Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 1, 2026

SOUND RECOGNITION FOR THE DETECTION OF DISTRESS SITUATIONS USING ACOUSTIC PATTERNS

Mahrukh Mansoor[*1], **Kahkashan Zeb**[2]

[*1,2]*Department of Computer Science University of Engineering and Technology (UET), Taxila*

[*1]mahrukh.mansoor@students.uettaxila.edu.pk, [2]kehkshanzb@gmail.com

DOI: https://doi.org/10.5281/zenodo.18374360

Keywords

Acoustic-LTP, Distress situation, Support vector machine (SVM), MFCC, LPC

Article History

Received: 01 December 2025
Accepted: 11 January 2026
Published: 26 January 202

Copyright @Author

Corresponding Author: *
Mahrukh Mansoor

Abstract

Recognizing distress situations is an essential area of research in health monitoring for elderly or frail individuals. When they are alone at home, the likelihood of a harmful fall is elevated. Throughout an accidental fall, the person remains on the floor without receiving any instant aid therefore well-timed reporting to the concerned caregivers is critical. The research community planned several fall detection methods, but certain limitations are still associated with these methods, i.e., computational complexity and high false alarm rate. To resolve these problems, we proposed a study on daily life's sound recognition to detect fall events at home. In this work, we used a benchmark RWCP dataset with another class of human screaming sounds (male and female) covering 100 sound instances. Primarily it is a novel approach based on 1-D Local Ternary Patterns (Acoustic-LTP) as feature extractors, along with conventional audio features such as MFCCs and LPCs. Three different types of features are combined to ensure efficient and robust extraction of features for characterizing the acoustics properties of the sound signals. The experimental findings indicate the efficacy of the proposed approach, as classification via SVM attained an accuracy of 98.1%, exceeding results reported in prior studies.

I. INTRODUCTION

The day-to-day sounds bred by the actions of humans and objects in the surroundings are a significant area of audio signal processing [1]. Daily sounds produced during distress situations such as (crying, screaming, fall event occurrence, or sounds from environmental objects) are a medium to capture information for elderly or frail people and their surrounding environment [2]. Currently, Distress situation detection for elderly or frail people in their habitat is an active research area [3] but still, it has not been able to attain a standard method or adequate performance. Besides, the accuracy and reliability of home monitoring systems are essential as a rapid response of the caretakers in any health-threatening situation can reduce the severity of injuries thus preventing prolonged illness[4]. Numerous studies have indicated that falls among elderly or frail individuals are among the most hazardous incidents that can occur at home[5]. In 2000, the immediate healthcare costs of falls among adults in the U.S. exceeded $19 billion, with around 28–35% of older adults experiencing one or more falls each year [6]. Henceforth, there is a sheer need for a continuous and reliable health care monitoring system in the homes of elderly or frail people to detect any distress situations. Various authors have explored assistive health monitoring technologies for elderly people immensely. The proposed methodologies are classified into two main categories: environment sensors and wearable sensors [7]. Environment sensors consist of passive infrared sensors (PIR), floor vibration sensors, video sensors, Doppler sensors, and microphones, whereas wearable sensors are mostly based on accelerometers and gyroscopes. The drawback of serviceable sensors is that aged users do not feel comfortable wearing them continuously. Besides, great power intake, restricted communication range, and data loss [8], [9] are some of the problems that resulted in the emergence of contactless fall recognition systems through environmental sensors. In [10], falls are detected using an arrangement of PIR and ground vibration sensors. PIR sensors face challenges such as line-of-sight requirements and limited coverage areas. Similarly, the floor vibration sensors performance is affected by the category of flooring, and their detection range is restricted. In [11], a wavelet transform-centered method was applied to detect human falls. The flaw of radar-based Doppler system is its inadequate applicability. Whereas video sensors are linked with several drawbacks including occlusions, invasion of user's privacy, and effect of variation in illumination.

Among the array of environmental sensor methods, acoustic analysis of daily sounds emerges as an effective substitute for tackling the limitations linked with both non-wearable and wearable solutions. Developing monitoring systems based on daily sounds not only reduces equipment costs but also yields pertinent information. Audio-based technologies prove non-intrusive and less susceptible to occlusion, offering broader coverage [12]. The daily sound recognition system, which utilizes acoustic sensors, generally follows a three-stage process. Firstly, sound acquisition takes place, followed by the assessment of raw audio data, rich in information about the daily lives of elderly individuals and their surroundings. Specifically, during distress situations, various sounds like glass breaking or screaming may be produced, allowing for classification and the detection of emergencies. In acoustic signal processing, feature extraction and classification are pivotal, with robust feature extraction significantly influencing performance [13],[14] Various methods enhance robustness in noisy environments, such as Local Binary Pattern (LBP) [15], Coding local spectrogram features (LSFs) temporally [16], and representation of sub-band power distribution (SPD) in image form [16]. Learnable Gamma-tone filter-bank layers [17] and Gabor filter-bank utilization for spectro-temporal modulation frequency features [18],[19] contribute to time-frequency representation. To tackle the challenges, more effective and efficient sound recognition techniques are necessary. This paper presents an acoustic-based distress situation detection framework that employs an innovative feature representation method by integrating three techniques: acoustic-LTP, MFCC, and LPC. This integration aims to capture the characteristics of sounds linked with falls in a better way, such as screams, in both indoor and outdoor environments with background noise. The extracted features are used to train an SVM for classifying distress situations. We evaluated the performance of our framework on two distinct datasets, i.e., RWCP-DB (Real-World Computing Partnership Sound Scene Database) and an additional class, encompassing male and female screams incorporated from the Daily Sound dataset. The key contributions of the suggested study are:

- We create a consistent and applicable structure for precisely detecting daily activities and distress situations by monitoring everyday sounds.
- We developed an innovative audio feature descriptor by combining three techniques. i.e., A-LTP, MFCC, and LPC capable of capturing the intricate acoustic structure inherent in daily sounds.
- Thorough experimentation was conducted on two distinct datasets to evaluate the effectiveness of the planned method compared to current distress situation detection methods.

The structure of this paper is as follows: Segment 2 reviews the literature, Segment 3 provides a detailed presentation of the proposed framework, Segment 4 outlines the experimental setup, Segment 5 presents the experimental results, and Segment 6 concludes the paper while suggesting directions for future research.

II. LITERATURE REVIEW

The examine society has investigated a range of fall detection and environmental sound classification methods. In this work, two main categories, vision-based and acoustic-based methods, are discussed in detail in the following sections.

2.1 Vison-Based Fall Detection Systems

In vision-based systems, one or more cameras are used to analyze images, allowing for the simultaneous identification of multiple events through posture detection. These systems can capture features such as shape changes during activities [20], body posture, and the 3D trajectory of head joints [21]. Functioning in genuine environment, they operate three dimensional cameras or devices to monitor individual happenings. Depth cameras, such as Kinect, incorporate infrared LEDs, ensuring operation in low-light conditions without the need for external lighting.

Pui Chau et al. [22] suggested a technique that centers at tracing body linkages by means of Kinect's infrared sensor to identify falls. An SVM classifier is applied, via the three-dimensional flight of the head joint as indication, to differentiate between fall and non-drop incidents. Nevertheless, the dependability of joint extraction raised doubts, and testing situations resulted in inaccurate outputs from the SVM. Watanapa et al. [23] devised a projected system that estimates the severity of falls while detecting real-time fall motion. The Euclidean distance between consecutive frames of Kinect video, mapped from normalized body joints, was used. Although the focus was measuring fall severity, future research addresses occlusions by incorporating multiple Kinect cameras.

Makris et al. [24] anticipated for a descent finding method based on the Kinect sensor, using a 3D bounding box process to determine the initial derivative (velocity) of height, width, and depth to decide whether a specific action is a fall. Shortcomings included lack of an object differentiation approach, and Kinect-based skeletal tracking features were not applied. Zhong et al. [25] projected a method to sense falls using Kinect and a LEGO Mind storms robot. The Kinect SDK was used for skeletal chasing to identify individuals and their actions. However, issues like signs of tracking and the failure of speech recognition in an unconscious state, coupled with the additional cost of the LEGO Mind storms robot, posed weaknesses.

Tian et al. [26] announced a spatio-temporal tracking method to explore 3D depth data captured with Kinect. The single Gaussian model (SGM) was applied to extract the silhouette, while the head position was determined based on the foreground coefficient of ellipses. Limitations included the fixed position of the Kinect and dependence on only four angles. Kalinga et al. [27] worked on instances of falls employing a preservation robot equipped with a Kinect sensor and an automated fall notification system based on a Q-learning algorithm. Body joint velocities were assessed to differentiate falls from daily activities. However, the use of an intelligent robot requires expertise and additional expenses for installation in a home environment. Denyer et al. [28] suggested a procedure to detect falls by analyzing height, velocity, and position of individuals captured by a depth camera in the Kinect sensor. Height was measured as the distance from the head to the ground, while activities were recognized by tracking variations in height during periods of abnormal movement speed.

2.2 Acoustic-based Fall detection and Environmental sound classification

2.2.1. Deep Learning-based Environmental Sound Classification and Fall Detection

Deep learning-based algorithms can perform high-level abstractions on data. The variability of audio signals can be traced. A deep learning approach presented by Kim et al. [29] conducted experiments using audio recordings including emergency sound events (screams, explosions) collected from signal person households. The Log-Scaled-Mel spectrograph is extracted as entered items later, grouping is carried out using CNN and LSTM. However, the deep learning model used in this study needs excessive computational training time and large training data. Similarly, Koerich et al.[30] classified environmental sounds by means of a 1D Convolutional neural network (CNN) which has three to five layers, according to the size of the input acoustic signal. Also, 1D-CNN learns the filters directly from the audio signal as an alternative to static filter banks used in the extraction of MFCC features. Yet, the filters learned in 1D-CNN midway convolutional layers failed to display dominant frequencies and were fairly noisy. Zhang et al. proposed [31] a convolutional recurrent neural network (ACRNN) based on the attention mechanism. Only semantically relevant frames and discriminative features are produced for environmental sound classification. The results attained for this model demonstrated that the attention for one of the sound classes focused on very few frames and robustness to sound is not quantified.

The intricate characteristics of environmental sounds are obscured by natural acoustic noise, making the classification of specific sounds even more challenging. To enhance the classification of environmental sounds Deigo et al.[32] applied a Siamese Neural network to the human fall detection system. The features are extracted using Log-Mel energy for audio analysis and the classification results are compared to machine learning approaches including SVM and OSVM. Though the Siamese network has a zero-miss rate still false alarm rate was higher than SVM. Another system also proposed by Droghini et al. [33] detected human falls by again using the Siamese auto encoders neural network. Firstly, Log-Mel feature extraction is carried out for audio signals and then one-shot learning is included. The proper selection of training pairs allowed mapping of simulated falls into real falls. The system is quite reliable, but it was annoying for some users, as three false alarms were raised for every two actual human falls.

The main problem of deep learning approaches is to provide an appropriate amount of data to train the model. Recently, generative adversarial networks (GAN) have allowed the creation of high-quality data samples. Koerich et al. [34]utilized weighted cycle-consistent generative adversarial networks (WCGANs) and k-Means++ for environmental sound recognition. The audial wave is converted into 2D space by discrete wavelet transform (DWT) later augmented by WCGANs. However, this system fails to work properly for 1D-to-1D audio waveforms as due to their high dimensionality it is very difficult to perform augmentation. Similarly, Birmingham et al.[35] combined recurrent neural networks (RNN) with CNN for environmental sound classification. The elimination of pitfalls associated with GANS, and high-quality data augmentation is accomplished by using a deep convolutional generative adversarial network (DCGAN). The results demonstrated that DCGAN can generate spectrograms, and the accuracy attained on the augmented data is higher.

2.2.2 Traditional Feature Extraction Techniques for Environmental Sound Classification and Fall Detection

Audio signals represented by handcrafted features such as MFCC help reduce both dimensionality and noise. An acoustic sensor equipped with a microphone depicts user movements, and MFCC features are mined to classify environmental sounds. Popescu et al. [36] employed MFCC features alongside the KNN algorithm to categorize actions as falls or non-falls. The strength of the audio signal was monitored expending a vertically aligned linear array of microphones. The false alarm rate was reduced by adding the sound source height. Yet, the limitations associated with this method were: the simulation of a realistic fall sound by stunt actors or fall dummies and acoustic properties' impact on height estimation accuracy.

The problem of generating a realistic sound was also solved by Popescu et al.[37] through the implementation of a one-class classifier. In this method, MFCC aspects from non-fall noises were used to train one-class SVM to classify happening as a fall. However, the mistaken alarm rate was not reduced and spatial information correlated with the sound source was ignored. Li et al. [38] addressed the issue of environmental effects on height estimation by using a circular array of 8 microphones. It provides a better 3-D evaluation of sound location in comparison to FADE versions [37] [38] but is confined to the usage of a limited experimental data set. In the system later proposed by Li et al. [39] the fixing complication of the circular array was avoided by using four Kinect microphone sensors. The system based on MFCC along with KNN handled the occlusion of a fall but attained less accuracy for strong interference. Therefore, Li et al. [40] introduced the bean forming (BF) technique to improve the required signal and shrink the strong intervention by environmental sound. The occlusion of a fall and practical world intrusion from unidentified foundations were still non tackled.

2.2.3 Hybrid Systems for Audio-based Fall Detection

The integration of multiple feature extraction techniques and hybrid systems was used by various authors to enrich the routine of audio-based fall detection systems. Hassan et al. [41] extracted 29-dim MFCC, spectral skewness, and GTCC features from complex environmental sound samples of non-fall and fall events. Various machine learning classifiers including Naïve Bayes, KNN, and decision trees were used to distinguish non-fall and fall incidents. Primarily, the performance of the system was enhanced by the combination of features extraction techniques with MFCC thus adding the limitation of increased hardware implementation cost. Furthermore, Wang et al. [42] presented a Transformer-based deep learning method to recognize a fall from audio input generated inside bathroom environments.

A broader class of audio is processed by using Log Mel spectrograms. Likewise, to increase the investigation area, the solution is implemented by using three far-field microphones located on a mobile robot. However, the drawback of using a mobile robot in healthcare applications is the higher cost. Additionally, Cheffena et al. [43] designed a hybrid approach based on radio frequency (RF) and audio signs to identify individual activity. The Mel-spectrogram feature of the audio data and the Doppler shift feature of RF measurements are extracted and given as input to six different classifiers. The results demonstrate that by using this hybrid approach the number of recognizable activities is increased. Yet, further analysis is anticipated for the elimination of background noise to improve the performance of procedure.

2.2.4 Floor-based Acoustic Fall Detection System

Floor sensor-based fall detection techniques apply the basic idea that a human fall always generates certain vibration patterns on the floor that are unlike the vibration produced by everyday activities and objects dropping on the surface. A system centered on ground vibration from the microphone and accelerometer was presented by Zigel et al. [44]. Special features like MFCC and shock response spectrum were used for classification purposes. The trials are performed on a concrete floor, and this may cause drastic effects on a signal when installed on other types of floor cover material, the system is also less sensitive towards low-impact falls. The challenges of noise foundation localization and surrounding noise interference were directed using a floor acoustic sensor (FAS).

Principi et al. [45] presented a fall detection system established on MFCC, super vectors, and SVM. Results attained indicated that the floor sensor performs better than the aerial microphone but this method requires optimization of FAS to make it more suitable for diverse floors. Another method proposed by Droghini et al. [46] also utilized floor acoustic features. Acoustic signals are first captured by FAS; then MFCC and Gaussian Mean Super vectors are extracted for identifying a fall or non-fall event. The performance of the algorithm was assessed via falls replicated by a human-mimicking doll consequently adding the restraint of simulation of a realistic fall sound. A different system centered on FAS was also presented by Droghini et al. [47] This solution was more suitable for realistic scenarios as the classifier is specifically designed to discriminate human falls from other events. However, the data set still requires expansion to detect a fall in varying scenarios.

2.2.5 Local Binary Patterns for Environmental Sound Classification

The Local Binary Pattern (LBP) can also be applied to one-dimensional audio signals. The combination of LBP with audio descriptors can attain improved accuracy for environmental sound classification. Toffa et al. [1] applied LBP directly on 1D audio signals and a strong characterization is incorporated by feature collaboration. This method signifies a better option when there is data inadequacy or negligible computing power. Though, it can be further upgraded by using a multichannel descriptor-based CNN covering a mixture of LBP and audio features. Another system based on LBP was presented by Kobayashi et al. [48] the discriminative power of LBP was improved by L_2-Hellinger normalization. The method was robust to noise and achieved good results for audio classification, but it is only applicable to 2D spectrograms.

While acoustic-based fall detection methods are more computationally proficient compared to vision-based or hybrid systems, they have certain limitations. These include reduced robustness to noise from background, dependence on microphone performance, hardware constraints for inserting, and the challenge of accurately distinguishing between screams and standard speech. To overcome these challenges, there is a demanding necessity to develop more strong and reliable acoustic-based fall detection systems.

III. Proposed Framework

Our proposed method aims to precisely differentiate amongst fall and non-fall occurrences by analyzing sounds recorded during daily activities, particularly in distress situations. This study introduces a novel approach to feature extraction by integrating three methods: acoustic-LTP, MFCC, and LPC. As a result, we generate a unique set of feature vectors, each containing 46 elements. We utilize this feature vector set to classify acoustic events and detect distress situations. The architecture of our technique is outlined in Figure 1.

i). Data acquisition from input audio signals
ii). Feature extraction based on MFCC, LPC, and A-LTP
iii). SVM-based Classification
iv). Accurate recognition of the daily sounds

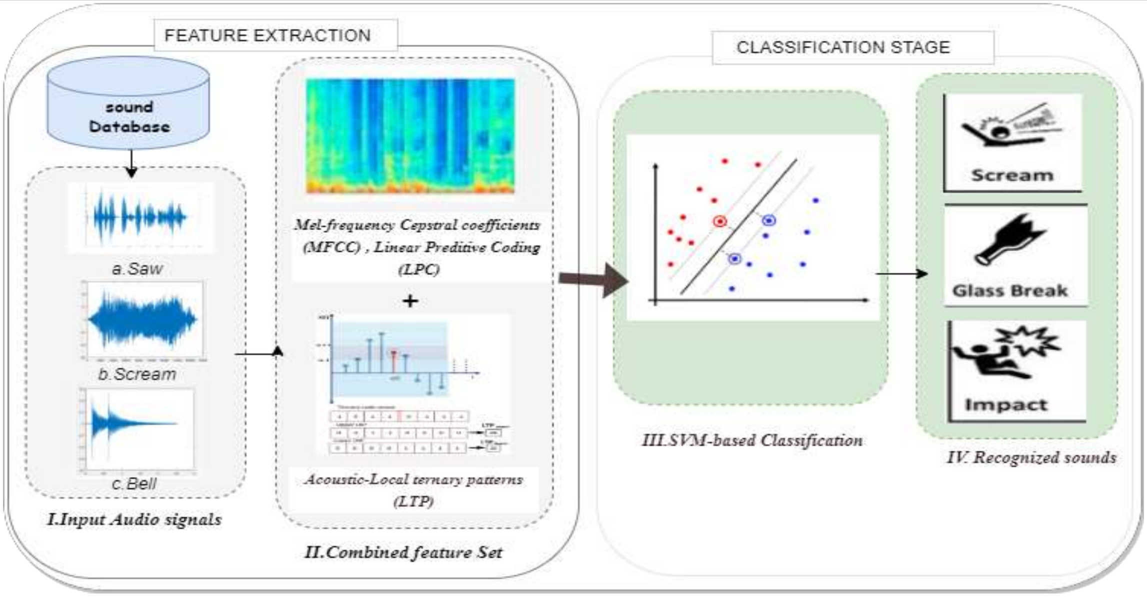

Fig 1. Proposed Framework Architecture

A. *Input Audio Data*

RWCP-DB (Real-World Computing Partnership Sound Scene Database) stands as a standardized dataset, composed of environmental sounds captured via a DAT recorder at 48 kHz along with a microphone[49]. This dataset encompasses 9722 instances, featuring 105 unique non-speech sounds free from room acoustics. The RWCP dataset is structured into three primary groups and 14 subgroups, spanning diverse sound categories. Each sound category, originating from 90 distinct sources, comprises approximately 100 sound recordings. Additionally, the daily sounds dataset contains all non-speech audio files in WAV format having a sampling frequency of 16 kHz. A total of 1,049 audio files were collected, either sourced from the internet or documented via a microphone and classified into 18 distinct sound categories.

Our study aims to classify falls by detecting distress situations using sounds from daily activities in the vicinity of elderly or frail individuals within a home setting. To detect fall events, we leverage the 16 daily activity-related sound classes from the RWCP dataset, each containing 100 sounds. An additional class, encompassing 100 samples of male and female screams in WAV arrangement with a sampling rate of 16 kHz is incorporated from the Daily Sound dataset to enhance performance. Table 1 provides examples of sound sources, several samples for each category, and their respective groups.

Table I. RWCP Database-Non-Speech Dry Source Sounds [2]

Group	Category	Sound source examples	No. of samples
Collision	Wood Metal Plastic Ceramic	Wood Board, Wood Stick Metal Board Metal Can Plastic Case Glasses, China	1187 1000 550 800
Action	Article Dropping Gas Jetting Rubbing Bursting/Breaking Clapping	Dropping Articles In Box Spray, Pump Sawing, Sanding Breaking Stick, Air Cap Hand Clap, Slamming	200 200 500 200 829

		Clap	
Category	Small Metal Articles Paper Instruments Electronic Sound Mechanical	Small Bell, Coin Dropping Book, Tearing Paper Drum, Whistle, Bugle Phone, Toy Spring, Stapler	1027 400 1079 705 1000

B. Working of Local Acoustic Ternary patterns

The proposed technique employs A-LTP to derive acoustic characteristics from the signal y[n]. Each frame Ωc of sound signal y[n] undergoes encoding to compute A-LTP locally. The ternary pattern is calculated by evaluating the magnitude difference between the central sample c and neighboring samples vi. A threshold value (t = 0.00008) is employed, and signal values within the width choice ± t around the central sample c are set to zero. Sample values below c-t are quantized to -1, while values above c + t are quantized to 1. Consequently, a function (x) reliant on three values is represented as:

$$x\left(v_i, c, t\right)=\begin{cases}+1, & v_i-\left(c+t\right)\geq 0\\ 0, & \left(c+t\right)<v_i<\left(c-t\right)\\ -1, & v_{i}\,_{-}\left(c-t\right)\leq 0\end{cases}$$

The representation of the acoustic signal via a ternary pattern, relying on three values, is designated as x (vi, c, t). To reduce the quantity of patterns, this representation is subdivided into upper xu(.) and lower xl(.) patterns. The upper arrangement is constructed by preserving only the values of +1, while all other values are assigned to zero.

$$x_u\left(v_i, c, t\right)=\begin{cases}1, & x\left(v_{i,}c, t\right) = +1\\ 0, & otherwise\end{cases}$$

Similarly, the lower pattern is created by preserving only the values of -1, with all other values being designated as zeros. The process of computing A-LTP is illustrated in Figure 2.

$$x_l\left(v_i, c, t\right)=\begin{cases}1, & x\left(v_{i,}c, t\right) = -1\\ 0, & otherwise\end{cases}$$

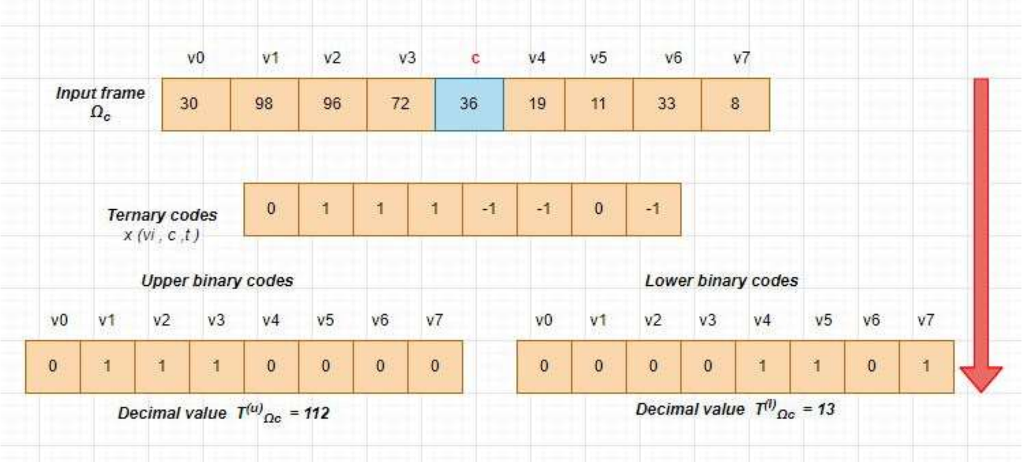

Fig.2. Computation of Acoustic Local Ternary Pattern

The decimal values of upper uniform patterns x_u^{uni}(.) and lower uniform x_l^{uni}(.) are used to perform encoding and computation. $T_{\Omega c}^{(l)}=\sum_{i-0}^{7} x_l^{uni}(vi,c,t).2^i$

The computation of histogram codes for both lower and upper segments contribute to generating the feature descriptor. The bins in the histogram, represented as K, are associated with even A-LTP codes. Each uniform pattern is assigned to a distinct bin, while all non-uniform patterns are grouped into a solo bin.

$$h_u(k) = \sum_{f=1}^{F} \delta\left(T_f^u, b\right)$$

$$h_l(k) = \sum_{f=1}^{F} \delta\left(T_f^l, b\right)$$

The variability within the data is encapsulated by the initial twenty-even patterns from both the upper and lower patterns. Consequently, the magnitude of the feature vector is double extent of each histogram. Concatenation of these two histograms gives rise to the formation of the 40-dimensional feature vector, denoted as z.

$$z = [h_u, h_l]$$

C. SVM-based Classification

Our system utilizes SVM to classify daily sounds. SVM operates based on two key conventions. Firstly, it tackles classification challenges by employing simple linear discriminant functions that transform data into higher-dimensional spaces. Secondly, SVM prioritizes training patterns near the decision boundary, deeming them to offer the most pertinent information for exact classification. The feature extraction module outputs a feature set with dimensions 1700*40, consisting of 1700 rows and 40 columns. This feature set serves as training data for SVM. However, before training SVM through this feature set, it is crucial to configure various parameters. Various approaches can be used to optimize SVM parameters for optimal performance. These methodologies include 10-fold cross-validation, leave-one-out cross-validation, or separating the training data into separate sets for testing and training. In this paper, we have opted for a 70%-30% split between training and testing.

For binary classification, we employ the One Versus One (OVO) classifier strategy. This approach involves constructing N(N-1)/2 classifiers, each dedicated to distinguishing between a specific duo of classes, i and j. The OVO classifiers were built to classify j as negative and i as positive. It is noteworthy that fji = -fij, ensuring consistent outcomes regardless of the class order. During classification, these OVO classifiers predict class labels for new instances, establishing class boundaries based on their learning. The final classification decision is typically made through majority voting or similar aggregation methods.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Feature Extraction

In this segment, the outcomes achieved on the RWCP dataset concerning the identification of distress situations are presented. The feature extraction process hinges on traditional acoustic feature extraction techniques, notably MFCC and LPC. This process yields the extraction of 13 MFCC and 13 LPC characteristics. Additionally, 40 features from Acoustic-LTP are incorporated for each sound sample, resulting in an overall of 66 structures developed for each sample.

The result of the feature extraction phase yields a feature set with dimensions of 1700 rows and 66 columns. In the classification module, this feature set serves as the training data for SVM. SVM training is conducted using a polynomial kernel on the extracted feature set. The dataset is divided into 70:30 ratios, allocating 70% for the educating set and the outstanding 30% for the testing set. This partitioning serves to validate the efficacy of the feature set derived from the amalgamation of Acoustic-LTP with MFCC and LPC.

In Figure 3. The confusion matrix over Medium Gaussian SVM of A-LTP, LPC, and MFCC features on the chosen dataset is represented. The recognition rate for phones, pumps, saw staplers, tears, whistles, spray, and wood is 100 %. Likewise, a high recognition rate of 97% is achieved for the remaining classes thus proving the reliability of the proposed feature extraction technique.

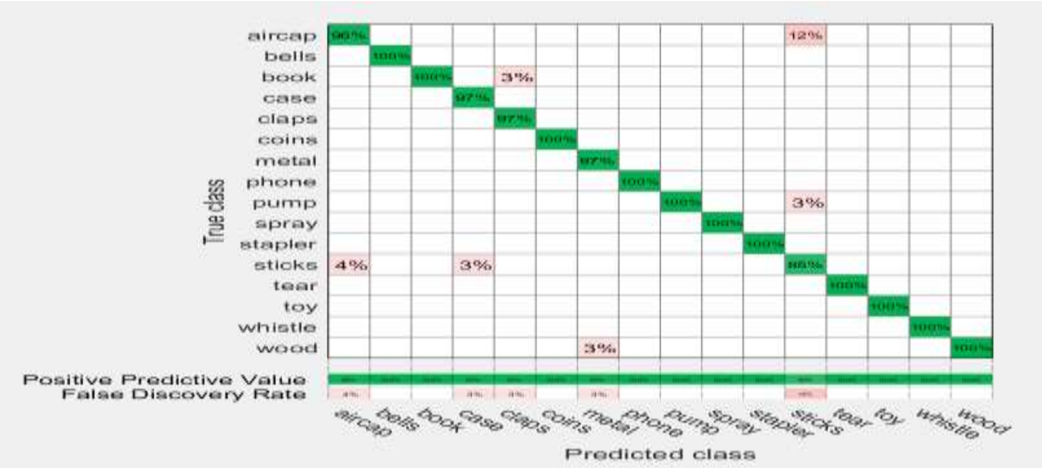

Fig.3. Confusion Matrix of Cubic SVM over RWCP dataset

B. Classification

In this investigation, the categorization is executed utilizing SVM. SVM boasts the capacity to classify through uncomplicated linear discriminative functions by converting data into a higher-dimensional realm. Additionally, SVM concentrates on training patterns proximate to the decision boundary, deeming them as the utmost pertinent information for categorization. In this undertaking, the polynomial kernel is employed to exploit the benefits of kernelized SVM. Before training SVM on the feature set several parameters were set and the feature set attained through a combination of acoustic-LTP, MFCC, and LPC is used as training data for SVM. The training of SVM is carried out by dividing the data in a ratio of 70:30: the training set is 70 percent while the remaining 30 percent is used as a testing set. The accuracy achieved by using different types of SVM is shown below in Table.

Table II. Recognition Accuracy using SVM

Kernel Function	Acc %
Medium Gaussian SVM	98.1
Linear SVM	97.1
Quadratic SVM	97.8
Cubic SVM	97.8
Fine Gaussian SVM	98.2
Coarse Gaussian SVM	93.9

C. *Performance Comparison with Existing Methods*

We assess the performance of our proposed approach by contrasting it with previously recommended methodologies outlined in references [2, 50, 51, and 52]. Shaukat et al. [2] devised a strategy for detecting daily sounds employing various ensemble techniques, which yielded superior sound recognition rates compared to individual classifiers. They reported an overall accuracy of 97.25% for their method. Similarly, Ye et al. [50] employed kernel discriminant analysis for daily sound recognition, extracting features via the Gabor transform. They accomplished a precision of 96.67% with 10-fold cross-validation on 105 sound categories sourced from the RWCP database. Chatlani et al. [51] introduced a novel 1-D LBP operator as a signal processing tool, achieving an overall accuracy of 97.4%. Finally, Sehili et al. [52]conducted distress situation detection based on daily sounds, using a combination of GMM and SVM on 18 sound classes with 16 MFCC features, resulting in an accuracy of 75%.

Results, presented in Table 3 and Figure 4, demonstrate that our recommended approach achieves the highest correctness associated to the techniques. Hence, we conclude that our method is more efficient in distress situation detection. Enhanced accuracy can be attributed to the generation of a comprehensive feature set based on A-LTP in conjunction with MFCCs and LPCs, amplifying the representation of acoustic event signals

Table III. Performance comparison with existing methods

Author	Proposed Method	Accuracy%
Shaukat et al.[2]	Non-spectral, MFCC, LPC and Ensemble method	97.2
Ye et al.[50]	Gabor spectrogram, Kernel-Fisher and discriminant analysis	96.6
Chatlani et al. [51]	1-D Acoustic LBP	97.4
Sehili et al.[52]	Gaussian mixture Model (GMM) and SVM	75.0
Proposed Work	**Acoustic-LTP, MFCC, LPC and SVM**	**98.1**

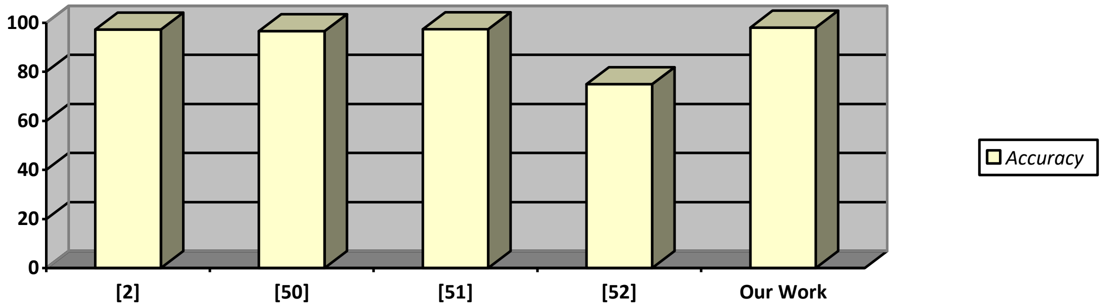

Figure.4. Comparison of results with other research methods

V. Conclusion

Daily sounds produced during distress situations such as (crying, screaming, fall event occurrence, or sounds from environmental objects) are a medium to capture information for elderly or frail people and their surrounding environment. In this work RWCP dataset of daily acoustics and an added class of (male, and female) screaming are used to detect a distress situation for elderly people. A novel approach of feature extraction based on A-LTP for daily sound recognition is presented. A-LTP objects capture the scattering of audio structure and in combination with LPC, MFCC enhances the retrieval of complex acoustic structures. Hence, this combination results in the extraction of a unique feature set. An accuracy of 98.1% is reached through SVM-based classification, which is higher than the existing results found in previous literature work.

VI. Future Work

While we have successfully identified daily sound activities for distress situation detection, continuous improvement remains a priority. The RWCP dataset, along with an extended scream sound class, has been utilized; however, there is potential for enhancement by incorporating additional classes related to distress situations (e.g., glass breaking, falls) to augment system reliability and dimensionality. Further advancements can be pursued in the feature extraction stage by exploring alternative variants of LTP. Specifically, the integration of local extrema patterns (LEP) can contribute to leveraging the complexities of acoustic structures, enhancing the overall effectiveness of the system. Feature selection is identified as another crucial aspect for improvement. The curse of dimensionality poses a challenge for classification algorithms, including SVM. To tackle this, our strategy involves delving deeper into feature choice techniques like Principal Component Analysis (PCA). This research avenue is expected to be pivotal in mitigating the challenges associated with the exponential growth of data requirements as the number of features increases. Future projects in distress situation detection will benefit from a more comprehensive exploration of these areas.

REFERENCES

1. Toffa, O.K. and M. Mignotte, *Environmental sound classification using local binary pattern and audio features collaboration.* IEEE Transactions on Multimedia, 2020. **23**: p. 3978-3985.
2. Shaukat, A., et al. *Towards automatic recognition of sounds observed in daily living activity*. in *2019 IEEE 18th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*. 2019. IEEE.
3. Noury, N., et al. *Fall detection-principles and methods*. in *2007 29th annual international conference of the IEEE engineering in medicine and biology society*. 2007. IEEE.
4. Tamura, T., et al., *A wearable airbag to prevent fall injuries.* IEEE Transactions on Information Technology in Biomedicine, 2009. **13**(6): p. 910-914.
5. Organization, W.H., *Good health adds life to years: Global brief for World Health Day 2012*. 2012, World Health Organization.
6. Mubashir, M., L. Shao, and L. Seed, *A survey on fall detection: Principles and approaches.* Neurocomputing, 2013. **100**: p. 144-152.
7. Nizam, Y., M.N.H. Mohd, and M.M.A. Jamil, *A study on human fall detection systems: Daily activity classification and sensing techniques.* International Journal of Integrated Engineering, 2016. **8**(1).
8. Habibipour, A., A.M. Padyab, and A. Ståhlbröst. *Social, ethical and ecological issues in wearable technologies*. in *AMCIS 2019, Twenty-fifth Americas Conference on Information Systems, Cancun, México, Augusti 15-17, 2019*. 2019. Association for Information Systems.
9. Adapa, A., et al., *Factors influencing the adoption of smart wearable devices.* International Journal of Human–Computer Interaction, 2018. **34**(5): p. 399-409.
10. Yazar, A., et al., *Fall detection using single-tree complex wavelet transform.* Pattern Recognition Letters, 2013. **34**(15): p. 1945-1952.
11. Su, B.Y., et al., *Doppler radar fall activity detection using the wavelet transform.* IEEE Transactions on Biomedical Engineering, 2014. **62**(3): p. 865-875.
12. Ren, L. and Y. Peng, *Research of fall detection and fall prevention technologies: A systematic review.* IEEE Access, 2019. **7**: p. 77702-77722.
13. Phan, H., et al., *Learning representations for nonspeech audio events through their similarities to speech patterns.* IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016. **24**(4): p. 807-822.
14. Zhang, H., I. McLoughlin, and Y. Song. *Robust sound event recognition using convolutional neural networks*. in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. 2015. IEEE.
15. Thwe, K.Z. and N. War. *Sound event classification using bidirectional local binary pattern*. in *2017 International Conference on Signal Processing and Communication (ICSPC)*. 2017. IEEE.
16. Dennis, J., H.D. Tran, and E.S. Chng, *Image feature representation of the subband power distribution for robust sound event classification.* IEEE Transactions on Audio, Speech, and Language Processing, 2012. **21**(2): p. 367-377.
17. Park, H. and C.D. Yoo, *CNN-based learnable gammatone filterbank and equal-loudness normalization for environmental sound classification.* IEEE Signal Processing Letters, 2020. **27**: p. 411-415.
18. Geiger, J.T. and K. Helwani. *Improving event detection for audio surveillance using gabor filterbank features*. in *2015 23rd European Signal Processing Conference (EUSIPCO)*. 2015. IEEE.
19. Schröder, J., S. Goetze, and J. Anemüller, *Spectro-temporal Gabor filterbank features for acoustic event detection.* IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015. **23**(12): p. 2198-2208.
20. Rougier, C., et al., *Robust video surveillance for fall detection based on human shape deformation.* IEEE Transactions on circuits and systems for video Technology, 2011. **21**(5): p. 611-622.
21. Shotton, J., et al. *Real-time human pose recognition in parts from single depth images*. in *CVPR 2011*. 2011. Ieee.
22. Bian, Z.-P., et al., *Fall detection based on body part tracking using a depth camera.* IEEE journal of biomedical and health informatics, 2014. **19**(2): p. 430-439.
23. Patsadu, O., et al., *Fall motion detection with fall severity level estimation by mining kinect 3D data stream.* Int. Arab J. Inf. Technol., 2018. **15**(3): p. 378-388.

24. Mastorakis, G. and D. Makris, *Fall detection system using Kinect's infrared sensor.* Journal of Real-Time Image Processing, 2014. **9**: p. 635-646.
25. Mundher, Z.A. and J. Zhong, *A real-time fall detection system in elderly care using mobile robot and kinect sensor.* International Journal of Materials, Mechanics and Manufacturing, 2014. **2**(2): p. 133-138.
26. Yang, L., et al., *New fast fall detection method based on spatio-temporal context tracking of head by using depth images.* Sensors, 2015. **15**(9): p. 23004-23019.
27. Kalinga, T., et al. *A fall detection and emergency notification system for elderly*. in *2020 6th international conference on control, automation and robotics (ICCAR)*. 2020. IEEE.
28. Nizam, Y., et al., *A novel algorithm for human fall detection using height, velocity and position of the subject from depth maps.* International Journal of Integrated Engineering, 2018. **10**(3).
29. Kim, J., et al., *Occupant behavior monitoring and emergency event detection in single-person households using deep learning-based sound recognition.* Building and Environment, 2020. **181**: p. 107092.
30. Abdoli, S., P. Cardinal, and A.L. Koerich, *End-to-end environmental sound classification using a 1D convolutional neural network.* Expert Systems with Applications, 2019. **136**: p. 252-263.
31. Zhang, Z., et al., *Attention based convolutional recurrent neural network for environmental sound classification.* Neurocomputing, 2021. **453**: p. 896-903.
32. Droghini, D., et al. *Few-shot siamese neural networks employing audio features for human-fall detection*. in *Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence*. 2018.
33. Droghini, D., et al., *Audio metric learning by using siamese autoencoders for one-shot human fall detection.* IEEE Transactions on Emerging Topics in Computational Intelligence, 2019. **5**(1): p. 108-118.
34. Esmaeilpour, M., P. Cardinal, and A.L. Koerich, *Unsupervised feature learning for environmental sound classification using weighted cycle-consistent generative adversarial network.* Applied Soft Computing, 2020. **86**: p. 105912.
35. Bahmei, B., E. Birmingham, and S. Arzanpour, *CNN-RNN and data augmentation using deep convolutional generative adversarial network for environmental sound classification.* IEEE Signal Processing Letters, 2022. **29**: p. 682-686.
36. Popescu, M., et al. *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*. 2008. IEEE.
37. Popescu, M. and A. Mahnot. *Acoustic fall detection using one-class classifiers*. in *2009 annual international conference of the ieee engineering in medicine and biology society*. 2009. IEEE.
38. Li, Y., et al. *Acoustic fall detection using a circular microphone array*. in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 2010. IEEE.
39. Li, Y., K. Ho, and M. Popescu, *Efficient source separation algorithms for acoustic fall detection using a microsoft kinect.* IEEE Transactions on Biomedical Engineering, 2013. **61**(3): p. 745-755.
40. Li, Y., K. Ho, and M. Popescu, *A microphone array system for automatic fall detection.* IEEE Transactions on Biomedical Engineering, 2012. **59**(5): p. 1291-1301.
41. Hassan, F., et al., *Comparative Analysis of Machine Learning Algorithms for Classification of Environmental Sounds and Fall Detection.* Science and Technology, 2022. **4**(1): p. 163-174.
42. Kaur, P., Q. Wang, and W. Shi, *Fall detection from audios with audio transformers.* Smart Health, 2022. **26**: p. 100340.
43. Mohtadifar, M., M. Cheffena, and A. Pourafzal, *Acoustic-and Radio-Frequency-Based Human Activity Recognition.* Sensors, 2022. **22**(9): p. 3125.
44. Litvak, D., Y. Zigel, and I. Gannot. *Fall detection of elderly through floor vibrations and sound*. in *2008 30th annual international conference of the IEEE engineering in medicine and biology society*. 2008. IEEE.

45. Principi, E., et al., *Acoustic cues from the floor: a new approach for fall classification.* Expert Systems with Applications, 2016. **60**: p. 51-61.

46. Droghini, D., et al., *A combined one-class SVM and template-matching approach for user-aided human fall detection by means of floor acoustic features.* Computational intelligence and neuroscience, 2017. **2017**.

47. Droghini, D., et al., *Human fall detection by using an innovative floor acoustic sensor.* Multidisciplinary approaches to neural computing, 2018: p. 97-107.

48. Kobayashi, T. and J. Ye. *Acoustic feature extraction by statistics based local binary pattern for environmental sound classification*. in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. 2014. IEEE.

49. Nishiura, T. and S. Nakamura. *An evaluation of sound source identification with RWCP sound scene database in real acoustic environments*. in *Proceedings. IEEE International Conference on Multimedia and Expo*. 2002. IEEE.

50. Ye, J., et al. *Kernel discriminant analysis for environmental sound recognition based on acoustic subspace*. in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013. IEEE.

51. Chatlani, N. and J.J. Soraghan. *Local binary patterns for 1-D signal processing*. in *2010 18th European signal processing conference*. 2010. IEEE.

52. Sehili, M.A., et al. *Daily sound recognition using a combination of GMM and SVM for home automation*. in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. 2012. IEEE.