NTNU
Norwegian University of
Science and Technology

Noman Zahid
NTNU Interview 2nd Round
Department of Computer Science

Task 2: Speaker Classification

Question 1 (a, b):- Create a Dataset

Answer: I have recorded 50 audio samples in my own voice and recorded same number of audio voice samples of our university's Chief Librarian (my mentor) and my colleague (female voice). Each audio sample is of 5 seconds containing different quotes and story lines recorded in English language. For the voice recording, "**.WAV"** format has been used.

Then, I have created 3 folders with the names:

1. Noman
2. Chief
3. Baji

And these sub folders has been placed in a main folder named "My_DataSet" and uploaded main folder at my google drive.

- For the speaker classification/identification, I used python programming language with the Google Colab's environment that runs in the browser using Google cloud.

Question 1 (c):-

Answer: Used "`librosa.load`" to convert the audio file (continuous time signal) into time series (discrete time signal) considering the 44100Hz sampling rate (sr) and extract the features of the audio file as data provided by the audio cannot be understand directly. After this merged all data/samples and created a labelled dataset with the speaker name such as, "Noman", "Baji", and "Chief".

Question 2:-

Answer: Here, I have used the "`train_test_split`" function for splitting (70/30 ratio) data arrays into two subsets: for training data and for testing data.

Question 3 (a):- Which conventional machine learning algorithm will you choose for the above task and why?

Answer: Artificial Neural Networks (ANNs) have been used for the classification/identification of audio recorded data as **ANN** is capable of learning any nonlinear function. ANNs use **Activation Functions** to introduce nonlinear properties to the network and that helps the network to learn any complex relationship between input

and output while for audio data this relationship is complex. Authors in [1] used ANN for the voice classification and feature vector voice signal has been obtained through the Mel Frequency Cepstral Coefficients (MFCC). While the results show that they got 99.96% accuracy. Thus, here I chose the ANN approach to classify/identify the voice of human being.

Question 3 (b):- Show the accuracy, F1-Score and confusion matrix for the algorithm you select in 3(a).

Answer: Results show **100**% accuracy. While the Fig. 1 and Fig. 2 show F1-Score and confusion matrix respectively.

Classification Report

	precision	recall	f1-score	support
Noman	1.00	1.00	1.00	15
Baji	1.00	1.00	1.00	19
Chief	1.00	1.00	1.00	11

Figure 1: F1-Score with 30 epochs.

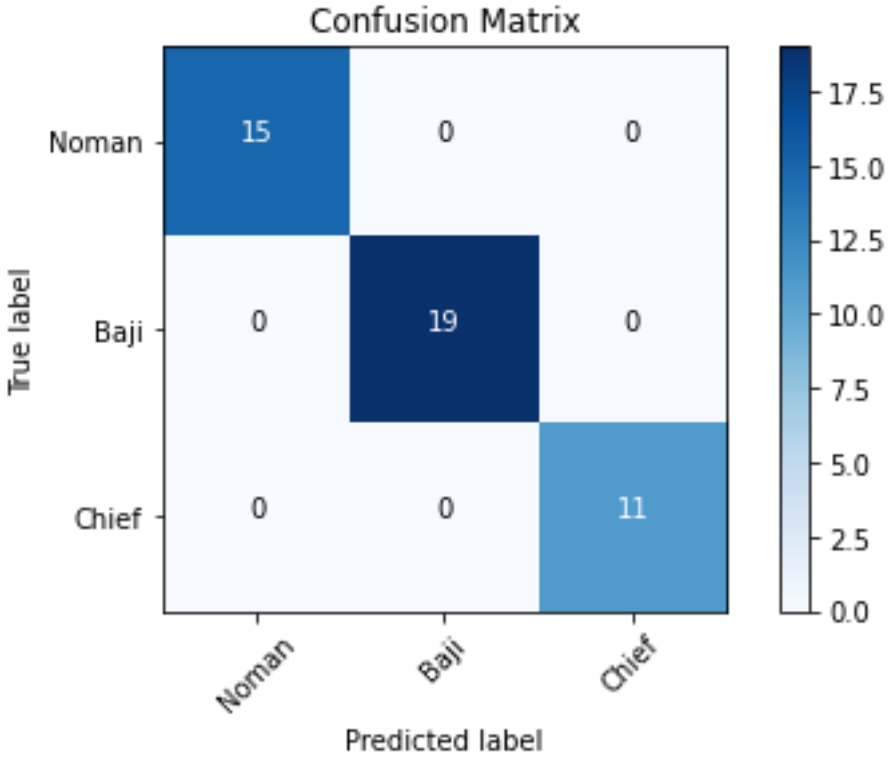

Figure 2: Confusion matrix.

Noman Zahid
NTNU Interview 2nd Round
Department of Computer Science

Question 3 (c):- Perform the same experiment with a Vanilla CNN model.

Answer: I have performed this experiment with the Vanilla CNN model while the results are:

i. The accuracy of Vanilla CNN model for this task is **96%.** While the Fig. 3 and Fig. 4 show F1-Score and confusion matrix respectively.

Classification Report

	precision	recall	f1-score
Noman	1.00	1.00	1.00
Baji	0.83	1.00	0.91
Chief	1.00	0.90	0.95

Figure 3: F1 score of Vanilla CNN model.

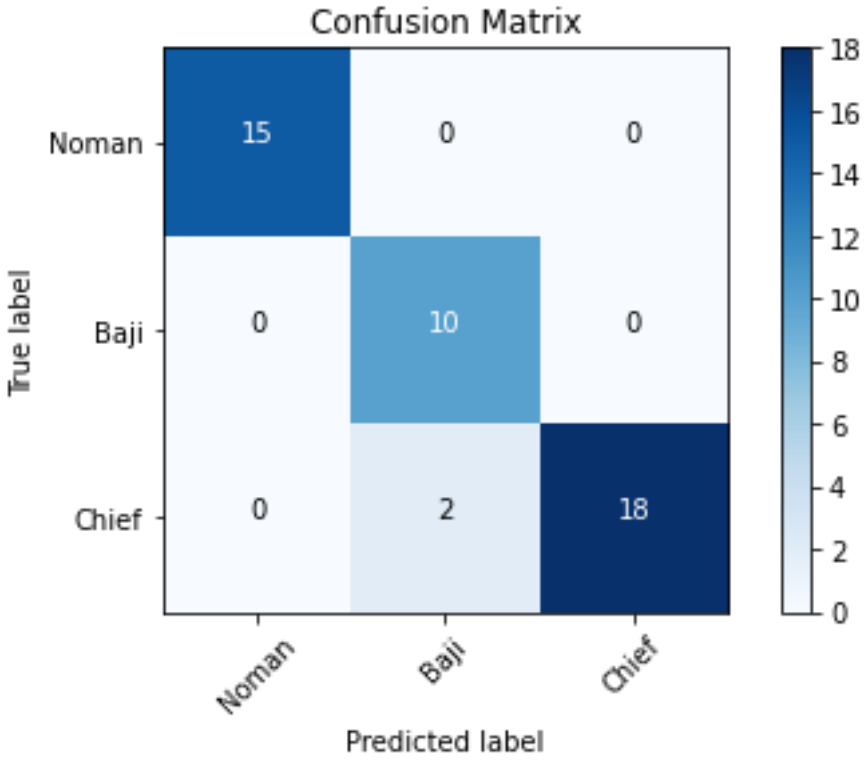

Figure 4: Confusion matrix of Vanilla CNN model.

NTNU
Norwegian University of
Science and Technology

Noman Zahid
NTNU Interview 2nd Round
Department of Computer Science

Question 3(d):- Which model performed better (b) or (c).

Answer: Results show that ANN model performed better as compared to the Vanilla CNN model.

Question 3(e):- Test your models' performance for (b) and (c) by recording 5 new samples in your voice and predict the speaker label. Report the accuracy obtained.

Answer: I have recorded 5 new samples in my own voice (Noman) and predict the speaker label by using both ANN and CNN models

Model	Correct	Incorrect
ANN	3	2
Vanilla CNN	3	2

References:

[1] Shetty, Surendra, Sarika Hegde, and Thejaswi Dodderi. "Classification of healthy and pathological voices using MFCC and ANN." In *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, pp. 1-5. IEEE, 2018.