Applied Acoustics 192 (2022) 108725

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Fall event detection using the mean absolute deviated local ternary patterns and BiLSTM

Ameen Banjar [a], Hussain Dawood [b,*], Ali Javed [c], Farman Hassan [c]

[a] Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Jeddah 23890, Saudi Arabia
[b] Department of Computer and Network Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 23890, Saudi Arabia
[c] Department of Software Engineering, University of Engineering and Technology Taxila 47050, Punjab, Pakistan

ARTICLE INFO

Article history:
Received 20 October 2021
Received in revised form 6 February 2022
Accepted 8 March 2022
Available online 19 March 2022

Keywords:
A3 Fall v2.0
BiLSTM
Fall events
Healthcare
MAD-local ternary patterns
The daily sound

ABSTRACT

Fall detection in elder persons may result in long-lasting injury that can have severe consequences for the rest of their lives. Additionally, prolonged delay in emergency treatment after the fall event escalates the chances of mortality. Hence, fall detection at early stage is critical in terms of providing timely aid with little complications and minimize hospitalization expenses. This work aims to provide an effective and efficient healthcare solution to determine the event of fall detection for elderly persons. We aim to address this fall detection problem for elder persons living lonely and encounter issues in case they fall and are unable to call for assistance. In this paper, we present a fall detection framework by proposing a novel feature space mean absolute deviated-local ternary patterns (MAD-LTP) to examine the environmental sounds and used these features to train the BiLSTM for fall events detection. Our proposed MAD-LTP features successfully address the limitations of existing features i.e., non-robust over dynamic pattern detection, brute force optimization, intolerance over non-uniform noise, etc., for fall detection. Performance of our system is evaluated on three diverse datasets i.e., The daily sounds, A3 Fall 2.0, and our in-house developed MSP-UET fall detection dataset. We compared the performance of the proposed framework against the state-of-the-art methods. We obtained an accuracy of 93.5%, 98.29%, and 98% for the daily sounds, A3 Fall 2.0, and our in-house developed MSP-UET fall detection dataset. Experimental findings indicate the reliability of our method for fall event detection.

1. Introduction

In current times, we have observed a tremendous rise in elderly population around the globe due to the improvement in public health, medication, economic and social development, disease control and injury prevention mechanisms, and reduction in premature death rate. According to the National Institute of Health report [1], 8.5% of the world's population was 65 years or elder in 2016 which is likely to rise to 17% of the globe's population by 2050. In US only, people with 65 and elder is estimated to rise from 48 to 88 million by 2050. Moreover, people with age 80 or more is expected to increase thrice from 2015 to 2050 while some Asian and Latin American countries might experience a rise of four times in the same period. Another report [12] published by the United Nations show the similar statistics with 1.5 billion aging people by 2050 across the globe. According to National Council of Aging [14], falls contribute 2.8 million injuries including 0.8 million hospitalizations with over 27,000 mortalities every year in the US. These falls cause head trauma, broken hips and legs, and delay in surgeries increases the risk of fatality rate from 7.3% to 8.7%. Elder persons living lonely face agony in case of falling on the floor and not been able to call for assistance. For elderly people, these falls specifically on the concrete floor can develop into a long-lasting and life-threatening injury that can ultimately results in permanent disability. Therefore, fall detection at early stage is vital to facilitate the patient for timely first aid, prevent complications, and reduces the chance of any long-term injury and treatment costs.

Current fall detection techniques can be categorized into sensor-based, vision-based, or audio-based. The sensor-based methods for fall detection use different sensors data which are acquired through either the wearable or environmental sensors [3–8]. Wearable devices at times are not convenient for elder persons or patients. A doppler radar-based fall detection technique was designed in [7] for human action recognition. Similarly, radar sensing fall detection method based on human motion detection was also proposed in [8]. Radar oriented doppler systems have

* Corresponding author.
E-mail address: hdaoud@uj.edu.sa (H. Dawood).

https://doi.org/10.1016/j.apacoust.2022.108725
0003-682X/

limited applicability. Video analysis-based methods [9,26] have also been proposed for fall detection. In [26], convolution neural networks-based approach was used for fall detection. In [9], three visual features consisting of movement details, variation of human shape, and histogram were used for falls detection. Video analysis-oriented methods suffer from different privacy issues and higher computational cost, thus, less suitable for real-time applications.

To address the above-mentioned issues associated with the wearable sensors-based or video analysis-based methods, acoustic-analysis based fall detection methods have also been proposed. In [11], Mel-frequency Cepstral Coefficients (MFCC) features were obtained from the acoustic signals to train the nearest neighbor (NN) classifier for fall detection. In [13], floor vibration waves were analyzed in combination of fall sounds to detect fall incidents. In [15], MFCC features were utilized to train NN, SVM, and Gaussian mixture model (GMM) models for fall detection. MFCCs are mostly used due to lower dimensions of features. However, acoustic data acquisition is affected due to the addition of environmental noise in the acoustic signal. MFCC features can be integrated with other acoustic features to reduce the effect of noise and performance improvement, however, at the cost of increased computational complexity.

To address the above-mentioned issues, more effective and efficient fall detection systems are required. This paper introduces an acoustic-based fall event detection framework that employs a novel feature representation method i.e., mean absolute deviated-local ternary patterns (MAD-LTP) to better capture the attributes of sounds associated with the fall i.e., scream, etc., under indoor and outdoor environments with the background noise. The MAD-LTP features are then utilized to train the BiLSTM architecture for classification of the fall as well as non-fall events. We assessed the performance of our framework on three different and diverse datasets i.e., The daily sounds [46], A3 Fall v2.0 [47], and our in-house developed MSP-UET fall detection dataset [48]. The main contributions of the proposed research work are as under:

- We develop an effective and efficient framework for fall event detection.
- We design a novel audio feature descriptor i.e., MAD-LTP for audio representation that is robust to non-uniform noise, rotation, dynamic pattern detection, and outdoor environments.
- We develop an in-house MSP-UET fall detection dataset that is captured under diverse environments to test the robustness of fall detection systems.
- Rigorous experimentation was performed on three diverse datasets to investigate the usefulness of the proposed method against existing contemporary fall detection systems.

The rest of the paper is organized as follows. Section 2 has presented the related work based on wearable devices, vision-based, and acoustic based fall detection systems. In Section 3, we present the proposed acoustic based fall detection system while in Section 4 we discuss the experimental setup and results in detail while the Section 5 concluded our work.

2. Related work

The research community has explored various fall detection techniques that can be divided into three distinct groups such as wearable devices/sensors-based techniques, vision-based techniques, and acoustic-based techniques. We discuss all of these three categories in detail in the subsequent sections. Moreover, summary of the existing wearable devices/sensor-based, vision-based, and acoustic-based fall detection systems are provided in Table 1.

2.1. Wearable devices/sensors-based fall detection systems

Research community has introduced numerous fall detection methods [16–25] established on wearable sensor technology, which are commonly employed to develop real-time fall detection systems. Instead of external sensors, body mounted sensors are attached to the body of person at various locations. Most common wearable sensors utilized for fall detection systems are accelerometer, gyroscope, magnetometer, heart rate sensor, and so forth. Tri-axial accelerometer sensors have three axes i.e., X, Y, and Z to determine the location and motion of person's body by determining the change in velocity. The fall event relies on the accidental increase in negative acceleration that is caused by change in the orientation from standing to the laying position horizontally. In [16], fall detection technique based on tri-axial accelerometer sensor was developed to detect the fall events for elderly people. Fall event was detected when tri-axial accelerometer sensor reported the surpass of the normal range of acceleration. In [17], threshold dependent based fall detection system was designed by collecting data from sensory devices such as cardiotachometer, smart sensor, and accelerometer. In [18], Convolutional neural network (CNN) was employed for fall incident detection. This method [18] reduces the preprocessing work by employing the CNN to extract features automatically from Gyroscope sensor data instead of feeding handcrafted features to the model. Research approaches combine multiple sensors when using gyroscope sensors for tracking the angular velocity and many use gyroscopes for fall events detection. In [19], fall detection system based on threshold reliant algorithm was designed using the data collected from bi-axial gyroscope sensor. This technique sets the threshold values for angular velocity, angular acceleration, and resultant values of the trunk angle. Alarms are triggered and fall events are detected if an event exceed the threshold values. In [20], multiple sensors, namely, accelerometer and gyroscope were utilized for detection of fall incidents. This approach used the posture information from sensors to reduce the false negatives (sitting on stairs) and false positives (sitting down fast) resulting in the improvement in accuracy for fall events detection along with minimum computational cost. In [21], body segments kinematics were explored to detect the fall events using two inertial sensors i.e., 2D gyroscope and 3D accelerometers. In [22], multimodal sensors i.e., light intensity, gyroscope, and tri-axis accelerometer, electroencephalograph helmet, cameras in lateral and front viewpoints, and infrared sensors were used for data acquisition. Distinct machine learning techniques such as SVM, RF, multilayer perceptron (MLP), and KNN were used for the classification of fall events and activities of daily living (ADL).

Existing methods [23–25] have also used the fusion of sensors such as micro-Doppler radar, context aware, depth camera, magnetometer, barometer, gyroscope, etc., to combine the statistics from numerous sensors for the development of fall detection systems. Multi sensors fusion often produces more accurate fall detection techniques than the single sensor-based techniques. In [23], sensors fusion based fall detection technique containing multiple sensors i.e., context aware sensors (e.g., depth camera and micro-Doppler radar) and tri-axial accelerometer was presented to enhance the classification accuracy of fall and non-fall events. In [24], fusion of multiple sensors i.e., magnetometer, barometer, and gyroscope were employed to detect the fall events with quaternion filter that extract acceleration relative to the frame of an earth. Thresholding approach was applied on different features such as angular velocity, acceleration, and altitude to detect the fall event. In [25], kinematic sensors were used for data acquisition. A feature selection algorithm based on the integration of J3 and

Table 1
Summary of the Existing Wearable Devices/Sensor Based, Vision-Based, and Acoustic-Based Systems.

Wearable devices/sensor based system				
Methods	Sensors	Classifiers/Technique	Dataset	Limitations
Lai CF et. al., [16]	Tri-axial accelerometer	SVM	Custom Dataset	sensors to be worn all the time
Wang J et. al., [17]	Cardio tachometer	SVM, Threshold-based method	Custom Dataset	Intrusive approach
Casilari E et. al., [18]	Gyroscope	CNN	SisFall Dataset	Detection dependent on short proximity of a range
Bourke AK el. al., [19]	Bi-axial gyroscope	Threshold-based method	Custom Dataset	Expensive equipment
Li Q et. al., [20]	Gyroscope and accelerometer	Threshold-based method	Custom Dataset	Complex syncing between devices, difficulties in equipment setup
Computer Vision-Based Systems				
Methods	Features	Classifiers	Dataset	Limitations
Núñez-Marcos A et. al., [26]	Optical flow image-based motion modeling system	Kernel-SVM	CASIA	false displacement vectors generation due to variations in the lighting conditions
Gunale K et. al., [27]	orientation angle, aspect ratio, threshold, and Motion History Image (MHI)	SVM, KNN, Stochastic Gradient Descent (SGD), decision trees (DT), and Gradient Boosting (GB)	Le2i Fall	Generalizability issue as a small amount of data was used
De Miguel K et. al., [30]	Kalman filtering, optical flow, and background subtraction	KNN	Custom Dataset	movement patterns of humans change due to health and age, which makes it difficult for the algorithm to detect the fall event correctly
Bian ZP et. al., [33]	Pose- invariant randomized DT	SVM	Custom Dataset	Human's privacy issues
Thuc HL et. al., [34]	5-dim feature vector using ellipse model	Hidden Markov Model (HMM)	NTURGB + DActionRecognition	High computational cost
Acoustic based fall detection systems				
Methods	Features	Classifiers	Dataset	Limitations
Collado-Villaverde A et. al., [35]	energy, zero crossing, spectral centroid, rolloff factor, and spectral flux	C4.5 (J48), nearest neighbor (NN), Logistic regression (LR), Naïve Bayes, PART, Random Forest, and SVM	War and non-war sounds clips	requirement of accurate segregation of scream and normal speech
Droghini D et. al., [36]	MFCC, and Gaussian mean super vectors (GMS)	One-class SVM (OCSVM)	A3 Fall v2.0	dependency on microphone quality, device limitations for embedding,
Li X et. al., [37]	MFCC and spectrogram	heterogenous ensemble learning (HEL)	A3 Fall v2.0	Complexity and computational cost
Khan MS et. al., [42]	MFCC	OCSVM	The Daily Sounds	High computational cost
Adnan SM et. al., [49]	Acoustic-LTP	SVM	The Daily Sounds	less robust to background noise

Fisher's discriminant ration criterion was created to pick the appropriate characteristics. Hierarchical classifier was trained on the selected features to detect the fall events. Although wearable devices-based fall detection systems are frequently used due to ease in availability of such sensors, however, there are certain limitations of wearable devices-based fall detection systems such as sensors to be worn all the times, intrusive approach, detection dependent on short proximity range, expensive equipment, complex syncing between devices, and difficulties in equipment setup.

2.2. Computer vision-based fall detection systems

To address the limitations of wearable devices-based fall recognition systems, existing approaches have also explored the computer vision-based techniques [26–34] to automatically detect the fall events. Vision based fall detection methods give more freedom to elderly people to perform their daily lives activities without wearing any sensor on the body. In [26], optical flow images-based motion modeling method with the CNN was employed for fall detection. However, this optical flow-based approach has a drawback of false displacement vectors generation due to variations in the lighting conditions. In [27], four visual features i.e., orientation angle, aspect ratio, threshold, and Motion History Image (MHI) were obtained and fed to the SVM, KNN, Stochastic Gradient Descent (SGD), decision trees (DT), and Gradient Boosting (GB) separately for fall detection. Experimental outcomes showed that DT performed well with minimum computational time over other classifiers. Some existing fall detection systems [28,29] depend on the posture classification techniques where all postures related to the fall event cannot be incorporated regardless of the amount of training data. In [30], several algorithms i.e., Kalman filtering, optical flow, and background subtraction were combined to develop a fall detection system for aged persons. However, movement patterns of humans change due to health and age, which makes it difficult for the algorithm to detect the fall event correctly. In [31], vision-oriented fall detection system using the depth images (3D) was designed. A dense spatio-temporal context algorithm was employed to trace the position of head. The centroid height of the human body, and distance between the floor plane and head were computed and compared with an adaptive threshold technique for fall detection. This approach has low computational complexity, but scale tracking process is highly unpredictable. In [32], a system for fall incidents detection based on the identification of abnormal velocity and position of subject was designed. The traced joints of the subject were utilized for measuring the velocity w.r.t the preceding position. In [33], a system for fall detection was presented that analyzed the traced key joints of person's body with help of a single depth camera. This approach is self-reliant on variations in the lightning and able to operate even in the low illumination conditions. Pose-invariant randomized DT was used to extract the key joints. SVM was employed for classification purpose that used two inputs i.e., head joint and 3D trajectory. In [34], a video-based fall detection technique was designed for aged persons. An adaptive-GMM was applied for human detection from the input frame and later transformed into a five-dimensional feature vector utilizing an ellipse model. Finally, this feature vector was employed for training the Hidden Markov Model (HMM) for fall event detection. Although vision-based fall event detection systems are effective, however, at the expense of increased computational complexity.

2.3. Acoustic based fall detection systems

Existing techniques [35–45] has investigated numerous fall incidents systems established on analyzing the sounds of people for fall event detection to address the high computational cost issue of vision-based systems. These fall detection systems require to identify the sounds people make while in pain after the fall event. Recently, wireless body area networks have been employed to develop large monitoring centers for acquiring data associated to falls and other hazardous conditions [43,44]. These techniques concentrate on observing large groups of elder persons at one time, collecting information of fall events, and then training the classification model on the collected data. Acoustic based fall detection systems use the microphone to record the sounds from environment. In [35], audio features comprising of energy, zero crossing, spectral centroid, rolloff factor, and spectral flux were employed to train different machine learning algorithms such as C4.5 (J48), nearest neighbor (NN), Logistic regression (LR), Naïve Bayes, PART, Random Forest, and SVM for fall detection. LR classifier performed the best among other classifiers. In [36], an acoustic based fall detection technique was developed utilizing a deep convolutional neural network autoencoder to detect the fall incidents. In [38], two spectral features i.e., MFCC, and Gaussian mean supervectors (GMS) were investigated for fall detection. One-class SVM (OCSVM) was utilized to classify fall and non-fall. In [37], an audio-based fall detection system using two spectral features i.e., MFCC and spectrogram and heterogenous ensemble learning (HEL) was developed to discriminate the fall and non-fall. In [39], a siamese neural network (SNN) was presented using an input audio for detection of fall incidents. In [42], MFCC features were utilized for the detection of fall events and OCSVM was employed to classify falls and non-falls. In our prior works [49,51], we employed the acoustic local ternary patterns and trained the SVM for classification of fall and non-fall incidents. This technique performed well in terms of an accuracy and precision, but the method was susceptible to noisy signals due to using the static threshold value. In [52], fall detection technique based on MFCC was designed for elderly people. Three classifiers i.e., SVM, KNN, and neural networks were employed for the classification purpose. In [53], two spectral features such as MFCC and linear predictive coefficients were utilized for training the ensemble classifier to detect fall event. In [54], four features such as MFCC, energy, zero crossing, and spectral flux were utilized to train a deep neural network for categorizing the fall and non-fall incidents. In [10], MFCC-based fall detection technique was developed and SVM was utilized as classifier. Although the acoustic-based fall detection systems are computationally efficient over the vision-based or hybrid systems, however, these fall detection systems based on acoustic features have certain limitations such as less robust to background noise, dependency on microphone quality, device limitations for embedding, requirement of accurate segregation of scream and normal speech, etc. To address these limitations, there exists a need to develop more robust acoustic-based fall detection systems.

3. Proposed methodology

This section provides the detailed working mechanism of our technique to detect fall incidents. The main objective of the proposed acoustic-based fall detection framework is to accurately differentiate between the fall and non-fall incidents. For this purpose, we developed a novel feature descriptor MAD-LTP to represent the input audio signal. The proposed system comprises of three stages i.e., preprocessing, features extraction and classification. In the first stage, we employed the HMM and Fast-ICA to suppress the silence segments from the audio. In the second stage, we extracted the 20 dimensional MAD-LTP features from input acoustic signal. In the third stage, MAD-LTP features are utilized to train a BiLSTM architecture for the classification purposes. The detailed working mechanism of our method is given in Fig. 1.

3.1. Preprocessing for silent segment suppression

The audio signal comprises of both the speech and silence segments. Since the silence zone do not hold any information that needs to be analyzed for fall event detection, therefore, we have eliminated the silent segments from the input audio to further reduce the content for processing. For this purpose, we employed the HMM model [2] and the FAST-ICA [41] techniques to segregate the low and high frequency signals. It is to be noted that the posterior probability of the acoustic events is larger than of the silence period. We used the FAST-ICA method to segment the audio frames with higher posterior probabilities from the input full-length signal. Hence, by suppressing the silent segments in the preprocessing stage helps to reduce the content needed to be analyzed for further processing.

3.2. Feature extraction

The key contribution of our work is the proposed novel feature extraction method MAD-LTP. The main objective of this new feature representation scheme is to address the issues associated with our prior acoustic-LTP features [49] that uses a fixed threshold value during the generation of LTP codes. Our existing acoustic-LTP features were proposed for indoor applications where it showed remarkable performance for fall detection event due to robustness over noise as compared to other acoustic features i.e., acoustic-LBP features. However, acoustic-LTP features have certain limitations due to this fixed threshold scheme, which are (a) not robust over dynamic pattern detection—spectral analysis of the fall-event audio under outdoor environment shows that the signal has dynamic repetitive patterns, which can be better captured via dynamic threshold scheme. However, the acoustic-LTP uses a fixed threshold for computing the LTP codes, therefore, there exists a need to improve the existing acoustic-LTP features for fall detection applications. (b) brute-force optimization—as in acoustic-LTP we need a brute-force scheme to optimize the threshold that makes it difficult to attain better peformance in real-time applications under diverse conditions. (c) intolerance over non-uniform noise— acoustic-LTP is robust against the consistent uniform noise that is available in the indoor audios, whereas, we experience the non-uniform noise in the outdoor environments, therefore, static threshold-based acoustic-LTP features are not robust under non-uniform noise and hence, not suitable for fall detection in outdoor environments. Therefore, we need more effective features robust to afore-mentioned limitations and can reliably detect the fall event in outdoor environments besides the indoor environments.

The feature computation process of MAD-LTP is shown in Fig. 2. To compute the MAD-LTP, we divide the suppressed audio signal $X[n]$ obtained after the preprocessing stage with N samples into non-overlapping frames of length $l = 9$. As we use 8 neighbors around a central sample c to compute the LTP code in our acoustic-LTP features, therefore, we also considered 8 neighbors around a center sample to compute the MAD-LTP codes. Hence, these 8 neighbors along-with the center neighbor makes a frame of 9 samples. Similar to acoustic-LTP, we quantize the sample

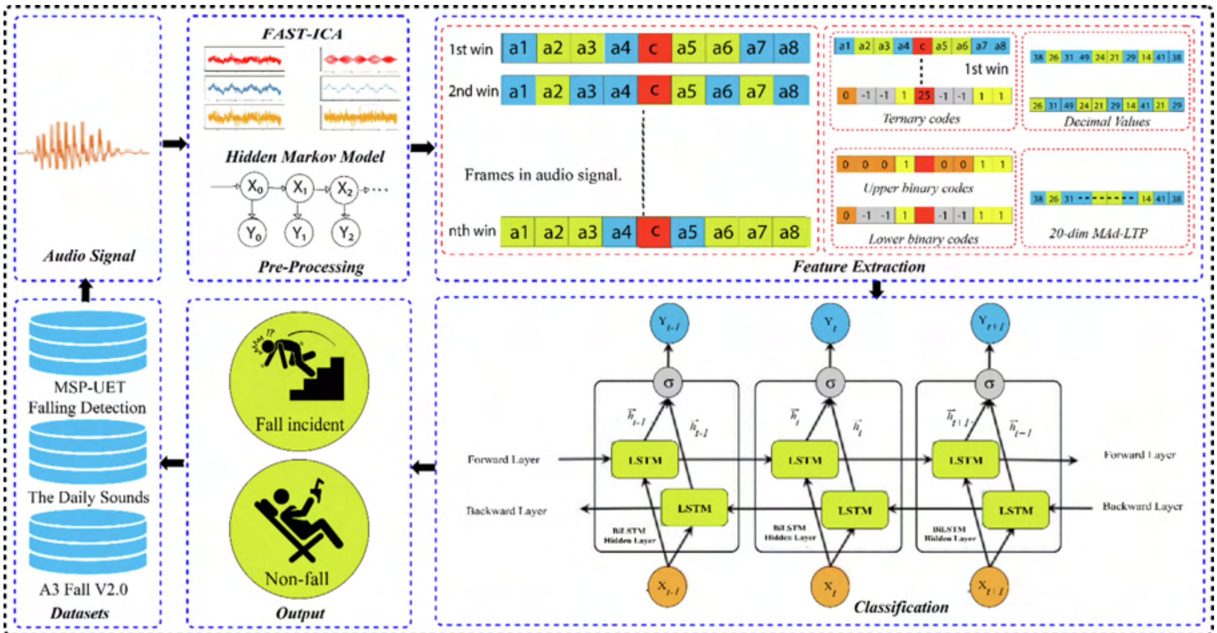

Fig. 1. Proposed system.

values to three codes i.e., -1, 0, and + 1 based on comparison of the neighboring samples against the center sample. More specifically, we assign a code of 0 when the neighboring sample lies in the range of $\pm\,\vartheta$around c. Next, we assign a code of + 1 when the neighboring sample is greater than or equals to the central sample +, and assign a code of -1 when the neighboring sample is less than or equals to the central sample - ϑ. We computed the ternary codes as follows:

$$MAD-LTP(s^i, c, \vartheta) = \begin{cases} +1, k^i \geq c + \vartheta \\ 0, |(k^i - \theta)| < \vartheta \\ -1, k^i \leq (c - \vartheta) \end{cases} \quad (1)$$

where *MAD-LTP* (s^i, c, ϑ) represents the adapted ternary codes, *c* is the central sample of the frame with k^i neighbors where *i* represents neighbor index and ϑ represents the threshold. In acoustic-LTP we used the fixed threshold, however, in our novel MAD-LTP, we propose an adaptive threshold calculation mechanism based on mean absolute deviation to assign the values of the threshold dynamically for each frame. More specifically, we computed the threshold as follows:

$$\vartheta = 1/l \sum_{i=1}^{l} |k_i - m(X)| \quad (2)$$

where *l* represents the length of the frame, k_i denotes values of the neighboring samples, *m* (*X*) represents the mean value of a single frame, and ϑ represents the mean absolute deviation of any given frame of our acoustic signal. Next, we transform the ternary codes into binary codes by splitting the MAD-LTP into upper ($MAD-LTP^U$) and lower ($MAD-LTP^L$) patterns. For ($MAD-LTP^U$), we quantize + 1 to 1 and rest values to 0, whereas, for ($MAD-LTP^L$) we quantize -1 to 1 and rest values to 0 as follows:

$$MAD-LTP^U(s^i, c, \vartheta) = \begin{cases} 1, if MAD-LTP(s^i, c, \vartheta) = +1 \\ 0, Otherwise \end{cases} \quad (3)$$

$$MAD-LTP^L(s^i, c, \vartheta) = \begin{cases} 1, if MAD-ALTP(s^i, c, \vartheta) = -1 \\ 0, Otherwise \end{cases} \quad (4)$$

As we already know from the computer vision domain that the uniform patterns contain important information in the signal over non-uniform patterns [40], therefore, we detect the uniform upper ($MAD-LTP^U_{uni}$) and lower ($MAD-LTP^L_{uni}$) patterns over all the extracted patterns to capture significant attributes from the acoustic signal and represent these patterns in decimal form as follows:

$$MAD-LTP^U_{uni}(s^i, c, \vartheta) = \sum_{i=0}^{7} 2^i \times MAD-LTP^U(s^i, c, \vartheta) \quad (5)$$

$$MAD-LTP^L_{uni}(s^i, c, \vartheta) = \sum_{i=0}^{7} 2^i \times MAD-LTP^L(s^i, c, \vartheta) \quad (6)$$

Finally, we compute the histogram of $MAD-LTP^U_{uni}$ and $MAD-LTP^L_{uni}$, where we use a separate bin for each uniform pattern and assign all non-uniform patterns to one bin that results in reduction of redundant information from the signal. We compute the histogram as:

$$H^U\left(MAD-LTP^U_{uni}, b\right) = \sum_{k=1}^{k} \left(MAD-LTP^U_k, b\right) \quad (7)$$

$$H^L(MAD-LTP^L_{uni}, b) = \sum_{k=1}^{k} \left(MAD-LTP^L_k, b\right) \quad (8)$$

where *b* is the histogram bin. From our experimental analysis, we found that the first ten uniform patterns each from the upper and lower class are enough to capture the salient attributes of the input audio. This makes our MAD-LTP features more efficient being a 20-dim feature descriptor than our prior acoustic-LTP features which comprise of 40-dim features. Finally, we concatenate both histograms to create a 20-dim feature descriptor as:

$$MAD-LTP = [H^U || H^L] \quad (9)$$

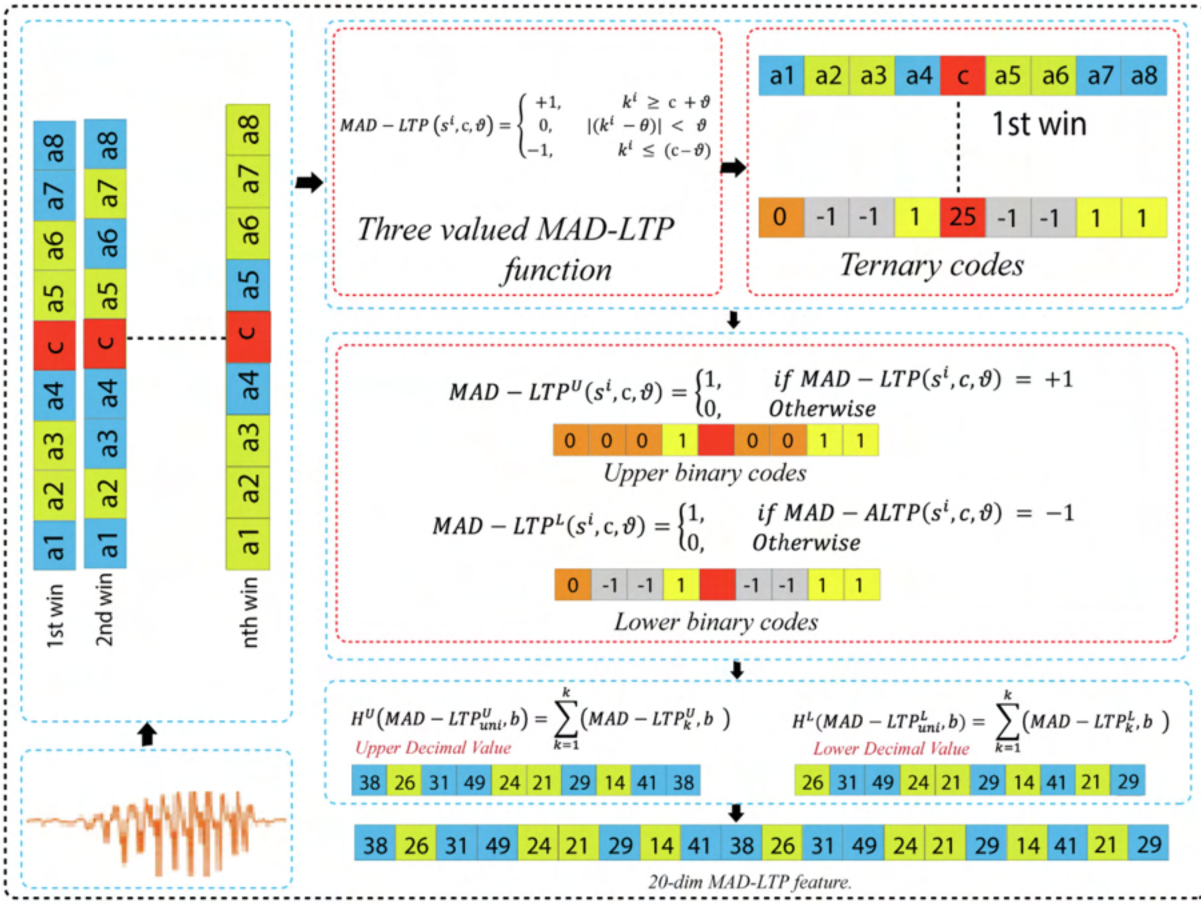

Fig. 2. Feature extraction.

Our analysis reveals that the audio signal for fall events contain the dynamic repetitive patterns and non-uniform noise under outdoor environment. Our prior static threshold based acoustic-LTP features are unable to reliably detect the fall events under these conditions due to employing a fixed threshold during the features extraction. The proposed MAD-LTP features by employing the mean absolute deviation for threshold computation better able to capture the dynamic repetitive patterns of the audios to distinguish between the fall and non-fall events even in the presence of non-uniform environmental noise.

3.3. Classification

As we know that the audio signal is a time-series data and BiLSTM being the recurrent network is well suited to classify sequential data. Moreover, LSTM is also used to classify sequential data, but it utilizes the preceding context only while BiLSTM overcome this problem by retrieving the data in two directions i.e., forward and backward. Therefore, in this work we utilized a BiLSTM architecture for classification purposes. To accomplish this, we extracted the proposed 20 dimension MAD-LTP features from the input audio to train the BiLSTM for the detection of fall incidents. We checked the performance of our approach by employing various configurations and tuned different parameters such as 25, 50, 75, and 100 hidden units, different optimizers as adam, sgd, adamax, rmsprop, sgdw, and nadam, different layers combination i.e., two layers, three layers, four layers, five layers, and so on up to ten layers, different mini-batch sizes as 16, 32, and 64, etc. After parameter tuning, we obtained the best results on the following: adam optimizer, max epoch to 15, 100 hidden units, 3 layers, and mini-batch size of 16. Thus, we used these parameters to tune our BiLSTM network for model training. Fig. 3. shows the detailed design of our BiLSTM model used for experiments in this work.

4. Experimental setup and results

In this section, we discuss different experiments performed to evaluate the performance of our system. Moreover, statistics of the datasets are also provided in this section. We measured the performance of our approach using the accuracy, precision, recall, and F1-score as also adopted by the comparative methods.

4.1. Dataset

We assessed the performance of the proposed system on the three publicly available datasets i.e., the daily sounds [46], A3 Fall v2.0 [47], and our in-house MSP-UET fall detection dataset [48].

The daily sounds dataset [46] comprises of a total of 1049 audio non-speech samples recorded at 16khz in a carpeted room of size 7 m by 12 m. This dataset consists of eighteen classes of sounds i.e., breathing, dishes, door clapping, electrical shaver, glass breaking, hair dryer, keys, paper tear, female scream, water falling, yawn, sneeze, male scream, laugh, hand clapping, female cry, door opening, and cough that are produced by different humans. Most of the sound's events are recorded at night-time to avoid the external interferences.

The A3 Fall v2.0 dataset [47] comprises of 720 events of fall and non-fall comprising of seven classes. Among the seven classes, six

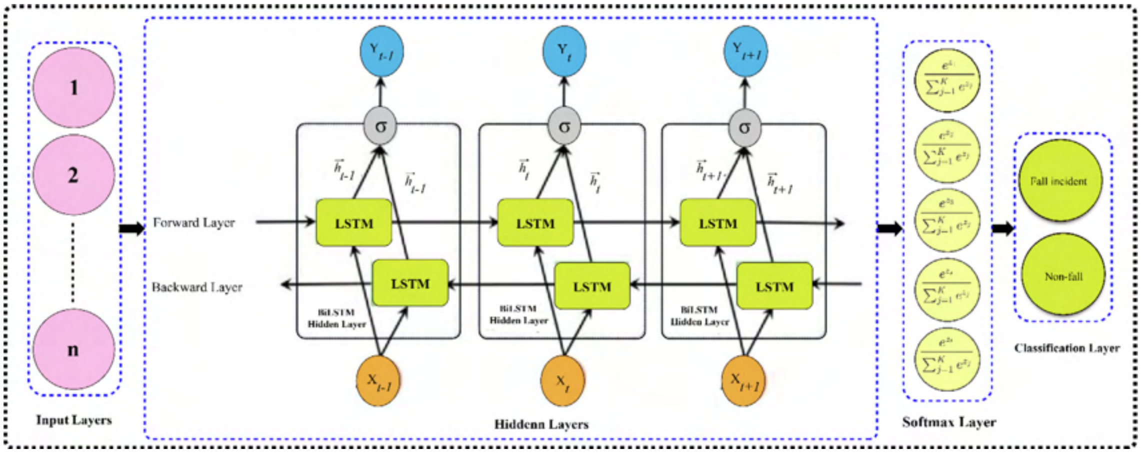

Fig. 3. BiLSTM architecture.

are non-speech audio such as ball, basket, chair, fork, book, and bag while the seventh class is of real human fall incidents. Audios were recorded using the three aerial microphones.

The third dataset MSP-UET fall detection [48] is our in-house dataset specifically designed to record the fall and non-fall events. We have used multiple human subjects of different age and gender to record this dataset that includes different tones of scream for fall events and normal conversation or silence for non-fall events. It is important to mention that we have recorded the samples on both the indoor and outdoor environments containing the background noise as well. This dataset has normal speech audio considered as non-fall events and voices of human screams during falls in indoor and outdoor environments as fall events that is recorded with multiple mobile devices i.e., Lenovo K6 note, infinix note 10 pro, Iphone 7, etc. The statistical details of these three datasets are given in Table 2.

4.2. Performance evaluation of the proposed method

The key purpose of this experiment is to assess the robustness of our technique on three different and diverse datasets for fall event detection. For this purpose, we designed a multi-stage experimentation for evaluating the performance of our approach (MAD-LTP-BiLSTM) on three different datasets. In the first stage, we evaluated our system on the daily sounds dataset for detecting incidents of fall and non-fall. To accomplish this, we employed 20 dimensional proposed MAD-LTP features from the audios of the daily sounds dataset [46] to train the BiLSTM network. We used 80% samples (839 audios) to train the model and 20% samples (209 audios) for model evaluation. We considered the fall and panic sounds i.e., female cry, female scream, male scream, and sneeze etc., for fall event. Whereas breathing, cough, dishes, door clapping, door opening, and electrical shaver for non-fall event. We achieved an accuracy of 93.5%, precision, recall, and F1-score of 100%, 91.59%, 95.61%, respectively, as shown in Fig. 4. These results show the usefulness of our approach to detect fall events on a diverse daily sounds dataset.

In the second phase of this experiment, we evaluated the performance of our approach on A3 Fall v2.0 dataset [47] for detection of fall incidents. To accomplish this, we extracted the proposed 20-dim MAD-LTP features from the audios of A3 Fall v2.0 dataset [47]. We split the dataset into 80–20 ratio to train the BiLSTM network for classification purpose. We used 80% of the data (576 samples) for training the model and remaining 20% (144 samples) for the testing. Fig. 4 illustrates the experimental results of the proposed system in terms of accuracy, precision, recall, and F1-score. Our approach obtained an accuracy, precision, recall, and F1-score of 98.29%, 97.53%, 100%, 98.75%, respectively. From the results reported in Fig. 4, we can observe that our method can reliably be used for the detection of fall events on A3 Fall v2.0 dataset.

In the third stage of this experiment, we measured the performance of our approach on our in-house developed MSP-UET fall detection dataset [48] to demonstrate the robustness of the proposed system on a diverse and challenging conditions. Again, we employed our proposed MAD-LTP features to train the BiLSTM network for classification of fall and non-fall events. We used 80% of data (408 samples) for training and rest 20% (100 samples) for evaluation and results are demonstrated in Fig. 4. From the above conclusions, we can examine that the proposed system achieved remarkable accuracy of 98%, precision of 100%, recall of 96.15%, and F1-score of 98.03%. Experimental results demonstrate that the proposed system has lowest false alarm rate of 4% for real human falls while 0% for non-fall events. This signifies the robustness of our system for fall event detection in the diverse indoor and outdoor environments.

Experimental outcomes of our approach on three different datasets using the computationally efficient MAD-LTP features signify that our method can reliably be used in real time and is able to capture the acoustic variations present in audio signals of fall events even in the presence of background noise. Moreover, our proposed system can be installed in cell phones and wearable devices i.e.,

Table 2
Statistics of daily sounds, A3 fall v2.0, and in-house fall detection datasets.

Dataset	Total no of samples	No of fall event samples	No of non-fall event samples	Training Samples	Testing Samples
MSP-UET Fall detection	508	234	274	408	100
The Daily sounds	1049	326	723	839	209
A3 Fall v2.0	720	154	412	576	144

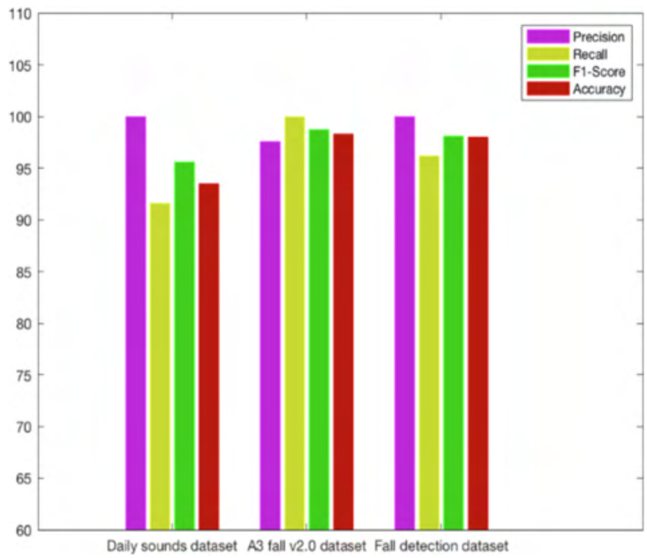

Fig. 4. Detection performance on The Daily Sounds, A3 fall v2.0 and proposed MSP-UET fall detection datasets.

smart watches, smart bracelet, etc., to successfully detect the fall and non-fall events.

4.3. Performance evaluation in multi-class environment

The key goal of this experiment is to assess the performance of our approach (MAD-LTP-BiLSTM) in a multi-class setting for detecting the panic, fall, and eco-friendly audios. There is significant correlation among all the classes of daily sounds dataset that becomes difficult to discriminate the fall incidents more accurately. Moreover, misclassification among various classes creates problem of class imbalance due to high correlation among multiple classes. To accomplish this, we extracted the MAD-LTP features from the audios of ten distinct classes i.e., breathing, dishes, door clapping, electrical shaver, glass breaking, hair dryer, keys, paper tear, female scream, and water falling of the daily sounds dataset. We utilized 80% (566 samples) of audios to train while the remaining 20% (140 samples) of the data for evaluating the trained model. From the outcomes illustrated in Table 3, we examined that our approach accurately detected the four classes i.e., breathing, dishes, electrical shaver, and keys with an accuracy, precision, recall, and F1-score of 100%. The proposed approach accomplished the second best performance on paper tear class and attained an accuracy of 98.72%, precision of 100%, recall of 96.55%, and F1-score of 98.24%. Our approach worked the worst on scream class and attained an accuracy of 89.16%, precision of 95.45%, recall of 100%, and F1-score of 97.67%. Overall, our system worked well and achieved an accuracy of 97.8%, precision of 98%, recall of 97.55%, and F1-score of 98.17% for all the ten classes. The detailed results of all the 10 classes are given in Table 3. Experimental outcomes signify that our approach is capable of detecting the fall events in multi-class environment.

4.4. Confusion matrix analysis

The key role of the confusion matrix is to illustrate the categorization assessment of approach. The accuracy metric alone can be misleading when there are unequal number of samples in each class. For example, we can achieve an accuracy of more than 90%, but this accuracy will not be considered good in case 90% of the samples belong to one class. Computing a confusion matrix can provide a better projection of the classification model in terms of correct or incorrect prediction for each class. Considering these facts, we designed a multi-stage experiment to visualize the classification performance of our system on the selected three datasets.

In the first stage, we created a confusion matrix to present the classification performance of our approach on the daily sounds dataset [46] as shown in Fig. 5. From the Fig. 5, we can see the projection of target (actual) and output (predicted) classes of fall and non-fall events. The true positives (TP) and true negatives (TN) are shown in the green diagonal blocks, whereas, the false negatives (FN) and false positives (FP) are shown in pink blocks. FP of zero indicates that the proposed system achieves 100% precision by not assigning any non-fall sample to fall sample, whereas, the FN of 10 among the total of 119 samples of fall event indicates that only 8% of the fall events are wrongly predicted as non-fall events. These outcomes demonstrate the reliability of our system to accurately detect the fall incidents.

Next, we provided the confusion matrix of our system for A3 Fall v2.0 dataset [47] and results are shown in Fig. 6. From this, we can see that our system obtained zero FNs that means 100% optimal recall rate. This recall rate signify that our system never detects the fall event as non-fall event. Moreover, FP of just 2 from the total true samples of 87 also show better precision rate of 97.53%.

In the third stage, we designed the confusion matrix of the proposed system on our own MSP-UET fall detection dataset as shown in Fig. 7. From this, we examine that FP rate is equal to 0% while FN rate is 4%. These results indicate that our approach effectively distinguish all the fall incidents and incorrectly classified 2 fall events as non-falls.

The lowest FN and FP rates on our MSP-UET fall detection dataset signifies the effectiveness of our approach for accurate fall events recognition on the indoor as well as outdoor environments.

In the last stage, we designed a confusion matrix of our method for multi-class environment of the daily sounds dataset for 10 classes such as breathing, dishes, door clapping, electrical shaver, glass breaking, hair dryer, keys, paper tear, female scream, water falling as shown in Fig. 8. From the confusion matrix on ten classes, we can observe that the FP values of seven classes such as breathing, dishes, door clapping, electrical shaver, key, paper tearing, and water falling are equal to 0 while FN values of seven classes such as breathing, dishes, electrical shaver, glass breaking, hair dryer, keys, and scream are equal to 0. Our method misclassified 3 glass breaking events into door clapping, 1 event of hair dryer into paper tear, and 1 event of scream into water falling event. These remarkable results demonstrate that our method is robust in multi-class environment problem and able to classify the complex and correlated environmental sounds with higher accuracy.

4.5. Performance evaluation on existing features

The key goal of this experiment is to assess the performance of our MAD-LTP features over the existing spectral features such as MFCC, GTCC, fusion of MFCC and GTCC (MFCC-GTCC), and acoustic-LTP on the same classifier. For this purpose, we extracted 14-dim MFCC, 14-dim GTCC, 28-dim features of MFCC and GTCC, 40-dim acoustic-LTP, and 20-dim MAD-LTP features from all the three datasets and used them to train the BiLSTM network separately for the classification of fall and non-fall events.

In the first phase of the experiment, we evaluated the performance of the proposed and existing features on the daily sounds dataset [46]. We extracted the MFCC, GTCC, MFCC-GTCC, acoustic-LTP, and the proposed MAD-LTP features and employed the BiSLTM network separately to distinguish between the fall and non-fall events. From the Table 4, we can observe that the MFCC-GTCC fusion performed the worst and achieved an accuracy

Table 3
Performance assessment in multi-class setting.

Classes	Accuracy%	Precision%	Recall %	F1-Score %
Breathing	100	100	100	100
Dishes	100	100	100	100
Door Clapping	93.08	100	92.10	95.89
Electrical Shaver	100	100	100	100
Glass Breaking	93.52	91.7	92.44	95.69
Hair Dryer	97.33	92.9	100	97.11
Keys	100	100	100	100
Paper Tear	98.72	100	96.55	98.24
Scream	89.16	95.45	100	97.67
Water Falling	92.2	100	94.44	97.14

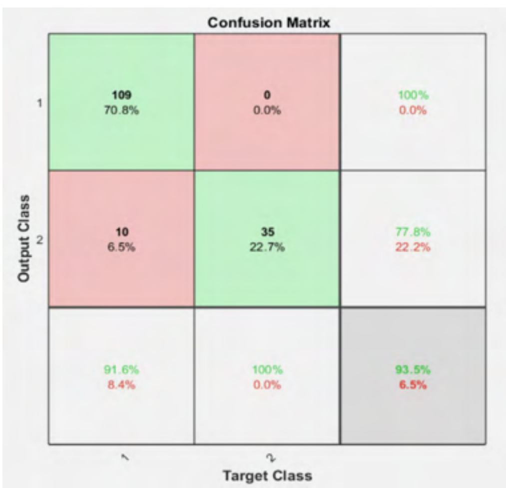

Fig. 5. Confusion matrix of the Daily Sounds dataset (1) Fall (2) Non-fall.

Fig. 6. Confusion matrix of Fall A3 v2.0 dataset (1) Fall (2) Non-fall.

of 35.8%, precision, recall and F1-score of 0%. This MFCC-GTCC features fusion based system detected all the fall events as non-fall. The MFCC + BiLSTM performed second best and achieved an accuracy of 81.7%, precision of 71.42%, recall of 100%, and F1-score of 83.33% while the proposed MAD-LTP achieved the best performance with an accuracy of 93.5%, precision of 100%, recall of 91.59%, and F1-score of 95.16%.

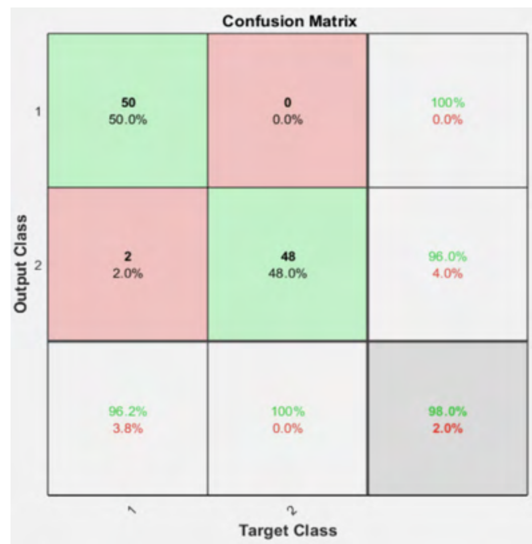

Fig. 7. Confusion matrix of MSP-UET fall detection dataset (1) Fall (2) Non-fall.

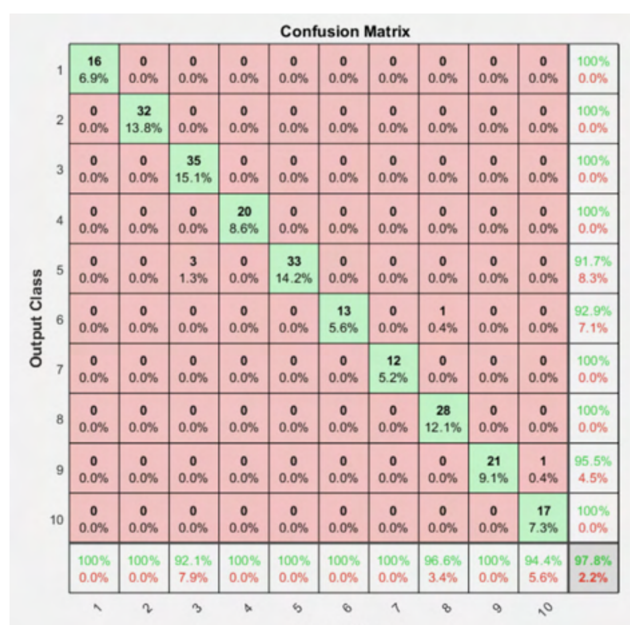

Fig. 8. Confusion matrix for 10 classes of The Daily Sounds dataset.

From this comparative analysis, we observed that our MAD-LTP features achieved 50.6% improved accuracy than our prior features acoustic-LTP on the BiLSTM classifier. The detailed results in terms of accuracy, precision, recall and F1-score are reported in Table 4. From the results of this comparative analysis on the daily sounds dataset, we concluded that our proposed MAD-LTP features are more robust for fall event detection over the comparative features on a diverse dataset.

In the second phase of this experiment, we checked the superiority of the proposed MAD-LTP features on A3 Fall v2.0 [47] against the comparative features reported in Table 4. For this purpose, we extracted the features to train BiLSTM network individually and results are shown in Table 4. From these results, we observed that fusion of MFCC and GTCC again, performed the worst by achieving an accuracy of 41%, precision, recall, and F1-score of 0%. Acoustic-LTP performed second-best by achieving the accuracy of 68%, precision of 100%, recall of 65.21%, and F1-score of 78.94% while our MAD-LTP features outperformed all features by achieving an accuracy of 98.29%, precision of 97.53%, recall of 100%, and F1-score of 98.75%. It is important to mention that our MAD-LTP features achieved an improved accuracy of 30.29% and 30.39% than our prior acoustic-LTP and GTCC features, respectively. Experimental results on A3 Fall v2.0 dataset illustrate the superiority of our MAD-LTP features for fall detection over the comparative features.

In the last phase of this experiment, we checked the effectiveness of the proposed MAD-LTP features against the comparative features on our in-house MSP-UET fall detection dataset [48]. Again, we employed all of these features to train the BiLSTM separately and provided the results in Table 4. From these results on our MSP-UET fall detection dataset, MFCC was the worst performer by achieving an accuracy of 72.7%, precision of 22.85%, recall of 100%, and F1-score of 37.20%. Our prior acoustic-LTP features performed second best by achieving an accuracy of 88.6%, precision of 100%, recall of 80.43%, and F1-score of 89.15% while our proposed MAD-LTP features again produced the best results with an accuracy of 98%, precision of 100%, recall of 96.15%, and F1-score of 98.03%. These results on our MSP-UET fall detection dataset proved that the proposed MAD-LTP features provide superior fall detection performance over the comparative features.

4.6. Comparison with contemporary methods

The key goal of this experiment is to make a comparison with existing approaches for fall detection techniques on the daily sounds [46] and A3 Fall v2.0 [47] datasets. We compared the performance of the proposed system with existing techniques [10,37,54] on A3 Fall v2.0 dataset and outcomes are reported in Table 5. From the Table 5, we observed that [10] achieved the worst results and yielded an accuracy of 85%, precision of 91.48%, recall of 91.23%, and F1-score of 91.04%. Our proposed method performed the best and achieved an accuracy of 98.29%, precision of 97.53%, recall of 100%, and F1-score of 98.75%. From this comparative analysis, we found that our system achieved an accuracy gain of 4.02%, 1.29% and 13.29% over the comparative methods [10,37,54] on the A3 Fall v2.0 dataset.

Next, we assessed the performance of the proposed system on the daily sounds dataset against the current techniques [42,49,50,53] and outcomes are stated in Table 5. From these results, we observed that [42] achieved the lowest results with an accuracy of 66%, precision of 100%, recall of 60%, and F1-score of 44%. Our prior work [49] was the second-best system and yielded an accuracy of 92%, precision of 91%, recall of 94%, and F1-score of 97% while our approach outperformed against all the comparative methods by achieving an accuracy of 93.5%, precision of 100%, recall of 91.59%, and F1-score of 95.16%. We observed that our method achieved 1.5% accuracy gain than the second-best performer [49]. This comparative analysis on two standard datasets demonstrates the significance of our approach over the contemporary techniques for accurate fall event detection on multiple and diverse datasets.

4.7. Discussion

This section aims to provide a more in-depth analysis of the experimental findings of the proposed fall detection method. We have three hypotheses that MAD-LTP features are capable to capture the dynamic repetition in the audio signal, robust to non-uniform noise in environmental sounds, and reliable for outdoor as well as indoor applications. Moreover, the findings of this study reveal that our prior work acoustic-LTP [49,51] has limitations such as non-robust to the dynamic audio signal and non-uniform noise present in environmental sounds that need to be addressed. In order to address the limitations, we proposed novel MAD-LTP features for audio representation to develop an effective system for the detection of environmental sounds. We conducted extensive experimentation on the three datasets such as the Daily Sounds, A3 Fall v2.0, and our own MSP-UET fall dataset. The daily sounds and A3 Fall v2.0 datasets contain the audios recorded in indoor environments, whereas, our MSP-UET fall detection dataset contains the audios of both the indoor and outdoor environments. The remarkable accuracy of more than 98% on The Daily Sounds and our MSP-UET datasets and above 93% on the A3 Fall v2.0 proves our first and third hypotheses by demonstrating the capability of our MAD-LTP features for better capturing the traits of dynamic repetitive patterns in the audios of both the indoor and outdoor environments for fall event detection.

Table 4
Performance comparison with existing features.

Dataset	Feature	Accuracy%	Precision%	Recall%	F1-score %
The Daily Sounds	MFCC	81.7	71.42	100	83.33
	GTCC	64.2	100	64.22	78.21
	MFCC-GTCC	35.8	0	0	0
	Acoustic-LTP	42.9	100	38.21	98.94
	MAD-LTP	**93.5**	**100**	**91.59**	**95.16**
A3 Fall v2.0	MFCC	48.7	13.04	100	23.07
	GTCC	67.9	100	64.78	78.63
	MFCC-GTCC	41	0	0	0
	Acoustic-LTP	68	100	65.21	78.94
	MAD-LTP	**98.29**	**97.53**	**100**	**98.75**
MSP-UET Fall Detection	MFCC	72.7	22.85	100	37.20
	GTCC	87.9	100	74.46	85.37
	MFCC-GTCC	73.7	25.71	100	40.90
	Acoustic-LTP	88.6	100	80.43	89.15
	MAD-LTP	**98**	**100**	**96.15**	**98.03**

Table 5
Performance comparison with contemporary systems.

Dataset	Authors	Method	Accuracy%	Precision%	Recall%	F1-score%
A3 Fall v2.0	Bin et al. [37]	MFCC-Spectogram + HEL	94.17	–	–	–
	Alex et al. [54]	MFCC-SF-SC + DNN	97	–	–	–
	Principi et al. [10]	MFCC + SVM	85	91.48	91.23	91.04
	Proposed	**MAD-LTP + BiLSTM**	**98.29**	**97.53**	**100**	**98.75**
The daily sounds	Adnan et al. [49]	Acoustic-LTP + SVM	92	91	94	97
	Tuncer et al. [50]	LBP-LTP + SVM	89.17	–	–	–
	Khan et al. [42]	MFCC + OCSVM	66	100	60	44
	Shaukat et al. [53]	LPC + DT	56	64	76	25
	Proposed	**MAD-LTP + BiLSTM**	**93.5**	**100**	**91.59**	**95.16**

There is a high correlation between the complex environmental sounds under the non-uniform noise in an outdoor environment that makes fall event detection a challenging task. As our MSP-UET fall detection dataset also contains the audios in outdoor environments besides the indoor that includes the non-uniform noise. Thus, by achieving the remarkable results of 98% accuracy, 100% precision, 96.15% recall, and 98.03% F1-score on MSP-UET dataset proves our second hypothesis by illustrating the robustness of our method for outdoor audios containing the non-uniform noise.

The two datasets such as A3 Fall v2.0 and MSP-UET Taxila Fall detection dataset have two classes i.e., fall and non-fall while the Daily Sounds dataset has multiple classes. To reveal the abilities of our method as a reliable environmental sounds detector, we have increased the evaluation scope by conducting experiments to check its performance in multi-class problems. In the multiclass scenario, the proposed method also performs well and obtained an average accuracy of 97.8% as shown in Fig. 8. We also conducted experiments on two classes such as fall and non-fall to check the superiority and generalizability of the proposed method. We achieved remarkable accuracy of 93.5%, 98.29%, and 98% on the Daily Sounds, A3 Fall v2.0, and our MSP-UET fall datasets. This experiment proves that the dynamic threshold-based MAD-LTP computed features are able to reliably detect the fall event in both the binary and multi-class classification scenarios.

5. Conclusion

In this work, we presented an effective and efficient fall detection method utilizing the MAD-LTP features and BiLSTM. We presented a novel audio features descriptor MAD-LTP to better capture the attributes of scream and pain voices. We also developed a diverse fall detection system to measure the performance of the proposed system under challenging indoor and outdoor environments. We evaluated the performance of the proposed system on three datasets to check the robustness of our method on multiple diverse datasets. Experimental results illustrate the reliability of the proposed system over the contemporary methods for fall detection. We conclude that the proposed system can be employed in various wearable machines to reliably monitor the patients in hospitals and elderly persons in houses to detect the fall incidents. In future, we intend to use our designed MAD-LTP features on cross dataset scenario to assess the generalizability of our approach under more challenging conditions such as multiple external interferences and significant amount of reverberation.

CRediT authorship contribution statement

Ameen Banjar: Formal analysis, Investigation, Resources, Methodology, Writing – original draft, Writing – review & editing, Funding acquisition. **Hussain Dawood:** Conceptualization, Formal analysis, Investigation, Resources, Methodology, Writing – original draft, Supervision, Writing – review & editing. **Ali Javed:** Conceptualization, Investigation, Project administration, Methodology, Software, Writing – original draft, Supervision, Writing – review & editing, Validation. **Farman Hassan:** Methodology, Software, Data curation, Writing – original draft, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-21-DR-109). The authors, therefore, acknowledge with thanks the University of Jeddah technical and financial support.

References

[1] Access on 9.12.2021, Available online at: https://www.nih.gov/news-events/news-releases/worlds-older-population-grows-dramatically.
[2] Bi Y, Lv M, Song C, Xu W, Guan N, Yi W. Autodietary: a wearable acoustic sensor system for food intake recognition in daily life. IEEE Sens J 2015;16(3):806–16.
[3] Doukas C, Maglogiannis I. Advanced patient or elder fall detection based on movement and sound data. In2008 Second International Conference on Pervasive Computing Technologies for Healthcare 2008 Jan 30 (pp. 103-107). IEEE. https://DOI: 10.1109/PCTHEALTH.2008.4571042 · Source: IEEE Xplore.
[4] Grassi M, Lombardi A, Rescio G, Malcovati P, Malfatti M, Gonzo L, Leone A, Diraco G, Distante C, Siciliano P, Libal V. A hardware-software framework for high-reliability people fall detection. InSENSORS, 2008 IEEE 2008 Oct 26 (pp. 1328-1331). IEEE.
[5] Bianchi F, Redmond SJ, Narayanan MR, Cerutti S, Lovell NH. Barometric pressure and triaxial accelerometry-based falls event detection. IEEE Trans Neural Syst Rehabil Eng 2010;18(6):619–27.
[6] Su BY, Ho KC, Rantz MJ, Skubic M. Doppler radar fall activity detection using the wavelet transform. IEEE Trans Biomed Eng 2014;62(3):865–75.
[7] Liu L, Popescu M, Skubic M, Rantz M, Yardibi T, Cuddihy P. Automatic fall detection based on Doppler radar motion signature. In2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops 2011 May 23 (pp. 222-225). IEEE. https://doi=10.1.1.457.2013.
[8] Wu M, Dai X, Zhang YD, Davidson B, Amin MG, Zhang J. Fall detection based on sequential modeling of radar signal time-frequency features. In2013 IEEE International Conference on Healthcare Informatics 2013 Sep 9 (pp. 169-174). IEEE.
[9] Albawendi S, Lotfi A, Powell H, Appiah K. Video based fall detection using features of motion, shape and histogram. InProceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference 2018 Jun 26 (pp. 529-536). https://dl.acm.org/doi/pdf/10.1145/3197768.3201539.
[10] Principi E, Droghini D, Squartini S, Olivetti P, Piazza F. Acoustic cues from the floor: a new approach for fall classification. Expert Syst Appl 2016;30(60):51–61.
[11] Yun Li, Ho KC, Popescu M. A microphone array system for automatic fall detection. IEEE Trans Biomed Eng 2012;59(5):1291–301.
[12] Access on 9.12.2021, Available online at: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing-2019-Report.pdf.

[13] Zigel Y, Litvak D, Gannot* I. A Method for automatic fall detection of elderly people using floor vibrations and sound—proof of concept on human mimicking doll falls. IEEE Trans Biomed Eng 2009;56(12):2858–67.
[14] Access on 9.12.2021, Available online at: https://www.ncoa.org/news/resources-for-reporters/get-the-facts/falls-prevention-facts/.
[15] Popescu M, Mahnot A. Acoustic fall detection using one-class classifiers. In2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2009 Sep 3 (pp. 3505-3508). IEEE. https://doi.org/10.1109/IEMBS.2009.5334521.
[16] Lai CF, Chang SY, Chao HC, Huang YM. Detection of cognitive injured body region using multiple triaxial accelerometers for elderly falling. IEEE Sens J 2010;11(3):763–70. https://doi.org/10.1109/JSEN.2010.2062501.
[17] Wang J, Zhang Z, Li B, Lee S, Sherratt RS. An enhanced fall detection system for elderly person monitoring using consumer home networks. IEEE Trans Consum Electron 2014;60(1):23–9.
[18] Casilari E, Lora-Rivera R, García-Lagos F. A study on the application of convolutional neural networks to fall detection evaluated with multiple public datasets. Sensors 2020;20(5):1466. https://doi.org/10.3390/s20051466.
[19] Bourke AK, Lyons GM. A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Med Eng Phys 2008;30(1):84–90. https://doi.org/10.1016/j.medengphy.2006.12.001.
[20] Li Q, Stankovic JA, Hanson MA, Barth AT, Lach J, Zhou G. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. In2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks 2009 Jun 3 (pp. 138-143). IEEE.
[21] Nyan MN, Tay FE, Murugasu E. A wearable system for pre-impact fall detection. J Biomech 2008;41(16):3475–81. https://doi.org/10.1016/j.jbiomech.2008.08.009.
[22] L. Martinez-Villaseñor H. Ponce Design and Analysis for Fall Detection System Simplification JoVE 158.
[23] Li H, Shrestha A, Fioranelli F, Le Kernec J, Heidari H, Pepa M, Cippitelli E, Gambi E, Spinsante S. Multisensor data fusion for human activities classification and fall detection. In2017 IEEE SENSORS 2017 Oct 29 (pp. 1-3). IEEE. https://doi.org/10.1109/ICSENS.2017.8234179.
[24] Pierleoni P, Belli A, Maurizi L, Palma L, Pernini L, Paniccia M, et al. A wearable fall detector for elderly people based on ahrs and barometric sensor. IEEE Sens J 2016;16(17):6733–44.
[25] He Y, Li Y. Physical activity recognition utilizing the built-in kinematic sensors of a smartphone. Int J Distrib Sens Netw 2013;9(4):. https://doi.org/10.1155/2013/481580481580.
[26] Núñez-Marcos A, Azkune G, Arganda-Carreras I. Vision-based fall detection with convolutional neural networks. Wireless Commun Mobile Comput 2017;6:2017. https://doi.org/10.1155/2017/9474806.
[27] Gunale K, Mukherji P. Indoor human fall detection system based on automatic vision using computer vision and machine learning algorithms. J Eng Sci Technol 2018;13(8):2587–605.
[28] Lotfi A, Albawendi S, Powell H, Appiah K, Langensiepen C. Supporting independent living for older adults; employing a visual based fall detection through analysing the motion and shape of the human body. IEEE Access 2018;6:70272–82.
[29] Oliver SA, Anuradha M, Justus JJ, Maheshwari N. Optimized low computational algorithm for elderly fall detection based on machine learning techniques. Biomedical Research (0970-938X). 2018 Nov 1;29(20).
[30] De Miguel K, Brunete A, Hernando M, Gambao E. Home camera-based fall detection system for the elderly. Sensors 2017;17(12):2864.
[31] Yang L, Ren Y, Hu H, Tian B. New fast fall detection method based on spatio-temporal context tracking of head by using depth images. Sensors 2015;15(9):23004–19. https://doi.org/10.3390/s17122864.
[32] Nizam Y, Mohd MN, Jamil MM. Human fall detection from depth images using position and velocity of subject. Proc Comput Sci 2017;1(105):131–7. https://doi.org/10.1016/j.procs.2017.01.191.
[33] Bian ZP, Hou J, Chau LP, Magnenat-Thalmann N. Fall detection based on body part tracking using a depth camera. IEEE J Biomed Health Inf 2014;19(2):430–9.
[34] Thuc HL, Van Tuan P. An effective video-based system for human fall detection. Int J Adv Res Comput Eng Technol 2014;3(8):2820–6.
[35] Collado-Villaverde A, R-Moreno MD, Barrero DF, Rodriguez D. Machine learning approach to detect falls on elderly people using sound. InInternational Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems 2017 Jun 27 (pp. 149-159). Springer, Cham.
[36] D. Droghini D. Ferretti E. Principi S. Squartini F. Piazza An end-to-end unsupervised approach employing convolutional neural network autoencoders for human fall detection. InItalian Workshop on Neural Nets, 14 2017 Jun Springer Cham 185 196.
[37] Li X, Li J, Lai J, Zheng Z, Jia W, Liu B. A Heterogeneous Ensemble Learning-Based Acoustic Fall Detection Method for Elderly People in Indoor Environment. InInternational Conference on Human-Computer Interaction 2020 Jul 19 (pp. 369-383). Springer, Cham. https://doi.org/10.1007/978-3-030-50334-5_25.
[38] Droghini D, Ferretti D, Principi E, Squartini S, Piazza F. A combined one-class SVM and template-matching approach for user-aided human fall detection by means of floor acoustic features. Comput Intell Neurosci 2017;2017:1–13.
[39] Droghini D, Vesperini F, Principi E, Squartini S, Piazza F. Few-shot siamese neural networks employing audio features for human-fall detection. InProceedings of the International Conference on Pattern Recognition and Artificial Intelligence 2018 Aug 15 (pp. 63-69). https://doi.org/10.1145/3243250.3243268.
[40] Xiaoyang Tan, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Trans on Image Process 2010;19(6):1635–50.
[41] Bingham ELLA, Hyvärinen AAPO. A fast fixed-point algorithm for independent component analysis of complex valued signals. Int J Neur Syst 2000;10(01):1–8.
[42] Khan MS, Yu M, Feng P, Wang L, Chambers J. An unsupervised acoustic fall detection system using source separation for sound interference suppression. Signal Process 2015 May;1(110):199–210. https://doi.org/10.1016/j.sigpro.2014.08.021.
[43] Fortino G, Di Fatta G, Pathan M, Vasilakos AV. Cloud-assisted body area networks: state-of-the-art and future challenges. Wireless Netw 2014;20(7):1925–38. https://doi.org/10.1007/s11276-014-0714-1.
[44] Gołosz M, Mrozek D. Exploration of Data from Smart Bands in the Cloud and on the Edge–The Impact on the Data Storage Space. InInternational Conference on Computational Science 2019 Jun 12 (pp. 607-620). Springer, Cham. https://doi.org/ 10.1007/978-3-030-22744-9_47.
[45] Wu T, Gu Y, Chen Y, Wang J, Zhang S. A mobile cloud collaboration fall detection system based on ensemble learning. InThe 22nd International ACM SIGACCESS Conference on Computers and Accessibility 2020 Oct 26 (pp. 1-7). https://doi.org/10.1145/3373625.3417010.
[46] Sehili MA, Istrate D, Dorizzi B, Boudy J. Daily sound recognition using a combination of GMM and SVM for home automation. In2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO) 2012 Aug 27 (pp. 1673-1677). IEEE.
[47] Access on 9.12.2021, Available online at: https://research.google.com/audioset/.
[48] MSP-UET fall detection dataset. Access on 9.12.2021, Available online at: https://github.com/farmanhassan555/MSP-UET-Fall-Detection.git.
[49] Adnan SM, Irtaza A, Aziz S, Ullah MO, Javed A, Mahmood MT. Fall detection through acoustic local ternary patterns. Appl Acoust 2018 Nov;1(140):296–300. https://doi.org/10.1016/j.apacoust.2018.06.013.
[50] Tuncer T, Dogan S. Novel dynamic center based binary and ternary pattern network using M4 pooling for real world voice recognition. Appl Acoust 2019;15(156):176–85. https://doi.org/10.1016/j.apacoust.2019.06.029.
[51] Irtaza A, Adnan SM, Aziz S, Javed A, Ullah MO, Mahmood MT. A framework for fall detection of elderly people by analyzing environmental sounds through acoustic local ternary patterns. In2017 ieee international conference on systems, man, and cybernetics (smc) 2017 Oct 5 (pp. 1558-1563). IEEE.
[52] Cheffena M. Fall detection using smartphone audio features. IEEE J Biomed Health Inform 2016;20(4):1073–80.
[53] Shaukat A, Ahsan M, Hassan A, Riaz F. Daily sound recognition for elderly people using ensemble methods. In2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) 2014 Aug 19 (pp. 418-423). IEEE. https://doi.org/ 10.1109/FSKD.2014.6980871.
[54] Alex JS, Kumar MA, Swathy DV. Deep Learning Approaches for Fall Detection Using Acoustic Information. In Advances in Smart Grid Technology 2021 (pp. 479-488). Springer, Singapore. https://doi.org/ 10.1007/978-981-15-7241-8_35.