Voice Operated Fall Detection System Through Novel Acoustic Std-LTP Features and Support Vector Machine

Usama Zafar, Farman Hassan, Muhammad Hamza Mehmood, Abdul Wahab, and Ali Javed

Abstract The ever-growing old age population in the last two decades has introduced new challenges for elderly people such as accidental falls. An accidental fall in elderly persons results in lifelong injury, which has extremely severe consequences for the remaining life. Furthermore, continued delay in the treatment of elderly persons after accidental fall increases the chances of death. Therefore, early detection of fall incidents is crucial to provide first aid and avoid the expenses of hospitalization. The major aim of this research work is to provide a better solution for the detection of accidental fall incidents. Most automatic fall event detection systems are designed for specific devices that decrease the flexibility of the systems. In this paper, we propose an automated framework that detects in-door fall events of elderly people in the real-time environment using a novel standard deviation local ternary pattern (Std-LTP). The proposed Std-LTP features are able to capture the most discriminatory characteristics from the sounds of fall events. For classification purposes, we employed the support vector machine (SVM) to distinguish the in-door fall occurrences from the non-fall occurrences. Moreover, we have developed our fall detection dataset that is diverse in terms of speakers, gender, environments, sample length, etc. Our method achieved an accuracy of 93%, precision of 95.74%, recall of 90%, and F1-score of 91.78%. The experimental results demonstrate that the proposed system successfully identified both the fall and non-fall events in various indoor environments. Subsequently, the proposed system can be implemented on various devices that can efficiently be used to monitor a large group of people. Moreover, the proposed system can be deployed in daycare centers, old homes, and for patients in hospitals to get immediate assistance after a fall incident occurs.

Keywords Fall event · Machine learning · Non-fall event · Std-LTP · SVM

U. Zafar · F. Hassan · M. H. Mehmood · A. Wahab · A. Javed (✉)
University of Engineering and Technology, Taxila, Pakistan
e-mail: ali.javed@uettaxila.edu.pk

S. Anwar et al. (eds.), *Proceedings of International Conference on Information Technology and Applications*, Lecture Notes in Networks and Systems 614,
https://doi.org/10.1007/978-981-19-9331-2_13

1 Introduction

The population of aged people around the world is increasing at a rapid pace because of the advancement that has been made in the medical field. As reported by United Nations World Aging Population Survey (World Population Ageing, United Nations, 2020), there were around 727 million people aged 65 years or more in 2020. It is expected that this figure will be doubled by 2050. One of the most prevalent causes of injuries is an accidental fall. Old age people are mostly affected by these accidental falls that happen due to various reasons such as unstable furniture, slippery floors, poor lighting, obstacles, etc. It is very common in most countries that elderly persons live alone in their homes without the presence of kids and nurses. The most devastating effect of the fall incident on elderly people is that they may lay on the floor unattended for a long interval of time. As a result, they develop long-lasting injuries that in some cases can even lead to death. Research study finds that these fall incidents of old people cost millions of Euros to the UK government [1]. Risk factors of accidental falling increase with old aged people and also cost more for their treatment and care [2]. According to one report [3], people aged 65 years or older are more likely to fall once a year and some of them may fall more than once as well. These statistics demand to develop reliable automated fall detection systems using modern-day technology to help reduce the after-effects of fall incidents and to provide immediate first-aid support to the concerned person.

The research community also explored motion sensors and acoustic sensors-based fall detection systems for elderly people using several techniques implemented in wearable devices i.e., smart watches, smart shoes, smart built, smart bracelets, and smart rings. In Yacchirema et al. [4], LowPAN, sensor networks, and cloud computing were used for the detection of fall events. Four machine learning classifiers i.e., logistic regression, ensemble, deepnets, and decision trees were employed for classification purposes, but the ensemble performed well using SisFall, sliding windows along with Signal magnitude area (SMA), and Motion-DT features. The healthcare professional receives the notification through a secure and lightweight protocol. In Giansanti et al. [5], a mono-axial accelerometer was utilized to calculate the acceleration of different parts of the body intending to detect any mishap. Acceleration is one of the important parameters that can be used to observe the motion of the body. A mono-axial accelerometer measures the vertical acceleration of a person's body to detect the fall event. However, elderly people need to wear the accelerometer every time which greatly affects their daily activities and routine lives. In a study by Muheidat [6], sensor pads were used and placed under the carpet, which monitors the aged persons. Computational intelligence techniques i.e., convex hull and heuristic were used to detect fall events. In a study by [7], instead of wearable devices, fall was detected using smart textiles with the help of a non-linear support vector machine. To analyze an audio signal, the Gabor transform was applied in the time and frequency domain to derive new features known as Wavelet energy. In the study by [8], a fall detection system was developed using an acoustic sensor that was placed on the z-axis to detect the pitch of the audio. However, this method has a limitation as only a single person is allowed in the locality. Moreover, elderly people are unable to carry the sensors all the time. This concern was addressed in [9] and two types of sensors were used i.e., body sensor and fixed sensor at home. Both the body sensor and fixed sensor were used at the same time. At home, fixed sensors can also work independently if a person is unable to carry body sensors. A mixed positioning algorithm was used to determine the position of the person that is used to decide the fall event.

The research community has also explored vision-based techniques for fall detection. In visual surveillance applying background, subtraction is a quite common approach to discriminate moving objects. In the study by Yu et al. [10], vision-based fall detection was proposed in which background subtraction was applied for extraction of the human body silhouette. The extracted silhouettes were fed into CNN for detecting both fall occurrences and non-fall occurrences. In the study by [11], the Gaussian mixture model was utilized for observing the appearance of a person during the video sequence. This approach detects the fall event in case of any deformation found in the shape of the concerned person. In the study by Cai et al. [12], a vision-based fall detection system was proposed in which hourglass residual units were introduced to extract multiscale features. SoftMax classifier was used for the categorization of both fall and non-fall events. In the study by Zhang et al. [13], the YOLACT network was applied to the video stream to distinguish different human activities and postures. A convolutional neural network (CNN) was designed for the classification of fall vs non-fall events. Vision-based fall detection systems are widely used; however, these fall detection systems have certain limitations i.e., privacy issues, real-life falls are difficult to detect because of the data set, high-cost because high-resolution cameras are required to cover the entire room, computational complexity due to the processing of millions of video frames.

The research community has also explored various machine learning and spectral features-based fall detection systems for the detection of fall occurrences and non-fall occurrences [14–17]. In the study by [14], the Hidden Markov model was used to determine the fall event. The Mel frequency cepstral coefficients (MFCC) features are capable to extract prominent information from the audio signals and are used for different research works [15, 18–22], respectively. In the study by [22], MFCC features were used to train Nearest Neighbor (NN) for the categorization of fall occurrences and non-fall occurrences. In the study by [23], MFCC, Gammatone cepstral coefficients (GTCC), and skew-spectral features were used for extracting features, and a decision tree was used for the classification of fall and non-fall events. In the study by Shaukat et al. [21], MFCC and Linear Predictive coding (LPC) were utilized for the voice recognition of elderly persons. An ensemble classifier was employed for classification purposes on daily sound recognition (AudioSet 2021) and RWCP (Open SLR 2021) datasets. In the study by [15], MFCC features were used with one class support vector machine method (OCSVM) for the classification of the fall and non-fall sounds. In our prior work [15], we proposed an acoustic-LTP features-based approach with the SVM classifier for fall event detection. This method was more effective than MFCC in terms of computational cost and also rotationally invariant. Although the above-mentioned sensors-based, acoustic-based, and computer vision-based fall detection systems achieve better detection of fall events. However, different restrictions are still present in modern methods i.e., fall detection systems can be implemented merely in wearable devices, some frameworks are only sensors-based which makes it difficult for elderly people to carry the body sensors all the time, and computer vision-based fall detection systems have privacy concerns, high computational costs, failure in fall detection in case server fails in a client–server architecture, etc. So, there is a need to develop automated fall event detection systems that are robust to the above-mentioned limitations. The major contributions of our study are as under:

- We present a novel audio feature i.e., Std-LTP that is capable of extracting the most discriminative characteristics from the input audio.
- We present an effective voice-operated fall detection system that can reliably be utilized for determining fall occurrences.
- We created our own in-house audio fall event dataset that is distinct respect of speakers, speaker gender, environment, etc.

The remaining paper is organized as follows. In Sect. 2, we discussed the proposed methodology. In Sect. 3, experimental results are discussed, whereas we conclude our work in Sect. 4.

2 Proposed Methodology

The main goal of the designed system is to identify fall occurrences and non-fall occurrences from audio clips. Extraction of features and classification of audios are the two steps that are involved in this proposed system. Initially, we extracted our 20-dimensional Std-LTP features from the audio input and then used all the 20-dimensional features to classify the fall occurrences and non-fall occurrences. For classification purposes, we employed the SVM. The flow diagram of the designed system is given in Fig. 1.

2.1 Feature Extraction

The extraction of features is critical for designing an efficient categorization system. The process of feature extraction of the proposed method is explained in the following section.

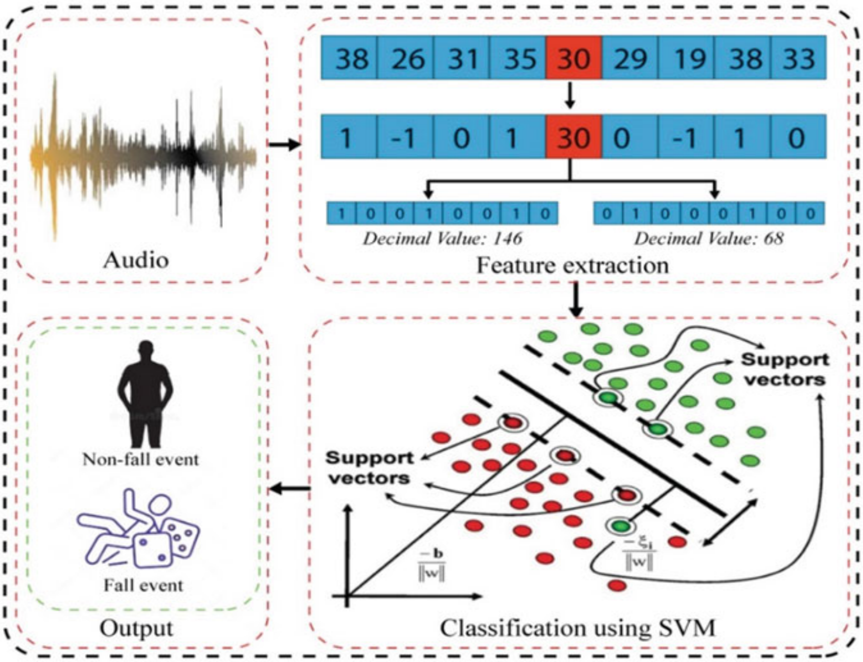

Fig. 1 Proposed system

2.2 Std-LTP Computation

In the proposed work, we presented the Std-LTP features descriptor to extract the characteristics of fall occurrences and non-fall occurrences from the audio. We obtained the 20-dimensional Std-LTP features from the audio signal y[n]. To extract features from the voice using Std-LTP, the audio signal is divided into multiple windows (Wc). We computed the Std-LTP by encoding each Wc of an audio signal y[n]. The total windows are obtained by dividing the samples by 9. Each Wc comprises nine samples which are used to generate the ternary codes. Initially, we computed the threshold value of each window Wc. In the prior study [22], acoustic-LTP utilized a static threshold value for each Wc which doesn't take into account the local statistics of the samples of each Wc. In this paper, we computed the value of the threshold using the local statistics of each sample around the central sample c in each Wc. We computed the threshold value by calculating the standard deviation of each Wc and multiplying it by scaling factor α, so, the threshold for each Wc varies. The standard deviation of each Wc is computed by using the following equation:

$$\sigma = \sqrt{\frac{\sum_{i=0}^{8}(q_i - \mu)^2}{N}} \tag{1}$$

where σ is standard deviation, q_i is the value of each sample in the Wc, μ is the mean of nine values of that Wc, N is the number of samples which is nine. The threshold is calculated as follows:

$$\sigma * \alpha \tag{2}$$

where α is the scaling factor and $0 < \alpha \leq 1$. We used $\alpha = 0.5$ in our work because we achieved the best results in this setting.

We compared the c with corresponding neighboring values. To achieve this purpose, we quantified the magnitude difference between c and the neighboring samples. Values of samples greater than *c+th* are set to 1 and those which are smaller than *c–th* are set to –1, whereas values between *c–th* and *c+th* are set to 0. Hence, we obtained the ternary codes as follows:

$$f(q_i, c, t) = \begin{cases} +1, & q_i \geq (c + (\sigma * \alpha)) \\ 0, & (c - (\sigma * \alpha)) < q_i < (c + (\sigma * \alpha)) \\ -1, & q_i \leq (c - (\sigma * \alpha)) \end{cases} \tag{3}$$

where $f(q_i, c, t)$ is the function representing the ternary codes.

For instance, consider the following frame having 9 samples as shown in Fig. 2. We computed the standard deviation of the Wc which is $\sigma \approx 6$ in this case. Next, we multiply the standard deviation value by the scaling factor of 0.5 to get the threshold value that is $\sigma * \alpha = 6 * 0.5 = 3$. So, values that are greater than 33 are set to +1, less than 27 are set to –1, and values between 33 and 27 are set to 0. In this way, the ternary code of the vector having nine values is generated.

Next, we compute the upper and lower binary codes. For upper codes, we set the value to 1 where the ternary code is +1 and values of 0 and –1 are set to zero as:

$$f_u(q_i, c, t) = \begin{cases} 1, & f(q_i, c, t) = +1 \\ 0, & otherwise \end{cases} \tag{4}$$

For lower codes, we set all values of –1 to 1 and 0 and +1 to 0 as:

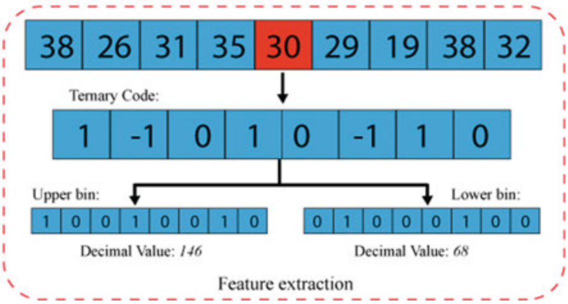

Fig. 2 Feature extraction

$$f_l(q_i, c, t) = \begin{cases} 1, & f(q_i, c, t) = -1 \\ 0, & otherwise \end{cases} \tag{5}$$

We transformed these upper and lower codes into decimal values as:

$$T^u_{\Omega_p} = \sum_{i=0}^{7} f_u^{uni}(q_i, c, t).2^i \tag{6}$$

$$T^l_{\Omega_p} = \sum_{i=0}^{7} f_l^{uni}(q_i, c, t).2^i \tag{7}$$

Histograms are calculated for the upper and the lower codes as follows:

$$h_u(k) = \sum_{w=1}^{W} \delta(T_W^{(u)}, k) \tag{8}$$

$$h_l(k) = \sum_{w=1}^{W} \delta(T_W^{(l)}, k) \tag{9}$$

where k shows the histogram bins. We used ten patterns for upper and lower binary codes to capture the characteristics of the sound involving the fall and non-fall events as our experiments provided the best results on ten patterns from both groups. We combined the ten upper and ten lower codes to develop a 20-dim Std-LTP descriptor as:

$$Std - LTP = [h_u||h_l] \tag{10}$$

2.3 Classification

The binary classification problems can be easily resolved by using a SVM, therefore we utilized SVM in our work for performing classification. The Std-LTP features are utilized to train SVM for categorizing fall occurrences and non-fall occurrences. We tuned different parameters for SVM and set the following values: box constraint of 100, kernel scale to 1, gaussian kernel, and outlier function to 0.05.

Table 1 Details of fall and non-fall dataset

No of samples	No of fall samples	No of non-fall samples	Training samples	Testing samples
508	234	274	408	100

3 Experimental Setup and Results Discussion

3.1 Dataset

We developed our fall detection dataset comprising audio clips of fall occurrences and non-fall occurrences recorded with two devices i.e., Lenovo K6 note and infinix note 10 pro. The dataset is specifically designed to detect fall events. We recorded the voices of different speakers for fall and non-fall incidents in various environments and various locations i.e., home, guest room, washroom, etc. The period of sound clips varies from 3 to 7 s in duration. Sound clips of fall occurrences comprise intense painful audio while the sound clips of non-fall occurrences consist of inaudible audio, conversations, TV being played in the background, etc. The dataset has 508 audio samples that comprise 234 samples of fall events and 274 samples of non-fall events. The audio categorization of the dataset is reported in Table 1.

3.2 Performance Evaluation of the Proposed System

This experiment is performed to evaluate the efficacy of the developed system for the detection of possible fall occurrences on our in-house fall detection dataset. For this experiment, we utilized the data up to 80% (408 samples to train the model and 20% data (100 samples) for testing purposes. More specifically, we used 234 fall event audios and 274 non-fall audios. We obtained the 20-dim Std-LTP features of all the sound clips and trained them on the SVM for the categorization of fall occurrences and non-fall occurrences. We obtained an accuracy of 93%, precision of 95.74%, recall of 90%, and F1-score of 91.78%. These results enhance the efficiency of the developed system for fall detection.

3.3 Performance Comparison of Std-LTP Features on Multiple Classifiers

We conducted an experiment to measure the significance of SVM with our Std-LTP features for fall detection. For this, we selected different machine learning classifiers i.e., Logistic regression (LR), Naïve Bayes (NB), K-nearest neighbour

Table 2 Performance comparison of multiple classifiers

Method	Kernel	Accuracy%	Precision%	Recall%	F1-score%
Std-LTP + LDA	Linear	90	90	91.83	90.90
Std-LTP + NB	Kernel NB	82	88	78.57	83.01
Std-LTP + KNN	Fine	92	92	92	92
Std-LTP + Ensemble	Subspace KNN	92	94	90.38	92.15
Std-LTP + DT	Coarse	78	58	95.45	72.15
Std-LTP + SVM	Fine Gaussian	93	95.74	90	91.78

(KNN), ensemble, Decision tree (DT), along with the SVM, and trained them using the proposed features and results are reported in Table 2. We can observe that Std-LTP performed best with the SVM by achieving an accuracy of 93%. The Std-LTP with the KNN and ensemble subspace KNN achieved the second-best results with an accuracy of 92%. The Std-LTP with DT was the worst and achieved an accuracy of 78%. We can conclude from this experiment that the proposed Std-LTP features with the SVM outperform all comparative classifiers for fall event detection.

4 Conclusion

In this research work, we have presented a better accidental fall detection approach for the elderly persons to provide first aid. Elderly people live in home alone, which need continuous monitoring and special care. Therefore, in this work, we designed a novel approach based on the proposed innovative acoustic Std-LTP features to obtain the prominent attributes from the screams of the accidental falls. Moreover, we developed our in-door fall incidents diverse dataset that has voice samples of screams and pain voices. We used our in-door fall events dataset for experimentation purposes and obtained 20-dimensional proposed Std-LTP features from the voice clips. We fed the extracted 20-dim Std-LTP features into SVM for distinguishing between the fall occurrences and non-fall occurrences. Experimental results show that the proposed method efficiently identifies fall occurrences with 93% accuracy and the lowest false alarm rate. Furthermore, it is possible to implement this system in actual environments i.e., in hospitals, old houses, nursing homes, etc. In the future, we aim to use the proposed Std-LTP features on other fall events datasets to check the effectiveness and generalizability of the proposed system. We also aim to send the location of monitored persons to caretakers where the fall is detected.

References

1. Scuffham P, Chaplin S, Legood R (2003) Incidence and costs of unintentional falls in older people in the United Kingdom. J Epidemiol Community Health 57(9):740–744
2. Tinetti ME, Speechley M, Ginter SF (1988) Risk factors for falls among elderly persons living in the community 319(26):1701–1707
3. Voermans NC, Snijders AH, Schoon Y, Bloem BR (2007) Why old people fall (and how to stop them). Pract Neurol 7(3):158–171
4. Yacchirema D, de Puga JS, Palau C, Esteve M (2019) Fall detection system for elderly people using IoT and ensemble machine learning algorithm. Pers Ubiquit Comput 23(5):801–817
5. Giansanti D, Maccioni G, Macellari V (2005) The development and test of a device for the reconstruction of 3-D position and orientation by means of a kinematic sensor assembly with rate gyroscopes and accelerometers 52(7):1271–1277
6. Muheidat F, Tawalbeh L, Tyrer H (2018) Context-aware, accurate, and real time fall detection system for elderly people. In: 2018 IEEE 12th international conference on semantic computing (ICSC). IEEE
7. Mezghani N, Ouakrim Y, Islam MR, Yared R, Abdulrazak B (2017) Context aware adaptable approach for fall detection bases on smart textile. In: 2017 IEEE EMBS international conference on biomedical & health informatics (BHI). IEEE, pp 473–476
8. Popescu M, Li Y, Skubic M, Rantz M (2008) An acoustic fall detector system that uses sound height information to reduce the false alarm rate. In: 2008 30th annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 4628–4631
9. Yan H, Huo H, Xu Y, Gidlund M (2010) Wireless sensor network based E-health system-implementation and experimental results. IEEE Trans Consum Electron 56(4):2288–2295
10. Yu M, Gong L, Kollias S (2017) Computer vision based fall detection by a convolutional neural network. In: Proceedings of the 19th ACM international conference on multimodal interaction
11. Rougier C, Meunier J, St-Arnaud A, Rousseau J (2011) Robust video surveillance for fall detection based on human shape deformation. IEEE Trans Circuits Syst Video Technol 21(5):611–622
12. Cai X, Li S, Liu X, Han G (2020) Vision-based fall detection with multi-task hourglass convolutional auto-encoder. IEEE Access 8:44493–44502
13. Zhang L, Fang C, Zhu M (2020) A computer vision-based dual network approach for indoor fall detection. Int J Innov Sci Res Technol 5:939–943
14. Tong L, Song Q, Ge Y, Liu M (2013) HMM-based human fall detection and prediction method using tri-axial accelerometer. IEEE Sens J 13(5):1849–1856
15. Khan MS, Yu M, Feng P, Wang L, Chambers J (2015) An unsupervised acoustic fall detection system using source separation for sound interference suppression. Signal Process 110:199–210
16. Younis B, Javed A, Hassan F (2021) Fall detection system using novel median deviated ternary patterns and SVM. In: 2021 4th international symposium on advanced electrical and communication technologies (ISAECT). IEEE
17. Banjar A et al (2022) Fall event detection using the mean absolute deviated local ternary patterns and BiLSTM. Appl Acoust 192:108725
18. Qadir G et al (2022) Voice spoofing countermeasure based on spectral features to detect synthetic attacks through LSTM. Int J Innov Sci Technol 3:153–165
19. Hassan F, Javed A (2021) Voice spoofing countermeasure for synthetic speech detection. In: 2021 International conference on artificial intelligence (ICAI). IEEE, pp 209–212
20. Zeeshan M, Qayoom H, Hassan F (2021) Robust speech emotion recognition system through novel ER-CNN and spectral features. In: 2021 4th international symposium on advanced electrical and communication technologies (ISAECT). IEEE
21. Shaukat A, Ahsan M, Hassan A, Riaz F (2014) Daily sound recognition for elderly people using ensemble methods. In 2014 11th international conference on fuzzy systems and knowledge discovery (FSKD). IEEE, pp 418–423

22. Li Y, Ho KC, Popescu M (2012) A microphone array system for automatic fall detection. IEEE Trans Biomed Eng 59(5):1291–1301
23. Hassan F, Mehmood MH, Younis B, Mehmood N, Imran T, Zafar U (2022) Comparative analysis of machine learning algorithms for classification of environmental sounds and fall detection. Int J Innov Sci Technol 4(1):163-174s