. Liu, S. J. Redmond, K. Wang, N. H. Lovell, and T. Shany, "Review: Are we stumbling in our quest to find the best predictor? Over-optimism in sensor-based models for predicting falls in older adults," *Healthc Technol Lett*, vol. 2, no. 4, pp. 79–88, Aug. 2015, doi: 10.1049/htl.2015.0019.
19. D. Hamacher, N. B. Singh, J. H. Van Dieen, M. O. Heller, and W. R. Taylor, "Kinematic measures for assessing gait stability in elderly individuals: a systematic review," *J R Soc Interface*, vol. 8, no. 65, pp. 1682–1698, Dec. 2011, doi: 10.1098/rsif.2011.0416.
20. J. Howcroft, J. Kofman, and E. D. Lemaire, "Review of fall risk assessment in geriatric populations using inertial sensors," *J Neuroeng Rehabil*, vol. 10, no. 1, p. 91, Aug. 2013, doi: 10.1186/1743-0003-10-91.
21. J. Howcroft, J. Kofman, and E. D. Lemaire, "Prospective Fall-Risk Prediction Models for Older Adults Based on Wearable Sensors," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1812–1820, 2017, doi: 10.1109/TNSRE.2017.2687100.
22. V. Cimolin *et al.*, "Do wearable sensors add meaningful information to the Timed Up and Go test? A study on obese women," *Journal of Electromyography and Kinesiology*, vol. 44, pp. 78–85, 2019, doi: 10.1016/j.jelekin.2018.12.001.
23. P. Tang, H. Wang, and S. Kwong, "G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition," *Neurocomputing*, vol. 225, pp. 188–197, Feb. 2017, doi: 10.1016/J.NEUCOM.2016.11.023.
24. H.-C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.
25. Q. Rao and J. Frtunikj, "Deep learning for self-driving cars," in *Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems - SEFAIS '18*, New York, New York, USA: ACM Press, 2018, pp. 35–38. doi: 10.1145/3194085.3194087.
26. P. Tamilselvan, P. W.-R. E. & S. Safety, and undefined 2013, "Failure diagnosis using deep belief learning based health state classification," *Elsevier*, Accessed: Jul. 19, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832013000574
27. O. Dehzangi, M. Taherisadr, and R. ChangalVala, "IMU-based gait recognition using convolutional neural networks and multi-sensor fusion," *Sensors (Switzerland)*, vol. 17, no. 12, 2017, doi: 10.3390/s17122735.
28. H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Information Fusion*, vol. 46, 2019, doi: 10.1016/j.inffus.2018.06.002.
29. J. Hannink, T. Kautz, C. F. Pasluosta, K. G. Gasmann, J. Klucken, and B. M. Eskofier, "Sensor-Based Gait Parameter Extraction with Deep Convolutional Neural Networks," *IEEE J Biomed Health Inform*, vol. 21, no. 1, 2017, doi: 10.1109/JBHI.2016.2636456.
30. D. Ravi, C. Wong, B. Lo, and G. Z. Yang, "A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices," *IEEE J Biomed Health Inform*, vol. 21, no. 1, 2017, doi: 10.1109/JBHI.2016.2633287.

31. A. Nait Aicha, G. Englebienne, K. S. van Schooten, M. Pijnappels, and B. Kröse, "Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry," *Sensors (Basel)*, vol. 18, no. 5, pp. 1–14, 2018, doi: 10.3390/s18051654.
32. V. Strini, R. Schiavolin, and A. Prendin, "Fall Risk Assessment Scales: A Systematic Literature Review," *Nurs Rep*, vol. 11, no. 2, pp. 430–443, Jun. 2021, doi: 10.3390/nursrep11020041.
33. C. Rezola-Pardo *et al.*, "Comparison Between Multicomponent Exercise and Walking Interventions in Long-Term Nursing Homes: A Randomized Controlled Trial," *Gerontologist*, vol. 60, no. 7, pp. 1364–1373, Sep. 2020, doi: 10.1093/geront/gnz177.
34. C. Rezola-Pardo *et al.*, "Comparison between multicomponent and simultaneous dual-task exercise interventions in long-term nursing home residents: the Ageing-ONDUAL-TASK randomized controlled study," *Age Ageing*, vol. 48, no. 6, pp. 817–823, Nov. 2019, doi: 10.1093/ageing/afz105.
35. C. Rezola-Pardo *et al.*, "A randomized controlled trial protocol to test the efficacy of a dual-task multicomponent exercise program in the attenuation of frailty in long-term nursing home residents: Aging-ONDUAL-TASK study," *BMC Geriatr*, vol. 19, no. 1, pp. 1–9, 2019, doi: 10.1186/s12877-018-1020-z.
36. A. G. Society, G. Society, A. A. Of, and O. S. P. On Falls Prevention, "Guideline for the Prevention of Falls in Older Persons," *J Am Geriatr Soc*, vol. 49, no. 5, pp. 664–672, May 2001, doi: 10.1046/j.1532-5415.2001.49115.x.
37. A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit*, vol. 91, 2019, doi: 10.1016/j.patcog.2019.02.023.
38. M. L. McHugh, "The Chi-square test of independence," *Biochem Med (Zagreb)*, vol. 23, no. 2, 2013, doi: 10.11613/BM.2013.018.
39. S. Mallat, "Group invariant scattering," *Commun Pure Appl Math*, vol. 65, no. 10, pp. 1331–1398, 2012.
40. M. L. McHugh, "The chi-square test of independence," *Biochem Med (Zagreb)*, vol. 23, no. 2, pp. 143–149, 2013.
41. M. Z. Alom *et al.*, "A state-of-the-art survey on deep learning theory and architectures," 2019. doi: 10.3390/electronics8030292.
42. C. Avilés-Cruz, A. Ferreyra-Ramírez, A. Zúñiga-López, and J. Villegas-Cortéz, "Coarse-fine convolutional deep-learning strategy for human activity recognition," *Sensors (Switzerland)*, vol. 19, no. 7, 2019, doi: 10.3390/s19071556.
43. F. J. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors (Switzerland)*, vol. 16, no. 1, 2016, doi: 10.3390/s16010115.
44. J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 8, 2013, doi: 10.1109/TPAMI.2012.230.
45. A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action Recognition in Video Sequences using Deep Bi-Directional LSTM with CNN Features," *IEEE Access*, vol. 6, pp. 1155–1166, 2017, doi: 10.1109/ACCESS.2017.2778011.