NITTE (Deemed to be University) | NMAM INSTITUTE OF TECHNOLOGY

Nitte(DU) established under Section 3 of UGC Act 1956 | Accredited with 'A+' Grade by NAAC

:

☎ 08258 - 281039 – 281263, Fax: 08258 – 281265

VOICE GENDER RECOGNITION

A Mini Project Report Submitted by

Meghana Bhat Pervaje
(4NM20AI025)

Kavana Pai
(4NM20AI020)

UNDER THE GUIDANCE OF

Ms. Disha D N
Associate Professor

Department of Artificial Intelligence and Machine Learning Engineering

In partial fulfillment of the requirements for the

MACHINE LEARNING (20AM502)

December 2022-23

NITTE
EDUCATION TRUST

N.M.A.M. INSTITUTE OF TECHNOLOGY
(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)
Nitte – 574 110, Karnataka, India
(ISO 9001:2015 Certified), Accredited with 'A' Grade by NAAC
☎ : 08258 - 281039 - 281263, Fax: 08258 - 281265

Department of Artificial Intelligence and Machine Learning Engineering

B.E. CSE Program Accredited by NBA, New Delhi from 1-7-2018 to 30-6-2021

CERTIFICATE

Certified that the mini project work entitled

"VOICE GENDER RECOGNITION"

is a bonafide work carried out by

Meghana Bhat Pervaje
(4NM20AI025)

Kavana Pai
(4NM20AI020)

in partial fulfilment of the requirements for the award of
Bachelor of Engineering Degree** in **Artificial Intelligence and Machine Learning Engineering
*prescribed by **Visvesvaraya Technological University, Belgaum***
during the year 2022-2023.

It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library.
The mini project report has been approved as it satisfies the academic requirements in respect of the mini project work prescribed for the Bachelor of Engineering Degree.

Signature of Guide

Signature of HOD

Evaluation

Name of the Examiners **Signature with Date**

1. Dr. Sharada Shenoy 12/12/22

2. Rakshitha 12/12/12

ACKNOWLEDGEMENT

We believe that our mini project will be complete only after we thank the people who have contributed to make this mini project successful.

First and foremost, our sincere thanks to our beloved principal, **Dr. Niranjan N. Chiplunkar** for giving us an opportunity to carry out our mini project work at our college and providing us with all the needed facilities.

I acknowledge the support and valuable inputs given by**, Dr.Sharada U Shenoy** the Head of the Department, Artificial Intelligence and Machine Learning Engineering, NMAMIT, Nitte

We express our deep sense of gratitude and indebtedness to our guide **Ms.Disha D N,** Assistant Professor Artificial Intelligence and Machine Learning Engineering, for his inspiring guidance, constant encouragement, support and suggestions for improvement during the course for our mini project.

We also thank all those who have supported us throughout the entire duration of our mini project.

Finally, we thank the staff members of the Department of Artificial Intelligence and Machine Learning Engineering and all our friends for their honest opinions and suggestions throughout the course of our mini mini project.

Meghana Bhat Pervaje

Kavana Pai

ABSTRACT

Automatic age and gender classification has become relevant to an increasing amount of applications, particularly since the rise of social platforms and social media. Nevertheless, performance of existing methods on real-world images is still significantly lacking, especially when compared to the tremendous leaps in performance recently reported for the related task of face recognition. In this paper we show that by learning representations through the use of deep-convolutional neural networks (CNN), a significant increase in performance can be obtained on these tasks. To this end, we propose a simple convolutional net architecture that can be used even when the amount of learning data is limited. We evaluate our method on the recent Adience benchmark for age and gender estimation and show it to dramatically outperform current state-of-the-art methods

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Nowadays, with the rapid spread in the use of internet resources and the huge amount of data shared on the internet, most of them were almost untreatable ,and were not recognized by people. Even a few People were afraid of those messages ,even some messages are leading to death of people.so as to detect them this project is used to detect the peoples who sent that message and classifies their age and gender. So, as we can control and stop these kind of works. Gender and age play a significant role in interpersonal interactions among people who live in communities. The use of smart gadgets has expanded as technology hasprogressed, and social media has begun to draw everyone's attention. Daily studies on gender and age prediction have grown in prominence, it increases the number of apps that use such techniques. In these applications, facial photographs are commonly employed since they contain useful information that may be used to extract human interaction. For gender detection and age prediction, Image processing, feature extraction, and classificationsteps are usually used. These steps may change based on the objective of the study and the characteristics to be used. The face images were processed using a variety of approaches, and calculations were performed based on the results of the investigations. For image processing, there are two basic and typical which we need to follow. Image enhancement is the process of improving an image so that the resultant image is of higher quality and can be used by other applications. The most popular technique for extracting information from an image is the other technique. The image is divided into a specified number of parts or objects in order to solve the challenge and this procedure is called Segmentation. Due to the accuracy of its classification technique, deep learning techniques are a variety of tasks such as classification, feature extraction, object recognition, and so on,it helps in gender and age prediction.

1.1 OVERVIEW

Speech Recognition(SR) is the process of converting input speech signals (voice signals) to precise or accurate digital signals which are used in methodologies and technologies. Speech recognition is achieved by programming using programming languages which performs the required task. The Speech Recognizing systems that depend on Human inputs or training are speaker dependent and those which do not require the training are called speaker independent. Here we use speaker dependent

SRs. The speech recognizers are programmed to find the age range of a person with different frequency and pitch. The application of this method can be used for Defensive Purpose , Bank security and various other defensive and investigative actions. The range of age of a person can be used to determine the actions that the person is capable of. The victim's age can be found which can provide a clue for the defensive officials to take action on.

1.2 PROBLEM STATEMENT

To design an automatic speech recognition system that gives best recognition results for both male and female speakers. Allow the users to identify themselves using nothing but their voices. The system should automatically recognize who is speaking on the basis of individual information included in speech waves. This can be much more convenient than traditional means of authentication which require carrying a key with you or remembering a PIN, also proving the identity of a recorded voice can help to convict a criminal or discharge an innocent in court.

There are various methods to solve the problem of accurately estimating age and gender. The first approach relies on manually extracting features such as the size of the head, position of eyes or length of the nose. Another approach is based on end-to-end deep learning models. The two methods can be combined to form a new mixed approach. Most modern methods use deep learning approach.

1.3 STUDY AREA

Speech recognition can be applied to voice authorization, typing, and remote health monitoring. It is an essential AT for people who need to convert their voice to text to communicate with others through writing via computers, smartphones, and the internet. In general, in the absence of noise and interference, a wearable microphone is the only sensor needed to record human voice accurately, along with readily available commercial or open-source software to recognize it

1.4 OBJECTIVE

The primary aim of this thesis is to detect the gender of a person from his/her Voice with decent accuracy. There has been a lot of work done in this field using various methods which all have their shortcomings. It ignores that some people are blind or does not have eyebrows. There are other methods that acknowledge these faults but fail to provide satisfactory results. This thesis has specifically targeted on the issue of establishing a new tactics which can help us to establish efficient operation of facial data extraction and gender classification techniques. The step-by-step procedure of this thesis is summarized here. This work marks the following issues:

- ♣ Firstly, we have used input audio to detect voice using Viola & Jones algorithm for robust and real-time extraction of voice.
- ♣ Then, the image has been processed to reduce noise using Adaptive Filtering and for adjusting contrast, Histogram Equalization.
- ♣ Finally, we have used DSP (Deformable Spatial Pyramid) to produce extremely accurate results and efficient computation to reduce computational time.

1.5 MOTIVATION

The motivation behind this thesis was to build an application for age and gender classification using a model that is suitable for real life predictions. Many models are focusing on datasets with constrained audio and are not suitable for in-the-wild estimation. In this thesis we will focus on deep learning end-to-end methods.

1.6 ORGANIZATION OF THE CHAPTERS

The project report has been organized under nine chapters, which are as follows:

Chapter I: Introduces to the main idea of the project. It gives a brief knowledge about the aim and methodology of the same.

Chapter II: It includes literature survey of related works.

Chapter III: Discusses the system requirements that are needed for the project. These include functional requirements, non-functional requirements, user requirements and hardware requirements.

Chapter IV: Includes the system design details which includes flowchart, sequence diagram.

Chapter V: Includes the implementation details of the project

Chapter VI: Deals with system testing concepts and the various test cases for the project.

Chapter VII: Includes the screenshots of the application.

Chapter VIII: Discuss the results of the project.

Chapter IX outlines conclusions and future work that can be done

CHAPTER-2 LITERATURE SURVEY

Automatic speech Recognition is a method by that a PC takes a speech signal and converts it into words . It is the method by which a PC acknowledges what a person said. Keyboard, though a standard medium, is not terribly convenient, as it needs a bound quantity of talent for effective usage .A mouse on the different hand needs a sensible hand eye co-ordination. Physically challenged individuals notice PC tough to use. Partly blind individuals notice reading from a monitor tough. All these constraints have to be eliminated. Speech interface facilitate us to tackle these issues. The objective is to entice human voice during a electronic computer and decipher it into corresponding text. Speech recognition will be outlined as the method of changing an acoustic signal, captured by a electro-acoustic transducer (microphone) or a telephone, to a group of words. When two individuals speak to one another, they each acknowledge the words and will be able to understand the meaning behind them. Computers, on the opposite hand, are solely capable of the initial thing: they can only acknowledge individual words and phrases; however they don’t extremely perceive speech within the same means as humans do. PC acknowledges the command and software system tells the PC what to try to once that command is recognized . Big data refers to information volumes within an extent of Exabyte (1018 B) and on the far side. Such volumes exceed the capability of current on-line storage and process systems. With characteristics like volume, velocity and variety massive information throws challenges to the normal IT sector. Computer aided innovation, real time information analytics, customer-centric business intelligence; trade wide higher cognitive process and transparency are potential blessings, to say few, of huge information. There are several problems with massive information that warrant quality assessment strategies. The problems are touching on storage and transport, management, and process. This paper throws lightweight into this state of quality problems associated with massive information. It provides valuable insights that may be accustomed leverage massive information science activities. [13] Big data and its analysis are at the middle of contemporary science and business. These information are generated from on-line transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social networking interactions, science information, sensors and mobile phones and their applications.

Machine learning algorithms are the programmes which will continuously learn from the given knowledge and improve from gained experience, while there is no human intervention needed. Learning tasks will include learning a mathematical function that maps input to the output, then the system learns the hidden structure in data that is not labelled; or ‘instance-based learning’, where a class label is provided for every new instance by rapidly comparing the fresh instance (row) to instances from the existing training data, which were stored in memory. Machine learning algorithms are divided into three categories according to the feature how they learn. They are as follows

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

“Supervised Learning: Supervised learning as itself the name indicates the presence of a supervisor as an educator. Essentially supervised learning may be a learning during which we have a tendency to teach or train the machine using knowledge which is well labelled which means some knowledge is already labelled with the proper answer. After that, the machine is supplied with a brand new set of examples (data) in order that supervised learning formula analyses the coaching knowledge (set of coaching examples) and produces an accurate outcome from labelled data. Supervised Learning algorithms are Regression, Decision Tree, Random Forest, KNN, Logistic Regression algorithms etc”. Classification is employed to predict the end result of a given sample once the output variable is within the style of classes. A classification model would possibly check out the input data and take a look at to predict labels like “medium” or “heavy.” Regression is employed to predict the end result of a given sample once the output variable is within the style of real values. For instance, a regression model would possibly process input data to predict the number of downfall, the peak of someone, etc. Ensembling is another sort of supervised learning. It suggests that combining the predictions of multiple machine learning models that square measure severally weak to provide a lot of correct prediction on a replacement sample. Unsupervised Learning: In this approach, we do not have any outcome variable to predict the situation. This is used for mainly clustering the population into different groups, where it is very widely used for customer segmentation into different groups for specific intervention or purposes. Examples of Unsupervised Learning algorithms are Apriori.

2.1 EXISTING SYSTEM

The process of proposed VRSML (Voice Recognition System through Machine Learning) is represented in the form of a flowchart as below. Voice which is given as the input is parallelly recorded as audio file and converted to text. Then the text is tokenized and passed to the analyzer, all the tokens are then categorized based on GSL (General Service List) by applying K-Nearest Neighbour algorithm of Machine Learning, finally a report is generated which describes the vocabulary used in the input speech. Below is the detailed step by step procedure followed by VRSML Step 1: An input is given to the system in the form of English speech which is generated by humans using microphones connected to the computer. The input can be either taken from Big data. Step 2: The human generated audio is recorded by the recorder as and then the user speaks. It can record the audio for unlimited amount of time. Step 3: The recorded audio is saved in .mp3 format for any further references and cross-checking the converted text. Step4: Convert the generated audio into text format using Speech to Text Converter tool. Step 5: Save the extracted text into a .txt file format. These files are further processed to extract the level of vocabulary used. Step 6: The whole text is given as input to a tokenizer and it tokenizes the whole text and extracts each and every single word from it accepting the character space as delimiter. Step 7: Using Recurrent Neural Networks we make the system learn from the given inputs and make an appropriate decision. Step 8: Using K-Nearest Neighbour algorithm of supervised learning approach of Machine Learning, the extracted tokens are classified into several categories such as repeated or not, general word or not. If repeated, then for how many times it was repeated and the number of words that are been used uniquely. In brief, summary of the vocabulary levels used in the speech is generated. Step 9: Textual summary as well graphical summary (Point Graph, Bar Graph) are used to show the vocabulary levels used in the speech generated. Step 10: With the use of the summary generated by the system, one can analyze their own vocabulary skills or can assess others as well

2.1 PROPOSED SYSTEM

In this Python Project, we will use Deep Learning to accurately identify the gender and age of a person from a single Voice . We will use the models trained by Tal Hassner and Gil Levi. The predicted gender may be one of ‘Male’ and ‘Female’, and the predicted age may be one of the following ranges- (0 – 2), (4 – 6), (8 – 12), (15 – 20), (25 – 32), (38 –43), (48 – 53), (60 – 100) (8 nodes in the final SoftMax layer). It is very difficult to accuratelyguess an exact age from a single audio because of factors like Volume (Loudness)

- Pitch (Rise and Fall)
- Pace (Rate)
- Pause (Silence)
- Resonance (Timbre)
- Intonation

And so, we make this a classification problem instead of making it one of regression.

The DFF Architecture

Deep NNs perform better than shallow ones. However, it is not always necessary to use a deep network. The choice will largely depend on the task you have at hand.

If you are working with many inputs, such as image data, then using a Deep Feed Forward (DFF) or a Convolutional Neural Network (CNN) would likely yield better results than a simple Feed Forward network.

However, suppose your task is to do some basic classification with a limited number of inputs. In that case, you may be better off using a simple FF network or even a tree-based algorithm such as XGBoost, Random Forest, or a single Decision tree.

The Dataset

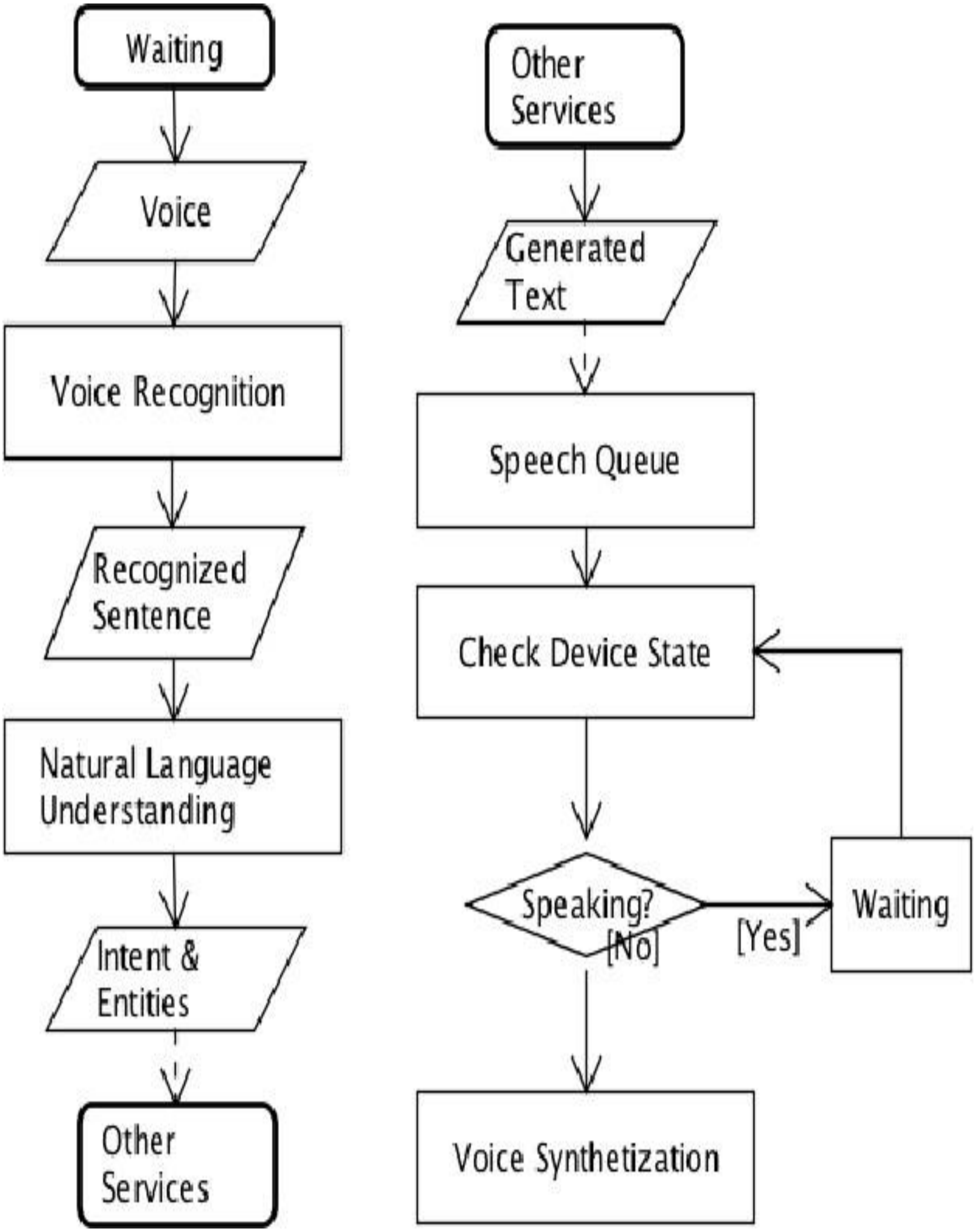
For this python project, we’ll use the Adience dataset; the dataset is available in the public domain and you can find it *here*. This dataset serves as a benchmark for face photos and is inclusive of various real-world imaging conditions like noise, lighting, pose, and appearance. The images have been collected from Flickr albums and distributed under the

Creative Commons (CC) license. It has a total of 26,580 photos of 2,284 subjects in eight age ranges (as mentioned above) and is about 1GB in size. The models we will use have been trained on this dataset.

install OpenCV (cv2) to be able to run this project. You can do this with pip- pip install

OpenCV-python

Other packages you’ll be needing are math and argparse, but those come as part of the standardPythonlibrary.


Waiting
Voice
Voice Recognition
Recognized Sentence
Natural Language Understanding
Intent & Entities
Other Services
Other Services
Generated Text
Speech Queue
Check Device State
Speaking?
[No]
[Yes]
Waiting
Voice Synthetization

CHAPTER 3 SYSTEM ANALYSIS AND REQUIREMENTS

3.1 SYSTEM ANALYSIS

3.1.1 Relevance of Platform

Neural networks, also known as Deep neural networks (DNNs) or simulated neural algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

Deep neural networks (DNNs) are comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer. Each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network.

Relevance of Programming Language

Python is a very popular general-purpose interpreted, interactive, object-oriented, and high-level programming language. Python is dynamically-typed and garbage-collected programming language. It was created by Guido van Rossum during 1985- 1990. Like Perl, Python source code is also available under the GNU General Public License (GPL).

Python supports multiple programming paradigms, including Procedural, Object Oriented and Functional programming language. Python design philosophy emphasizes code readability with the use of significant indentation.

This tutorial gives a complete understanding of Python programming language starting from basic conceopts to advanced concepts. This tutorial will take you through simple and practical approaches while learning Python Programming language.

3.2 REQUIREMENT ANALYSIS

Scope and Boundary

Requirements are during early stages of a system development as a specification of what should be implemented or as a constraint of some kind of on the system. They may be a user level facility description, a detailed specification of expected system behaviour, a general system property, a specific constraint on the system, and information on how to carry out some computation or a constraint on the development of the system. The end product of the requirement analysis phase is a requirement specification. The requirement specification is a reconstruction of the result of this analysis phase. Its purpose is to communicate this result to others. System requirements are more detailed descriptions of the user requirements. They may serve as the basis for a contract to the implementation of the system and should therefore be a complete and consistent specification of the whole system. In principle, the system requirements should state what the system should do and not how it should be implemented. However, at the level of detail required to specify the system completely, it is virtually impossible to exclude all design information.

FUNCTIONAL REQUIREMENTS

Software Requirements:

Software: 1. Python Software

Hardware Requirements

Operating system: Windows 9 and above.

RAM : 4GB and above.

Processor : Intel® Core(TM)2 duo CPU T6500.

Processor speed : 2.67 GHz.

CPU : 64-bit operating system.

NON-FUNCTIONAL REQUIREMENTS:

In systems engineering and requirements engineering, a non-functional requirement (NFR) is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. Non-functional requirements are conditions under which the system must be able to function and the quality the system must have. It defines how a system is supposed to be.

- ♣ Performance
- ♣ With ideal condition , response should be fast and error free.
- ♣ Flexibility: This code will be easy to learn and use. Is able to analyze and give the output as quickly as possible.
- ♣ User-friendly: The users should be able to find their age easily.The multiple features should be self-explanatory.
- ♣ Response Time: The selected video should load and display quickly without consuming much buffer time.
- ♣ Understandability: All users can learn to operate the website because of its simplicity.

CHAPTER 4 SOFTWARE APPROACH

4.1 DEEP LEARNING

Deep learning can be considered as a subset of machine learning. It is a field that is based on learning and improving on its own by examining computer algorithms. While machine learning uses simpler concepts, deep learning works with artificial neural networks, which are designed to imitate how humans think and learn. Until recently, neural networks were limited by computing power and thus were limited in complexity. However, advancements in Big Data analytics have permitted larger, sophisticated neural networks, allowing computers to observe, learn, and react to complex situations faster than humans. Deep learning has aided image classification, language translation, speech recognition. It can be used to solve any pattern recognition problem and without human intervention.

Artificial neural networks, comprising many layers, drive deep learning. Deep Neural Networks (DNNs) are such types of networks where each layer can perform complex operations such as representation and abstraction that make sense of images, sound, and text. Considered the fastest-growing field in machine learning, deep learning represents a truly disruptive digital technology, and it is being used by increasingly more companies to create new business models.

4.2 PYHTON

Python is consistently rated as one of the world's most popular programming languages. Python is fairly easy to learn, so if you are starting to learn any programming language then Python could be your great choice. Today various Schools, Colleges and Universities are teaching Python as their primary programming language

Python is a MUST for students and working professionals to become a great Software Engineer specially when they are working in Web Development Domain. I will list down some of the key advantages of learning Python:

- **Python is Interpreted −** Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive −** You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

- **Python is Object-Oriented** − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language** − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

4.3 JUPYTER NOTEBOOK

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.

Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

4.CLASSIFICATION ALGORITHM

Based on training data, the Classification algorithm is a Supervised Learning technique used to categorize new observations. In classification, a program uses the dataset or observations provided to learn how to categorize new observations into various classes or groups. For instance, 0 or 1, red or blue, yes or no, spam or not spam, etc. Targets, labels, or categories can all be used to describe classes. The Classification algorithm uses labeled input data because it is a supervised learning technique and comprises input and output information. A discrete output function (y) is transferred to an input variable in the classification process (x).

In simple words, classification is a type of pattern recognition in which classification algorithms are performed on training data to discover the same pattern in new data sets.

4.5 OPEN CV

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in commercial products. Being an Apache 2 licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

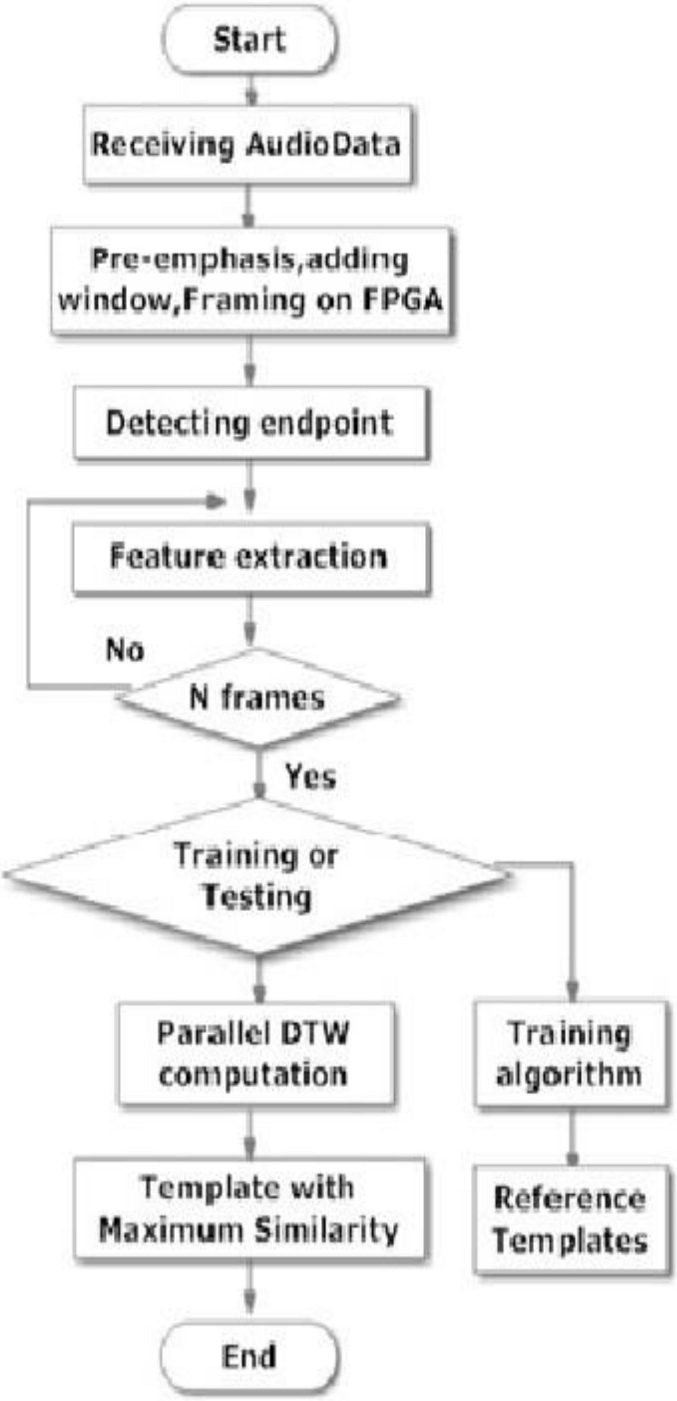
The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 18 million. The library is used extensively in companies, research groups and by governmental bodies.

Along with well-established companies like Google, Yahoo, Microsoft, Intel, IBM, Sony, Honda, Toyota that employ the library, there are many startups such as Applied Minds, VideoSurf, and Zeitera, that make extensive use of OpenCV. OpenCV’s deployed uses span the range from stitching street view images together, detecting intrusions in surveillance video in Israel, monitoring mine equipment in China, helping robots navigate and pick up objects at Willow Garage, detection of swimming pool drowning accidents in Europe, running interactive art in Spain and New York, checking runways for debris in Turkey, inspecting labels on products in factories around the world on to rapid face detection in Japan.

It has C++, Python, Java, and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available. A full-featured CUDA and OpenCL interfaces are being actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. OpenCV is written natively in C++ and has a templated interface that works seamlessly with STL containers.

CHAPTER 5 SYSTEM DESIGN

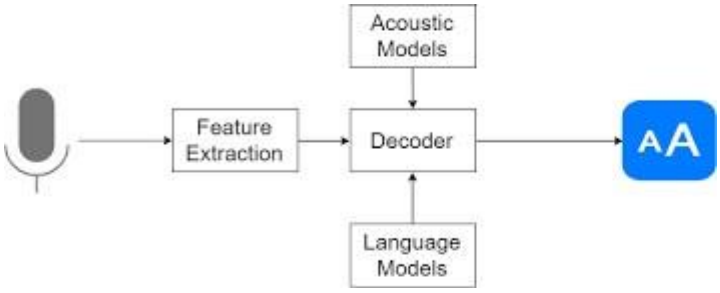
ARCHITECTURE



5.1 LOW LEVEL DESIGN ARCHITECHTURE

5.1.1 Sequence Diagram /DFD

A sequence diagram shows object interaction arranged in time sequence. It describes interactions among classes in terms of an exchange of messages over time. It is also called as event diagram. A sequence diagram is a good way to visualize and validate various run time scenarios. These can help to predict how a system will behave and to discover responsibilities a class may need to have in the process of modelling the new system. Messages are arrows that represent communication between the objects. Lifelines are vertical dashed lines that indicate the object presence over time.



CHAPTER 6 SYSTEM IMPLEMENTATION

6.1 SOFTWARE APPROACH

6.1.1 Working

1. To get started, install the following libraries using pip:

```
pip3 install numpy pandas tqdm sklearn tensorflow pyaudio librosa
```

2. To follow along, open up a new notebook and import the modules

```
import pandas as pd
import numpy as np
import os
import tqdm
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM, Dropout
from tensorflow.keras.callbacks import ModelCheckpoint, TensorBoard, EarlyStopping
from sklearn.model_selection import train_test_split
```

3. Now to get the gender of each sample, there is a CSV metadata file (check it here) that links each audio sample's file path to its appropriate gender

4. Check the number of samples of each gender:

a large number of balanced audio samples, the following function loads all the files into a single array; we don't need any generation mechanism as it fits the memory (since each audio sample is only the extracted feature with the size of 1KB)

5. function is responsible for reading that CSV file and loading all audio samples in a single array, this will take some time the first time you run it, but it will save that bundled array in `results` folder, which will save us time in the second run.

6. `label2int` dictionary simply maps each gender to an integer value; we need it in the `load_data()` function to translate string labels to integer labels.

7. Now, this is a single array, but we need to split our dataset into training, testing, and validation sets

8. sklearn's train_test_split() convenient function, which will shuffle our dataset and split it into training and testing sets. We then rerun it on the training set to get the validation set

9. Building the Model

10. We're using a 30% dropout rate after each fully connected layer; this type of regularization will hopefully prevent overfitting on the training dataset. Check this tutorial to learn more about Dropout.

An important thing to note here is we're using a single output unit (neuron) with a sigmoid activation function in the output layer; the model will output the scalar 1 (or close to it) when the audio's speaker is a male, and female when it's closer to 0

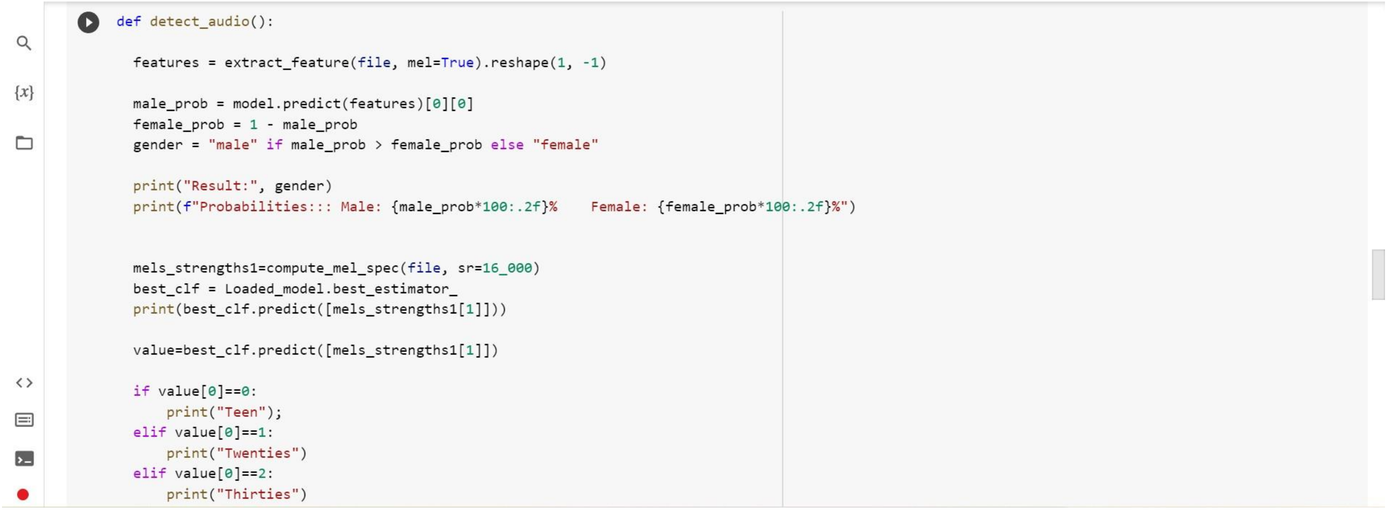
Also, we're using binary cross-entropy as the loss function, as it is a special case of categorical cross-entropy when we only have two classes to predict

- The first is the tensorboard; we will use it to see how the model goes during the training in terms of loss and accuracy.
- The second callback is early stopping; this will stop the training when the model stops improving. I've specified a patience of 5, which means it will stop training after 5 epochs of not improving, setting restore_best_weights to True will restore the optimal weights recorded during the training and assign them to the model weights.Since the model now is trained and the weights are optimal

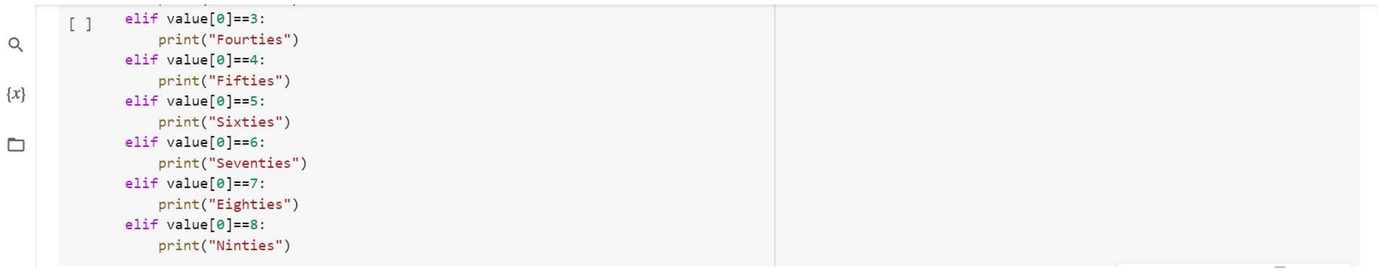
CHAPTER 7 SYSTEM TESTING

7.1 INTRODUCTION

OpenJupyter Notebook, run our script with the option and specify an audio toclassify



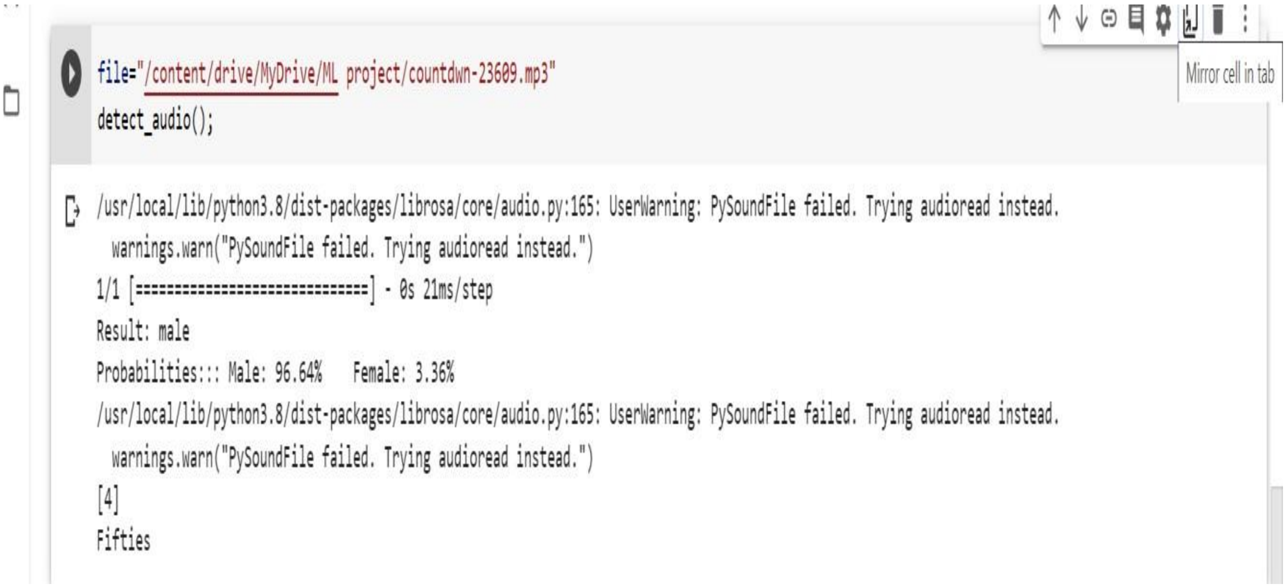
```
def detect_audio():

    features = extract_feature(file, mel=True).reshape(1, -1)

    male_prob = model.predict(features)[0][0]
    female_prob = 1 - male_prob
    gender = "male" if male_prob > female_prob else "female"

    print("Result:", gender)
    print(f"Probabilities::: Male: {male_prob*100:.2f}%    Female: {female_prob*100:.2f}%")


    mels_strengths1=compute_mel_spec(file, sr=16_000)
    best_clf = Loaded_model.best_estimator_
    print(best_clf.predict([mels_strengths1[1]]))

    value=best_clf.predict([mels_strengths1[1]])

    if value[0]==0:
        print("Teen");
    elif value[0]==1:
        print("Twenties")
    elif value[0]==2:
        print("Thirties")
```



```
[ ]    elif value[0]==3:
           print("Fourties")
       elif value[0]==4:
           print("Fifties")
       elif value[0]==5:
           print("Sixties")
       elif value[0]==6:
           print("Seventies")
       elif value[0]==7:
           print("Eighties")
       elif value[0]==8:
           print("Ninties")
```

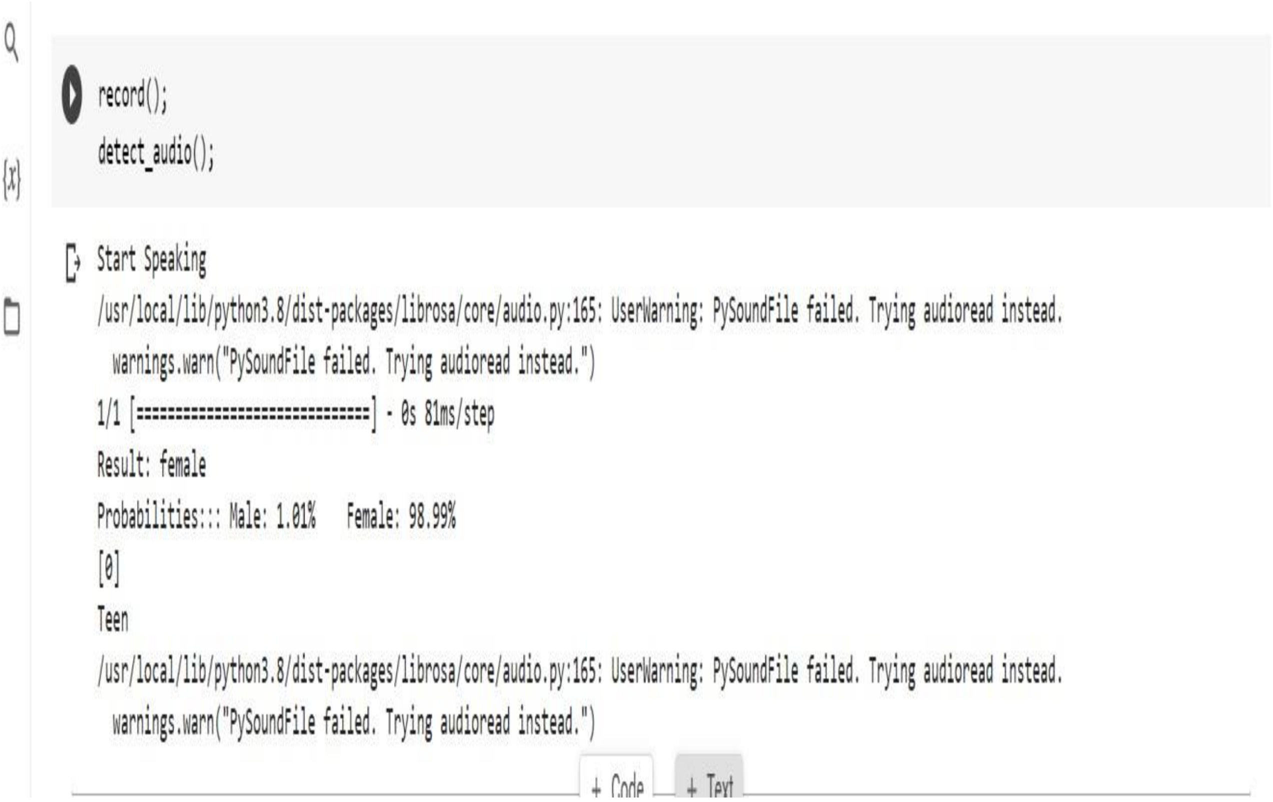
CHAPTER 8 RESULTS AND DISCUSSSIONS

RESULT



```
file="/content/drive/MyDrive/ML project/countdwn-23609.mp3"
detect_audio();
```

```
/usr/local/lib/python3.8/dist-packages/librosa/core/audio.py:165: UserWarning: PySoundFile failed. Trying audioread instead.
  warnings.warn("PySoundFile failed. Trying audioread instead.")
1/1 [==============================] - 0s 21ms/step
Result: male
Probabilities::: Male: 96.64%    Female: 3.36%
/usr/local/lib/python3.8/dist-packages/librosa/core/audio.py:165: UserWarning: PySoundFile failed. Trying audioread instead.
  warnings.warn("PySoundFile failed. Trying audioread instead.")
[4]
Fifties
```



```
record();
detect_audio();
```

```
Start Speaking
/usr/local/lib/python3.8/dist-packages/librosa/core/audio.py:165: UserWarning: PySoundFile failed. Trying audioread instead.
  warnings.warn("PySoundFile failed. Trying audioread instead.")
1/1 [==============================] - 0s 81ms/step
Result: female
Probabilities::: Male: 1.01%    Female: 98.99%
[0]
Teen
/usr/local/lib/python3.8/dist-packages/librosa/core/audio.py:165: UserWarning: PySoundFile failed. Trying audioread instead.
  warnings.warn("PySoundFile failed. Trying audioread instead.")
```

8.1 DISCUSSIONS

There are many possibilities in age and gender estimation research. A n immediate idea would be to look more deeply into training models with integral images as additional colour channels using more varied neural network architectures. Another idea would be to use more varied neural network architecture specifically for gender prediction. Many current tools use the same architecture for both age and gender prediction.

Age and gender prediction from images is an important application of computer vision. There are many approaches to solve this problem. We evaluate three different methods. We combine publicly available datasets and one manually labelled dataset into a large set and train the best method. We further extend the data by adding a colour channel to the images and train the best method. We show that training a network with a large dataset improves the performance, however adding additional colour channel does not. Based on our results we develop an application for age and gender classification.

CHAPTER 9 CONCLUSION AND FUTURE WORK

9.1 CONCLUSION

Age and Gender Classification are two of the most essential resources for getting information from an individual. Human faces contain enough information to be useful for a variety of purposes. Human age and gender classification are critical for reaching the right audience. We attempted to replicate the process using standard equipment. The algorithm's efficiency is determined by a number of factors, but the major goal of this study is to make it as simple and quick as possible while maintaining the highest level of accuracy. Work is being done to improve the algorithm's efficiency. Future enhancements include discarding faces for nonhuman objects, adding more datasets for people of other ethnic groups, and giving the computer more granular control over its workflow. Deep learning and CNN could be used to improve this prototype's ability to reliably identify a person's gender and age range out of a single image of their face. From this study, we can conclude with two important conclusions. First, despite the limited availability of age and gender-tagged photos, CNN can be used to improve age and gender detection outcomes. Second, by employing additional training data and more complex systems, the system's performance can be slightly increased.

9.2 FUTURE WORK

Using other Deep Convolution Neural Network architecture in the same problem in place of VGGNet that we have used in this project. I would have the wished to supplant the different completely associated fully connected layers at the end part of this architecture with just a single layer and rather than this moved those parameters over to extra convolutional layers. By a wide margin, the most troublesome region of this undertaking was building up the preparation foundation with the appropriate division of the information into folds, prepare every classifier, performing cross-approval, what’s more, join the different resulting classifiers into a test-prepared classifier

REFERENCES

- Amit Dhomne, Ranjit Kumar and Vijay Bhan, “Gender Recognition Through Face Using Deep Learning”, International Conference on Computational Intelligence and Data Science (ICCIDS 2018).
- Tivive FHC, A.Bouzerdoum(Sep 2006) ”A gender recognition system using shunting inhibitory convolutional neural networks” In: International Joint Conference on Neural Networks; Vancouver, Canada. New York, NY, USA: IEEE. pp.5336-5341.
- Gil. Levi and T. Hassner(June 2015) ”Age and gender classification using convolutional neural networks.In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops