

**COMSATS UNIVERSITY ISLAMABAD**

VEHARI CAMPUS

DEPARTENT OF COMPUTER SCIENCE

**Object Sound Detection**

*Undertaken By:*

**Tahir Siddique**

Reg. No. CIIT/FA18-BSE-036/Vhr

**Muhammad Sohail Amjad**

Reg. No. CIIT/FA18-BSE-063/Vhr

*Supervised By:*

**M. Jawad Rafeeq**



DEPARTMENT OF COMPUTER SCIENCES

COMSATS UNIVESITY ISLAMABAD, VEHARI CAMPUSVEHARI

SESSION 2018-2022

**SCOPE DOCUMENT REVSION HISTORY**

|  |  |  |
| --- | --- | --- |
| **No.** | **Comment** | **Action** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Supervisor Signature**

**Date:**

**Table of Contents**

Contents

[Abstract 5](#_Toc86058381)

[1. Introduction 5](#_Toc86058382)

[2. Problem Statement 6](#_Toc86058383)

[3. Problem Solution for Proposed System 6](#_Toc86058384)

[4. Related System Analysis/Literature Review 6](#_Toc86058385)

[5. Advantages/Benefits of Proposed System 7](#_Toc86058386)

[6. Scope 7](#_Toc86058387)

[7. Concept 7](#_Toc86058388)

[8. System Limitations/*Constraints* 7](#_Toc86058389)

[9. Software Process Methodology 8](#_Toc86058390)

[10. Tools and Technologies 8](#_Toc86058391)

[11. Project Stakeholders and Roles 8](#_Toc86058392)

[12. Team Members Individual Tasks/Work Division 9](#_Toc86058393)

[13. Data Gathering Approach 9](#_Toc86058394)

[14. Concepts 9](#_Toc86058395)

[15. Mockups 10](#_Toc86058396)

[16. Conclusion 10](#_Toc86058397)

## Abstract

In every aspect of human life, sound plays an important role. From personal security to critical surveillance, a sound is a key element to develop automated systems for these fields. Few systems are already in the market but their efficiency is a concerned point for their implementation in real life scenarios. The learning capabilities of Deep learning architectures can be used to develop sound classification systems to overcome efficiency issues of traditional systems. Our aim is to use deep learning networks for classifying environmental sounds based on the generated spectrograms of these sounds. We will use spectrogram images of environmental sounds to train Convolutional Neural Network. To train and test CNN we will use Urban8kSounds Dataset.

# Introduction

# Problem Statement

# Classifying the urban sound although an inherently subjective task comes quite easy to the human’s ear. The classification of environmental sounds, specifically focusing on the identification of particular urban sounds. When given an audio sample in a computer readable format (such as a .wav file) of a few seconds duration, we want to be able to determine if it contains one of the target urban sounds with a corresponding**Classification Accuracy score**. So we need to classify the urban sounds to make it easier to get the sound from a sound.

# Problem Solution for Proposed System

In the world of modern society there are different urban sounds that a human ear can classify. But in the world of technology the machines are working to do many of tasks that a human was performing. In the past it was a hard task to differentiate between thousands of urban sounds according to their type. But in the world of technology we need the work to be done as fast as possible. To make it possible we will use CNN (Convolutional Neural Network) Algorithmthat is fast enough and responsive for the results. Although there are multiple algorithms that can help but this algorithm is more responsive and easy to use than other algorithms. **Convolutional Neural Network** is a Deep Learning algorithm which can take in an input, assign importance (learnable weights and biases) to various aspects/objects in the input and be able to differentiate one from the other.

# Related System Analysis/Literature Review

Most works on the abovementioned architectures are commonly restricted to a single audio classification task, and a few studies have focused on a uniform framework to solve different audio classification problems. To compare with our experiments on the same datasets, we review some task-independent models. Med-hat et al. suggest a binary-masked CNN model that adopts a controlled systematic sparseness such as embedding a filter bank-like behavior within the network to preserve the spatial locality of features in the process of training weights. The experiments on GTZAN and UrbanSound8K have achieved the accuracies of 85.1% and 74.2%, respectively. In their model, they introduce a set of hidden layers in which each neuron establishes a connection with the input by the activation function through the influence of distinct active regions in the feature vector, so the spatial information of the learned feature is saved as these active weights’ locations which are fixed in position. In order to exploit the inter-frame relations in a sequential signal, the network allows a concurrent exploration of a range of feature combinations to find the optimum combination of features through an exhaustive manual search. This causes the increased complexity and training difficulties, especially when the number of features increased substantially.

# Advantages/Benefits of Proposed System

The advantage of the systems is that we:

* Can differentiate between two sounds of object.
* Can be helpful for the persons having listening problem
* Can detect the sound in real time.

It will provide many of other advantages and can be implemented to provide solution to the world.

# Scope

The scope of this work is focused on a urban sounds classification. The elaboration of this proposal is organized as follows describe the research field of Urban. As an important processing step, feature extraction plays a critical role that will significantly affect the final classification performance. We try to break through the performance bottleneck, using novel feature sets extracted with image information retrieval techniques. It describes the experiments applying convolutional neural network (CNN), a state-of-the-art image digit recognition algorithm, to automatic extraction of sound pattern features. The system architecture, the characteristics of CNN and the classification performance are explained. studies the invariance of the widely used MFCC feature set to Urban Sounds.

# Concept

* Convolutional neural networks

# System Limitations/*Constraints*

* Limited only on given sounds

# Software Process Methodology

* MLLC (Machine Learning Life Cycle )
  + Gathering Data
  + Data Preparation
  + Data Wrangling
  + Data Analysis
  + Train Model
  + Test Model
  + Deployment

# Tools and Technologies

**Table 2Tools and Technologies for Proposed Project**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tools**  **And**  **Technologies** | **Tools** | **Version** | **Rationale** |
| Spyder | 2018 | IDE |
| MS Word | 2010 | Documentation |
| MS Power Point | 2010 | Presentation |
| **Technology** | **Version** | **Rationale** |
| Python | 3.7 | Programming language |

# Project Stakeholders and Roles

**Table 3Project Stakeholders for Proposed Project**

|  |  |
| --- | --- |
| **Project Sponsor** | COMSATS University, Islamabad |
| **Stakeholder** | * Tahir Siddique & Muhammad Sohail Amjad * Project Supervisor Name: Mr. Jawad Rafeeq * Final Year Project Committee: Evaluation of project |

# Team Members Individual Tasks/Work Division

|  |  |  |
| --- | --- | --- |
| **Student Name** | **Student Registration Number** | **Responsibility/ Modules** |
| Tahir Siddique | FA-18/BSE/036 | Documentation Design and Implementation |
| Muhammad Sohail Amjad | FA-18/BSE/063 | Documentation Design and Implementation |

# Data Gathering Approach

Our study is based on the Million Song Dataset, which is a freely-available collection of audio features and metadata for a million contemporary popular music tracks. We will use **Urban8K** dataset provided on **Kaggle**.

# Concepts

We will develop the system by using the Machine Learning Life Cycle (MLLC) Methodology. It’s a cyclical process to go through to build and manage good quality models. The Broad Phases of Machine Leaning Life Cycle are as follows:

1. Business Requirements

2. Gathering and Data Preparation

3. Exploratory Analysis

4. Training

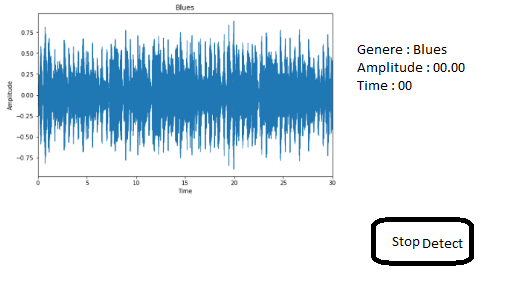
5. Model Selection and Verification

6. Deployment

7. Monitoring

The Developing process of the System that is based on the Machine Learning and Deep Learning is differ from other software developing like desktop, or web based or Android base system. Software system wouldn’t fail once deployed as long as the requirements are no change but in Machine and Deep Learning system the underlying characteristics of the data might be change and your models may not be giving the accurate result.

# Mockups



# Conclusion

The conclusion of this document is basically to build a system that will classify the Urban Sounds in different categories and will help in differentiating the Urban Sounds .

**17. References**

[1] \*