IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IDENTIFICATION OF FRAUDULENT AUDIO POSTS ON SOCIAL MEDIA

[1]N. Pavani, [2] K. Shyamala

[1]Research Scholar, [2]Professor, Dept. of CSE, UCE, Osmania University

[1]Dept. of CSE, UCE

[1]Osmania University, Hyderabad, India

Abstract: This work focuses on the identification of fraudulent audio posts on social media. There are numerous posts posted on social media every day. It is infeasible to check the credibility of all posts. Audio posts look like spoken words and seem to be reliable to the public. Normal public cannot distinguish between real and tampered audio posts. This work presents a novel approach to detect fake audio on social media. The work uses techniques like MFCC to extract audio features and then the features are used to train an ANN model which finally classifies the posts. The work achieved 99.7% accuracy with real and fake audio dataset from Kaggle.

Index Terms **– MFCC, ANN, Audio, CNN, Librosa.**

I. INTRODUCTION

Numerous amounts of voice recordings are transmitted over the social media. It is challenging to identify the authenticity of audio posts. The fraudulent audio includes spoofing voices of individuals, organization heads, politicians and government bodies. Fake audio is used by fraudsters to manipulate public opinion for propaganda, defamation, or even terrorism [1]. On 27th March 2020 three people were arrested in Nagpur for making fake audio clip reeling about '59 positive' cases of COVID-19 and suspecting 200 more cases. The culprits released a 4.52 minute audio post which created panic in the region [2]. In audio posts, the receiver only knows the content of the message, but some characteristics like rhythm, intonation and genre are easily manipulated to create fake audio [3]. The Deepfake audio was first developed for good reasons like helping people with speech problems and to create audiobooks. But, the use of this technology in a wrong way creates havoc when it addresses the voices of public representatives and leaders. Therefore, it is very much needed to authenticate the audio posts before it is distributed or forwarded to avoid spread of misinformation.

Librosa is a Python package used for audio analysis. Librosa helps to visualize the audio signals. It provides a bundle of functions to work with sound information like automatic speech recognition, music analysis. It consists of various sub-modules like beat, util, core, segment, display, feature etc. Feature sub-module provides extraction of audio features through chromatograms, Mel-spectogram and MFCC[7][8].

The proposed work uses Librosa package to extract audio features and Mel Frequency Cepstral Coefficients (MFCC) features are also extracted from audio. Audio signals cannot be used as input in its raw form due to noise. It is always better to extract features from audio signals and use that as input to neural networks for efficient classification. MFCC is a technique for extracting features from the audio input. MFCC technique generates features from each sample of audio signals; these features are taken as input for the ANN model for classifying audio yielding excellent results in identifying authentic and forged audio.

II. LITERATURE SURVEY

In **Shilpa Lunagaria et al.** [4] authors explain about speaker recognition and explains that it includes both identifying and verifying the speaker. The authors explain that Automatic Speaker Verification (ASV) system verifies the veracity of the speaker, whereas identification only identifies the speaker. Some types of spoofing attacks are Replay, Synthetic Speech (SS), Twins, Voice Conversion (VC) and Impersonation which can be used to create tampered audio posts or modify existing posts. The authors focus on voice generation using Text-To-Speech(TTS) and Voice Conversation (VC) to generate fake voice, Deepfake voice detector pre-processes the voice samples and converts them to spectrograms, finally the input dataset Google ASV Spoof dataset is trained and tested to achieve 85% accuracy.

Bismi Fathima Nasar et. al. [5] authors proposed detection of fake audio, video and images. The paper first processes the images and uses deep learning to detect fraudulent content. Data Preparation, Image Enhancement, CNN Model Generation, and Testing or Detection are the four segments of the proposed work. The video posts are converted to a sequence of frames using OpenCV and the audio posts are converted into spectrogram images using Matplotlib, which are then enhanced using Librosa. finally, the frames are trained and tested using CNN. VidTIMIT, DeepfakeTIMIT and Face Forensics++ datasets were used and achieved 99%, 85% and 90% respectively. The propagation of fake poses create threat to the law, society and privacy.

Tianyun Liu et al. [6] authors proposed two different models to detect fake audio 1) Support Vector Machines(SVM) and Mel-frequency Cepstral Coefficient(MFCC) features combined to classify real and fake audio and 2) a Convolutional Neural Network(CNN) for clssification. The SVM Classifier is trained using the MFCC coefficients (features) extracted from the training set to identify fraudulent audio. The second classifier consists of two CNN models that are trained using the stereo audio posts and the fusion of results is taken to classify the real and fake audio. These two classification results are combined to obtain a final detection result. An accuracy of 99% is achieved on FILM and MUSIC datasets.

III. PROPOSED WORK

Due to technological advancements, generating fake audio has become an easy task for even normal public. The propped work extracts MFCC features from the audio posts and then classifies them using ANN, this improves the accuracy of the model.

3.1 Data Processing

The audio data in its original form contains noise and is not suitable for classification. Hence, the audio posts should be pre-processed before classification. Librosa package is used to pre-process the audio clips. The audio clips are converted into a sorted list of audio files with a uniform sampling rate and resample type as Kaiser_fast. The next step is to extract the MFCC features. The Mel frequency Cepstral coefficients (MFCCs) of a signal are set of features that model the characteristics of audio. They describe the shape of a spectral envelope. A total of 39 MFCC features are extracted from audio clips, which are then scaled by calculating the mean of the features. Finally these features are appended along with labels (real and fake) to a dataframe. This is taken as input to train the ANN model.

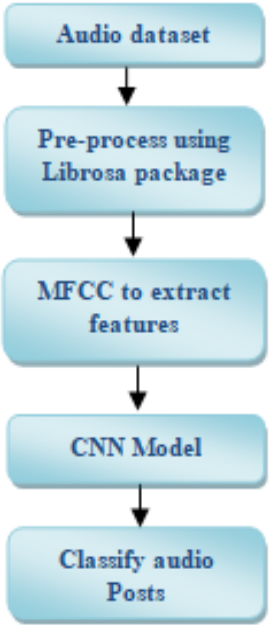

Fig. 1. Framework for Audio classification

Figure 1 describes the framework that is used in the proposed work to classify audio posts. Two datasets, Fake Audio from kaggle and a custom dataset is created with a combination of fake audio taken from ASVSpoof dataset and real audio take from VidTIMIT dataset. Experiments were performed on the datasets to obtain best accuracy. Figure 2 and figure 3 lists MFCC Scaled features of the top 5 and bottom 5 audio posts in the custom dataset along with their class (1 for real and 0 for fake audio).

MFCC coefficients present the change in rate of different spectrum bands. The positive values of cepstral coefficients are depicted by majority of the spectral energy concerted in the low-frequency regions and a negative value represents that most of the spectral energy is concentrated at high frequencies. Mean of these values is considered as an array to predict audio labels.

Experiments proved that choosing the MFCC features as input to the neural network model gives better results compared to raw audio being processed as input. Librosa provides various functions to pre-process the audio.

	feature	class
0	[-696.7611, 53.995308, -3.358449, 18.37832, -0...	1
1	[-638.3319, 50.888287, -13.244806, 11.880238, ...	1
2	[-607.6682, 55.69361, -7.136121, 16.152592, -4...	1
3	[-680.3341, 38.99618, -11.000199, 12.335847, -...	1
4	[-696.3558, 67.93551, 2.273451, 22.290327, 1.2...	1

Fig. 2. MFCC scaled features of original audio

	feature	class
6312	[-340.39563, 101.611664, 19.855385, 46.5141, 9...	0
6313	[-293.92764, 118.926506, 5.021471, 24.076345, ...	0
6314	[-361.59113, 92.15747, 22.426678, 39.82647, 5....	0
6315	[-334.4247, 96.258385, 21.512108, 44.863884, 9...	0
6316	[-324.55823, 92.876785, 23.013445, 38.795315, ...	0

Fig. 3 MFCC scaled features of fake audio

3.2 Developing and training the model

The proposed work uses Artificial Neural Network (ANN) for classification of the real and fake audio. The ANN model developed consists of a combination of Dense layer, Relu activation function and a Dropout; this trio is repeated three times, followed by a flatten layer and an output dense layer. The relu activation function is used in the hidden layers and the output layer used the sigmoid activation function which is best for binary classification. Through experiments, it is finalized that the combination of the layers proved best for audio classification. VisualKeras is an open-source python library which helps us depict the neural network model in a graphical form. This gives a better understanding of the layers and their combinations in the neural networks. Figure 4 presents the ANN model that is generated using VisualKeras which is a colorful, attractive and easy to understand representation of the ANN model.

Fig. 4 Visualization of the optimum ANN model using visualkeras.

Audio classification is not as complex as image or video classification. It takes less computational power and requires a simple model with lesser number of layers to classify the audio input. Thus, by experiments, it was understood that an ANN model is enough and a CNN model is not much required to classify audio. Finally the above ANN model was finalized for audio classification. The proposed model gave better accuracy than state of art models, with the MFCC features taken as input for classification by the ANN model. The figure 5 presents the architecture of ANN model that is derived using the plot_model() method of keras.utils. The architecture depicts the input, output, type of layer and number of units in each layer. This depiction gives a clear understanding of the neural network used.

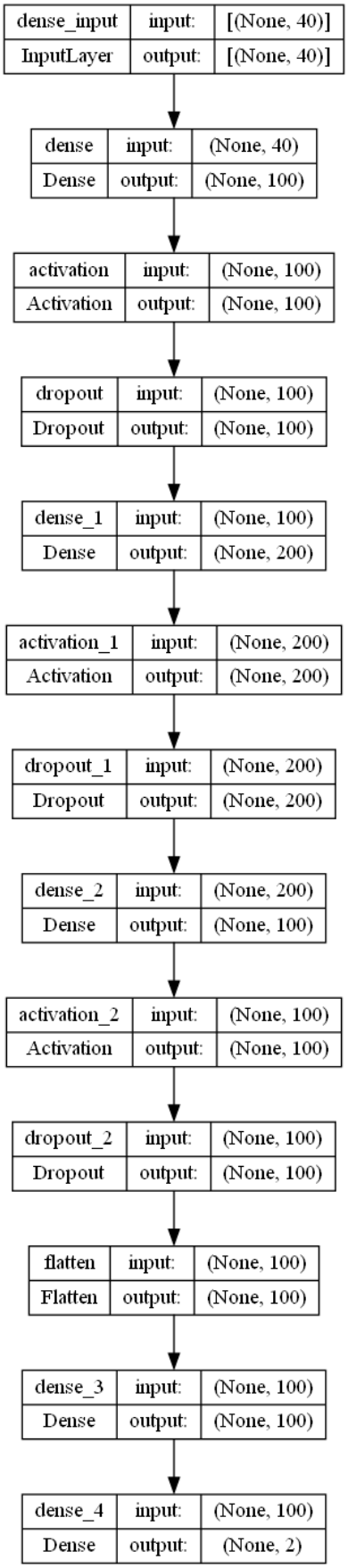

Fig. 5 Architecture of the proposed model

IV. DATASETS

The model is experimented on two datasets. The first is the Real and fake audio dataset from kaggle consisting of fifty original and fifty fake audio clips are considered. each clip of 10 seconds. The second is the custom dataset created with real audio from VidTIMIT and fake audio from ASVSpoof2015. The custom dataset consists of 4635 fake audio clips and 4483 real audio clips; the audio clips vary in size and are in wav file format ranging from 1 to 5 seconds. The ANN model gave best results for real and fake audio detection. Table 1 presents the results obtained by the proposed model with two datasets.

V. RESULTS

Audio Classification included all the techniques like Librosa and MFCC techniques for feature extraction which are fed into the model consisting of 5 Dense layers, 5 activation layers, 3 Dropout layers and one flatten layer. The model is experimented on two datasets: An accuracy of 99.7% and 99.8% was achieved respectively.

Table 1. presents a comparison of the proposed model with the existing works. The model performed well and gave better results even with a simplified model. The table presents the accuracy obtained on test dataset for two datasets used in the proposed work.

Table 1: Comparison of results with existing literature

S.No	Paper	Dataset Size	Dataset	Methodology	Accuracy
1	Shilpa et. al. [4]	25000 audio clips	ASV Spoof	Spectrogram, CNN	99%
2	Tianyun et. al.[6]	29,284 audio clips on 1sec.	FILM and MUSIC corpus	MFCC, SVM, CNN	99%
3	Proposed _Model	Real-4483 Fake-4635	VidTimit & ASV Spoof	MFCC, ANN	99.8%
4	Proposed _Model	Real-2200, Fake- 2200	Real and Fake audio dataset- Kaggle	MFCC, ANN	99.7%

The present work achieved very good results even with a very simple model with few layers. 99.8% accuracy was achieved in just 10 epochs. Figure 6 presents the training accuracy, validation accuracy and training loss, validation loss that the proposed model achieved with the test dataset of the custom dataset used.

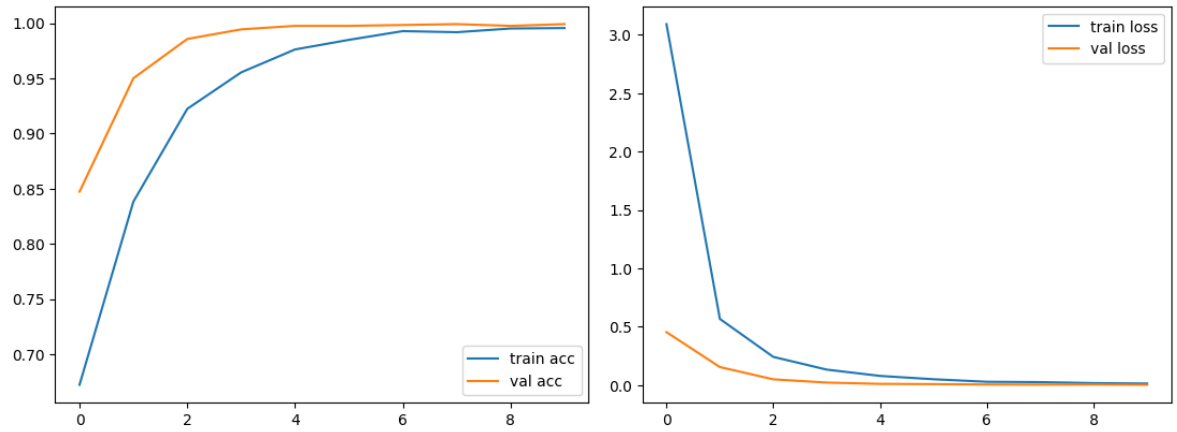

Fig. 6 Training, Validation accuracy and Training, Validation loss of custom dataset.

VI. CONCLUSION

The proposed model first pre-processes the audio data for removal of noise. It then extracts features from the audio using MFCC. Mel-Frequency Cepstral Coefficients (MFCC) helps retrieve best features that are helpful in identifying fraudulent content. Then a simple ANN model with Dense and Dropout layers gave very good accuracy. The proposed work worked on two datasets. Real and Fake audio from kaggle and achieved 99.7% accuracy. A custom dataset was created with real audio from VidTIMIT and fake audio from ASVSpoof and achieved an accuracy of 99.8%.

REFERENCES

[1] Almutairi, Zaynab, and Hebah Elgibreen. "A review of modern audio deepfake detection methods: challenges and future directions." Algorithms 15, no. 5 (2022): 155.

[2] https://timesofindia.indiatimes.com/videos/city/mumbai/mumbai-social-media-influencer-arrested-for-faking-his-suicide/videoshow/84723993.cms, last accessed 2023/07/29.

[3] Ballesteros, Dora M., Yohanna Rodriguez-Ortega, Diego Renza, and Gonzalo Arce. "Deep4SNet: deep learning for fake speech classification." Expert Systems with Applications 184 (2021): 115465.

[4] Shilpa Lunagaria, F., Mr. Chandresh Parekh, S.: FAKE AUDIO SPEECH DETECTION. IJIRT, Volume 7 Issue 1, ISSN: 2349-6002 (2020).

[5] Bismi Fathima Nasar, F., Sajini T, S., Elizabeth Rose Lason, T.: Deepfake Detection in Media Files - Audios, Images and Videos. IEEE Recent Advances in Intelligent Computational Systems (RAICS) | December 03-05 (2020).

[6] Tianyun Liu, F., Diqun Yan, S., Rangding Wang, T., Nan Yan, and Gang Chen.: Identification of Fake Stereo Audio Using SVM and CNN. Information 12, no. 7: 263 (2021).

[7] Pavel Korshunov, F., Sébastien Marcel, S.: Speaker Inconsistency Detection in Tampered Video. 26th European Signal Processing Conference (EUSIPCO), (2018).

[8] Mcuba, Mvelo, Avinash Singh, Richard Adeyemi Ikuesan, and Hein Venter. "The Effect of Deep Learning Methods on Deepfake Audio Detection for Digital Investigation." Procedia Computer Science 219 (2023): 211-219.

[9] Reimao, Ricardo, and Vassilios Tzerpos. "For: A dataset for synthetic speech detection." In 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), pp. 1-10. IEEE, 2019.

[10] S.Anitha Jebamani, S.Gomathi, Dr.Soma Prathibha, Deepika Sree D , Preetha S. " Detection Of Fake Audio", Ilkogretim Online - Elementary Education Online, 2020; Vol 19 (Issue 4): pp. 6813-6819 http://ilkogretim-online.org doi: 10.17051/ilkonline.2020.04.765085.

[11] Almutairi, Zaynab, and Hebah Elgibreen. "A review of modern audio deepfake detection methods: challenges and future directions." Algorithms 15, no. 5 (2022): 155.

[12] Hamza, Ameer, Abdul Rehman Rehman Javed, Farkhund Iqbal, Natalia Kryvinska, Ahmad S. Almadhor, Zunera Jalil, and Rouba Borghol. "Deepfake audio detection via MFCC features using machine learning." IEEE Access 10 (2022): 134018-134028.