ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

A HARDWARE BASED GUNSHOT SOUND DETECTION SYSTEM

M.Sc. THESIS

Mustafa Koray AKÇOCUK

Department of Electronics and Communication Engineering

Electronics Engineering Programme

JULY 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

A HARDWARE BASED GUNSHOT SOUND DETECTION SYSTEM

M.Sc. THESIS

Mustafa Koray AKÇOCUK
504171214

Department of Electronics and Communication Engineering

Electronics Engineering Programme

Thesis Advisor: Prof. Dr. Ece Olcay GÜNEŞ

JULY 2020

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DONANIM TABANLI SİLAH SESİ TESPİT SİSTEMİ

YÜKSEK LİSANS TEZİ

Mustafa Koray AKÇOCUK
504171214

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Elektronik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Ece Olcay GÜNEŞ

TEMMUZ 2020

Mustafa Koray AKÇOCUK, a M.Sc. student of İTU Graduate School of Science Engineering and Technology student ID 504171214, successfully defended the thesis/dissertation entitled "A HARDWARE BASED GUNSHOT SOUND DETECTION SYSTEM", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Ece Olcay GÜNEŞ** İstanbul Technical University	.....
Jury Members :	**Prof. Dr. İsmail Serdar Özoğuz** İstanbul Technical University	.....
	Dr. Nihan Kahraman Yıldız Technical University	.....

Date of Submission : 15 June 2020
Date of Defense : 14 July 2020

To my family

FOREWORD

First of all, I would like to express my gratitude to my consultant Ece Olcay GÜNEŞ for sharing her guidance and support.
I would like to thank my family for their support during the thesis process.
Finally, I would like to thank Tubitak BİDEB for the scholarship support they provided during my gradute studies.

June 2020

Mustafa Koray AKÇOCUK
(Electronics Engineer)

TABLE OF CONTENTS

Page

ABBREVIATIONS

ADC **:** Analog to Digital Converter
ANN **:** Artificial Neural Networks
ASIC **:** Application Spesific Integrated Circuit
BSID **:** Blind System Identification and Deconvolution
CNN **:** Convolutional Neural Networks
DCT **:** Discrete Cosine Transform
ESC **:** Environmental Sound Classification
FFT **:** Fast Fourier Transform
FPGA **:** Field Programmable Gate Array
GMM : Gaussian Mixture Model
GPRS : General Packet Radio Service
GPS : Global Positioning System
HDL **:** Hardware Description Language
HMM **:** Hidden Markov Model
IIR **:** Infinite Impulse Response
IP **:** Intellectual Property
k-NN **:** K-Nearest Neighbor
LPC **:** Linear Predictive Coding
MFCC **:** Mel Frequency Cepstral Coefficients
SNR **:** Signal to Noise Ratio
SVM **:** Support Vector Machine
USA **:** United States of America

SYMBOLS

$\boldsymbol{\omega}$	**:** Angular Frequency
sgn	**:** Signum function
e	**:** Natural logarithm
$\boldsymbol{F_s}$	**:** Sampling frequency
$\boldsymbol{\pi}$	**:** Pi number
log	**:** Logarithm function
cos	**:** Cosine function
$\boldsymbol{m}$	**:** Coefficient number for m_{th} MFCC
$\boldsymbol{L}$	**:** Total band number
$\boldsymbol{l}$	**:** Band number (l=1,2,3,…..,L)
$\boldsymbol{w^T}$	**:** Normal vector to the hyperplane
$\boldsymbol{x_i}$	**:** i_{th} feature vector
$\boldsymbol{y_i}$	**:** Class Label
$\boldsymbol{\varepsilon_i}$	**:** Slack variable

LIST OF TABLES

Page

LIST OF FIGURES

Page

A HARDWARE BASED GUNSHOT SOUND DETECTION SYSTEM

SUMMARY

With the development of semiconductor technology, embedded systems' capacity of operations also increases day by day. In this way, small-sized devices are able to perform complex works. As a result, people take advantage of embedded systems in a wide variety of areas to enhance the life quality of living. As a result of technological developments, the use of tools that assist law enforcement officers in crime detection is also increasing. Available gunshot detection systems mainly focuses on preventing illegal hunting, decreasing crime rates in public space, and detecting gunshot direction in battlefield areas. When the literature is examined, it is seen that machine learning methods are used in the studies used in gunshot sound detection. However, the number of hardware-based systems used in gunshot sound detection is quite a few and mostly simple methods such as cross-correlation threshold, edge detection are implemented.

In this work, it is aimed to realize a gunshot sound detection system on hardware. In this context, it is aimed to select the system that uses the advantages of machine learning methods and is the most suitable for implementation on the hardware. When the literature is examined, it has been observed that the mel coefficients, signal energy and zero crossing properties perform well in determining the gunshot sound. For this reason, the mentioned features were obtained from the audio signal and used in k-nearest neighbors (k-NN) and support vector machines (SVM) classification algorithms. An accuracy rate of 96.1538% was obtained with the k-NN classifier and 91.3462% with the SVM classifier. Within the framework of the results, k-NN classifier was selected for hardware implementation.

Gunshot sounds were obtained from the Still North Media firearm sound effects library, while other sounds such as bark, bird sounds, footsteps were collected from online web pages. In the simulations, in order to obtain accuracy rates in classifying gunshot sounds, 104 tests and 389 training sounds were selected to create the data set.

Implementation of the simulated system on the hardware was done using Matlab Simulink System Generator application. The test dataset features were obtained by

simulation and used in hardware implementation. The sound used for the test is given to the system implemented on hardware, mel coefficients, energy and zero transition properties are obtained and by using the k-NN classifier, it is determined which class the sound belongs to. Since the same method is used with simulation, the accuracy rates obtained in simulation are also valid for hardware implementation.

DONANIM TABANLI SİLAH SESİ TESPİT SİSTEMİ

ÖZET

Yarıiletken teknolojisinin gelişimi ile beraber gömülü sistemlerin işlem yapabilme kapasiteleri de artmaktadır. Böylece küçük boyutlu cihazlar, daha büyük çapta işler yapılabilir hale gelmektedir. İnsanlar yaşam kalitelerini artırmak için birçok farklı alanda gömülü sistemlerden faydalanmaktadırlar.

Teknolojik gelişmelerin bir sonucu olarak suç tespitinde kolluk kuvvetlerine yardımcı olan araçların kullanımı da artmaktadır. Mevcut silah sesi tespit sistemleri genel olarak yasadışı avlanma tespiti, kamusal alanda yaşanan suç oranlarını düşürme ve savaş alanlarında silah sesinin yönünü tespit etme amaçlı kullanılmaktadır. Literatür incelendiğinde silah sesi tespitinde kullanılan çalışmaların genelinde, makine öğrenmesi yöntemlerinin kullanıldığı görülmektedir. Ancak silah sesi tespitinde kullanılan donanım tabanlı sistemlerin sayısı oldukça azdır ve genellikle çarpaz korelasyon ve kenar tespiti gibi basit yöntemler için kullanmaktadır.

Bu çalışmada, silah sesi tespit sisteminin donanım üzerinde gerçekleştirilmesi hedeflenmektedir. Bu kapsamda, makine öğrenmesi yöntemlerinin avantajlarını kullanan ve donanım üzerinde gerçeklemeye en uygun sistemin seçilmesi amaçlanmıştır. Literatür incelendiğinde mel katsayıları, sinyal enerjisi ve sıfır geçisi özelliklerinin silah sesini tespitinde iyi bir performans gösterdikleri gözlemlenmiştir. Bu sebeple, bahsedilen özellikler ses sinyalinden elde edilerek k-en yakın komşuluk (k-NN) ve destek vektör makineleri (DVM) sınıflandırma algortimalarında kullanılmıştır. k-NN sınıfılandırıcısı ile %96.1538, SVM sınıflandırıcısı ile %91.3462 doğruluk oranı elde edilmiştir. Sonuçlar çerçevesinde donanım ile gerçekleme için k-NN sınıflandırıcısı seçilmiştir.

Silah sesleri Still North Media ateşli silah ses efektleri kütüphanesinden elde edilmiştir, havlama, kuş sesleri, ayak sesleri gibi diğer sesler ise çevrimiçi web sayfalarından toplanmıştır. Simülasyonlarda, ateşli silah seslerini sınıflandırmada doğruluk oranları elde etmek için, veri setini oluşturmak üzere 104 test ve 389 eğitim seti seçilmiştir.

Benzetimi yapılan sistemin donanım üzerinde gerçeklenmesi Matlab Simulink System Generator uygulaması kullanılarak yapılmıştır. Sınıflandırma veritabanına ait özellikler simulasyon ile elde edilerek, donanım gerçeklemesinde de kullanılmıştır. Test için kullanılan ses, donanım üzerinde kurulan sisteme verilerek mel katsayıları, enerji ve sıfır geçişi özellikleri elde edilir ve k-NN sınıflandırıcısı kullanılarak, sesin hangi sınıfa ait olduğu bulunur. Benzetim ile aynı yöntem kullanıldığı için benzetimde elde edilen doğruluk oranları donanım gerçeklemesi için de geçerlidir.

1. INTRODUCTION

With technological developments, smart systems have been increasing and machine learning applications help human beings to make life easier. Numerous automated systems are available, from the entertainment industry to healthcare services.

Technological developments in the last century have also affected the defense industry in a way that countries with better defense technology are called “powerful countries”. Automated defence systems such as gunfire incident detectors that detect the gun sound and inform related people are used widely.

Detecting a gunshot fire location on battlefields can be significantly important to prevent the counterattack. Also, in civilian life gunfire detectors can save lives and assist law enforcement officers to locate offenders. According to researches, there have been approximately 10,000 murders committed yearly in the USA with handguns [1]. Authorities can reduce murder rates using gunshot detection systems. In 2015, gunshot detectors were set up in San Francisco. Therefore, homicide rates decreased by 34.6 percent in comparison to 2015 [2].

Moreover, gunshot sound detection systems are used in civilian applications. The systems are widely impelented to prevent illegal hunting. Controlling illegal hunting can be challenging due to big and rugged rainforests. Detectors warn officers about illegal hunting and assist them to save animals. Morever, low power consumption and high connect ability (GPRS, GPS, etc..) are the highest demands for systems which are used in rainforests [3].

Besides, the ShotSpotter system is utilized to prevent fish bombing in Malaysia [4].

Field Programmable Gate Array (FPGA)s are widely used in military applications to provide security and speed conditions.

1.1 Literature Overview

1.1.1 Commercial systems

Gunshot detector and locator systems are commonly used by authorities to prevent crime.

The most known commercial gunshot locator is ShotSpotter which aims to decrease crime rates in the cities and on the campuses [5]. ShotSpotter co-uses multiple acoustic sensors placed on several buildings to detect gunshot location. Time of arrival to acoustic sensors are utilized to predict gunshot firing point.

Other commercial gunshot detectors and locators are Boomerang III and Boomerang Warrior X were developed by Raytheon Technologies. Boomerang III gunshot locators are mounted on Army of USA vehicles to locate gun firing directions [6]. However, Boomerang III systems cannot be carried by soldiers because of their large size and weight. To solve this mobility problem, Boomerang Warrior X has been developed. According to Raytheon, Boomerang Warrior X provides the same features with Boomerang III but extremely smaller and lighter than Boomerang III [7].

There is no available detailed information about the working mechanism of commercial gunshot sound detection systems because of security reasons.

1.1.2 Gunshot sound detection methods

Since the beginning of World War I, researchers have been trying to discover the fastest and the most accurate way of gunshot sound detection.

In 2007, a scream and gunshot sound detection and localization system was proposed by Valenzise G. et al. The system extracts 50 features which include zero-crossing rate, 30 mel frequency cepstral coefficients (MFCCs) and the spectral slope from the sound signal. Sounds were categorized using the gaussian mixture model (GMM) classifier and a detection rate of %93 is achieved. In this classification, gunshot sounds were taken from movies but scream and noise sounds were recorded [8]. However, the size of the dataset was not specified in the paper.

"Shooter localization and weapon classification with soldier-wearable networked sensors" was proposed by Volgyesi P. et al. in 2007. The system uses zero-crossing detection and rising and falling edge detection to identify muzzle blast and shockwave.

Obtained accuracy rates for detecting the sound of different types of guns vary from %23 to %100. The accuracy rates were achieved by using 196 gunshot sound signal samples [9].

In addition, Ahmed T., Uppal M., and Muhammad A. suggested a new technique to improve the efficiency and reliability of gunshot sound detection systems in 2013. The system achieved %99.702 true positive rate by implementing linear predictive coding (LPC) and cross-correlation with a linear SVM classifier. Researchers recorded 332 "gunshot" labeled and 102 "other" labeled sound signals to attain accuracy rates [10].

Besides, in 2015 Kiktova E. et al. designed a system to distinguish gun types using gunshot sounds. Features such as MFCCs, spectral roll-off and zero-crossing rate were obtained in this study. After that, a dimensionality reduction technique was implemented using Min-Redundancy Max-Relevance Mutual Information Base algorithm. As a result, accuracy rates between %80 and %89 were achieved for different types of guns.This system is not suitable for hardware implementation due to the lots of extracted features and the complex dimensionality reduction algorithm. Researches obtained the sounds from the "Joint Database of Acoustic Events and background sounds from Technical University of Kosice". 372 training and 43 testing gunshot sounds were selected to create the dataset [11].

A hardware-based gun-shot sound detection system that takes advantage of the cross-correlation method was suggested by Tobin J. D., in 2015. The gained accuracy rates by using the system change from %40 to %93 for different types of guns. Gunshot sounds were selected from The Free Firearm Sound Effect Library. 280 gunshot sounds were used in order to obtain the accuracy rates. The system was realized on the Xilinx Artix-7 XC7A100T FPGA board [12].

Also in 2015, Salazar-García C. et al implemented Quadrature Mirror Third Order IIR Cauer Filters to obtain features. Besides, Hidden Markov Model (HMM) was realized to classify data on Xilinx Artix XC7A100T-CSG324 [13]. However, the accuracy rates and the used database were not specified in the paper.

Arslan Y. and Güldoğan B. compared the accuracy rates using SVM and multilayer perceptron classifiers in 2015. MFCCs were extracted from sound signals to supply input to classifiers. As a result, this research proved that SVM demonstrates better performance with a %94.7 true positive rate comparing to %92.6 true positive rate of

multilayer perceptron classifier. Gunshot sounds were recorded using 25 microphones in order to create dataset. Moreover, outsider signals such as traffic, train and plane sounds were collected from online resources [14].

A microcontroller-based gunshot sound detection system that extracts LPC, energy and correlation features from a sound signal obtained a %82 true positive rate in 2016. Researchers recorded more than 50 samples for each category to create their own database which consists of gunshot and animal sounds [3]. However, the true positive rate of the system is lower than the true positive rates of the systems mentioned before.

In 2016, a method to weaken the impacts of reverberation was proposed by Shiekh A. A., Tahir M., and Uppal M.. The system uses blind system identification and deconvolution to weaken the impacts of reverberation. Features such as MFCC, LPC and cross-correlation were extracted from sound signals and transmitted to the SVM classifier. Researchers recorded 332 "gunshot labeled" sounds and 102 "outsider labeled" sound signals which consist of claps, ticks, and human speech. As a result, %99 true positive rate was achieved. [15].

Besides, a system that obtains LPC from the sound signal and implements random forest classifier has proposed by Grama L. and Rusu C. in 2017. 1597 audio signals which include gunshot, chainsaw, bird, human voice, and tractor sounds were recorded to create the dataset. Therefore, the system obtained a %95,05 true positive rate for gunshot sound classification [16].

Khamparia A. et al. attained a %85 accuracy rate using convolutional neural networks in 2018. ESC-10 and ESC-50 sound databases which were created by Piczak K.J were used to obtain the accuracy rate. ESC-10 sound library consists of 400 sound signals which include firecracker, dog barking, rain, baby cry, sneezing, chainsaw and clock sounds. ESC-50 sound library includes 2000 environmental sound signals from 50 categories [17]. On the other hand, the obtained accuracy rate is lower than comparing to the systems which implement SVM as a classifier.

Galangque C. M. J. and Guimaldo S. A. introduced a new idea to classify and locate sounds using artificial neural networks (ANN) in 2019. Two ANN classifiers were implemented to classify ballistic shockwave and muzzle blast. Researchers created their own database which includes 864 gunshot and 53 firecracker recordings. The

system achieved a %99 accuracy rate to differentiate M16 guns from ambient noise and firecrackers[18].

1.2 Motivation of the Thesis

Nowadays, high processing capability and low power consumption are the essential requirements of customer-centered electronic design. Existing gunshot sound detection systems are generally run on an operating system. Nevertheless, satisfying low power consumption and high processing capability requirements in this situation is challenging to accomplish. It is also a desired feature to have the system portable.For this purpose, a hardware-based gunshot sound detector was proposed.

In the thesis, the mathematical definitions of the utilized features and classifiers are given in 2nd section.

Besides, the simulation details, achieved results and comparison with other studies that are not realized on hardware is given in the 3rd section.

Furthermore, the hardware structure, hardware implementation details of the system and comparison with the existing hardware-based gunshot sound detection systems are given in the 4th section.

2. SOUND PROCESSING

A sound signal consists of the sum of sinusoidal signals whose frequencies are between 20 and 20,000 Hz. A mechanical sound wave can be transformed into an electrical signal using a microphone. Thereafter, appropriate analog or digital signal processing methods are chosen to process the sound signal. Frequently, an analog signal is filtered and then transformed into a digital signal. Thus, the digital signal can be processed using various methods such as machine learning algorithms. Satisfactory accuracy rates can be achieved in voice recognition applications by using machine learning algorithms. Therefore, a machine learning algorithm was implemented in this study. The implementation details are given in this section.

2.1 Features

Features allow researchers to discriminate the characteristics of the signals. Therefore, combining more than one feature for classification may increase the obtained accuracy rate. In this research, time and frequency domain features are co-used to achieve higher accuracy rates.The mathematical definitions of the extracted features are given in following sections.

2.1.1 Energy

As shown in Eqn. 2.1, the energy sum of the sound signal is defined by the sum of squares of the signal's magnitude [19].

$$E = \sum_{-\infty}^{\infty} (|x|)^2 \tag{2.1}$$

2.1.2 Zero-crossing rate

The zero-crossing rate is defined by the number of zero transitions in a frame. Furthermore, as described in equation 2.2 the zero-crossing rate is formulated as the sum of the magnitude of the difference of the signs of successive points. The speech

sound signals can be separated from background noise by checking zero-crossing rates. The number of zero transitions decreases if there is a speech or meaningful sound in the environment.

$$ZC = \sum_{i=1}^{n-1} | sgn(x_i) - sgn(x_{i+1})| \quad (2.2)$$

2.1.3 Mel frequency cepstral coefficients

Mel scale representation provides the non-linear human ear perception of the sound. Satisfying accuracy rates are obtained using mel frequency cepstral coefficients in speech recognition applications. Extracting steps of MFCCs are shown in Figure 2.1. Moreover, the mathematical definition of the MFCCs is given below.

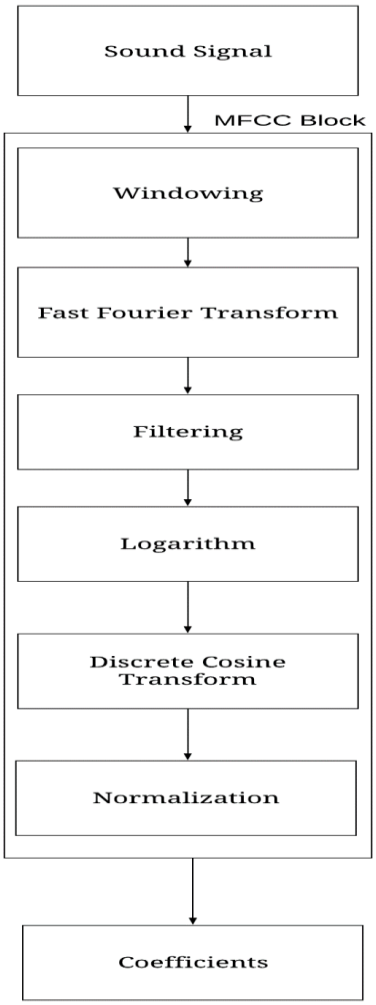

Figure 2.1 : Extracting steps of MFCCs

2.1.3.1 Windowing input stream

Sound streams are separated into small parts using windowing methods before the processing step. Different window lengths are chosen for different systems to meet the application requirements. The sound signal stream is windowed with overlapping windows to prevent data loss. A sample of overlapping windows is presented in Figure 2.2.

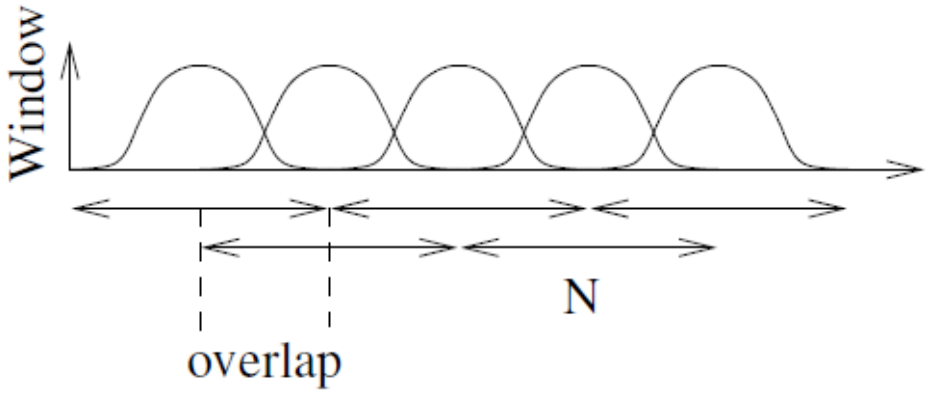

Figure 2.2 : Overlapping window samples [20]

The window length of 1500 milliseconds was selected to get full characteristics of gunshot sound signal.

2.1.3.2 Fast Fourier transform

The fast fourier transform of the windowed sound stream is taken to get the frequency domain representation of the signal [19]. The discrete fourier transform is defined by the Eqn. 2.3.

$$F[j\omega] = \sum_{n=0}^{N-1} f[n]e^{-(j\omega n)} \tag{2.3}$$

2.1.3.3 Filtering

Triangular filters are implemented to transform frequency domain representation into mel scale representation.

Triangular windows are defined by the formula (2.4) [21].

$$h_l[k] \triangleq \begin{cases} 0, & 0 \le k\,{}^{F_s}\!/_{N} < f_c(l-1) \\ a_l \dfrac{k\,{}^{F_s}\!/_{N} - f_c(l-1)}{f_c(l) - f_c(l-1)}, & f_c(l-1) \le k\,{}^{F_s}\!/_{N} < f_c(l) \\ a_l \dfrac{k\,{}^{F_s}\!/_{N} - f_c(l+1)}{f_c(l) - f_c(l+1)}, & f_c(l) \le k\,{}^{F_s}\!/_{N} < f_c(l+1) \\ 0, & f_c(l) \le k\,{}^{F_s}\!/_{N} < f_c(l+1) \end{cases} \tag{2.4}$$

Where, F_s defines the sampling frequency of the sound signal, N defines the sound signal length, l defines the filter number and k defines a point on the l-th filter.

The magnitude factor which only depends on to band level is defined by the formula 2.5 [21].

$$a_l \triangleq \begin{cases} 0.015 & ,1 \le l \le 14 \\ \dfrac{2}{f_c(l+1) - f_c(l-1)} & ,15 \le l \le 48 \end{cases} \tag{2.5}$$

Center frequencies of the filters are defined by the formula 2.6 [21].

$$f_c[l] \triangleq \begin{cases} 66.\overline{66}l & ,1 \le l \le 14, \\ 1073.4(1.0711703)^{(l-14)} & ,15 \le l \le 48 \end{cases} \tag{2.6}$$

2.1.3.4 Logarithm

The human ear responds to the loudness of sound logarithmically. Therefore, the logarithm of the filtered data is taken as shown in equation 2.7 [21].

$$\log\left(\sum_{k=0}^{N-1} |x[k]h_l[k]\right) \tag{2.7}$$

2.1.3.5 Discrete cosine transform

Overlapping triangular filters are preferred to prevent data loss in the filtering step. As a result, repetitive information comes up in different MFCCs. Therefore, a discrete cosine transform of the data is taken using the Eqn. 2.8 to remove repetitive data from the sample [21].

$$\text{dct (m)}=\sum_{l=1}^{L} \log\left(\sum_{k=0}^{N-1} |x[k]h_l[k] \right)\cos\left(\frac{m\pi}{L}\left(l-\frac{1}{2}\right)\right) \quad (2.8)$$

Where, m defines the MFCC coefficient number, l defines the filter number and k defines a point on the filter.

2.1.3.6 Normalization

MFCCs are obtained by normalizing discrete cosine transform result using the Eqn. 2.9.

$$cc_x(m) \triangleq B_l(m)dct(\text{m}) \quad (2.9)$$

Where the normalization factor B_l is defined by the formula (2.10) [21].

$$B_l(m) \triangleq \begin{cases} \sqrt{1/L}, & m = 0 \\ \sqrt{2/L}, & m > 0 \end{cases} \quad (2.10)$$

Where, L defines the total filter number and m defines the MFCC number.

The 0th MFCC is generally ignored by researchers since it only includes information about the signal energy [21].

In this study, sound signals are classified using a feature set including energy, zero-crossing rate and 12 MFCCs. Besides, the features are divided by the highest feature value that is extracted from the classification dataset. In this way, the feature values are scaled to have a maximum value of 1 so they can have equal weight.

2.2 Classification Techniques of Sound

Various classification techniques are preferred by the researchers for sound recognition applications. In existing studies, satisfactory accuracy rates were obtained for gunshot sound detection using SVM and k-NN classifier. Hence, the obtained accuracy rates using support vector machines (SVM) and k-nearest neighbors (k-NN) classifiers were compared in this study. The mathematical background of k-NN and SVM algorithms are specified following.

2.2.1 Support vector machines

Support vector machines are supervised, non-probabilistic learning models that are used for classification or regression. SVM classifier fits a hyperplane to separate categories. The hyperplane parameters are calculated to separate classes with the widest gap. The working mechanisms of linear and non-linear support vector machines are explained in the following section.

2.2.1.1 Linear SVM

Linearly or non-linearly separable classes can be classified using a linear or non-linear SVM classifier. To satisfy application requirements, soft or hard margin SVMs are preferred.

Assume that two classes that are labeled as $y_i = -1$ and $y_i = 1$ wanted to be classified. In this case, linear soft margin SVM is defined by the formula (2.11) [22].

$$\begin{aligned} w^T \cdot x_i - b \geq 0 \; y_i = 1 \\ w^T \cdot x_i - b < 0 \; y_i = -1 \end{aligned} \tag{2.11}$$

Linear hard margin SVMs are defined by the formula (2.12) [22].

$$\begin{aligned} w^T \cdot x_i - b \geq 1 - \varepsilon_i \;\; y_i = 1 \, , \varepsilon_i \geq 0 \\ w^T \cdot x_i - b \leq -1 + \varepsilon_i \;\; y_i = -1 \; \varepsilon_i \geq 0 \end{aligned} \tag{2.12}$$

Where, ε_i defines the margin gap between classes, w^T defines the normal vector to the hyperplane, x_i defines i_{th} feature vector, y_i defines the class label and b defines a constant.

The linear soft margin SVM defines a hyperplane to separate classes in contrast to linear hard margin SVM which defines two hyperplanes to separate classes.

The visual representation of soft and hard margin hyperplanes is shown in Figure 2.3.

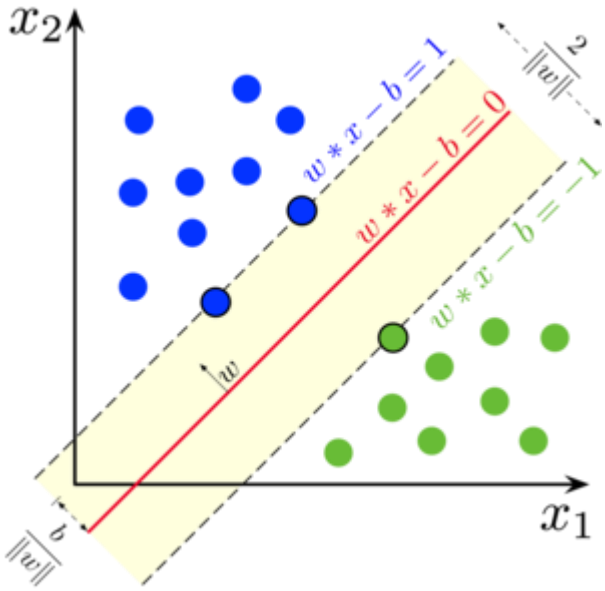

Figure 2.3 : Linear hard and soft margin SVM [23]

Blue and yellow hyperplanes define a hard margin SVM classifier with a $\frac{2}{\|w\|}$ margin. Moreover, the red hyperplane identifies a soft margin SVM classifier. The noise tolerance of soft margin SVM is higher than hard margin SVM.

2.2.1.2 Nonlinear SVM

In some cases, linear SVM is not sufficient to separate categories. In this case, the kernel trick is applied to data to distinguish non-linearly separable classes. The kernel trick transforms the feature space in a way that a linear hyperplane can separate the categories. The visual representation of the kernel trick is given in Figure 2.4.

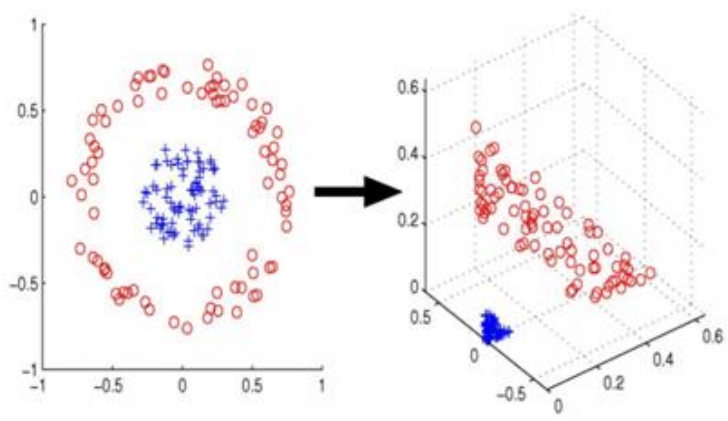

Figure 2.4 : Nonlinear SVM [24]

2.2.1.3 Loss function

To find the hyperplanes that maximize the margin $\frac{2}{\|w\|}$, $\frac{\|w\|}{2}$ must be minimized. "w" and "b" values that maximize the margin are calculated. Calculated "w" and "b" values define maximum margin hyperplane for classification.

2.2.2 K-Nearest neighbors (k-NN)

The k-nearest neighbors algorithm is a simple and efficient supervised machine learning algorithm that is used for classification and regression. All training features $X_1, X_2 \dots \dots X_n$ where $X_{(i)} \in \mathbb{R}^m$ and the testing feature $X \in \mathbb{R}^m$ should be given to the k-NN classifier. Then, k-NN classifier calculates the norm $n_i = \|X_i - X\|$ where $\|\cdot\|$ describes a norm on $\mathbb{R}^m$. After that, the classifier sorts norms by ascending order and the smallest "k" norms $n_1, n_2, \dots, n_k$ are selected.

The test signal is labeled with the class label which is more in number among the smallest "k" norms [25]. It is suggested that to choose "k" as an odd number.

In addition, classification time increases proportionally with the training dataset size because the distances between the test sample and every training sample are calculated.

In Figure 2.5, the purple circle is desired to be classified using 3-nearest neighbors algorithm. Since two green triangles and one red rectangle obtained as 3-nearest neighbors of the purple circle, the circle is classified as a green triangle.

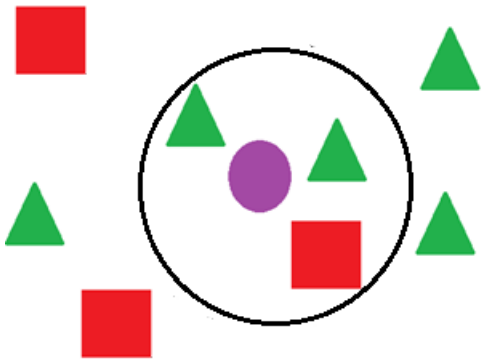

Figure 2.5 : k-NN

3. GUNSHOT SOUND CHARACTERISTICS

Gunshot sound signals consist of various acoustic components including muzzle blast and ballistic shockwave. The muzzle blast comes up when heated and rapidly expanding gases create an acoustic blast. Moreover, if the bullet exceeds the speed of sound, a supersonic explosion occurs. As a result of the explosion, a ballistic shockwave is formed [26]. An example of a gunshot sound signal is given in Figure 3.1.

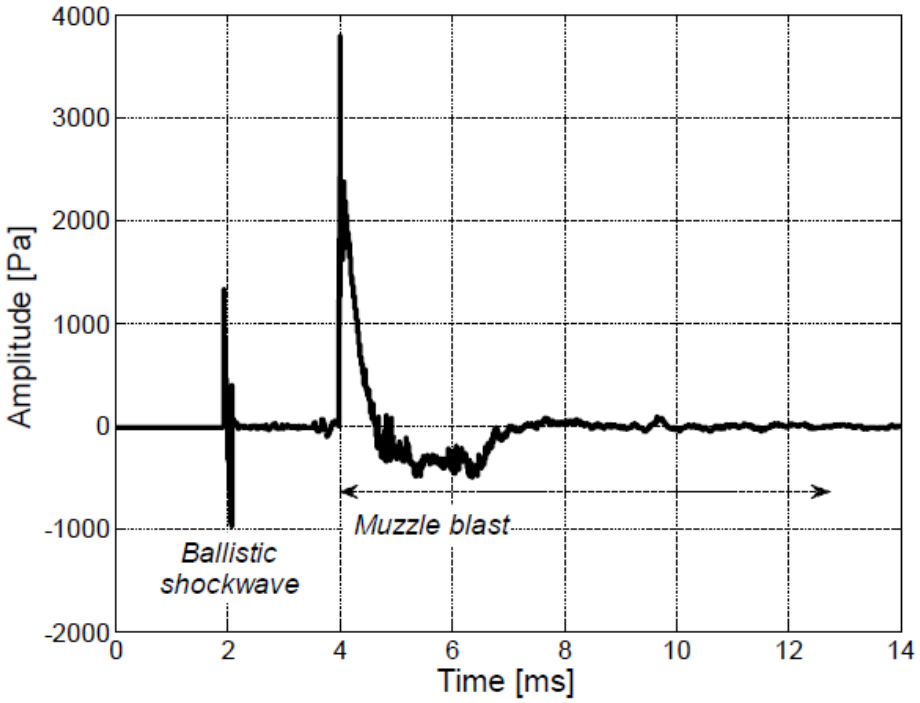

Figure 3.1 : Gunshot sound signal characteristics [26]

3.1 Database

104 testing and 389 training sound signals were collected from online sources to create dataset. Gunshot sound signals were taken from Still North Media firearm sound effects library [27]. Because, this library is the only free gunshot library which includes noise-free gunshot sounds . In addition, wind, footsteps, human voice, birds singing

and dog barks sounds were collected from different online free sound sources. While the used classification sound signals samples are specified in Table 3.1, testing sound signals samples are given in Table 3.2.

Table 3.1 : Classification samples

Classification samples	
Sound	**Number of the samples**
Gunshot	236
Dog bark	23
Bird singing	36
Footsteps	18
Human voice	28
Metal	28
Environmental Noise	8
Wind	12
	389

Table 3.2 : Testing samples

Testing samples	
Sound	**Number of the samples**
Gunshot	73
Barking	3
Bird singing	4
Footsteps	4
Human voice	4
Metal	5
Environmental Noise	8
Wind	3
	104

3.2 Simulations

This section contains the simulation steps and results. The proposed system was simulated using MATLAB to obtain accuracy rates and check whether the system is appropriate for hardware implementation. A gunshot sound sample of "Ruger single six" type was selected, which is shown in Figure 3.2, from the testing dataset to demonstrate simulation steps. The explanation of the simulation steps is given below.

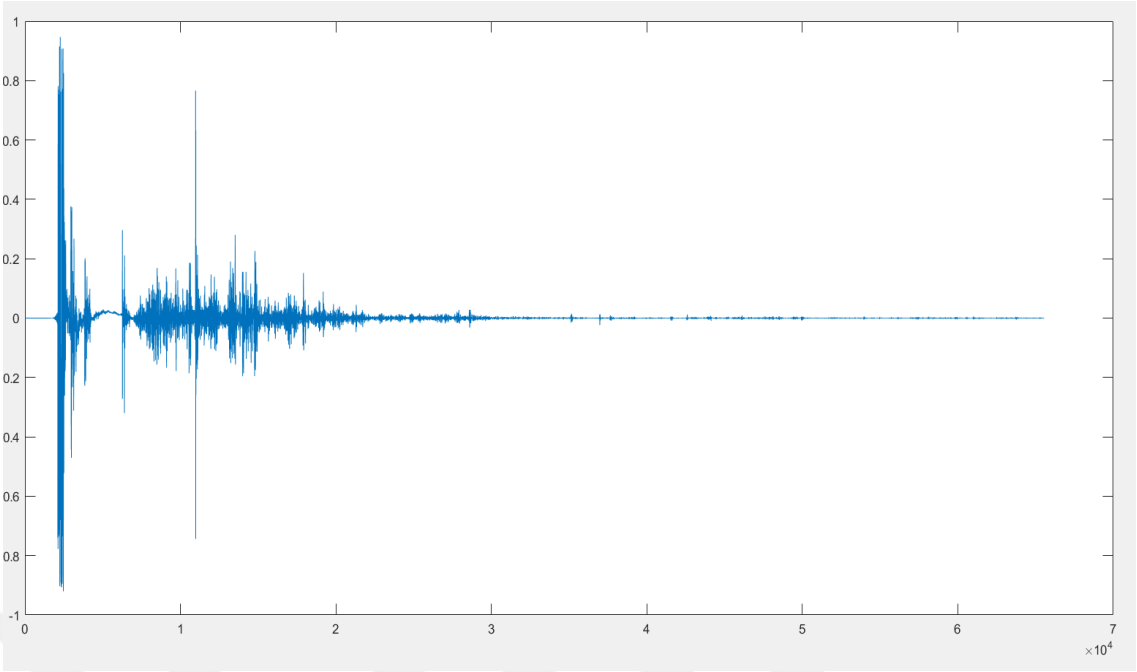

Figure 3.2 : Gunshot sound signal

3.2.1 Energy

The first feature extracted from the sound signal is the sum of the signal energy. As shown in Figure 3.3, the square of the sound signal is calculated to obtain the energy of it. As shown in Figure 3.3, the calculated square of the sound signal is used to obtain the energy of it.

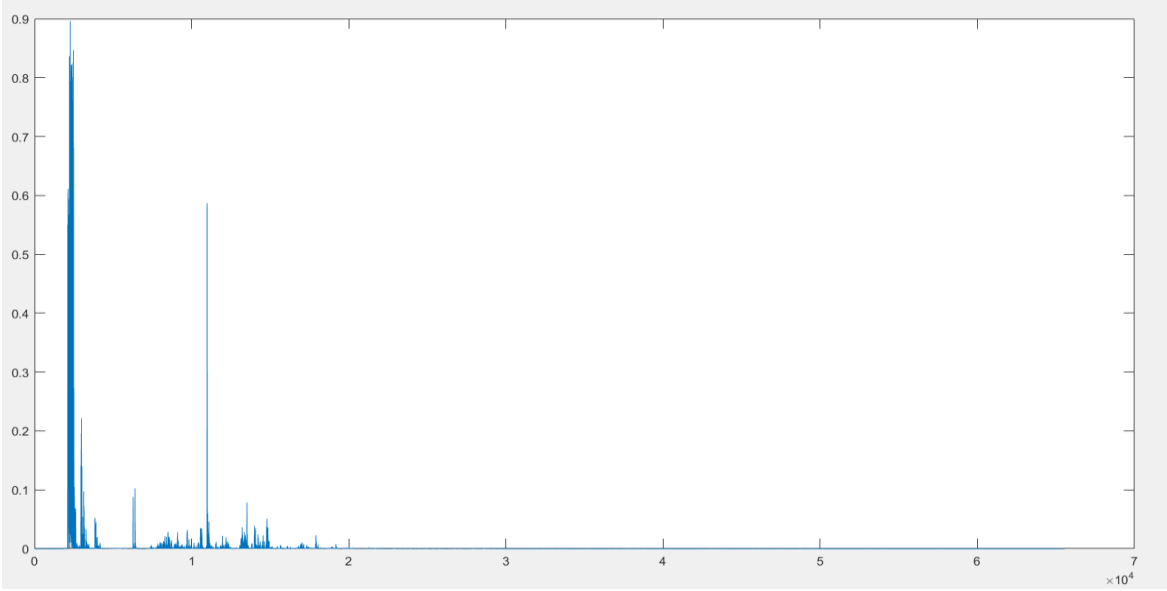

Figure 3.3 : Energy of the sound signal

3.2.2 Zero-crossing rate

The second feature is the zero-crossing rate of the signal. There are three levels of zero transitions which are 0,1 and 2 according to Eqn. 2.2. As can be seen in Figure 3.4, the zero-crossing rate is lower where the signal has a higher amplitude.

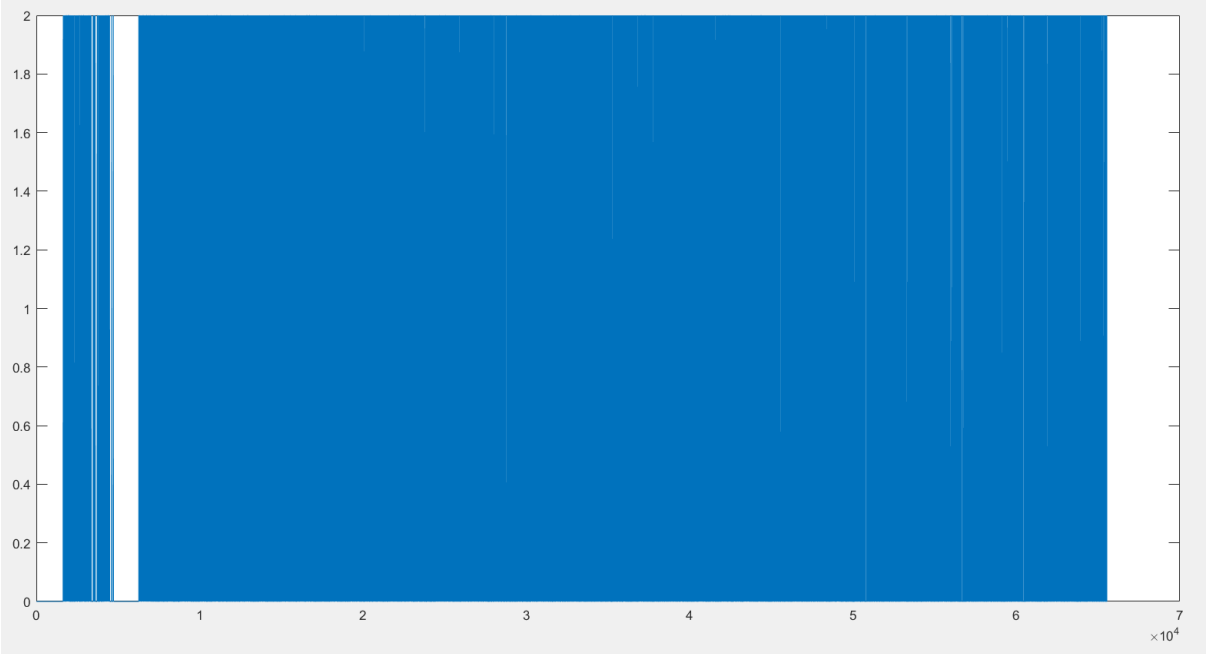

Figure 3.4 : Zero cross of the sound signal

3.2.3 MFCC

Generally, 13 MFCCs are preferred by researchers in order to optimally adjust the trade-off between accuracy rate and implementation complexity [28]. In this research, 12 MFCCs are obtained from the sound signal to reduce computation complexity. The extraction steps of the MFCCs are given below.

3.2.3.1 Fast fourier transform

MFCCs are obtained by transforming frequency domain representation of the signal into Mel scale representation. In this study, sound signals are sampled using 48000-hertz sampling frequency and 65536-point sound signals were chosen to get the full characteristic of the gunshot sound signal. Since Xilinx FFT module creates 65536-points output signal for 65536-points input signal, 65536-points frequency domain representation is used to obtain MFCCs.

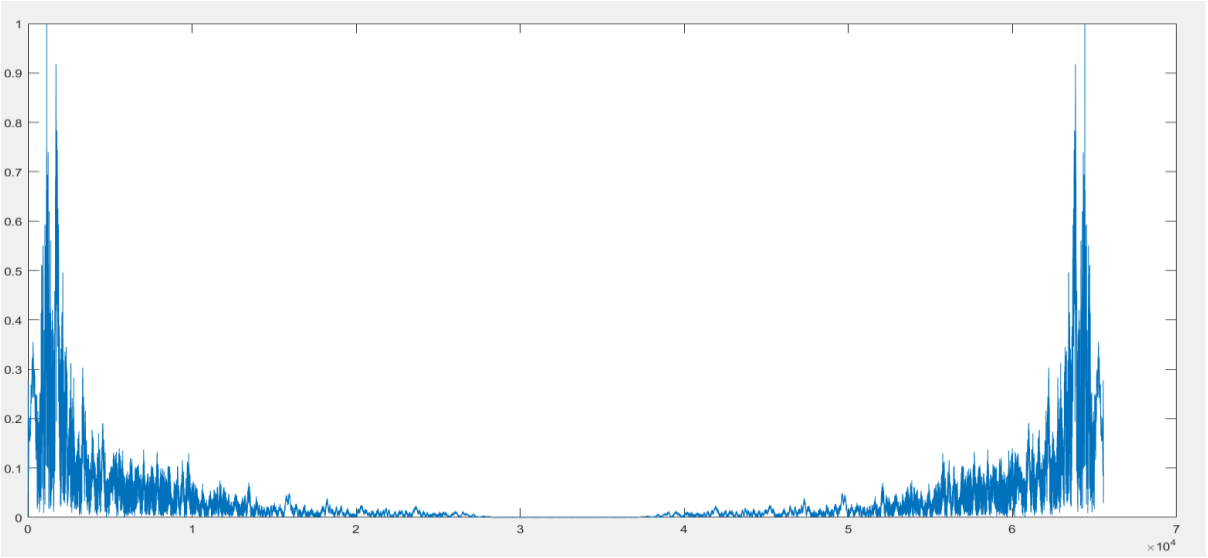

Figure 3.5 : Frequency representation of the gunshot sound signal

As can be seen in Figure 3.5, the testing gunshot sound signal has a higher amplitude in lower frequencies. Hence, lower frequency bands are more important to separate the gunshot sound signals from the other sound signals.

The first 128-points of the frequency representation of the signal are utilized to obtain Mel scale representation, since transforming 65536-point frequency domain representation into the Mel scale requires too many hardware resources.

The first 128-points of the FFT is shown in Figure 3.6.

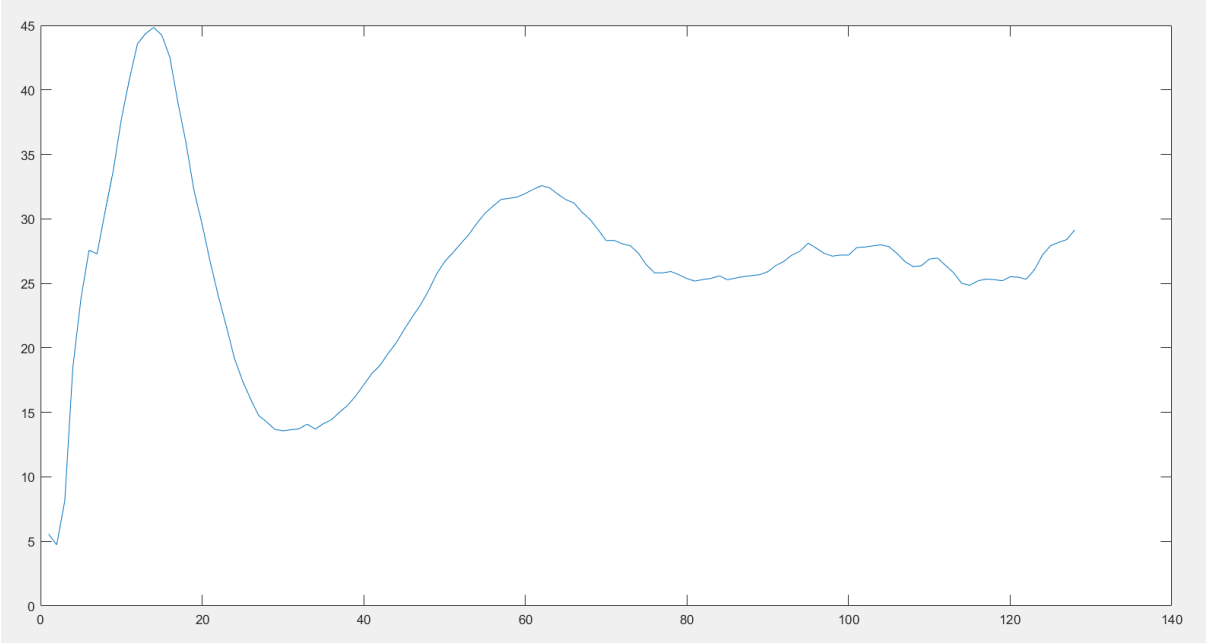

Figure 3.6 : First 128 samples of fourier transfrom of gunshot sound

3.2.3.2 Filtering

The frequency response is filtered with 48 overlapping triangular windows that are created using the Eqn. 2.4. Since 48-points provides a sufficient representation of the signal and the calculation complexity is lower than the systems that implement more filters, 48 filters were chosen for implementation. The designed filters are shown in Figure 3.7.

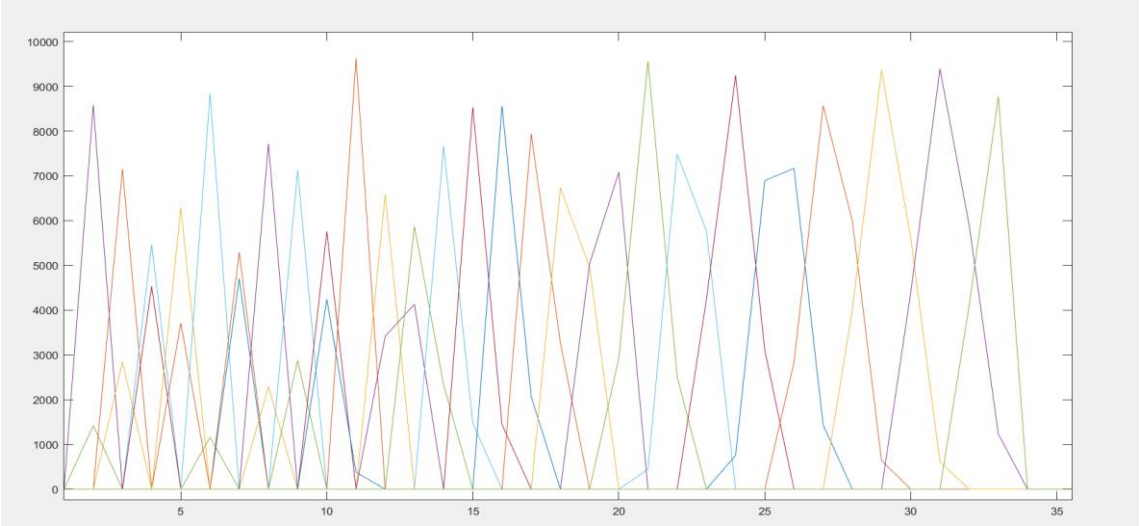

Figure 3.7 : 48 triangular filters

As mentioned before, the MFCCs mimics the human ear sound perception behavior. Since the human ear better distinguishes frequency changes at lower frequencies, the filters become narrower at lower frequencies. The 48-point Mel scale representation shown in Figure 3.8 is obtained by summing up the filter outputs.

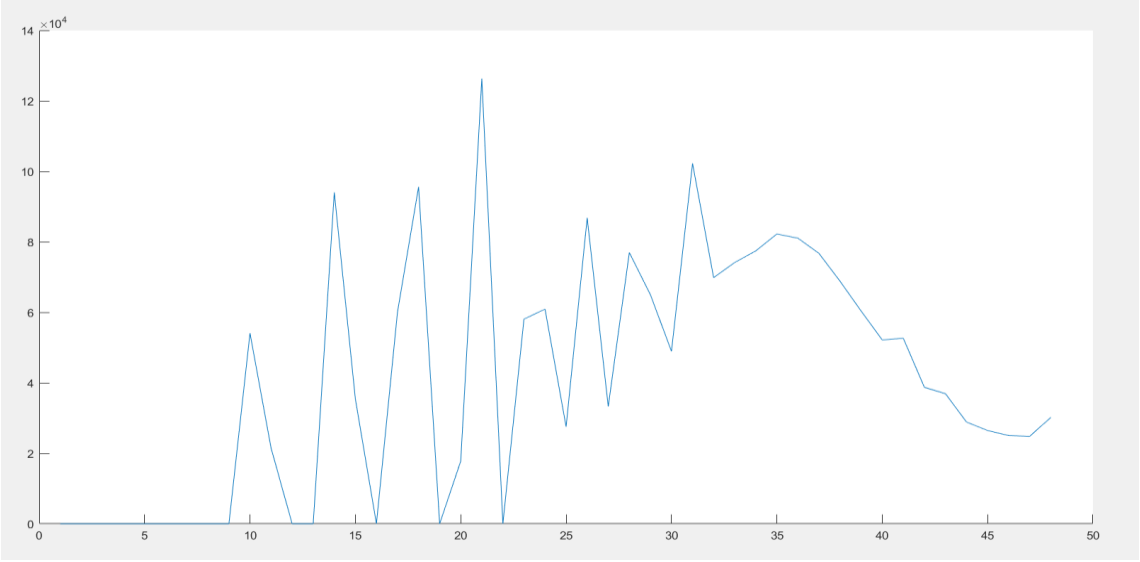

Figure 3.8 : Filtered frequency representation

3.2.3.3 Logarithm

The logarithm of the filtered data is taken using the Eqn. 2.7. The logarithm output is shown in Figure 3.9.

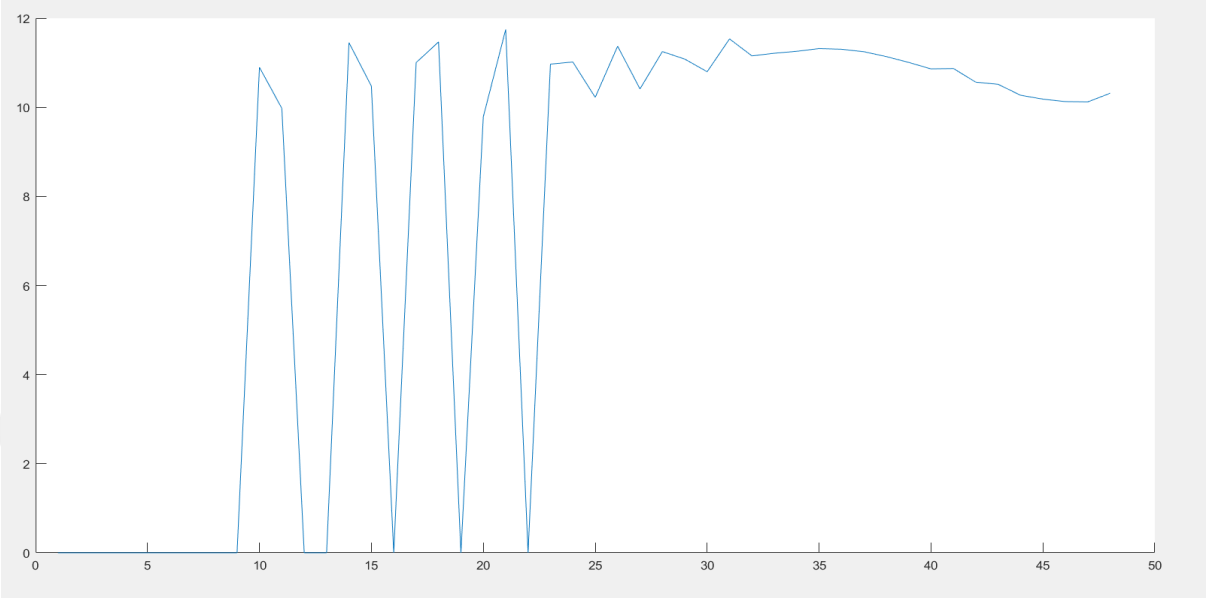

Figure 3.9 : Logarithm of filter output

Since the human ear hears the loudness level in a logarithmic scale, the logarithm of the filtered frequency representation is taken.

3.2.3.4 Discrete cosine transform

The discrete cosine transform of the logarithm output is taken using Eqn. 2.8. The discrete cosine transform output for m=0 is shown in Figure 3.10.

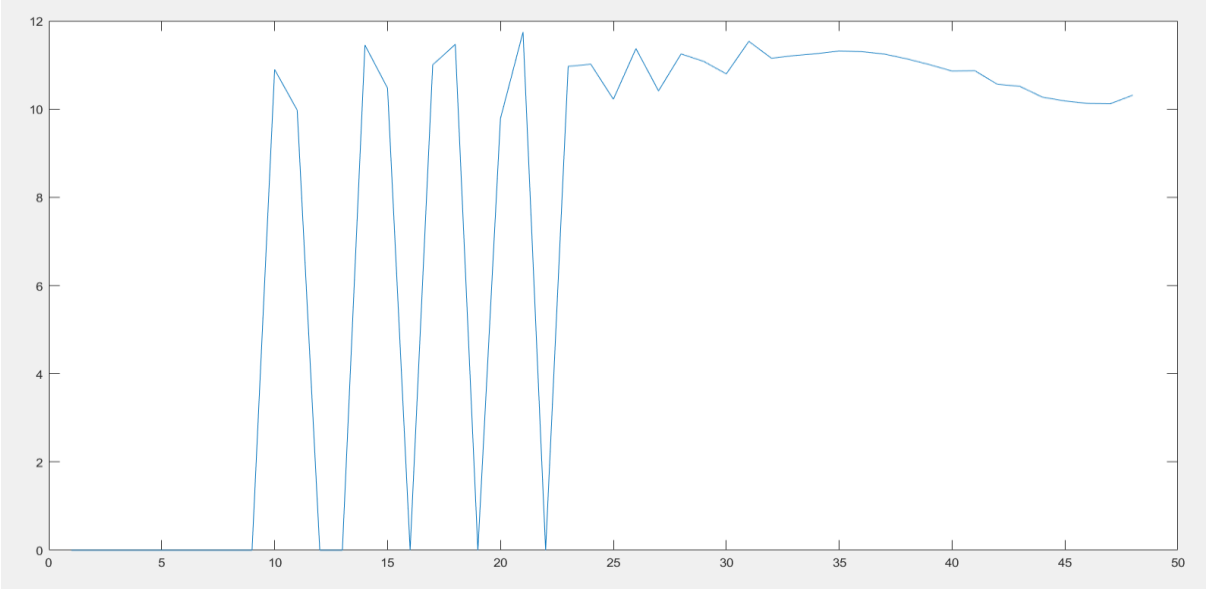

Figure 3.10 : Discrete cosine transform of the logarithm output

The features extracted from “Ruger single six” type testing sound sample are shown in Table 3.3.

Table 3.3 : Features obtained from simulation

Simulation Features	
Feature Name	**Feature Value**
Energy	0.0226
Zero-Crossing Rate	0.1107
MFCC 1	0.4714
MFCC 2	0.1001
MFCC 3	0.1796
MFCC 5	0.3991
MFCC 5	0.6789
MFCC 6	0.2897
MFCC 7	0.7657
MFCC 8	0.7513
MFCC 9	0.6301
MFCC 10	0.5090
MFCC 11	0.7446
MFCC 12	0.8361

3.3 Classification

3.3.1 Visualization of Training Samples

In this part of the thesis, training features are visualized. The 14-dimensional features space is visualized using seven 2-dimensional graphs. In the graphs, “gunshot” labeled sound samples and “other” labeled sounds are shown with the symbols below.

- ○ : gun sample
- × : other sample

The seven 2-dimensional graphs are explained below.

3.3.1.1 Energy and zero-crossing rate

Energy and zero-crossing rate features obtained from classification samples are shown in Figure 3.11. As can be seen in the figure, the energy sums of the "gunshot" labeled sounds are relatively higher than the "other" labeled sounds. Comparing to the zero-crossing rate, the energy sum of the sound provides better performance for separating classes.

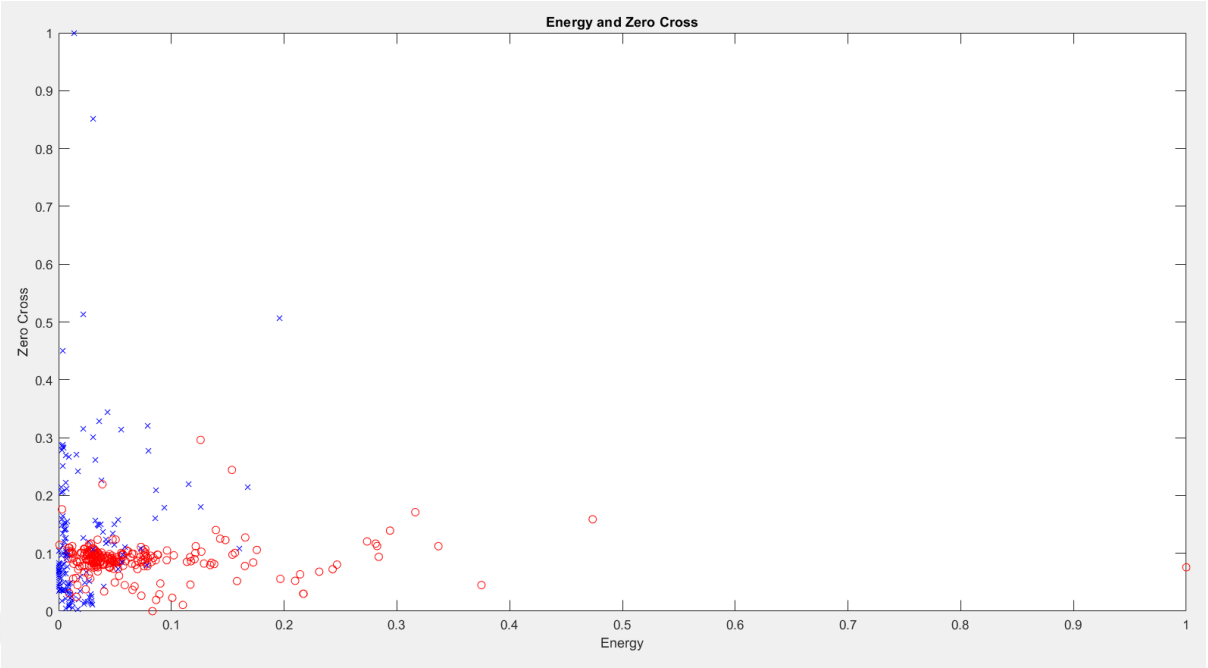

Figure 3.11 : Energy and zero cross features of classification dataset

3.3.1.2 MFCC1 & MFCC 2

As can be seen in Figure 3.12, not only MFCC 1 but also MFCC2 do not properly distinguish “gunshot” labeled sound from “other” labeled sounds. However, the use of MFCC1 and MFCC2 features together provides a better result for classification.

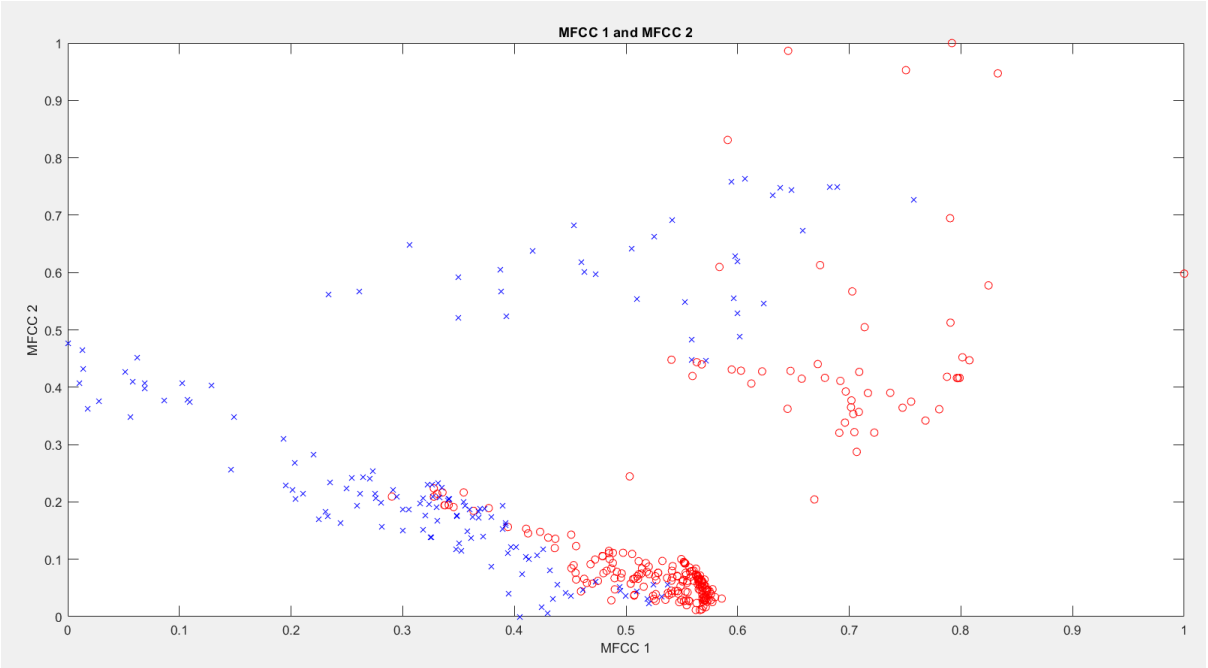

Figure 3.12 : MFCC 1 & MFCC 2 features of classification dataset

3.3.1.3 MFCC 3 & MFCC 4

As shown in Figure 3.13,MFCC 3 provides better performance in terms of separation classes by comparison with MFCC 4. However, the use of MFCC 3 & MFCC 4 coefficients together provides a better results in terms of accuracy.

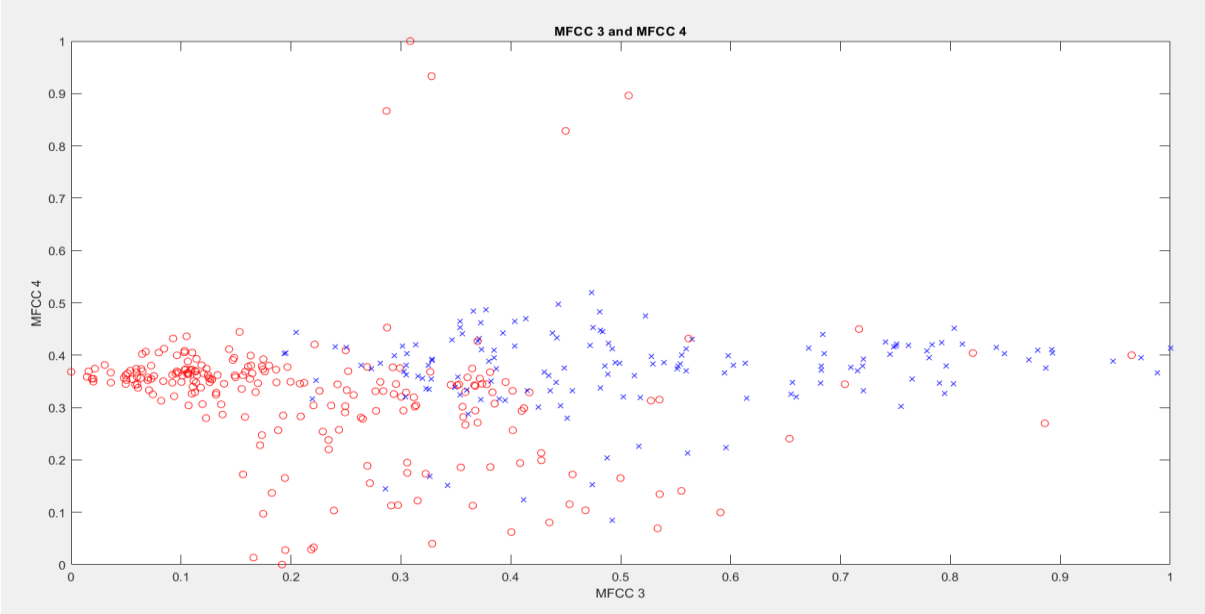

Figure 3.13 : MFCC 3 & MFCC 4 features of classification dataset

3.3.1.4 MFCC 5 & MFCC 6

As illustrated in Figure 3.14, MFCC 5, MFCC 6 and combination of them do not provide satisfactory results for classification. Since data distribution is nonlinear, a non-linear classification method can be used to increase the accuracy rate.

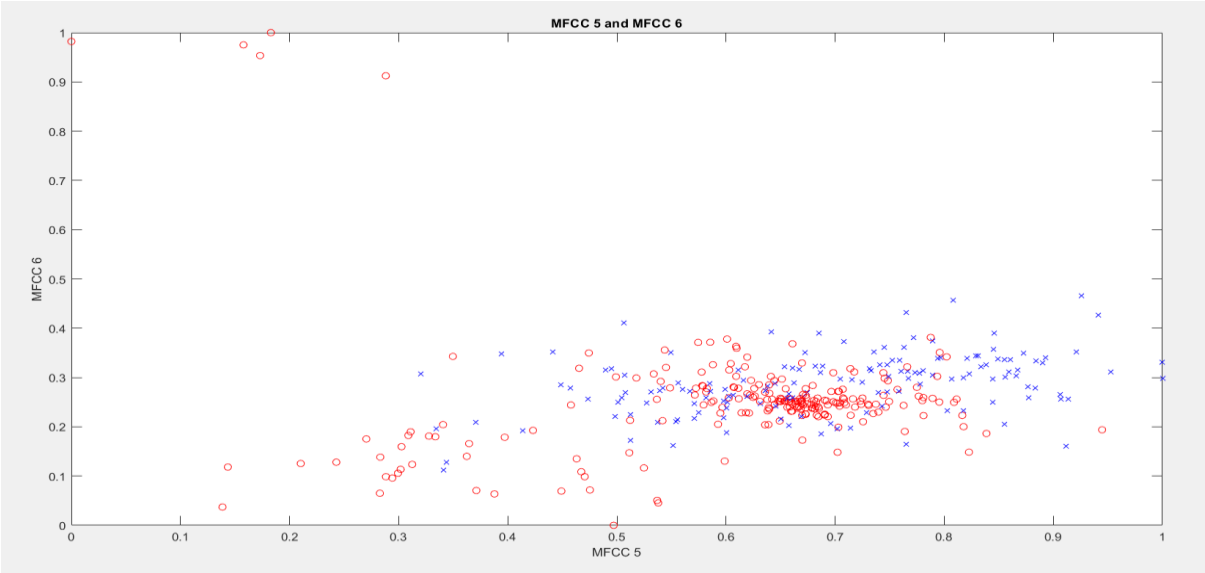

Figure 3.14 : MFCC 5 & MFCC 6 features of classification dataset

3.3.1.5 MFCC 7 & MFCC 8

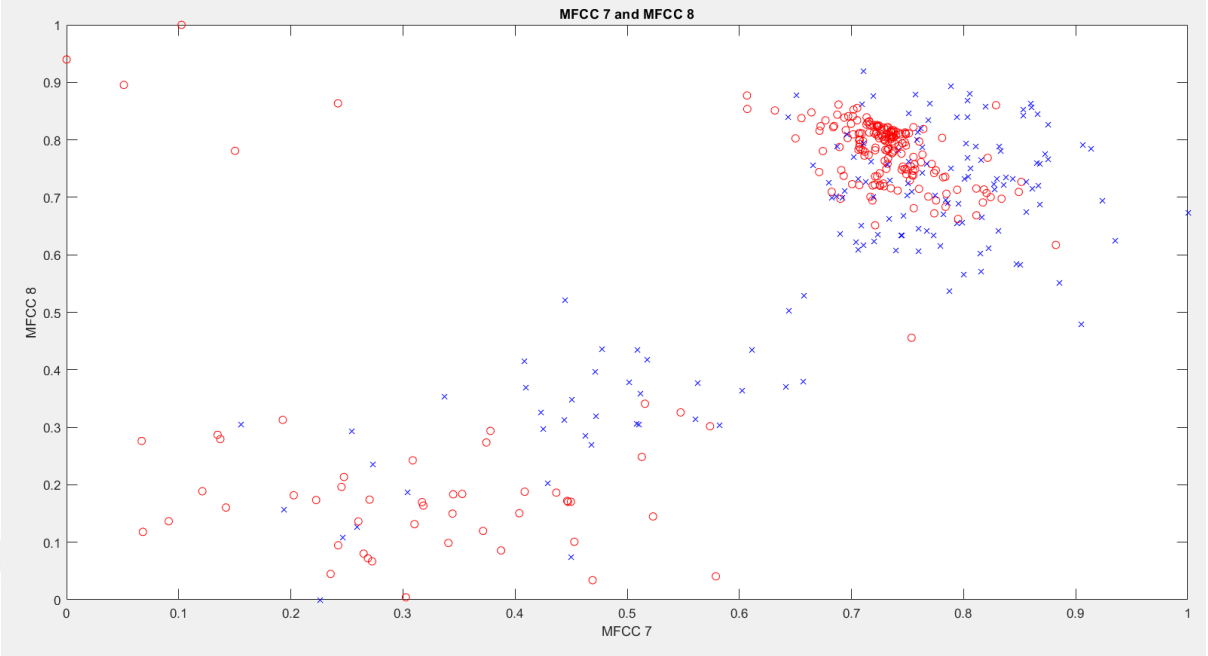

Figure 3.15 : MFCC 7 & MFCC 8 features of classification dataset

As can be seen in Figure 3.15, the "gunshot" labeled and "other" labeled samples are overlapped. However, a non-linear classifier can be used to separate classes.

3.3.1.6 MFCC 9 & MFCC 10

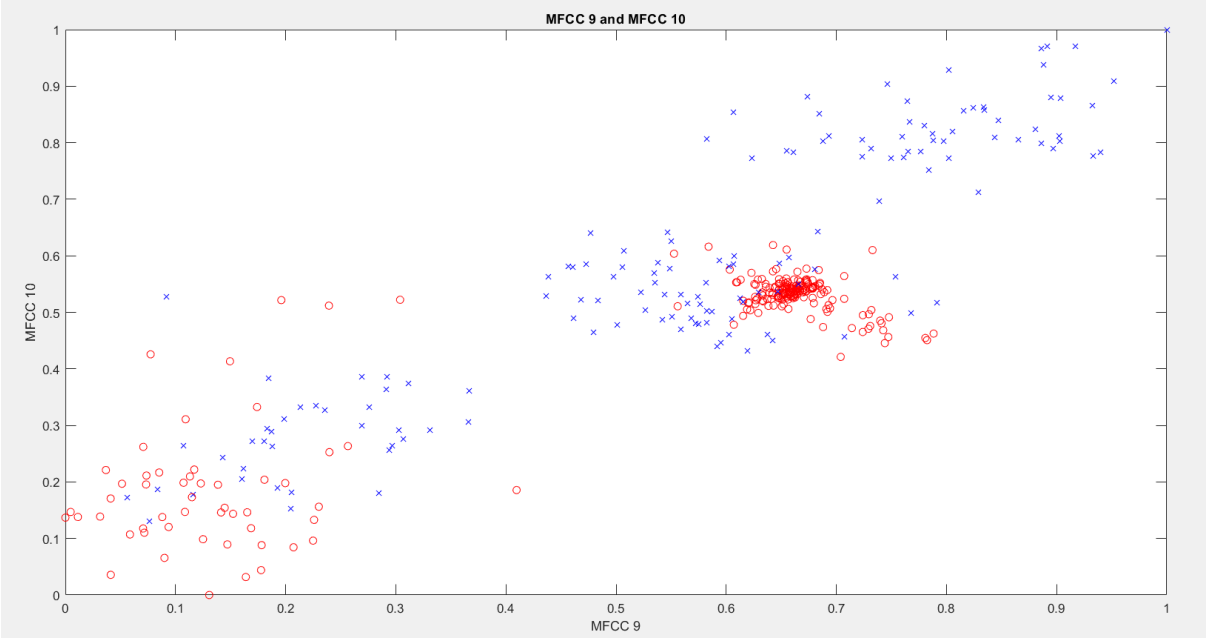

Figure 3.16 : MFCC 9 & MFCC 10 features of classification dataset

As can be seen in Figure 3.16, "gunshot" and "other" labeled samples can be classified using MFCC 9 and MFCC 10 features and a non-linear classifier.

3.3.1.7 MFCC 11 & MFCC 12

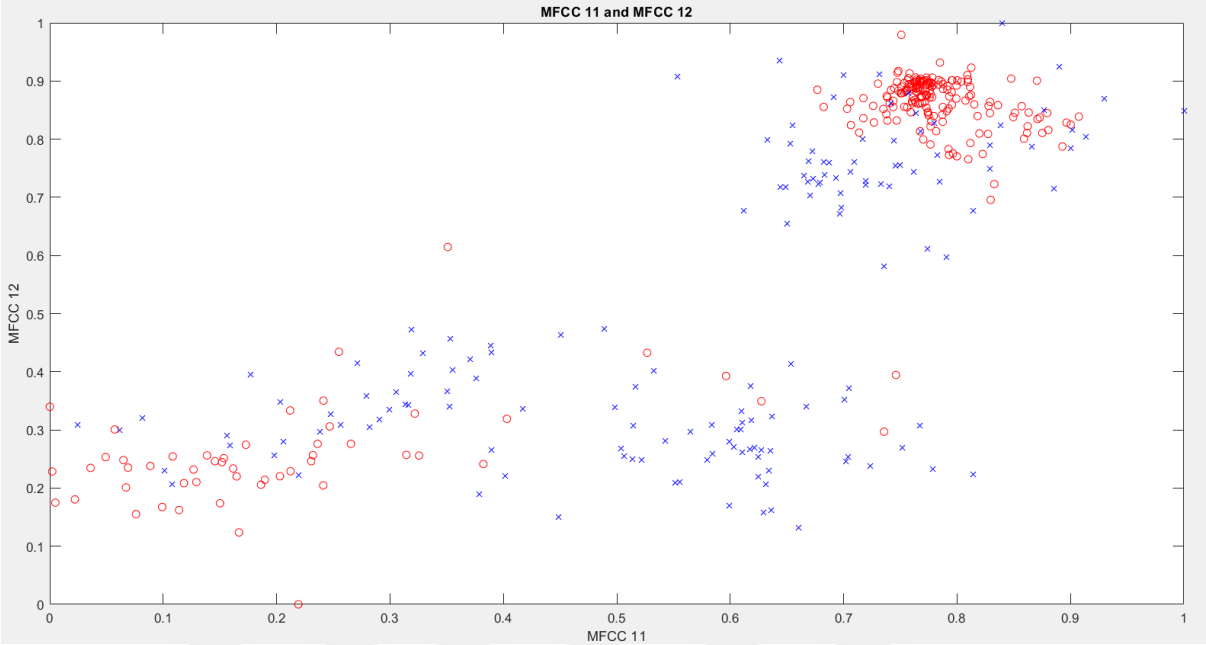

Figure 3.17 : MFCC 11 & MFCC 12 features of classification dataset

As indicated in Figure 3.17, "gunshot" and "other" labeled sounds cannot be separated using a linear classifier. However, a non-linear classifier can be utilized to classify "gunshot" and "other" labeled sounds using MFCC 11 and MFCC 12 features.

As shown in the figures above, the features used do not give satisfactory results by themselves to label gunshot sounds. However, using 14 features together to label gunshot sounds significantly improves accuracy.

3.4 Test Results

The achieved accuracy rates using k-NN and SVM classifiers were compared for choosing the most appropriate classifier for hardware implementation. The attained accuracy rates using noiseless and white-noise added test samples are shown below.

3.4.1 k-NN and SVM Comparison

k-NN and SVM are commonly used in machine learning applications. Besides, SVM classifier is preferred to identify gunshot sounds in a significant number of researches. Hence, achieved accuracy rates using k-NN and SVM classifiers were compared for selecting the appropriate classifier.

Table 3.4 : k-NN accuracy rates table

Guess	Truth		
		Gun	Other
	Gun	71	2
	Other	2	29
Accuracy	% 96.1538		

Table 3.5 : SVM accuracy rates table

Guess	Truth		
		Gun	Other
	Gun	68	4
	Other	5	27
Accuracy	% 91.3462		

As shown in Table 3.5 and Table 3.6, a higher accuracy rate was achieved using a k-NN classifier. Also, it is easier to realize a k-NN classifier on FPGA compared to an SVM classifier. For this reason, the k-NN classifier was chosen to implement the system.

3.4.2 Testing data with white noise:

To analyze the behavior of the proposed system against the noise, white noise was added to test samples and changes in accuracy rate were observed. The attained accuracy rates are given in following section.

3.4.2.1 100dbm SNR white noise

Table 3.6 : Accuracy rates with 100dbm white noise

Guess	Truth		
		Gun	Other
	Gun	72	2
	Other	1	29
Accuracy	% 97.1154		

If noise is added to the test samples, the accuracy rate is expected to decrease. Contrary to expectations, As indicated in Table 3.6, an increase in accuracy rate was observed in this case.To investigate this unexpected behavior, that specific "gunshot" labeled sample was observed. While classifying a noiseless sample, two "other" labeled sounds and one "gunshot" labeled sound were obtained as the 3 closest neighbors. However, there were more "gunshot" labeled samples in 10 nearest neighbors. The

addition of the white noise changed the sample in a way that 3 nearest neighbors consist of two "gunshot" labeled sounds and one "other" labeled sound.

3.4.2.2 80dbm SNR white noise

Table 3.7 : Accuracy rates with 80dbm white noise

	Truth		
Guess		Gun	Other
	Gun	65	2
	Other	8	29
Accuracy	% 90.3846		

As demonstrated in Table 3.7, adding 80 dBm white noise to testing samples were not decrease the accuracy rates significantly. However, the false-negative rate increased considerably.

3.4.2.3 50dbm SNR white noise

Table 3.8 : Accuracy results with 50dbm white noise

	Truth		
Guess		Gun	Other
	Gun	41	1
	Other	32	30
Accuracy	% 68.2692		

As can be seen in Table 3.8, the addition of 50 dBm noise reduced the accuracy rate and significantly increased the false negative rate.

Addition of the white noise decreases the accuracy rates since mostly frequency-dependent features are used in this study.

3.5 Comparison with the other works in the literature

In this part of the thesis, the presented studies are compared with the studies not performed on the hardware in the literature.

A scream and gunshot sound detection and localization system has been proposed by Valenzise G. et al.. The number of extracted features is high for the system implemented using limited resources. In addition, the %93 detection rate obtained is lower than our study. However, since different training and test sets were used in this study, it is not meaningful to compare with the detection rates in our study. In addition,

this work uses gunshot sound effects taken from the movies for the classification dataset [8]. But it is a known fact that they are artificial sounds.

In Ahmed T., Uppal M., and Muhammad A.'s work, the obtained accuracy rates using LPC, MFCC, and cross-correlation features were compared to reach the system that achieves the best performance. A %99.702 true positive rate was achieved using LPC and cross-correlation features together with a linear SVM classifier. However, %97.321 true positive rate was attained by using MFCCs and a linear SVM classifier [10]. Although the sound database used is shared in the article, the link is currently unavailable. Again, the comparison of true positive rates is not appropriate since the same database is not used in our work.

Besides, MFCC, spectral roll-off, zero-crossing rate and many other features have been obtained in "gun-type recognition from gunshot audio recordings". In addition, a complex dimensionality reduction system was implemented. A maximum accuracy rate of %89 has been achieved in this system [11]. Comparing to our research, this system obtains a larger number of features and realizes a complex dimensionality reduction technique. Hence, the proposed technique is unsuitable to implement on a limited resourced system . Accuracy rates were not compared to our work since they use their own records.

Arslan Y. and Güldoğan B. compared SVM and multilayer perceptron classifier performances using MFCC features. And %94.7 true positive rate was achieved by combining MFCCs and SVM [14]. Since they recorded gunshot sounds to create their own database, accuracy rates should not be compared with our work. In our work, k-NN classifier was chosen for the realization of the system because better results were observed using the k-NN classifier.

Hrabina M. and Sigmund M. proposed a microcontroller-based gunshot sound detection system. However, the true positive rate of the system which is %82 is lower than our work [3].

Shiekh A. A., Tahir M., and Uppal M. implemented blind system identification and deconvolution to reduce the reverberation effect. Features such as MFCCs, LPC, and cross-correlation were extracted from sound signals and given into the SVM classifier. As a result, a %99 true positive rate was achieved [15]. Nonetheless, the number of extracted features is large in number to realize a hardware-based system. Also,

accuracy rates were not compared since the recorded gunshot database is not reachable.

In addition, the system proposed by Grama L. and Rusu C. achieved the correct classification rate of 95.05% for gunshot sound. [16]. However, the classification technique used is very complex compared to k-NN in terms of its implementation as a hardware-based system.

In Khamparia A. et al.'s work, the accuracy rate of %85 was attained using CNNs [17]. However, the accuracy rate is lower comparing to our study. Besides, CNN is harder to implement on FPGA comparing to k-NN classifier.

Galangque C. M. J. and Guimaldo S. A.'s work provides a %99 accuracy rate. However, the system uses two artificial neural network classifiers that are very complex for hardware-based implementation. Furthermore, the %99 accuracy rate was achieved only for differentiating M16 guns from ambient noise and fireworks [18].

The researches mentioned in this section have not been implemented on hardware. The methods used were compared with our studies and it was mentioned whether the research is suitable for hardware implementation.

4. REALIZATION OF THE GUNSHOT SOUND DETECTION SYSTEM USING FPGA

Faster and more secure systems can be realized on FPGAs comparing to the microprocessors. Moreover, FPGAs provide less development time and fewer development expenses comparing to ASIC designs. Therefore, it is decided to implement the system using FPGA. In this section, the implementation steps are mentioned and the study is compared with other studies that are realized on hardware.

4.1 FPGA

Field programmable gate arrays are designed as re-configurable integrated circuits to update the design in the field. FPGAs are configured using hardware description languages (HDLs) which are also used to configure ASIC designs. Thus, FPGAs are used to test ASIC designs before realization. There are programmable logic blocks and interconnects in FPGAs. Using them, complex or simple logic functions can be accomplished.

4.2 System Description and Requirements

The diagram of the system is shown in Figure 4.1.

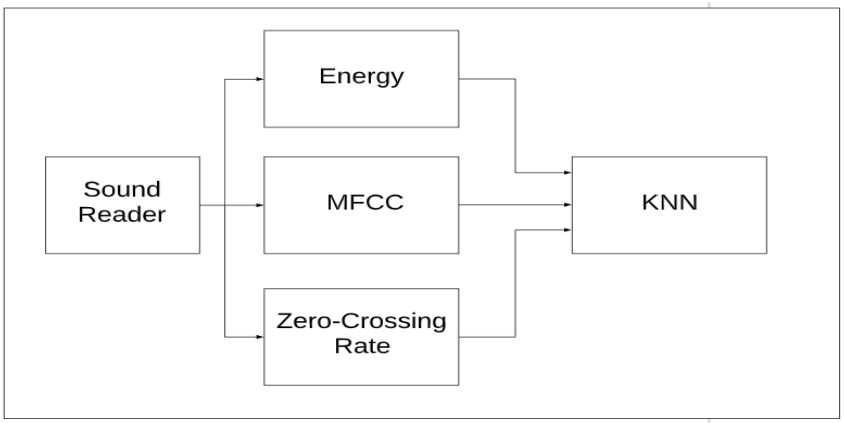

Figure 4.1 : System diagram

In this study, sound files were read from MATLAB Simulink. After that, a hardware implementation was made for extracting energy, zero-crossing rate and 12 MFCCs features.

4.3 Realization with System Generator

The system was implemented on FPGA to satisfy security, high speed and mobility needs. Matlab System Generator tool was used to realize the system on the Kintex Ultrascale+ KCU116 Evaluation Platform 1.0.

As demonstrated in Figure 4.2, the system was divided into 4 parts. Energy, zero-crossing, MFCC and k-NN submodules were created. Training was completed in MATLAB simulations and obtained training features were transferred into FPGA to classify sound signal input.

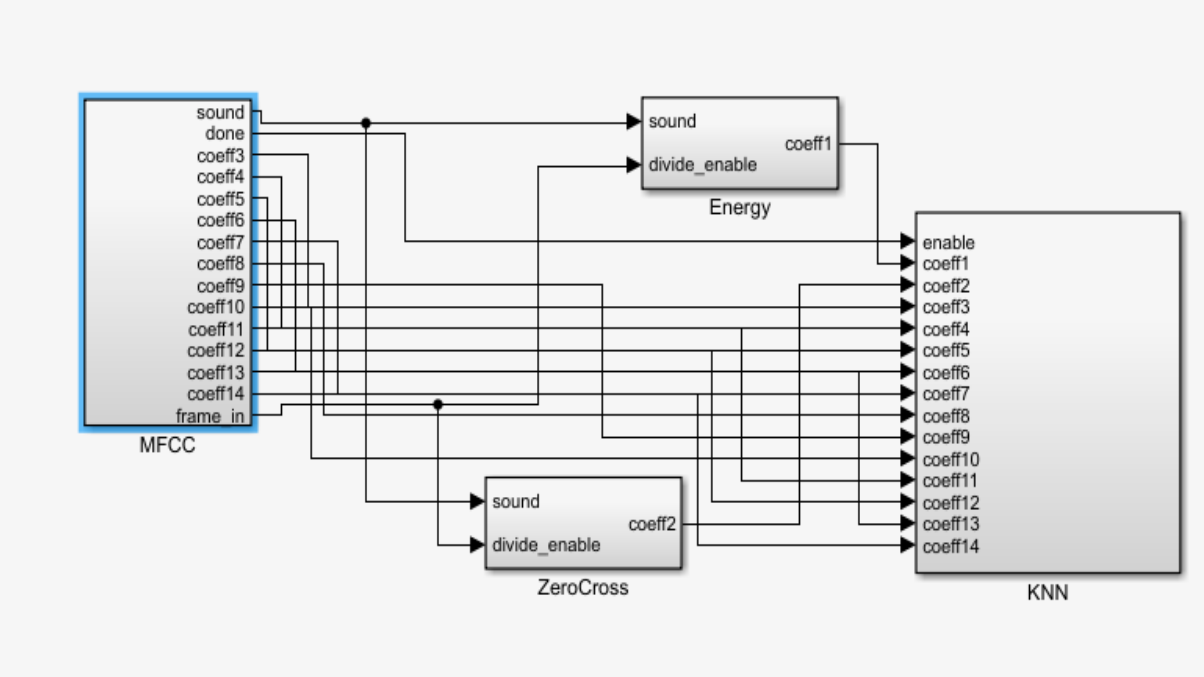

Figure 4.2 : System generator realization

4.3.1 Sound

First, sound signal is read using Matlab System Generator and given to FPGA.

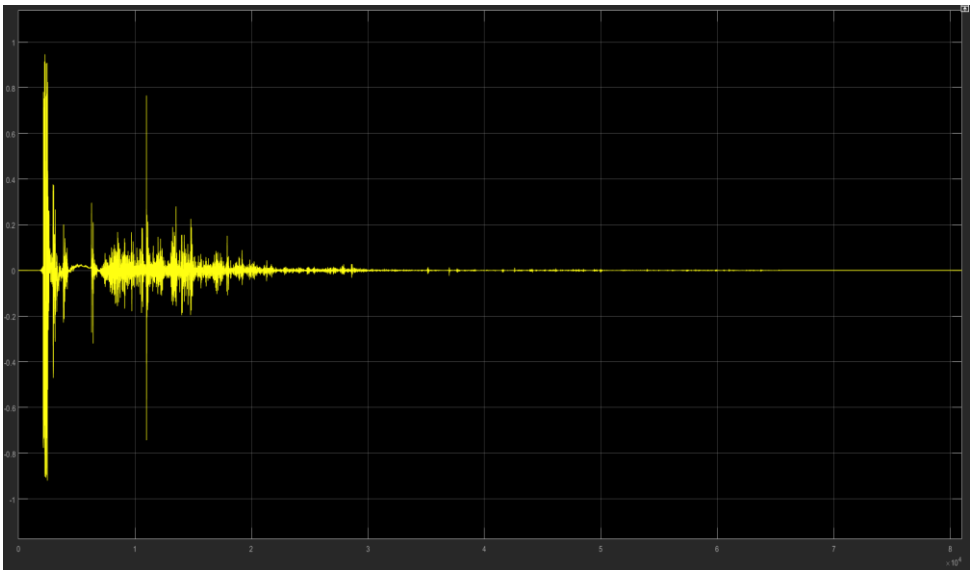

Figure 4.3 : Gunshot sound signal

The testing sample used in simulations was transmitted into the System Generator to realize the system. The testing sample is illustrated in Figure 4.3.

4.3.2 Energy

The sum of the signal energy is calculated using Eqn. 2.1.

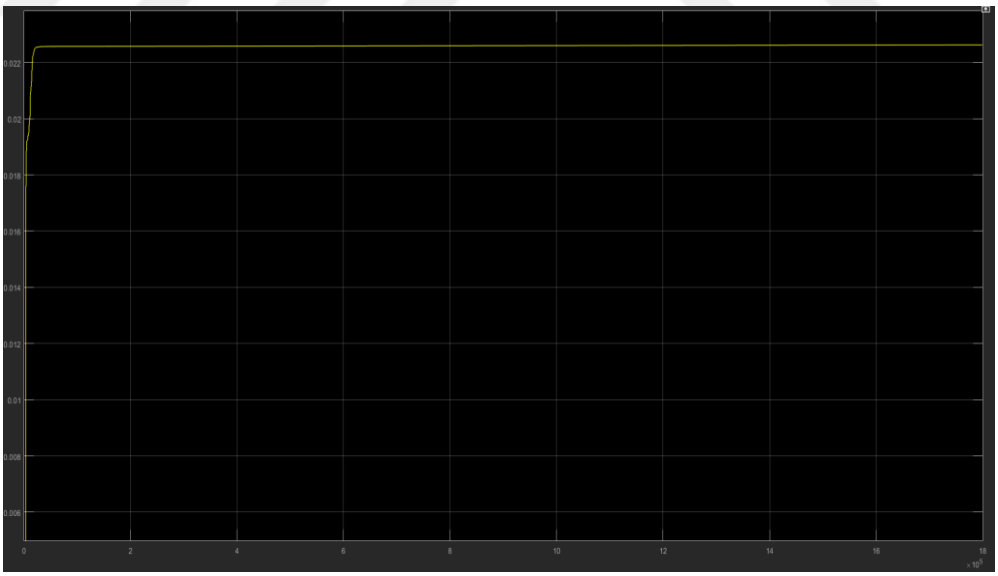

Figure 4.4 : Energy sum of the sound

As demonstrated in Figure 4.4, the energy sum of the sound signal is calculated in 65630 clock cycles of the system.

4.3.3 Zero-crossing rate

The zero-crossing rate of the sound signal is calculated using Eqn. 2.2.

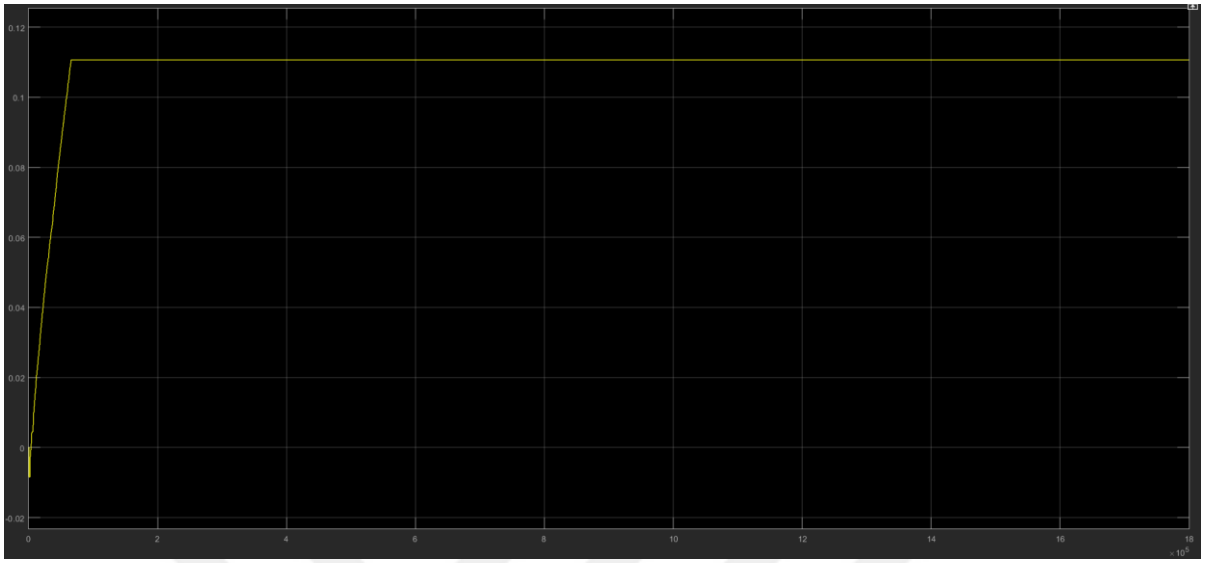

Figure 4.5 : Zero transitions sum of the sound

As can be seen in Figure 4.5, the zero-crossing rate of the signal is calculated in 65630 clock cycles of the system.

4.3.4 MFCC

In this part of the thesis, extracting steps of MFCC features are specified.

4.3.4.1 FFT

65536-point FFT of the sound signal is taken using Xilinx FFT Block. The magnitude of the obtained FFT result using FPGA is different compared to the attained FFT result in simulations. Hence, the FFT result values were scaled to have values between 0 and 1. The frequency domain representation of the signal is shown in Figure 4.6.

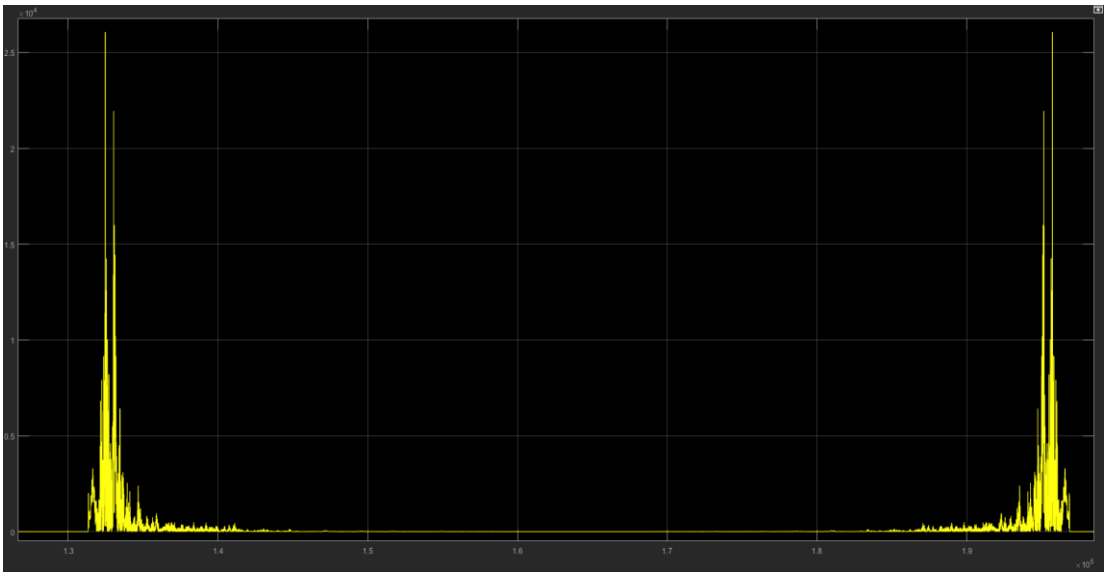

Figure 4.6 : Frequency domain representation of the sound

Only the first 128-points of the frequency representation of the signal are used to calculate MFCCs. The first 128 points of the FFT signal is shown in Figure 4.7.

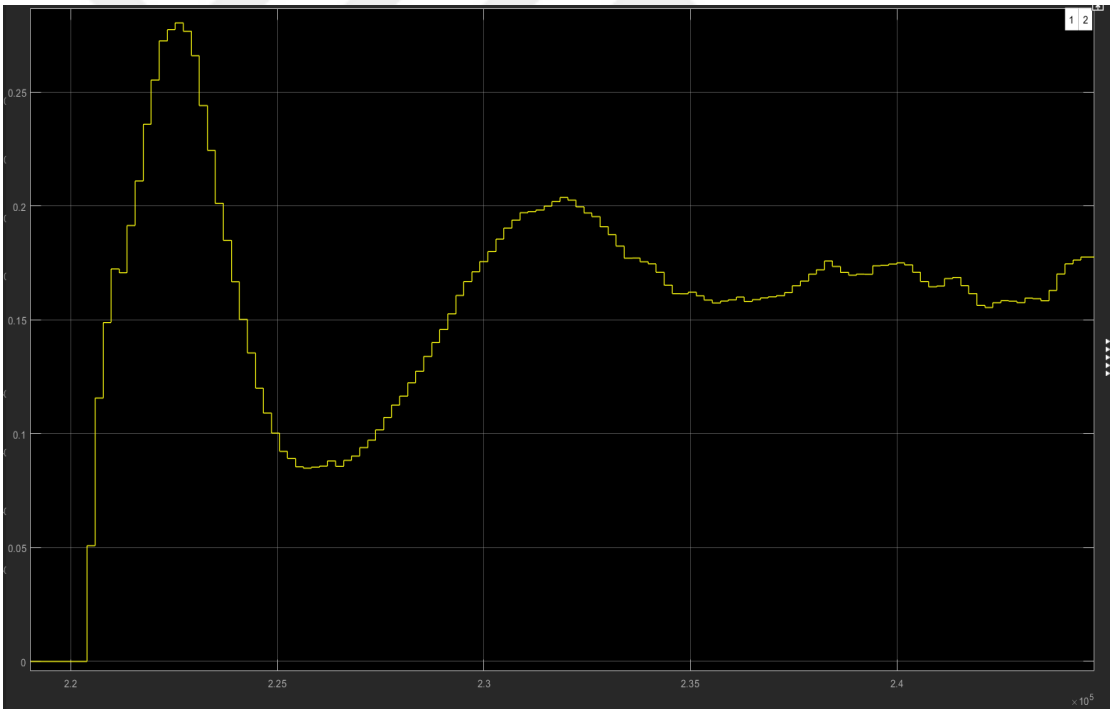

Figure 4.7 : First 128 points of FFT

As shown in Figure 4.8, the difference of the first 128 points of FFTs obtained in simulations and using FPGA is negligible.

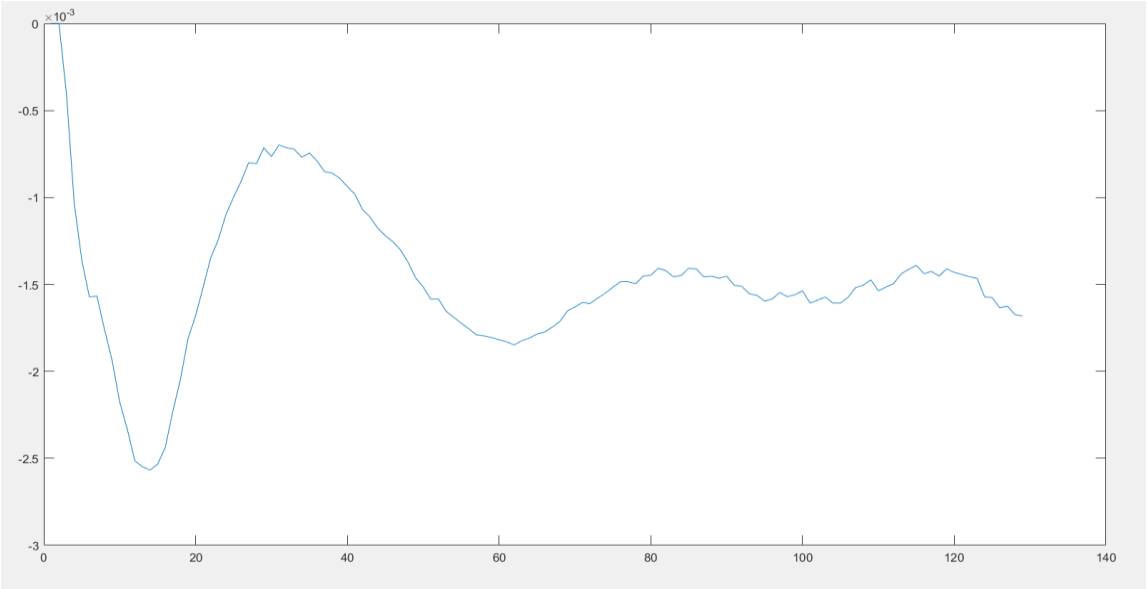

Figure 4.8 : The difference of FFTs

4.3.4.2 Filtering

Frequency representation of the signal is filtered using 48 triangular filters and 48-points Mel spectrum is obtained as demonstrated in Figure 4.9.

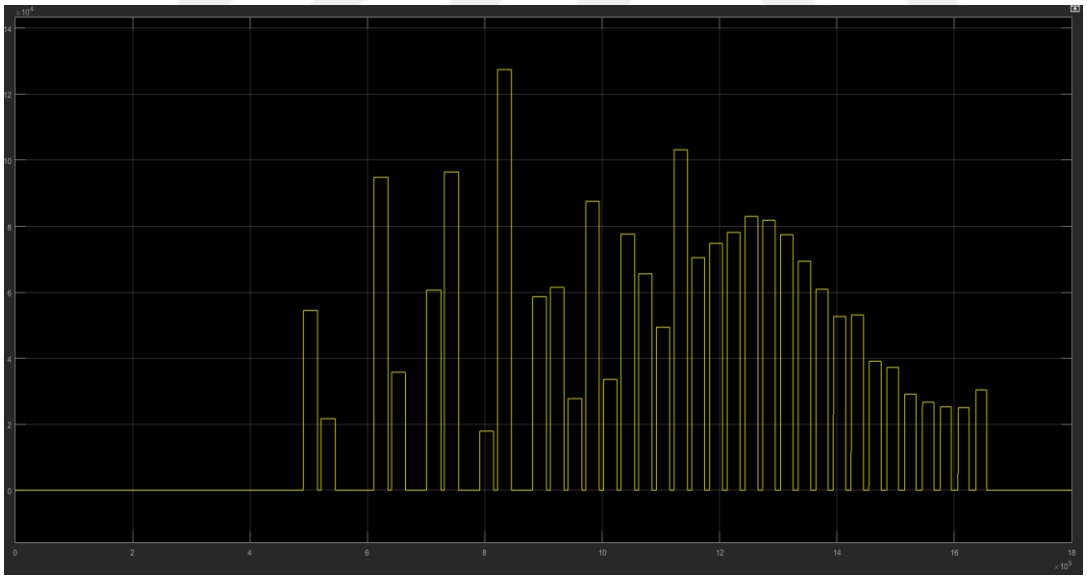

Figure 4.9 : Filter output

4.3.4.3 Logarithm

Logarithm of filtered data is taken using Eqn. 2.7. Besides, the logarithm is shown in Figure 4.10.

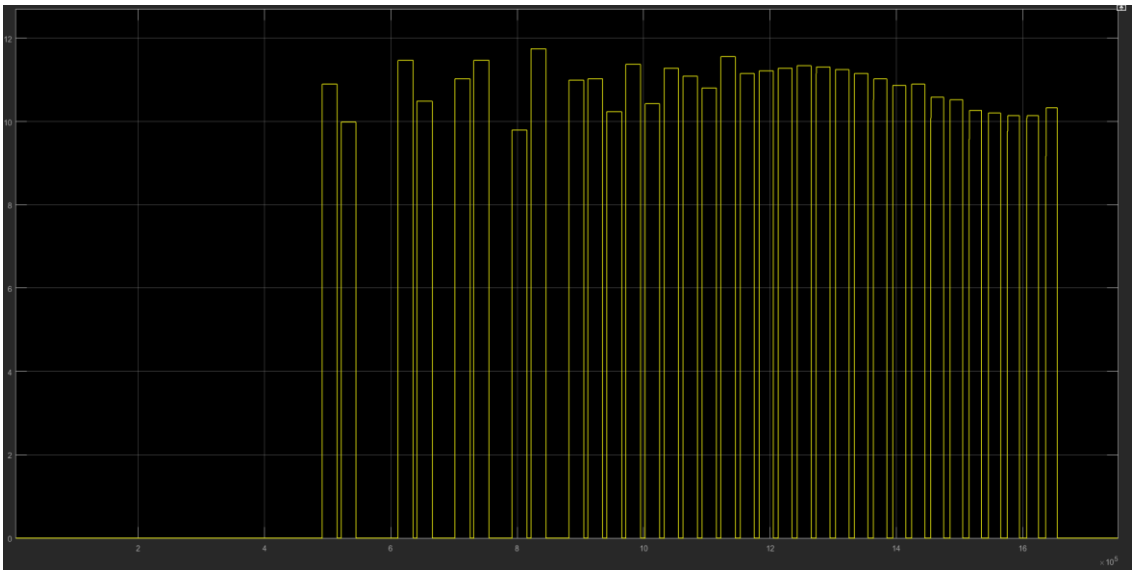

Figure 4.10 : Logarithm of filter output

4.3.4.4 Discrete cosine transform

The discrete cosine transform of the logarithm output is taken using Eqn. 2.8. Furthermore, the sum of the discrete cosine transform output is shown in Figure 4.11.

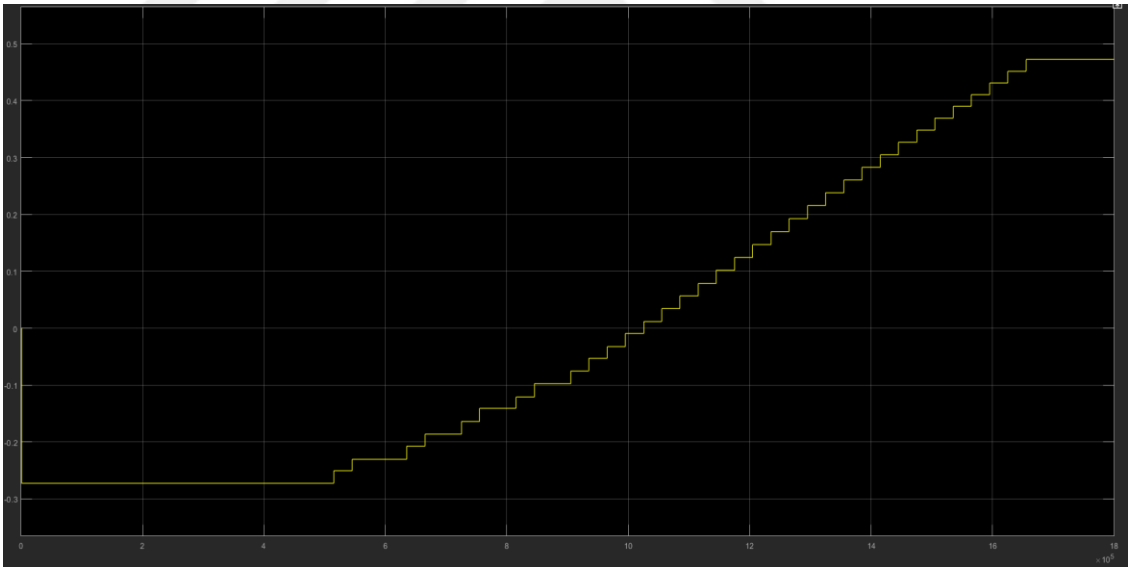

Figure 4.11 : Sum of discrete cosine transform

The obtained features is shown in Table 4.1.

Table 4.1 : Features obtained from FPGA realization

Features	
Feature Name	**Feature Value**
Energy	0.02262
Zero Cross	0.1107
MFCC 1	0.4722
MFCC 2	0.09931
MFCC 3	0.1791
MFCC 5	0.3998
MFCC 5	0.6788
MFCC 6	0.2891
MFCC 7	0.7658
MFCC 8	0.7518
MFCC 9	0.6304
MFCC 10	0.5082
MFCC 11	0.7440
MFCC 12	0.8355

The difference between the features obtained on FPGA and the features obtained with simulations results from precision differences. The difference of features resulting from presicion differences are shown in Table 4.2.

Table 4.2 : The difference of features obtained from FPGA and simulation

Features	
Feature Name	**Feature Value**
Energy	0.00002
Zero Cross	0
MFCC 1	0.0008
MFCC 2	0.00079
MFCC 3	0.0005
MFCC 5	0.0007
MFCC 5	0.0001
MFCC 6	0.0006
MFCC 7	0.0001
MFCC 8	0.0005
MFCC 9	0.0003
MFCC 10	0.0008
MFCC 11	0.0006
MFCC 12	0.0006

4.3.5 Classification

Obtained features are trasmitted into k-NN classifier and the sound signals are classified.

As shown in Figure 4.12, for this spesific sample, the testing sample was classified as "gunshot" labeled sound which is represented in FPGA as “1”.

Figure 4.12 : Classification result

4.4 Resource Utilization

The used resources are given in Table 4.3 for every module of the designed system. As can be seen from Table 4.3, the MFCC module consumes more resources compared to the other modules. Moreover, the module consumes more DSP resources since the k-NN module contains intense sum and multiplication operations.

Table 4.3 : FPGA resource utilization

Module Name	BRAMs	DSPs	LUTs	Registers
ZeroCross	0	0	6635	18497
MFCC	258	199	167882	363596
k-NN	0,5	192	2419	2411
Energy	0	4	6644	17180
Total	258,5	395	183580	401684

As can be seen from Table 4.4, extracting MFCCs is the most time-consuming part of the system. A parallel processing architecture was implemented to extract MFCCs,

zero-crossing rate, and energy features. After extracting the features, the features are transmitted into the k-NN classifier to determine the class of the testing sample.

Table 4.4 : FPGA timing table

Module Name	Timing (clock cycle)
ZeroCross	65630
MFCC	1660195
k-NN	1661
Energy	65630
Total	1661856

4.5 Comparison with the Other Systems in the Literature

In this section, our study is compared with the researches that are realized on hardware.

In Tobin J. D's work, machine learning methodologies were not used. Only cross-correlation and a threshold mechanism were implemented on FPGA to detect gunshot sounds. This project uses the same dataset with the study that we proposed. However, accuracy rates vary from %40 to %93 according to the gun type[12]. Besides, the threshold mechanism is not convenient for realization, since the magnitude of the gunshot sound changes with distance.

Salazar-García C. et al. implemented quadrature mirror third order IIR cauer filters and hidden markov model to classify sound signals. The system was realized on Xilinx Artix XC7A100T-CSG324 [13]. However, accuracy rates and used database were not specified in the paper. MFCCs, energy and zero-crossing rate features which are commonly used in machine learning applications were preferred for implementation in our work. Furthermore, k-NN classifier which is better than the hidden markov model for realization using limited resources was chosen.

Volgyesi P et al. used the zero-crossing rate and sound signal edge detection techniques to detect muzzle blast and ballistic shockwave. Machine learning methods were not implemented to detect gunshot sounds. However, the proposed system's

accuracy rates change from %23 to % 100 for different types of guns. A %73 average success rate was obtained which is lower than our study. Hardware IPs and 32 bit microcontroller were used together to design the system. [9]. But, only hardware IPs were implemented in our research.

As mentioned above, hardware-based gunshot detection systems are limited. Furthermore, in the majority of the researches, simple methods such as edge detection and threshold mechanism are applied. It is observed that there are not any papers about realizing MFCCs and k-NN on hardware to detect gunshot sounds in the literature. Besides, the accuracy rate of the this new design is relatively higher comparing the other hardware-based gunshot sound detectors.

5. CONCLUSION AND DISCUSSION

In this work, a gunshot detection system is realized on FPGA.

Firstly, commercial gunshot sound detection systems were examined. Then, gunfire detection methods were reviewed. And, the most appropriate methods for hardware implementation were determined. Available commercial systems and the existing gunshot sound detection methods are mentioned in the introduction section.

12 MFCCs, energy and zero-crossing rate features were chosen to compare k-NN and SVM classifier performances. And, the mathematical background of the chosen features and classifiers are given in the sound processing section.

The system was simulated in MATLAB. Obtained accuracy rates using k-NN and SVM classifiers were compared. As a result, a higher accuracy rate was achieved using the k-NN classifier. Simulation details and results are presented in the gunshot characteristics section.

Furthermore, the system was realized on Kintex Ultrascale+ KCU116 Evaluation Platform 1.0. The sound signal was classified using 12 MFCCs, signal energy and zero-crossing rate features and k-NN classifier. The details and results of the realization are mentioned in the realization of the gunshot detection system using FPGA section.

As a consequence, a %96.1538 accuracy rate was achieved. The previous methods mentioned in the 4th section achieved lower accuracy rates than our work and generally, machine learning methodologies were not implemented. In this study, a hardware-based machine learning application is realized and higher accuracy rate was achieved.

5.1 Future Works

The available gunshot sound detection systems generally implement sound source locater algorithms in order to determine gunshot fire direction. To detect the gunshot fire direction , a microphone array can be mounted to our system and a gunfire location detection algorithm can be realized.

In addition, a reverberation reduction algorithm may be used to reduce misleading reverberation effects.

REFERENCES

[1] "Homicide Trends In The U.S.," *Bureau of Justice Statistics Homicide trends in the U.S. – Weapons used*. [Online]. Available: https://web.archive.org/web/20130123074121/http:/bjs.ojp.usdoj.gov/content/homicide/weapons.cfm. [Accessed: 31-Mar-2020].

[2]"Shot-Detection Tech Reduces Gunfire in San Francisco by 35 Percent," Government Technology State & Local Articles - e.Republic. [Online]. Available: https://www.govtech.com/dc/articles/Shot-Detection-Tech-Reduces-Gunfire-in-San-Francisco-by-35-Percent.html. [Accessed: 31-Mar-2020].

[3] M. Hrabina and M. Sigmund, "Implementation of developed gunshot detection algorithm on TMS320C6713 processor," 2016 SAI Computing Conference (SAI), London, 2016, pp. 902-905.

[4] *Stop Fish Bombing USA*. [Online]. Available: https://www.sfbusa.org/. [Accessed: 31-Mar-2020].

[5] *Home - ShotSpotter*. [Online]. Available: https://www.shotspotter.com/. [Accessed: 31-Mar-2020].

[6] *Boomerang III | Raytheon*. [Online]. Available: https://www.raytheon.com/capabilities/products/boomerang. [Accessed: 31-Mar-2020].

[7] *Boomerang Warrior-X | Raytheon*. [Online]. Available: https://www.raytheon.com/capabilities/products/boomerang_warriorx. [Accessed: 31-Mar-2020].

[8] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance systems," 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, London, 2007, pp. 21-26.

[9] P. Volgyesi, G. Balogh, A. Nadas, C. B. Nash, and A. Ledeczi, "Shooter localization and weapon classification with soldier-wearable networked sensors," *Proceedings of the 5th international conference on Mobile systems, applications and services - MobiSys 07*, 2007.

[10] T. Ahmed, M. Uppal and A. Muhammad, "Improving efficiency and reliability of gunshot detection systems," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013, pp. 513-517.

[11] E. Kiktova, M. Lojka, M. Pleva, J. Juhar and A. Cizmar, "Gun type recognition from gunshot audio recordings," *3rd International*

Workshop on Biometrics and Forensics (IWBF 2015), Gjovik, 2015, pp. 1-6, doi: 10.1109/IWBF.2015.7110240.

[12] Tobin, Jacob Daniel, "A Hardware Based Audio Event Detection System. " Master's Thesis, University of Tennessee, Knoxwille,2015.

[13] C. Salazar-García *et al*., "Digital integrated circuit implementation of an identification stage for the detection of illegal hunting and logging," *2015 IEEE 6th Latin American Symposium on Circuits & Systems (LASCAS)*, Montevideo, 2015, pp. 1-4.

[14] Y. Arslan and B. Güldoğan, "Impulsive sound detection and gunshot recognition," *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, Malatya, 2015, pp. 511-514.

[15] A. A. Shiekh, M. Tahir and M. Uppal, "Accurate gunshot detection in urban environments using blind deconvolution," 2017 International Multi-topic Conference (INMIC), Lahore, 2017, pp. 1-4.

[16] L. Grama and C. Rusu, "Audio signal classification using Linear Predictive Coding and Random Forests," 2017 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Bucharest, 2017, pp. 1-9.

[17] A. Khamparia, D. Gupta, N. G. Nguyen, A. Khanna, B. Pandey and P. Tiwari, "Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network," in IEEE Access, vol. 7, pp. 7717-7727, 2019.

[18] C. M. J. Galangque and S. A. Guirnaldo, "Gunshot Classification and Localization System using Artificial Neural Network (ANN)," 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 2019, pp. 1-5.

[19] R. Baraniuk, *Signals and systems*. [Place of publication not identified]: Orange Grove Books, 2009, pp. 9,175.

[20] G. Heinzel, A. Rudiger and R. Schilling, "Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows.", Max-Planck-Institut fur ¨ Gravitationsphysik, Hannover, 2020.

[21] B. L. Sturm, M. Morvidone and L. Daudet, "Musical instrument identification using multiscale Mel-frequency cepstral coefficients," 2010 18th European Signal Processing Conference, Aalborg, 2010, pp. 477-481.

[22] V. Kecman, *Support Vector Machines – An Introduction*. Virginia, 2020, pp. 21-23.

[23]“Support-vector machine,” *Wikipedia*, 02-Mar-2020. [Online]. Available: https://en.wikipedia.org/wiki/Support-vector_machine. [Accessed: 03-Apr-2020].

[24] M. Jordan and R. Thibaux, "The Kernel Trick -CS281B/Stat241B: Advanced Topics in Learning & Decision Making", UC Berkeley EECS, 2020.

[25] Shalev-Shwartz, S. and Ben-David, S., 2020. *Understanding Machine Learning: From Theory To Algorithms*. [ebook] Cambridge University Press, pp.258-259. Available at: <https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf> [Accessed 3 April 2020].

[26] R. Maher, "Advancing Audio Forensics of Gunshot Acoustics", U.S Department of Justice, 2018.

[27] "Sound Libraries," *Still North Media*. [Online]. Available: https://www.stillnorthmedia.com/libraries. [Accessed: 29-May-2020].

[28] L. R. Rabiner and B.-H. Juang, *Fundamentals of speech recognition*. Delhi: Pearson Education, 2005.

CURRICULUM VITAE

Name Surname **: Mustafa Koray AKÇOCUK**

Place and Date of Birth **: Alanya / May 25,1994**

E-Mail **: k.akcocuk@gmail.com**

EDUCATION **:**

- **B.Sc.** **:** 2017, Istanbul Technical University, Faculty of Electrical and Electronics Engineering, Departmant of Electronics and Communication Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2017-2018 Arçelik, Electronics Engineer
- 2018-2020 Token Financial Technologies, Electronics Engineer