An evaluation of supervised, novelty-based and hybrid approaches to fall detection using Silmee accelerometer data

Aneta Lisowska, Gavin Wheeler, Victor Ceballos Inza, Ian Poole
Toshiba Medical Visualization Systems Europe Ltd.
Bonnington Bond, 2 Anderson Place, Edinburgh, UK
ALisowska@tmvse.com, GWheeler@tmvse.com, victor@ed-alumni.net, IPoole@tmvse.com

Abstract

Elderly people often experience a fear of falling. A reliable fall detector could increase their confidence in receiving prompt help after a fall, thus reducing their mental distress. A wearable sensor such as Toshiba's Silmee device can gather accelerometer data, which can be used to detect falls. We collected data from 20 volunteers wearing Silmee during simulated falls and activities of daily living (ADL). This gave 168 fall and 375 ADL recordings. We used these recordings in three experiments conducted to compare the performance of machine learning techniques for the detection of falls from accelerometer data. These experiments evaluate supervised methods, novelty based fall detection techniques, and finally our proposed hybrid techniques which use supervised methods for feature learning but can be applied in the context of novelty detection. We found that the best performing supervised method was the Convolutional Neural Network (CNN) and the best performing unsupervised method was the one-class Nearest Neighbour Classifier. The best performing hybrid approach resulted from a combination of the CNN and the one-class Support Vector Machine. It draws on the strengths of the CNN (appropriate feature learning) and may offer more accurate real world fall identification.

1. Introduction

In an ageing society, falls are one of the major public health problems. They may have dangerous physical consequences, and can also impact on the psychological well being of the elderly. An older person may develop a fear of falling as a result of past falls, which can hinder their mobility and independence [7], reducing quality of life. In a retrospective study, Wild *et al*. [24] found that, of the patients who lay on a floor for a period of time longer than an hour after falling, half died within six months of the fall. These findings underline the need for immediate intervention after a fall. Therefore automatic systems which could reliably detect falls continue to be widely researched. Such systems have a dual role in expediting the arrival of medical intervention and increasing a person's confidence in receiving prompt help, thus reducing their mental distress.

Numerous studies have used accelerometer data to detect falls. In recent years there has been particular interest in smart phone fall detection applications [1, 11, 16], nevertheless Igual *et al*. [9] pointed out that the elderly with low technical skills might not easily adapt to these systems. An alternative is wearable sensors such as tri-axial accelerometers attached to the waist[5], wrist or head [10]. In this study, we used a Silmee [22], which is a chest attached device with the capacity to record both heart beat and acceleration data.

The simplest fall detection systems are based on thresholding. If an event has a peak acceleration magnitude above the set threshold value it is identified as a fall [4]. More sophisticated and precise fall detection methods rely on machine learning. A brief review of commonly used machine learning approaches to fall detection is given in section 2.

In this study, we compare different supervised and unsupervised machine learning approaches and we propose a hybrid approach that we describe in section 3. We selected algorithms which have previously been applied to the accelerometer based fall detection problem, based on their prevalence and performance. We also selected two algorithms, which to our knowledge have not yet been applied to this problem: the 1D Convolutional Neural Network and the Replicatory Neural Network, a novelty detector proposed by Hawkins *et al*. [8].

2. Fall Detection Algorithms

2.1. Supervised Approaches

Supervised learning approaches are trained on fully labelled data from both activities of daily living (ADL) and falls. Once the training phase is complete, the classifier is able to identify the category to which a new, unknown event

978-1-4673-9711-7/15 $31.00 © 2015 IEEE
DOI 10.1109/ICCVW.2015.60

belongs.

Albert *et al*. [1] evaluated five supervised classifiers in terms of fall detection accuracy on accelerometer data gathered using a mobile phone and wearable accelerometer. The authors considered support vector machines (SVM), sparse multinomial logistic regression (SMLR), Naive Bayes, k-nearest neighbours (K-NN), and decision trees. SVM and SMLR were found to be the best performing methods in this study, both identifying simulated falls with 98% accuracy.

Lusterek *et al*. [13] evaluated eight algorithms on an activity recognition task. The algorithms used body location and velocity data. The authors concluded that the SVM offers the best performance and is closely followed by the random forest classifier.

Artificial neural networks (ANN) have also been frequently employed in daily activity classification tasks (e.g [2, 6, 12, 14]), but as far as we are aware they have not been used in the fall detection problem. Most of the ANN based systems for activity recognition from accelerometer data use multilayer perceptron classifiers (a broader review can be found in: [21]). Nevertheless, in the recent years, convolutional neural networks (CNN) have gained popularity in the computer vision community, due to their impressive performance in object recognition tasks [23]. We intend to investigate whether a simple 1D versions of CNN, applied to accelerometer data, could match SVM performance on the fall detection task.

The key weakness of supervised methods is that it is difficult to gather data from real falls, and unethical to ask elderly patients to provide simulated fall data. As a result, the classifiers are generally trained on falls simulated by healthy young people, which may not be representative of real-life falls by elderly people.

2.2. Novelty Detection Approaches

Novelty detection methods do not require simulated falls for training, which make them more suitable for the identification of real falls in an older population. The detector is trained purely on ADL data. New inputs are classified as falls if they are very different from the ADL data on which the detector was trained. A review of novelty detection methods by Pimentel *et al*. can be found in [20].

Zhang *et al*. [25] was the first to propose using a novelty detector in the fall identification problem. The authors used one-class SVM and in their study, 96.7% of the test activities were classified correctly.

More recently, Medrano *et al*. [15] compared three novelty based fall detectors: k-means, One-class Nearest Neighbour and One-class SVM and reported that Nearest Neighbour(1NN) performed best. In their study they also compare the best unsupervised novelty detector (1NN) with one of the supervised methods (SVM). Even though the SVM marginally outperformed 1NN, the authors were optimistic about the real world usability of the novelty detection methods.

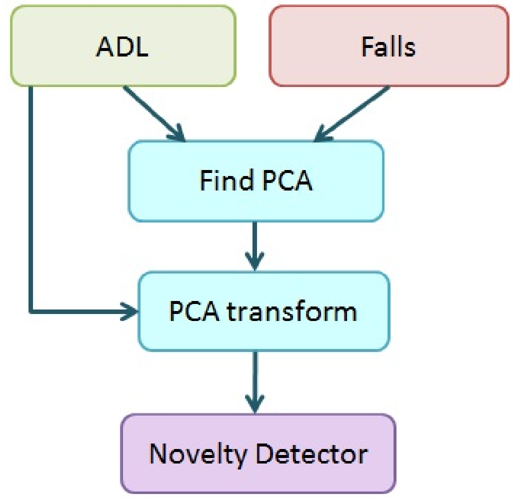

Figure 1. PCA-based hybrid architecture. The PCA transformed ADLs are the input to the novelty detector.

One of the novelty based methods not evaluated by Medrano *et al*. is the replicatory neural network (RNN)[8]. The idea here is the following: during training on ADL data, the neural network learns to replicate its inputs. In the testing phase, a new event is fed through the network, and the replication produced by the network is compared with the input. If there is a large difference, then the event is classified as abnormal.

The key strength of the novelty detection approach is that detectors can be easily personalised, since the data is continuously gathered from the user, with no requirement for the user to perform numerous fall simulations.

3. Novelty Hybrid

One of the potential problems of novelty detectors is that they might detect novelties which are unconnected with the target task (for instance, the person started hobbling). We propose to identify the dimension in which detector should look for novelties, by finding the directions of maximal variance in data which is a mixture of both ADL and fall examples. As shown in figure 1, PCA is fitted to profiles of both types of activities: normal (ADL) and abnormal/novel (falls), but the novelty detector is subsequently trained on only the normal activities, projected onto the space of the maximal variation. During the test phase, activities are transformed using fitted PCA and they are then fed to the abnormality detector, which classifies the activity.

The alternative way of finding an appropriate feature space is to train a convolutional neural network, which hierarchically extracts features. The final classification layer of the CNN can then be replaced with an abnormality detector. The CNN is trained as usual on ADL and fall data. Then, similarly to the PCA + novelty detector hybrid, only normal data is fed through the CNN to train an abnormality detector, which uses as input the features discovered by the CNN.

For this approach, two population cohorts are required: a group of healthy young people who perform falls in the controlled environment (as for supervised approaches) and a second group of elderly patients, from whom normal data is collected. Feature discovery is performed on the data from the simulated falls, but the training of the novelty detector is performed on the ADL data from the elderly group. Using this method, the fall detector is personalised to the user, with the aim of improving performance for real data.

4. Experiments

This work consists of three comparative studies. The first study evaluates four supervised methods of fall detection against a simple threshold based classifier (ST). The algorithms selected are: Support Vector Machine (SVM), K-Nearest Neighbour (K-NN), Random Forest (RF) and Convolutional Neural Network (CNN).

The second study focuses on novelty based fall detection techniques. A Replicatory Neural Network (RNN) is compared with a widely used 1-class SVM (1SVM) and 1-class Nearest Neighbour classifier (1NN), recently reported to be the best performing novelty detector in fall detection task [15].

Finally, we propose six novelty hybrid techniques which draw benefits from the supervised methods (feature/dimension selection) but can be applied in a novelty detection context.

4.1. Dataset

The data was collected from 20 volunteers (22-49 years old) in four data gathering sessions. Each participant was asked to perform activities of daily living (ADL) and twelve different types of falls as proposed by Noury *et al*. [17]. During all activities volunteers wore a Silmee device placed just below their clavicle. All falls were completed on a crash mat in a controlled environment. We gathered in total 641 ADLs and 168 falls. We are interested in discriminating between ADLs, which are above 1.6g threshold and falls. This threshold was chosen to eliminate sedentary ADL [18]which are very easily distinguishable from falls. Interesting ADLs are those which are harder to differentiate from accelerometer data as normal events e.g. sitting down heavily. We have considered 375 ADLs, that are above the 1.6g threshold, for training and testing the algorithms (see Figure 2).

The Silmee device that was employed was a Silmee Bar type. This has an online battery life of 8 hours and a maximum sample rate of 125Hz. The sample rate employed was between 16Hz (64ms) and 64Hz (16ms).

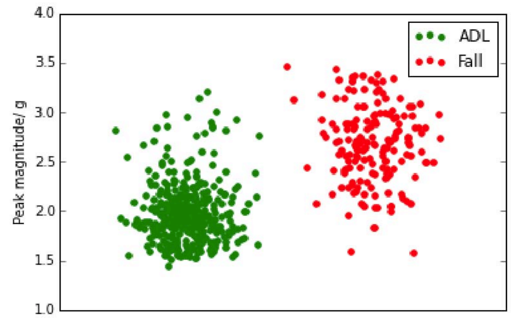

Figure 2. Jitter plot of the peak magnitude of the extracted ADL and Falls (jittering applied in the x-axis direction). A fixed peak magnitude threshold cannot perfectly separate ADL from Falls.

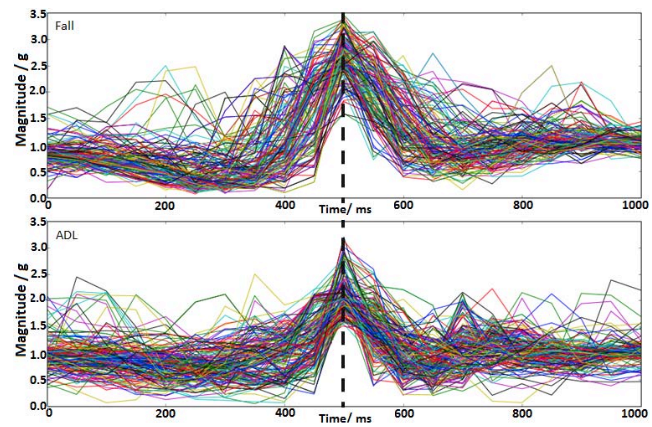

Figure 3. The curves represent extracted one-second long events. Each extracted event is a 21 feature long vector, with a sample at the peak magnitude and 10 samples before and after the peak. The top figure shows falls and the bottom figure shows ADLs.

4.2. Experimental setup

For each activity, an acceleration magnitude vector was computed from the acceleration in the x, y and z directions. The resulting magnitude vector was interpolated and re-sampled at 50ms rate to ensure that any inconsistency in sampling rate between sessions was removed. In each event, the peak magnitude was located, 500ms before and after this peak was extracted, resulting in a 1-second long acceleration magnitude feature vector of 21 samples (see Figure 3).

The algorithms were trained on data from three sessions and tested on the fourth session in 4-fold cross validation. All methods were implemented using Python. We used the scikit-learn package [19] implementation of most of the novelty methods and most of the supervised algorithms evaluated in our study. CNN and RNN were implemented using the Theano Python library [3]. The parameters for all scikit-learn based algorithms were selected using grid search to maximise precision.

The CNN was built from the two pairs of convolutional and pooling layers. The first convolutional layer has 30 nodes and the second has 15 nodes. The filter size is 4 for both and the pool size is 2. The fully connected layer has 6 nodes and it is followed by a soft-max classification layer, or a novelty detector in the CNN based novelty hybrid implementation. The CNN uses L2 regularisation with a penalty of 0.002. All parameters were chosen empirically.

The replicatory neural network has 3 hidden layers with 70, 40 and 70 nodes respectively. The number of input features is equivalent to the number of output nodes. Each one-second extracted data example has 21 features. The feature vectors after PCA transformation are shorter and are equal to the number of the principal components with an additional feature, which is the peak magnitude of the extracted activity (as shown in figure 1). All neural network based approaches use the rectifier activation function.

The problem with interpreting the results from most of the studies mentioned in section 2 is that accuracy scores are usually given, instead of than the separate false positive and true positive rates. This information might be important because the test data usually contains more ADLs than falls, and even if the falls were always identified incorrectly as ADL, the classification accuracy results might be high. Therefore in this study we evaluate all algorithms in terms of the area under the ROC curve (AUC). Following Medrano *et al*. study and Noury *et al*. guidelines for evaluation of fall detector, we also provide specific value of sensitivity(SE) and specificity(SP), which are selected at the point maximising their geometric mean $\sqrt{SE \times SP}$.

4.3. Results and Discussion

Supervised approaches. Figure 4 shows the area under the ROC curve for the supervised methods. The AUC of all of the selected algorithms in the comparison is higher than the AUC of the baseline simple threshold classifier. The best performing method in terms of AUC is CNN. Table 1 presents additional information about the results of sensitivity (SE) and specificity (SP) of the methods. The best results for each measure are shown in bold text. The SVM has the best SE, whereas the K-NN and CNN have the highest SP (CNN is slightly better, but there is higher variance in the results). The best result of the geometric mean of SE and SP is obtained by K-NN, closely followed by CNN.

Novelty detection approaches. Table 2 presents a comparison of novelty detection based fall identification techniques. All of the novelty based methods have lower AUC results than any of the supervised algorithms. However, this is not a like-for-like comparison since the test and training populations differ. The best performing novelty detector is 1NN (see figure 5) and the poorest performance is given by the RNN, which has also the highest variance. In terms of SP, 1SVM performs the best.

Supervised approaches				
Method	AUC	$\sqrt{SE \times SP}$	SE	SP
ST	0.910	0.852	0.872	0.833
CNN	**0.946**	0.895	0.877	**0.911**
SVM	0.934	0.881	**0.917**	0.845
RF	0.927	0.854	0.811	0.899
K-NN	0.921	**0.907**	0.909	**0.903**

Table 1. Comparison of supervised methods for fall detection. RF and CNN are not deterministic, therefore we report the mean and standard deviation of the results over 5 runs of the experiment. RF AUC $\sigma = 0.005$, RF SE and SP $\sigma = 0.027$, CNN AUC $\sigma = 0.004$, CNN SE and SP $\sigma = 0.019$.

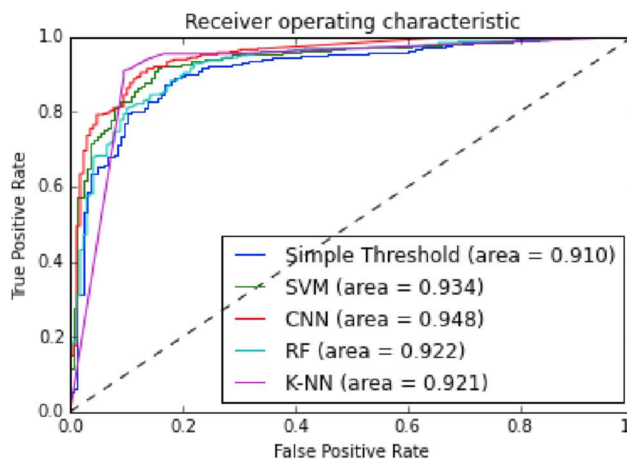

Figure 4. ROC curves for supervised methods.

Novelty detection approaches				
Method	AUC	$\sqrt{SE \times SP}$	SE	SP
RNN	0.856	0.784	0.710	0.871
1SVM	0.888	0.850	0.828	**0.872**
1NN	**0.915**	**0.855**	**0.858**	0.853

Table 2. Comparison of novelty detection methods for fall detection. The experiment was run 5 times for each method and the mean results are presented in the table. The variance due to randomness for 1NN and 1SVM was 0, these methods are deterministic. RNN AUC $\sigma = 0.016$, RNN SE and SP $\sigma = 0.085$

Hybrid approaches. The novelty hybrid methods are compared in table 3. For PCA hybrids, the number of PCA components resulting in the best AUC is reported. CNN extracts six features, as the fully connected layer has six nodes. The lowest AUC is given by the combination of the PCA and RNN. The highest AUC is achieved by the combination of CNN and one-class SVM. Interestingly this hybrid performs even better than CNN on its own.

Overall the novelty hybrid methods perform better than

Novelty Hybrids				
Method	AUC $[\sigma]$	$\sqrt{SE \times SP}$ $[\sigma]$	SE $[\sigma]$	SP $[\sigma]$
PCA(13 components) + RNN	0.864 [0.018]	0.786 [0.021]	0.748 [0.060]	0.841 [0.024]
PCA(13 components) + 1SVM	0.882 [0.000]	0.843 [0.000]	**0.852** [0.000]	0.833 [0.000]
PCA(13 components) + 1NN	**0.928** [0.000]	**0.869** [0.000]	0.814 [0.000]	**0.927** [0.000]
CNN + RNN	0.889 [0.028]	0.806 [0.017]	0.786 [0.049]	0.826 [0.047]
CNN + 1SVM	**0.958** [0.002]	**0.921** [0.012]	**0.946** [0.014]	**0.912** [0.016]
CNN + 1NN	0.876 [0.037]	0.847 [0.032]	0.918 [0.033]	0.818 [0.036]

Table 3. Comparison of novelty hybrid methods for fall detection. The experiment was run 5 times for each method, mean results and the standard deviation for each method are presented in the table. Note that the standard deviation for PCA based hybrid with 1NN and 1SVM is zero, because these methods remain deterministic.

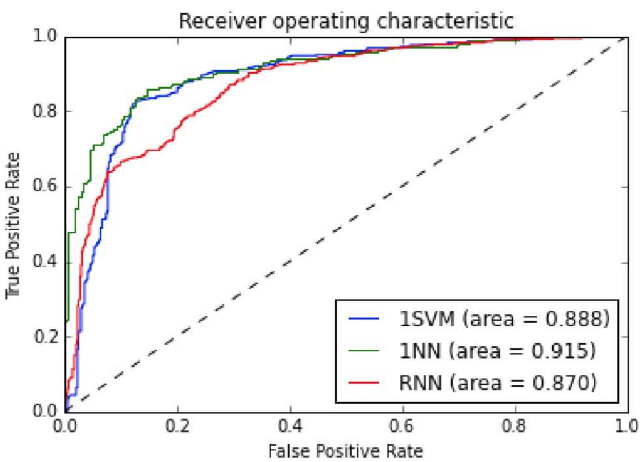

Figure 5. ROC curves for novelty detection methods.

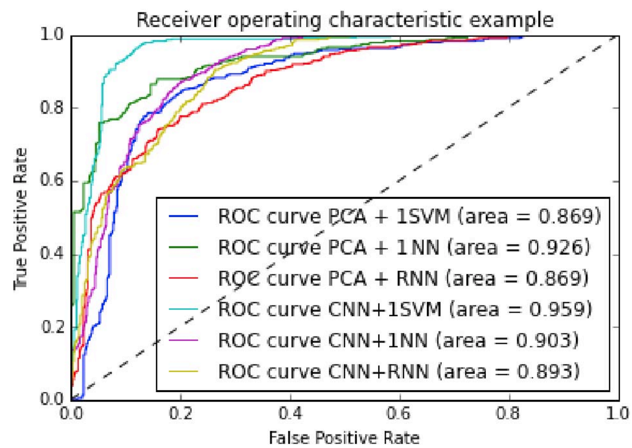

Figure 6. Novelty Hybrid ROC Curves. PCA based hybrid here with 16 components. Novelty hybrids performance vary between 0.869 to 0.959

the novelty detectors, which may suggest that selection of appropriate feature space is crucial for these approaches. The one-class SVM is the detector whose performance is most improved by the new feature space definition. CNN based novelty hybrids give better results than PCA based hybrids (see figure 6), indicating that features extracted by the CNN are more meaningful than features resulting from dimensionality reduction. The exception is the 1NN, which is boosted by PCA, but its performance varies when it uses CNN extracted features.

Figure 7 shows how the AUC varies with the number of PCA components. If fewer than 9 principle components are selected, then information needed for discrimination between falls and ADL is partially lost. Interestingly 6 principal components are less meaningful to a novelty detector than 6 features extracted by CNN. A subject for future work is to investigate the difference between those features and attempt to explain why CNN extracted features lead to better performance of novelty detectors.

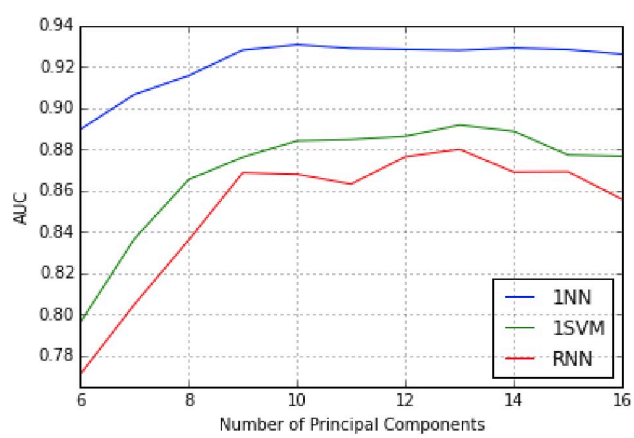

Figure 7. Graph showing the relationship between the number of PCA components and the AUC for PCA based novelty hybrid methods.

5. Conclusion

We have presented an experimental comparative study of the performance of five supervised, three unsupervised and six hybrid methods for fall detection. We found that the best performing supervised method was the CNN and the best performing novelty detector was the 1NN.

While the novelty detection methods might be more suited to detect real-life falls based on personalised training data, their performance is not comparable with the supervised techniques on the simulated falls. Nevertheless, as mentioned previously, the performance of supervised approaches is expected to drop when applied to the data collected from elderly people.

In the future it would be interesting to test these methods on a distinctly different population cohort. In such a situation the supervised methods may be found to be less advantageous.

The best performing hybrid approach is the combination of CNN and 1-class SVM. It draws on the strengths of the CNN (appropriate feature selection) and may offer more accurate real-life fall identification.

Acknowledgements

The authors wish to thank Alison O'Neil and Shiko Yo for their valuable advice and discussion.

References

[1] M. V. Albert, K. Kording, M. Herrmann, and A. Jayaraman. Fall classification by machine learning using mobile phones. *PloS one*, 7(5):e36556, 2012. 1, 2

[2] J. Baek, G. Lee, W. Park, and B.-J. Yun. Accelerometer signal processing for user activity detection. In *Knowledge-Based Intelligent Information and Engineering Systems*, pages 610–617. Springer, 2004. 2

[3] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio. Theano: A cpu and gpu math compiler in python. In *Proc. 9th Python in Science Conf*, pages 1–7, 2010. 3

[4] A. Bourke, J. Obrien, and G. Lyons. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait & posture*, 26(2):194–199, 2007. 1

[5] J. Chen, K. Kwong, D. Chang, J. Luk, and R. Bajcsy. Wearable sensors for reliable fall detection. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 3551–3554. IEEE, 2006. 1

[6] M. Ermes, J. Parkka, J. Mantyjarvi, and I. Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Information Technology in Biomedicine, IEEE Transactions on*, 12(1):20–26, 2008. 2

[7] S. M. Friedman, B. Munoz, S. K. West, G. S. Rubin, and L. P. Fried. Falls and fear of falling: which comes first? a longitudinal prediction model suggests strategies for primary and secondary prevention. *Journal of the American Geriatrics Society*, 50(8):1329–1335, 2002. 1

[8] S. Hawkins, H. He, G. Williams, and R. Baxter. Outlier detection using replicator neural networks. In *Data warehousing and knowledge discovery*, pages 170–180. Springer, 2002. 1, 2

[9] R. Igual, C. Medrano, and I. Plaza. Challenges, issues and trends in fall detection systems. *Biomed. Eng. Online*, 12(66):1–66, 2013. 1

[10] M. Kangas, A. Konttila, P. Lindgren, I. Winblad, and T. Jämsä. Comparison of low-complexity fall detection algorithms for body attached accelerometers. *Gait & posture*, 28(2):285–291, 2008. 1

[11] R. Y. Lee and A. J. Carlisle. Detection of falls using accelerometers and mobile phone technology. *Age and ageing*, page afr050, 2011. 1

[12] N. Lovell, N. Wang, E. Ambikairajah, and B. G. Celler. Accelerometry based classification of walking patterns using time-frequency analysis. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 4899–4902. IEEE, 2007. 2

[13] M. Luštrek and B. Kaluža. Fall detection and activity recognition with machine learning. *Informatica*, 33(2), 2009. 2

[14] J. Mäntyjärvi, J. Himberg, and T. Seppänen. Recognizing human motion with multiple acceleration sensors. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, volume 2, pages 747–752. IEEE, 2001. 2

[15] C. Medrano, R. Igual, I. Plaza, and M. Castro. Detecting falls as novelties in acceleration patterns acquired with smartphones. *PloS one*, 9, 2014. 2, 3

[16] C. Medrano, R. Igual, I. Plaza, M. Castro, and H. M. Fardoun. Personalizable smartphone application for detecting falls. In *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*, pages 169–172. IEEE, 2014. 1

[17] N. Noury, P. Rumeau, A. Bourke, G. ÓLaighin, and J. Lundy. A proposal for the classification and evaluation of fall detectors. *Irbm*, 29(6):340–349, 2008. 3

[18] O. Ojetola. *Detection of Human Fal ls using Wearable Sensors*. PhD thesis, Coventry University, 2013. 3

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011. 3

[20] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014. 2

[21] S. J. Preece, J. Y. Goulermas, L. P. Kenney, D. Howard, K. Meijer, and R. Crompton. Activity identification using body-mounted sensorsa review of classification techniques. *Physiological measurement*, 30(4):R1, 2009. 2

[22] T. Suzuki, H. Tanaka, S. Minami, H. Yamada, and T. Miyata. Wearable wireless vital monitoring technology for smart health care. In *Medical Information and Communication Technology (ISMICT), 2013 7th International Symposium on*, pages 1–4. IEEE, 2013. 1

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014. 2

[24] D. Wild, U. Nayak, and B. Isaacs. How dangerous are falls in old people at home? *BMJ*, 282(6260):266–268, 1981. 1

[25] T. Zhang, J. Wang, L. Xu, and P. Liu. Fall detection by wearable sensor and one-class svm algorithm. In *Intelligent Computing in Signal Processing and Pattern Recognition*, pages 858–863. Springer, 2006. 2