



Footstep Recognition Using Mel Frequency Cepstral Coefficients and Artificial Neural Network

Thasya Nurul Wulandari¹, Hilal H. Nuha², Rahmat Yasirandi³

^{1, 2, 3}Informatics, School of Computing, Telkom University

¹thasyanurulwulandari@student.telkomuniversity.ac.id, ²hilalnuha@telkomuniversity.ac.id,

³batanganhitam@telkomuniversity.ac.id

Abstract

Footstep recognition is relatively new biometrics and based on the learning of footsteps signals captured from people walking on the sensing area. The footstep signals classification process for security systems still has a low level of accuracy. Therefore, we need a classification system that has a high accuracy for security systems. Most systems are generally developed using geometric and holistic features but still provide high error rates. In this research, a new system is proposed by using the Mel Frequency Cepstral Coefficients (MFCCs) feature extraction, because it has a good linear frequency as a copycat of the human hearing system and Artificial Neural Network (ANN) as a classification algorithm because it has a good level of accuracy with a dataset of 500 recording footsteps. The classification results show that the proposed system can achieve the highest accuracy of validation loss value 57.3, Accuracy testing 92.0%, loss value 193.8, and accuracy training 100%, the accuracy results are an evaluation of the system in improving the foot signal recognition system for security systems in the smart home environment.

Keywords: biometric, footstep, recognition system, mel frequency cepstral coefficients (MFCCs), artificial neural network (ANN)

1. Introduction

Every human being has different characteristics, one of which is his footsteps. Every human footstep has differences, between one human and another. The differences come many features, like the swing of his footsteps, the rhythm of his footsteps, and the length of his footsteps. A lot of size-based technologies that exist in the human body or commonly referred to as biometrics such as fingerprints, eye corneas, and faces, are used in the field of security to prevent acts of crime.

According to 2017 crime statistics, the number of crime incidents in Indonesia has not decreased significantly. In 2013 there were an estimated 140 people who were at risk of crime, each 100,000 population, 131 people in 2014, and 140 people in 2015 [1]. From 2013 to 2015, the national police registration data revealed that crime incidents in Indonesia tended to change or were not constant. In 2013 there were around 341 thousand cases of crime. However, in 2014 it decreased to approximately 325 thousand cases. Then, in 2015 there were about 353 thousand cases. The number of crime victims in 2013, from around 2.43 million people in Indonesia, increased by approximately 2.66 million people in 2014. For about 2.64 million people in 2015

[2], data shows that the occurrence of crime continues every year. The rise of these actions caused unrest, which made the community must be wary of home security. Therefore a security system is needed to prevent crime. The footstep recognition system is one of the ways that homeowners can do it.

Previously, many studies have been conducted to build a recognition system using human footsteps. Difference systems have been built before. Riwurohi et al [3] conducted a study for a recognition system that focuses on data acquisition in applying scenarios to obtain various variations of footsteps on existing media with recordings to recognize the footsteps of people from three sides. The sides are left, middle, and right sides (for general/unspecified person) by applying the MFCCs feature extraction method and ANN backpropagation algorithm for classification. The ANN architecture uses five input neurons, two hidden, each of which has 30 neurons, and ten output neurons (as a result of the representation of the number of people involved in the study). The dataset used changed ten people consisting of 5 men and 5 women using wooden or other footwear of the same type. The research dataset was taken in a closed room with a recording sample using four microphones with a total of 720 footsteps. At the time of

testing the distribution of training data and testing, data amounted to 70:30. The results obtained in that study were 98.8% for the recognition of people who walked on the left side of the track, 98.8% for the middle side of the track, and 95% for the recognition of people who ran on the right side of the track. While the system that will be built now is the writer will build a footstep signal recognition system as the identity of the homeowner that is applied to the security system in a smart home environment (people are more specific/in accordance with the characteristics of his footsteps) by applying the MFCCs feature extraction method and ANN algorithm (backpropagation and feed-forward) for classification. Accuracy results that will be obtained in the form of system accuracy can recognize the signal of the smart home owner's footsteps.

The development of digital technology has made a lot of security systems that come from manuals into digital forms. In research [4], a security system study was conducted on smart homes by applying footstep signals as a security system by obtaining an error accuracy of 13% using SVM as a classification algorithm. In research [5], a security system study was conducted to identify people in the room using structural footsteps that have an accuracy of 83% using uncertain footsteps. When footsteps are unknown, identification accuracy increases to 96.5%. Furthermore, research [6] conducted a study of the body's safety system by detecting the severity of asthma based on breath voices using the extraction of the Mel Frequency Cepstral Coefficients (MFCCs) and K-Nearest Neighbor (KNN) features as a classification algorithm. Accuracy results reached an average of 97.5%. In [7], a security system for advanced driver assistance (ADAS) has been carried out by classifying the driver's intention to change paths based on information that has been measured and estimated using machine learning-based estimation techniques using Artificial Neural Network (ANN) and Support Vector Machine (SVM). Accuracy results reach 90% under any surface conditions.

One classification algorithm that can be used to carry out the classification process is ANN. ANN has been used for various classification problems. In research [8][9] a comparison of ANN and SVM was conducted. The results obtained, ANN had better results with the achievement of the accuracy of 88.4% compared to SVM, which reached 78.2%. Furthermore, in research [10][11] ANN algorithm and MFCCs feature extraction were used for 96.4% Malayalam speech recognition. In research [12][13] a study was carried out using MFCCs feature extraction to translate speech into digital text with an accuracy of around 87.5%.

Therefore, this research proposes the feature extraction of MFCCs and ANN algorithms to be used for footstep signals recognition classification. The focus in this research includes testing of footstep signals recognition

to improve the footstep signals recognition system for security systems in the smart home environment.

2. Research Method

This research is an experimental based research. This section explains research methods, biometric theory, MFCCs, and ANN.

2.1. Research Methodology

The system is designed to consist of a process to the data split or data separation process. Next, the process is divided into two parts, namely the training process and the testing process, as illustrated in Figure 1.

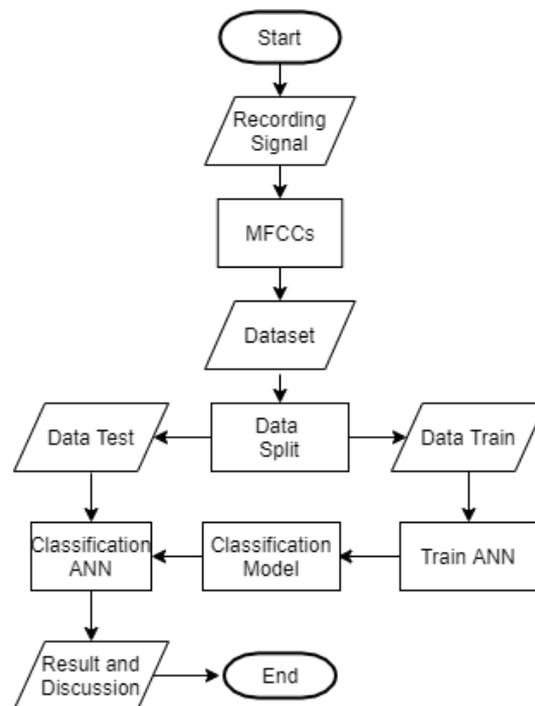


Figure 1. Research Methodology

Figure 1. In the first process carried out is to convert the footstep signal using the MFCCs feature extraction by dividing the signal into several frames and changing it into a cepstrum to get the vector value. The results of MFCCs will be labeled and used as a dataset in this study. In the next process, the separation of the dataset is grouped into two parts, namely training data and testing data. In the training process, the system will be trained using training data by applying the ANN algorithm to get the best classification model for footstep recognition. Furthermore, after getting the model, then the testing process is carried out using test data by applying ANN classification that have been obtained in the training process. After carrying out these tests it will get the results of the accuracy of the footstep recognition system process.

2.2. Biometric

According to research [14][15], biometrics is the science that identifies a person based on unique features related to the human body, such as fingerprints, voice, face, retina, and footsteps. Good biometrics are the ones with certain characteristics, such as unique to everyone. Biometric features can be obtained quickly, have high-resolution capabilities, and features obtained from biometric extraction will not change over time or can be said to be fixed. The biometric features in research [16] stated that biometrics are very important to be applied for safety factors. The application of biometrics as a security system can be applied in various environments. In [4], biometrics is applied in the smart home environment. It also can be applied in smart parking areas [17]. In addition to security systems, biometric factors can be used as recognition systems [18].

2.3. Mel Frequency Cepstral Coefficients (MFCCs)

MFCCs are variations of the human hearing bandwidth with frequencies below 1000 Hz to copy the mechanism of human hearing. The characteristics of MFCCs are having linear frequencies below 1000 Hz and logs above 1000 Hz. The advantages of MFCCs are having a high level of performance in hearing recognition accuracy, capturing the characteristics of a sound, and having a low complexity [19], that the characteristics and performance of MFCCs can be applied as a signal translator system [20].

The way MFCCs work is the following steps. First, the frequency signal is divided into several frames and applies hamming windowing to each frame. After the process, to get the wave set from each frame, we use the formula in equation (1).

$$Y(k) = \sum_{n=0}^{N-1} y(n) e^{-\frac{j2\pi kn}{N}}, 0 < k < N \quad (1)$$

where $y(n)$ is the windowing frame and $Y(k)$ is the Discrete Fourier Transforms (DFT) of each frame. After getting a collection of waves, the next process is the calculation of tones that play a different scale (mel scale). Next in equation (2) the formula for calculating the estimate of mels.

$$mel(f) = 2595 \times \log_{10}(1 + f/700) \quad (2)$$

Where f is the frequency (Hz) [21], then the next step is to filter the results of the mel scale using the mel filterbank. The results of the filtering process will take the output value for each filter. In the process of taking it, we use the Discrete Cosine Transform (DCT) on the log mel obtained by the spectrum. The following formula for the calculation of DCT in equation (3).

$$c[n] = \sum_{m=0}^{F-1} s[m] \cos \frac{\pi n(m+0.5)}{F} \quad (3)$$

where $s[m]$ is the result of the previous stage, and F is the number of filters contained in the filterbank. The

final result of the calculation is referred to as the coefficient of MFCCs with human hearing system. It is sensitive to the static and dynamic characteristics of the signal [8].

2.4. Artificial Neural Network (ANN)

ANN is one of the classification algorithm structures to get data accuracy and has a good level of performance [10]. The use of ANN can be applied to classify various features of the recognition system [22]. The structure of the ANN algorithm consists of three layers, namely the input, hidden, and output layers. Each layer consists of several neurons that are interconnected with other neurons. Each connection represents the weight that contributes to each relationship with the right activation function. It also represents a combination that can be optimized by predicting the weights in equation (4) as follows.

$$Y = \sigma \left(\sum_{i,j}^n w_{ij} x_i + b \right) \quad (4)$$

where w_{ij} is the representation of the connection weight value, x_i represents the input variable, b represents the bias, and σ is an activation function.

Development of ANN models based on knowledge, training, and testing. The training process is the process of learning from a database and being trained using datasets that were not previously used in the training process. ANN training consists of database preparation and selection of a large number of variables to reduce errors in the training process, and the range of data must also be broad so that the training gets the optimal value. To validate a trained ANN, it is necessary to test the model using a data set that is not used during the training process. From the input results, the test dataset can be compared with the actual data when testing using the calculation of the root mean square error (RMSE) in equation (5) as follows.

$$RSME = \sqrt{\sum_i^n \frac{(P_i - A_i)^2}{n}} \quad (5)$$

Where P_i is the predicted value of ANN, A_i is the actual value, and n is the number of datasets used [23].

2.5. Dataset

The dataset is obtained from the conversion of footstep signals using the MFCCs feature extraction. In the process of MFCCs, the data used were 500 footstep records consisting of three people (first 400 records and two other 50 records each). Each footstep has recorded by using a standard microphone that is duration 2-3 seconds in duration. A Recording is done on the floor without using footwear with the microphone beside the feet which are about 10-15 cm apart. The standard microphone used is one. To get footstep signal

conversion results, here are the *pseudo-code* from MFCCs.

```

Program of MFCCs


---


Input: Footstep Signals
Output: MFCCs vector // MFCCs of recording signal
Algorithm :
Function MFCCs (parameters)
    Initialization parameters.
    The division of signal into several frames.
    Apply Hamming windowing to each frame.
    Getting a set of waves by applying Fast Fourier Transforms (FFT) to the entire frame.
    Getting a matrix for mel-spaced filterbank.
    Change the wave set to mel-spectrum.
    Getting a MFCCs vector from each frame by applying Discrete Cosine Transform (DCT).
end function
    
```

The *pseudo-code* above is a way of working to convert and get vector results from footstep signals using the MFCCs feature extraction.

The next step is how the dataset works. The following *pseudo-code* from dataset.

```

Program of dataset


---


Input: Footstep Signals
Output: Vector collection // Not yet labeled
Initialization i, n, result
result [0..n]
for i = 0 to n do
    result[i] = Mfccs(Footstep Signals)
    i = i + 1
end for
    
```

From the *pseudo-code* above, the conversion results of each footstep signal are converted into one frequency, as in Table 1.

Table 1. Footstep Signals Conversion Results (Dataset Sample)

Footstep Signals	Results				
1st	-792.8	-792.8	...	4.6828	4.6770
2st	-865.0	-865.0	...	5.1063	3.7309
3st	-863.8	-863.8	...	12.009	15.551
4st	-804.9	-804.9	...	14.464	6.8209
5st	-814.8	-814.8	...	15.832	12.002
.
.
.
440th	-7.593	-7.553	...	-5.275	-5.217
441th	-8.563	-8.563	...	-4.952	-5.160
442th	-7.682	-7.682	...	-5.514	-5.554
443th	-7.490	-7.490	...	-4.202	-5.024
444th	-7.518	-7.518	...	-5.723	-5.741

Table 1. is the result of footstep signals conversion that does not yet have a label, in other words the actual data. From the conversion results, each data has 1260 columns. In this process, the researcher will categorize each of the results of signal conversion. The results after categorization can be displayed in Table 2.

Table 2. is the data after being labeled by showing label number 1 is data that will be recognized by the system as actual data and label number 0 is data that is not true (as a comparison when doing classification). The results

of categorization are a dataset that will be used during footstep signal training and footstep signal testing.

Table 2. Categorization Results From the Dataset Sample

Footstep signals	Results					Label
1st	-792.8	-792.8	...	4.6828	4.6770	1
2st	-865.0	-865.0	...	5.1063	3.7309	1
3st	-863.8	-863.8	...	12.009	15.551	1
4st	-804.9	-804.9	...	14.464	6.8209	1
5st	-814.8	-814.8	...	15.832	12.002	1
.
.
.
440th	-7.593	-7.553	...	-5.275	-5.217	0
441th	-8.563	-8.563	...	-4.952	-5.160	0
442th	-7.682	-7.682	...	-5.514	-5.554	0
443th	-7.490	-7.490	...	-4.202	-5.024	0
444th	-7.518	-7.518	...	-5.723	-5.741	0

2.6. Split Data

In this process, the dataset that was previously obtained will be divided into two parts, namely the data used for the training process and the testing process. In this process, the dataset is divided into 75% for training data and 25% for testing data. The following *pseudo-code* of split data.

```

Program of Split Data


---


Input: Dataset // Assumptions are already labeled
Output: Data Train and Data Test
Initialization i, data_train, data_test
for i = 0 to  $(\text{length}(\text{dataset}) * 0.75)$  do
    data_train[i] = random(dataset)
    eliminate(dataset by random attribute)
    sort dataset by index
    i = i + 1
end for
for i = 0 to dataset do
    data_test[i] = dataset
    i = i + 1
end for
    
```

2.7. Training and Testing

The following in Figure 2. Diagram Schema of Training and Testing is a schematic diagram for the footstep signals training process and footstep signals testing.

Figure 2. Diagram Schema of Training and Testing, shows the flow process of the footstep training data and footstep signal testing data. In the first process, the dataset that has been previously processed will be separated into two parts, namely training data and testing data. The data will be classified into the input layer, then the results issued by the input layer are forwarded to the hidden layer for computing. Furthermore, the results of the computing will be processed at the output layer and output from the previous process.

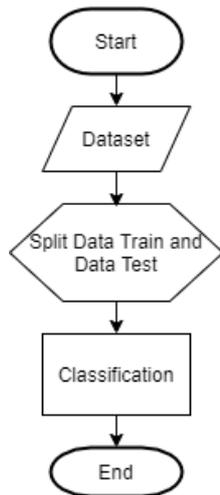


Figure 2. Diagram Schema of Training and Testing

2.7.1. Training

Footstep signals training data will be used to build the model. In this process, ANN will study all training data and optimize the classification process. The ANN classification module uses three-layer structures namely, at the input layer and hidden layer, there are 50 neurons each with *relu* activation and at the output layer, there is one neuron with *sigmoid* activation. The total training data used was 375 data, as illustrated Figure 3. Layer Structure of ANN Classification.

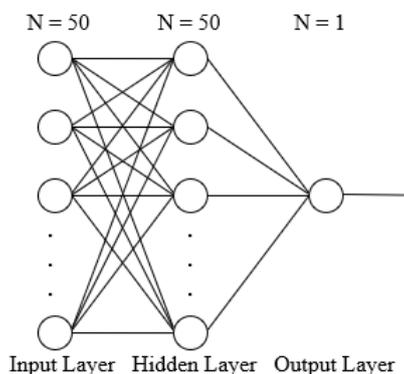


Figure 3. Layer Structure of ANN Classification

Figure 3. Layer Structure of ANN Classification with the input layer will receive training data and supply vector inputs to the network, then, the input entered computes the hidden layer, then the results of the hidden layer are used as input from the output layer. After that, the output layer will distribute output signals from network processing.

2.7.1. Testing

After conducting the training process, the test data serves to test and evaluate the model that has been built. In this process, the test will be carried out by classifying data that has not been seen and doing accuracy calculations. The total test data used was 125 data.

3. Result and Discussion

Testing were carried out using Python language with the Google Collaboratory platform. The device used in this study uses a laptop with AMD Ryzen 5 2500U processor specifications, 4GB of RAM, 128GB hard disk capacity (SSD) + 1TB, and Windows 10 as the operating system.

3.1. Cepstrum Result of MFCCs

Footsteps recording is converted into a vector collection form that is processed using MFCCs feature extraction. In the MFCCs process, the recording is partitioned into several frames and each frame is applied Hamming Windowing. From each frame, you will get a collection of waves by applying FFT. From this wave group, matrix and cepstrum are obtained to be converted into vector. Following is the cepstrum results using MFCCs feature extraction.

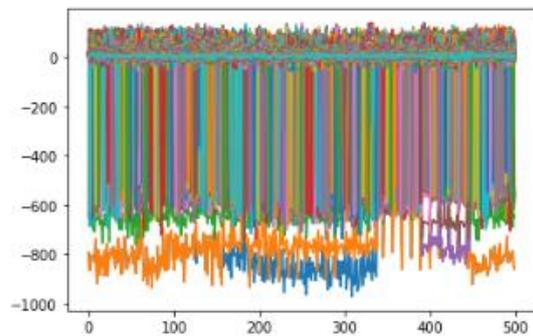


Figure 4. Cepstrum Results of MFCCs

Figure 4. represents the cepstrum results using MFCCs with matrix 432 x 288. The picture shows the results of the cepstrum with the amount of data as much as 500. The results of the cepstrum will be used as a dataset in testing.

3.2. Classification Test Result of ANN

The partition is done in a partition to determine the amount of training set and testing data set. Then the data is conducted training and testing to obtain the highest accuracy results as shown in Figure 5.

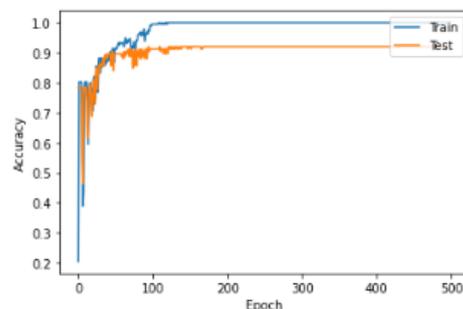


Figure 5. Accuracy of Training and Testing

Figure 5. indicates the x and y axes. The x-axis represents the number of experiments carried out in the

testing process between the range 0-500. then on the y-axis represents the accuracy value obtained in each experiment. This accuracy graph means, each accuracy value from all footstep signal training obtained by the train data and test data.

Furthermore, there are results of accuracy loss that occurs in the data train and data test, as illustrated in Figure 5. Loss of Training and Testing.

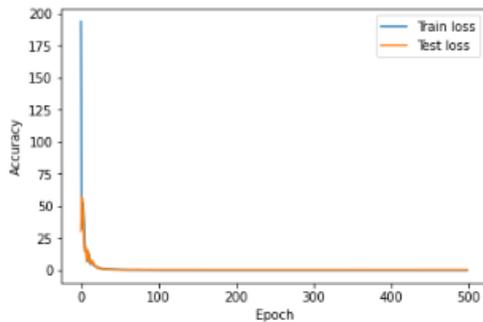


Figure 6. Loss of Training and Testing

Figure 6. Loss of Training and Testing, it can be seen each result of loss value from all training obtained by the train data and test data. From all the training conducted, the highest accuracy was obtained, as in Table 3.

Table 3. Highest Accuracy Results

Name	Epoch	Accuracy
Validation Loss	1	57.3
Validation Accuracy	94	92.0%
Loss	0	193.8
Accuracy	108	100%

In Table 3. based on the results of the classification using the ANN algorithm, the validation loss value (error when testing) was obtained at Epoch-1 with the highest value of 57.3. Furthermore, validation accuracy on the 94th epoch with the highest accuracy of 92.0%. Then, loss (Error during training) at epoch 0 with the highest value of 193.8. Finally, accuracy (Training Accuracy) on the 108th epoch with the highest accuracy of 100%.

4. Conclusion

In this research, the implementation of the Mel Frequency Cepstral Coefficients (MFCCs) feature extraction and Artificial Neural Network (ANN) algorithm for footstep signals recognition systems has been carried out. This research uses a footstep signals conversion result dataset containing 500 records consisting of three people (first 400 records, two others 50 records each). Each footstep was recorded using a standard microphone that was approximately 2-3 seconds in length. The recording was performed on the floor without using footwear with the microphone beside the feet, which are about 10-15 cm apart and was added with a label category with the number 1 or 0 as the output label. The results showed that MFCCs feature extraction and ANN algorithm were successfully used to classify the recognition of footstep signals with a validation loss

value of 57.3, accuracy (testing accuracy) of 92.0%, loss value (training error) 193.8, and accuracy (training accuracy) 100%. From the classification results, it is expected to be used in further research in implementing security systems.

References

- [1] M. Gunawan, T. Aditya, D. T. Geodesi, F. Teknik, and U. G. Mada, "Analitik geovisual pola kejadian dan pergerakan kriminalitas berbasis data media sosial 1,2," *Pros. Semin. Nas. GEOTIK*, pp. 335–345, 2019.
- [2] B. P. Statistik, *Statistik Kriminal 2016*, vol. 111, no. 479. 1965.
- [3] J. E. Riwurohi, J. E. Istiyanto, K. Mustofa, and A. E. Putra, "People Recognition through Footstep Sound Using MFCC Extraction Method of Artificial Neural Network Back Propagation," *Int. J. Comput. Sci. Netw. Secur.*, vol. 18, no. 4, pp. 28–35, 2018.
- [4] R. V. Rodríguez, R. P. Lewis, J. S. D. Mason, and N. W. D. Evans, "Footstep recognition for a smart home environment," *Int. J. Smart Home*, vol. 2, no. 2, pp. 95–110, 2008.
- [5] S. Pan and N. Wang, "Indoor Person Identification through Footstep Induced Structural Vibration," *Proc. 16th Int. Work. Mob. Comput. Syst. Appl.*, pp. 81–86, 2015.
- [6] S. M. Shaharum, S. Aniza, K. Sundaraj, R. Palaniappan, and K. Helmy, "Classification of asthma severity levels by wheeze sound analysis," *Proc. - 2016 IEEE Conf. Syst. Process Control. ICSPC 2016*, no. December, pp. 172–176, 2017.
- [7] I. H. Kim, J. H. Bong, J. Park, and S. Park, "Prediction of driver's intention of lane change by augmenting sensor information using machine learning techniques," *Sensors (Switzerland)*, vol. 17, no. 6, pp. 1–18, 2017.
- [8] T. M. Rajisha, A. P. Sunija, and K. S. Riyas, "Performance Analysis of Malayalam Language Speech Emotion Recognition System Using ANN/SVM," *Procedia Technol.*, vol. 24, pp. 1097–1104, 2016.
- [9] H. H. Nuha, A. Balghonaim, B. Liu, M. Mohandes, M. Deriche, and F. Fekri, "Deep Neural Networks with Extreme Learning Machine for Seismic Data Compression," *Arab. J. Sci. Eng.*, 2020.
- [10] M. Moneykumar, E. Sherly, and W. S. Varghese, "Isolated Word Recognition System for {M}alayalam using Machine Learning," *Proc. 12th Int. Conf. Nat. Lang. Process.*, pp. 158–165, 2015.
- [11] H. Nuha, A. Balghonaim, M. Mohandes, B. Liu, and F. Fekri, "Seismic data compression using deep neural network predictors," in *SEG International Exposition and Annual Meeting 2019*, 2020.
- [12] S. Sen, A. Dutta, and N. Dey, "Speech processing and recognition system," *SpringerBriefs Appl. Sci. Technol.*, pp. 13–43, 2019.
- [13] H. Nuha, M. Deriche, and M. Mohandes, "Bilateral filters with elliptical Gaussian kernels for seismic surveys denoising," in *2019 5th International Conference on Frontiers of Signal Processing, ICFSP 2019*, 2019.
- [14] D. Bhattacharyya, R. Ranjan, F. Alisherov, and C. Minkyu, "Biometric Authentication: A Review Biometric Authentication: A Review," *Int. J. Trend Res. Dev.*, vol. 2(4), no. September 2009, pp. 113–123, 2014.
- [15] M. Mohandes, A. Hussain, P. Khan, and H. Nuha, "Non-Vital Privacy Compromise for Improved Services in Pilgrimage Seasons," in *2019 IEEE RFID Conference*, 2019.
- [16] T. K. Ghazali and N. H. Zakaria, "Security, comfort, healthcare, and energy saving: A review on biometric factors for smart home environment," *J. Comput.*, vol. 29, no. 1, pp. 189–208, 2018.
- [17] I. Mulia Insan, Parman Sukarno, Rahmat Yasirandi, "Multi-Factor Authentication Using a Smart Card and Fingerprint (Case Study: Parking Gate)," *Indones. J. Comput.*, vol. 4, no. August, pp. 55–66, 2019.
- [18] X. Wang, T. Yang, Y. Yu, R. Zhang, and F. Guo, "Footstep-Identification System Based on Walking Interval," *IEEE Intell.*

- Syst., vol. 30, no. 2, pp. 46–52, 2015.
- [19] S. Narang and M. Divya Gupta, “Speech Feature Extraction Techniques: A Review,” *Int. J. Comput. Sci. Mob. Comput.*, vol. 4, no. 3, pp. 107–114, 2015.
- [20] F. Liu and Q. Jiang, “Research on recognition of criminal suspects based on foot sounds,” *Proc. 2019 IEEE 3rd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2019*, no. Itnec, pp. 1347–1351, 2019.
- [21] P. J. Chaudhary and K. M. Vagadia, “A Review Article on Speaker Recognition with Feature Extraction,” *Int. J. Emerg. Technol. Adv. Eng.*, vol. 5, no. 2, pp. 94–97, 2015.
- [22] M. Ariyanto *et al.*, “Finger movement pattern recognition method using artificial neural network based on electromyography (EMG) sensor,” *Proc. 2015 Int. Conf. Autom. Cogn. Sci. Opt. Micro Electro-Mechanical Syst. Inf. Technol. ICACOMIT 2015*, pp. 12–17, 2016.
- [23] H. Li, Z. Zhang, and Z. Liu, “Application of artificial neural networks for catalysis: A review,” *Catalysts*, vol. 7, no. 10, 2017.