Fall Detection System Using Novel Median Deviated Ternary Patterns and SVM

Babar Younis
University of Engineering and Technology-Taxila, Punjab, Pakistan.
hmbarkhan@gmail.com.

Ali Javed*
University of Engineering and Technology-Taxila, Punjab, Pakistan.
ali.javed@uettaxila.edu.pk.

Farman Hassan
University of Engineering and Technology-Taxila, Punjab, Pakistan.
farmanhassan555@gmail.com.

Correspondence: Ali Javed (ali.javed@uettaxila.edu.pk)

2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT) | 978-1-6654-3773-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISAECT53699.2021.9668520

Abstract--- **In recent years, we have noticed exponential growth in the elderly population of the world due to the advancement in the medical field that necessitates proper care and more attention of elderly people. Accidental falls can be life threatening and can cause severe head trauma, bone fractures, and internal bleedings. Moreover, the most devasting problem of accidental fall incident is that the person remains on the floor for a long time without getting any immediate assistance and response. Research community proposed various fall detection systems but still there exist certain limitations of the existing methods i.e., computational complexity, expensive sensors, unable to wear wearable sensors, and associated privacy issues. To address these issues, we proposed a novel feature descriptor median deviated ternary patterns (MDTP) for audio representation to effectively capture the discriminatory traits of fall and non-fall events. We used the proposed MDTP features to train the support vector machine (SVM) to classify the fall and non-fall incidents. Our proposed method is evaluated against two datasets i.e. A3 fall 2.0 dataset and the MSP-UET fall detection dataset. Our proposed method achieved remarkable accuracy of 98% and 97%, precision of 100% and 96%, recall of 97% and 96%, and F1-score of 98% and 96% on the A3 fall 2.0 and MSP-UET fall detection datasets respectively. Experimental results signify the effectiveness of the proposed system for reliable monitoring of elderly people for fall detection.**

Keywords-- ***fall events, median deviated ternary patterns, non-fall, support vector machine.***

I. INTRODUCTION

In many countries, the average life expectancy of people has grown significantly with improved healthcare systems. Advancement in the medical system increases the aging population of the world that needs more attention and proper care of elder people. However, in most of the countries, it is difficult to provide proper care due to some reasons such as in newly industrialized and developed countries frequent workplace changes, improved mobility of young generations greatly affects the housing and family relationships. Subsequently, this leads to increasing the average distance among family members that reduces the chance of getting timely assistance when elderly people found themselves in an inconvenient situation. Elderly people have significant issues of falling incidents and are prone to falls as they cannot control themselves easily due to their weak muscles. Recently, a recent report [1] published by World Health Organization (WHO) has reported about 646,000 casualties annually due to falls making it the second major cause of accidental deaths around the globe. Report from WHO also suggests that the individuals of age 60 or above are more vulnerable to falling which makes higher death rates in such age groups. For an elderly person, fall incidents can be extremely dangerous and might result in severe health problems i.e., head trauma, bone fracture, internal bleeding. The injured person needs to get immediate assistance after the fall incident. Often, people after falling are unable to rise up without any assistance and need urgent medical attention. These challenges necessitate the development of automated fall detection systems to ensure quality care and timely assistance for elderly people in case of fall event.

Research community has explored various fall detection systems [2-5] based on wearable devices/sensors i.e., accelerometer, gyroscope, magnetometer, or ambient-based sensors i.e., microphones, Microsoft Kinect, pressure sensors or fusion of both wearable and environmental sensors for the detection of fall incidents. In [2], an IoT-based fall event detection system was designed using various sensors i.e., gyroscope, magnetometer, and tri-axial accelerometer. Fiver features i.e., standard deviation, mean, minimum, maximum, and acceleration were used while Alex-Net was employed for classification purposes. The proposed method used principal component analysis to reduce the dimensionality of features. In [3], a multiple sensors-based fall detection system was designed comprising of three sensors i.e., accelerometer, depth camera, Kinect sensor. Kinect was used for fall alert and accelerometer to indicate a potential fall incident. In [4], a self-learning technique based on a depth camera to detect fall incident was designed. The proposed system analyzed eye status and posture to address the issue of misdetection. In [5], a fall detection system was proposed using a 2D camera. CNN and support vector machine (SVM) were employed to classify daily life activities, occlusion, and fall incidents. However, there exist some limitations of the above-mentioned methods i.e., privacy issues, wearable devices are expensive, and difficult for elderly people to use sensor as it affects their daily life activities.

Research community has also explored various fall detection systems [6-12] using acoustic features i.e., MFCC, acoustic-LTP, acoustic-local binary pattern (acoustic-LBP). In [6], the method captured acoustic waves traveling through the floor. SVM was used to classify fall and non-fall events. In [7], a fall event detection system was designed using the Mel-frequency cepstral coefficients (MFCC) and one-class SVM for fall detection. Similarly in [8], MFCC features were used with the SVM to discriminate between the fall and non-fall events. Few works [9, 10] explored deep learning to detect fall and non-fall incidents. In [9], MFCC features were fed to SVM, KNN, and neural networks to classify the fall and non-fall events. In [10], MFCC, energy, zero crossing, and spectral flux were used with the deep neural network for classification of fall and non-fall events. In [11], MFCC and linear predictive coefficients were used to train an Ensemble classifier for fall detection. In [12], acoustic local ternary pattern (acoustic-LTP) features were used to train the SVM for fall event detection. This technique [12] employed a static threshold for LTP calculation that made this approach unable to perform well under noisy conditions.

Acoustic-LBP [13], acoustic-LTP [12], and MFCC are powerful acoustic feature extraction methods. However, these techniques are prone to noise present in an audio signal this might be due to using a fixed threshold value [12], taking the central value as threshold value [13], or due to the effects of environmental factors on MFCC [7].

Even the above-mentioned fall detection systems have achieved reasonable performed, however, still there exist certain limitations i.e., elderly people can't wear sensors for a long time, computational complexity, less robustness over environmental noise, privacy issues, etc. To better address these concerns, we develop a robust voice-based fall detection system using a novel MDTP feature and SVM. Moreover, the proposed system is computationally efficient and can be easily implemented in IoT devices to monitor the patients and elder people. The main contributions of our work are as follows:

- We propose a novel feature descriptor MDTP to better capture the most discriminatory traits from an audio signal.
- Our MDTP features are robust to noise present in an audio signal.
- To validate our approach, we conducted rigorous experiments on the A3 Fall 2.0 and MSP-UET fall detection dataset to accurately detect the fall incidents and environmental sounds.

The remaining paper is organized as follows. In Section 2, we discussed the proposed system. Section 3 includes the details of experimental results and discussion. Finally, Section 4 presents the conclusion of our work.

II. PROPOSED SYSTEM

The main objective of the proposed system is to detect two events i.e., fall and non-fall. The proposed system comprises of two different steps i.e., feature extraction and classification. For this purpose, we extracted 20-dim MDTP features from the input audio signal. We used all the 20-dim MDTP features to train the SVM and classify the fall and non-fall events.

A. Feature extraction

The first stage of the proposed system is the feature extraction from input audio. The reliable feature extraction is crucial to develop an effective classification system. The details of the proposed MDTP feature extraction process is discussed in the subsequent section.

B. MDTP Computation

In this paper, we proposed a novel feature descriptor MDTP for extracting 20-dim features from an audio signal *y[n]*. To extract the features from an audio signal, the MDTP analyzes an audio signal *y[n]* and divides it into multiple frames (Ωp). Each Ωp is further divided into 9 overlapping samples. Initially, we compute the threshold (θ) value for all the Ωp of an audio signal. In our previous work acoustic-LTP [12], we used a fixed θ value that was not robust to noise present in an audio signal. In this work, we compute the θ based on the local neighborhood of an audio Ωp which is robust to noise and better captures the local structural information of the audio. We compute the θ value of each Ωp by taking median absolute deviation (MAD) of full Ωp of an audio signal. The MAD of each Ωp is computed as:

$$\theta = Median(|vi - \acute{m}|) \quad (1)$$

where θ is MAD, vi is the value of each neighboring sample, and $\acute{m}$ is the median value of the full Ωp.

To compute the MDTP, we compute the magnitude difference of the central sample (p) by subtracting the vi from p and then compared it with the θ value. The value of θ is based on the local statistics of the vi around the p. Magnitude differences of values greater than θ are quantized to 1 and magnitude differences of values ranges from 0 to θ are quantized to 0, whereas magnitude difference of values smaller than θ are quantized to -1, hence, a three value $MDTP\ (vi, p, \theta)$ is computed as:

$$MDTP(vi, p, \theta) = \begin{cases} 1\ if\ vi - p\ \geq\ \theta \\ 0\ if\ |vi - p| <\ \theta \\ -1\ if\ vi - p\ \leq -\theta \end{cases} \quad (2)$$

where the function $MDTP\ (vi, p, \theta\ )$ represents the MDTP codes.

For illustration, consider a single Ωp comprising of nine samples as shown in Fig. 2. Initially, we compute the MAD of complete Ωp that is equal to 2.19. For the computation of a single Ωp, we subtract the value of v_0 =25 from p=55 and compared its magnitude difference with the value of θ = 2.19. The magnitude difference is -30 which is smaller than 0, so, we quantized it to -1. In the second step, we subtracted the value of v_1=80 from p=55 and its magnitude difference is 25 that is greater than 2.19, so, we quantized it to 1. We computed the magnitude difference of v_3 = 53 from p = 55

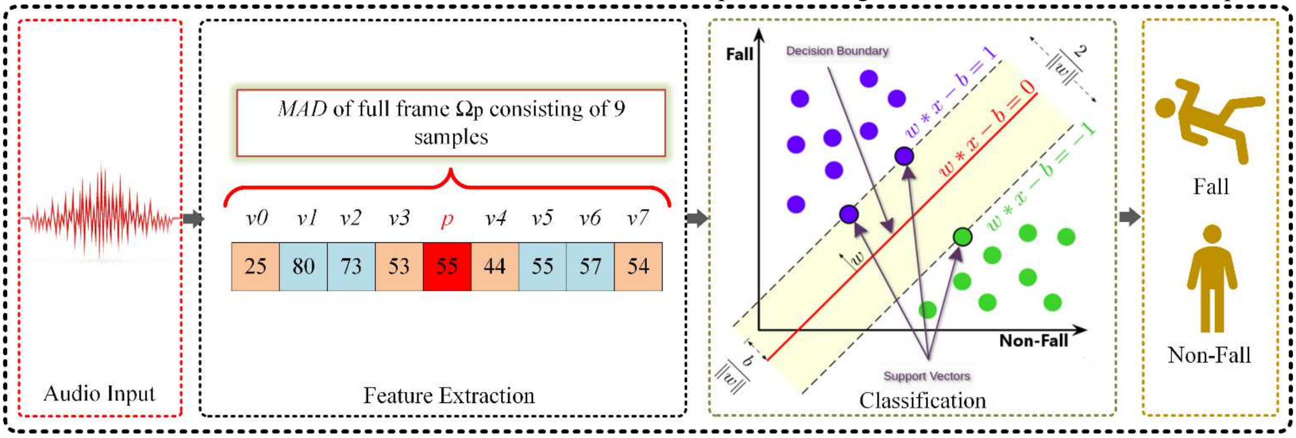

Fig. 1. Proposed system.

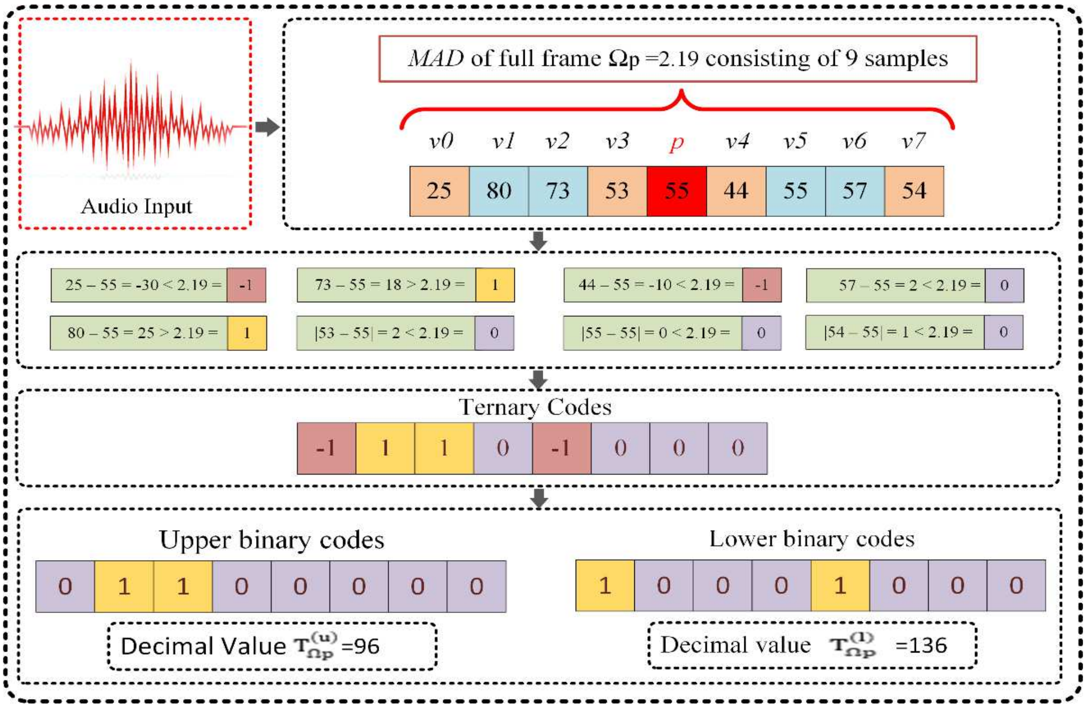

Fig. 2. Feature extraction process.

that is equal to 2, so, we quantized it to 0. In this way we computed the remaining MDTP of the Ωp.

The number of patterns is further divided into upper MDTP$_u$ and lower MDTP$_l$ patterns. Only +1 values are retained in the upper patterns as:

$$MDTP_u\,(vi, p, \theta) = \begin{cases} 1, & MDTP(vi, p, \theta) = +1 \\ 0, & Otherwise \end{cases} \quad (3)$$

Similarly, in the lower patterns, -1 values are retained as 1.

$$MDTP_l\,(vi, p, \theta) = \begin{cases} 1, & MDTP(vi, p, \theta) = -1 \\ 0, & Otherwise \end{cases} \quad (4)$$

The decimal values of the MDTP$_u$ and MDTP$_L$ patterns are encoded as:

$$T^{u}_{\Omega P} = \sum_{i=0}^{7} MDTP_u\,(vi, p, \theta).\,2^i \quad (5)$$

$$T^{l}_{\Omega P} = \sum_{i=0}^{7} MDTP_l\,(vi, p, \theta).\,2^i \quad (6)$$

Next, histograms of the $T^{u}_{\Omega P}$ and $T^{l}_{\Omega P}$ patterns are computed by 7 and 8, where a single bin is assigned to each uniform pattern, and all non-uniforms patterns are assigned to one bin.

$$h_u(f) = \sum_{h=0}^{H} \delta\left(T_w^{(u)}, f\right) \quad (7)$$

$$h_l(f) = \sum_{h=0}^{H} \delta\left(T_w^{(l)}, f\right) \quad (8)$$

where f denotes the histogram bin, and δ is the Kronecker delta function. After experiments, we found that the first ten uniform patterns from the upper and lower classes each are sufficient to capture variations in the acoustic data. As a result, concatenating the two histograms generates a twenty-dimensional feature vector FV as follows:

$$FV = [h_u || \, h_l] \quad (9)$$

C. Classification

In the proposed work, we employed the SVM for classification purpose. We used our 20-dim MDTP features to train the SVM model. We tuned the following parameters for SVM: kernel scale=1, outlier function=0.05, box constraint=100. Moreover, we trained the SVM using the Gaussian radial basis function (RBF) kernel.

III. EXPERIMENTAL SETUP AND RESULTS DISCUSSION

A. Dataset

We used two standard datasets i.e., A3 Fall 2.0 [8] and MSP-UET Fall detection dataset for experimentation purpose that is publicly available. The A3 Fall 2.0 dataset consists of fall events from daily life which include (Ball, Basket, Chair, Fork, Book, and Bag) and a human mimic doll "Rescue Randy". It is professional equipment used to emulate humans in emergency conditions. The Rescue Randy has been used here to impersonate human falls and is being dropped from several positions both forward and backward from a chair for a total of 44 events.

The MSP-UET Fall detection dataset has a total number of 508 audio samples recorded with Infinix note 9 and Lenovo k6. The length of the audio samples ranges from 3 to 7 sec. The audio samples were recorded by different people in different places i.e., washroom, bedroom, living room, etc. The dataset was created to distinguish between the fall and non-fall events. Audio samples of fall events have loud, harsh voices, while the non-fall audio samples are silent with minimum background voices. The details of the datasets are given in Table I.

TABLE I: DETAILS OF DATASETS.

Dataset Name	Total Samples	Total Classes	Fall	Non-Fall
A3 FALL 2.0	460	8	44	416
MSP-UET Fall detection dataset	508	2	234	274

B. Performance evaluation on A3 Fall Dataset

We designed this experiment to evaluate the performance of the proposed method (MDTP +SVM) on the A3 fall 2.0 dataset to detect the fall and non-fall events. For this purpose, we extracted our 20-dim MDTP features from the audio samples of eight different classes of the A3 fall 2.0 dataset to train the SVM model. For the A3 fall dataset, we considered environmental sounds classes as non-fall and Rescue Randy (Doll) which has been used to mimic human falls as the fall event class. We used 80% data for training and 20% data for testing. The classes used as fall in our experiments are Doll and Doll (chair). A doll is used to simulate a real-life human fall. We selected fall samples from the noisy version of the dataset since there were fewer clean fall audio samples in the A3 fall 2.0 dataset. The background music and noise have been added to the noisy version of the dataset to obtain realistically fall events.

For environmental sounds, we consider following classes i.e., Book, Bag, Fork, Chair, Basket, and Ball. We were able to classify between the fall and non-fall events with remarkable accuracy of 98.3%. Experimental results of the proposed system signify that our method is reliable to be used in real-time and can capture the variations present in the audio signal that can successfully detect the two incidents i.e., fall and non-fall incidents.

C. Performance evaluation on MSP-UET Fall detection dataset

We designed this experiment to measure the performance of the proposed method (MDTP +SVM) on our own dataset i.e., MSP-UET fall detection dataset. The dataset contains two classes i.e., fall and non-fall. We extracted 20-dim MDTP features from the audio samples of both fall and non-fall classes. We used 80% data for training and 20% to evaluate the trained model. The fall class contains the audio samples of real humans panic voices after fall incidents. In our experiments, we considered these voices as fall class and other background noises and silent audio samples as a non-fall class. We achieved a remarkable accuracy of 97% using our proposed method. These results show the effectiveness of our method for fall detection on a diverse dataset.

D. Performance evaluation in multiclass environment

We designed this experiment to evaluate the performance of the proposed (MDTP+SVM) method in multiclass environment to detect the fall and environmental sounds. There is a significant correlation between the classes of the A3 fall 2.0 dataset which makes it more challenging to detect between the fall and non-fall event. Misclassification between different classes leads to a class imbalance problem. For this purpose, we extracted our MDTP features from the audio samples of six different classes i.e., ball, basket, chair, fork, book, bag of the A3 fall 2.0 dataset. We used 80% of the data for training and rest 20% of the data for testing. From the results reported in Table II, we observed that our method performed remarkably well in a multiclass environment. Our method achieved the maximum accuracy, precision, recall, and F1-score of 100% for the ball and basket classes. Whereas, unable to provide better detection performance for the chair class by achieving an accuracy of 83.89%, precision of 77.77%, recall of 90.32%, and F1-Score of 83.58%. The proposed method performed well for all other classes i.e., fork, book and bag. The detailed results in terms of accuracy, precision, recall, and F1-score are given in Table II. Experimental results illustrate that the proposed system can effectively be used to detect the fall events in multiclass environment.

TABLE II: PERFORMANCE EVALUATION ON MULTI-CLASS ENVIRONMENT.

Classes	Accuracy %	Precision %	Recall %	F1-Score %
Ball	100	100	100	100
Basket	100	100	100	100
Chair	83.89	77.77	90.32	83.58
Fork	95.78	94.44	97.14	95.77
Book	93.15	94.44	91.89	93.14
Bag	87.33	93.93	81.57	87.32

E. Performance comparison with existing methods

This experiment is designed to compare the performance of the proposed system with the existing state-of-the-art methods [8, 10] on the A3 fall 2.0 dataset. From the results reported in Table III, Principi et al. [8] employed the MFCC, energy, zero crossing and spectral flux features with the DNN and achieved the lowest accuracy of 91%. Alex et al. achieved 97% accuracy by combining various spectral features i.e., MFCC, spectral flux, zero crossing rate, energy, and spectral centroid. Experimental results on the A3 fall 2.0 dataset demonstrate that our system offered superior fall detection performance over the comparative approaches.

TABLE III: PERFORMANCE COMPARISON WITH EXISTING METHODS

Dataset	Author	Proposed Method	Accuracy%
A3 Fall 2.0	Alex et al[10]	Spectral Features + DNN	97%
	Principi et al[8]	MFCC+SVM	91%
	Proposed	MDTP+SVM	98%

IV. CONCLUSION

This paper has presented a reliable and automated fall detection system using a novel feature descriptor MDTP and SVM. The usage of median absolute deviation for threshold calculation in our MDTP features make them robust to environmental noise. The proposed system is capable of accurately detecting the fall events and multiple sounds classes. We used two datasets A3 fall 2.0 and the MSP-UET fall detection dataset for experiments. Experimental results show that the proposed system successfully detect the fall events with an accuracy of 98% on the A3 fall 2.0 dataset and 97% on the MSP-UET fall detection dataset. The proposed method based on low cost MDTP features and SVM can efficiently be used in a real-time environment for fall detection. In future, we aim to test the generalizability of the proposed system by cross dataset evaluation.

ACKNOWLEDGMENT

This work was supported by the Multimedia Signal Processing Research Lab at UET Taxila.

REFERENCES

[1]. WHO. (2018, Falls. *No 1(February 29,2020),* 10. Available: https://www.who.int/en/news-room/fact-sheets/detail/falls

[2]. A. Alarifi and A. Alwadain, "Killer heuristic optimized convolution neural network-based fall detection with wearable IoT sensor devices," *Measurement,* vol. 167, p. 108258, 2021.

[3]. B. Kwolek and M. Kepski, "Improving fall detection by the use of depth sensor and accelerometer," *Neurocomputing,* vol. 168, pp. 637-645, 2015.

[4]. X. Kong, L. Meng, and H. Tomiyama, "A self-learning fall detection system for elderly persons using depth camera," *International Journal of Advanced Mechatronic Systems,* vol. 8, pp. 16-25, 2020.

[5]. F. Lezzar, D. Benmerzoug, and I. Kitouni, "Camera-Based Fall Detection System for the Elderly With Occlusion Recognition," *Applied Medical Informatics.,* vol. 42, pp. 169-179, 2020.

[6]. D. Droghini, E. Principi, S. Squartini, P. Olivetti, and F. Piazza, "Human fall detection by using an innovative floor acoustic sensor," in *Multidisciplinary Approaches to Neural Computing*, ed: Springer, 2018, pp. 97-107.

[7]. M. S. Khan, M. Yu, P. Feng, L. Wang, and J. Chambers, "An unsupervised acoustic fall detection system using source separation for sound interference suppression," *Signal processing,* vol. 110, pp. 199-210, 2015.

[8]. E. Principi, D. Droghini, S. Squartini, P. Olivetti, and F. Piazza, "Acoustic cues from the floor: a new approach for fall classification," *Expert Systems with Applications,* vol. 60, pp. 51-61, 2016.

[9]. M. Cheffena, "Fall detection using smartphone audio features," *IEEE journal of biomedical and health informatics,* vol. 20, pp. 1073-1080, 2015.

[10]. J. S. R. Alex, M. A. Kumar, and D. Swathy, "Deep Learning Approaches for Fall Detection Using Acoustic Information," in *Advances in Smart Grid Technology*, ed: Springer, 2021, pp. 479-488.

[11]. A. Shaukat, M. Ahsan, A. Hassan, and F. Riaz, "Daily sound recognition for elderly people using ensemble methods," in *2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2014, pp. 418-423.

[12]. S. M. Adnan, A. Irtaza, S. Aziz, M. O. Ullah, A. Javed, and M. T. Mahmood, "Fall detection through acoustic local ternary patterns," *Applied Acoustics,* vol. 140, pp. 296-300, 2018.

[13]. N. Chatlani and J. J. Soraghan, "Local binary patterns for 1-D signal processing," in *2010 18th European Signal Processing Conference*, 2010, pp. 95-99.