5th International Conference on Electronics and Sustainable Communication Systems (ICESC 2024)
IEEE XPlore Part Number: CFP24V66-ART; ISBN: 979-8-3503-7994-5

Recognition of Deep Fake Voice Acoustic using Ensemble Bagging Model

Shaik Abdul Yunus Basha
Research Scholar,
Department of Computer Science and Engineering,
Vels Institute of Science Technology & Advanced Studies,
yunusbasha@gmail.com

Dr.K.Ulaga Priya
Associate Professor,
Department of Computer Science and Engineering,
Vels Institute of Science Technology & Advanced Studies)
Chennai, Tamilnadu, India
upriya.se@velsuniv.ac.in

***Abstract*—As artificial speech generates technological evolution; fake voice information has become an increasingly prevalent avenue for fraud. Numerous investigations have been carried out into machine learning techniques, signal processing techniques, and the Automatic Speaker Verification System to address such issues. Since there are increasing methods to create fake information, techniques must be developed to determine when audio recordings employed as electronic proof are corrupted. Such recordings contain the potential to be manipulated. Hence, this research work identified the data regarding audio systems as either fake or real by i) extracting features from audio file metadata using oversampling techniques and ii) Developing an Ensemble Bagging Model appropriate for analyzing input audio signal followed by classification of fake (fraudulent) and real (authentic) audio. Our experimental results reveal that the ensemble bagging model attained an accuracy of 99.5%, precision of 99%, 99.6% recall, and 99.9% F1-measure in the prediction of the audio system compared with traditional approaches.**

Keywords— Ensemble Bagging Model (EBM), Zero Crossing Rate (ZCR), Validation accuracy and loss, Sequential model, Automatic Speaker Verification, Audio Signal Processing, Data Efficiency

I. INTRODUCTION

Several researchers identified deep fakes on audio and video frames using deep learning algorithms. From that, numerous instances are observed in which they have been exploited to distribute deceptive data. Fakes have been purposefully utilized to disseminate false information for several years now. The latest developments in automated speech, including voice translation techniques [1] that make it more feasible to recreate speech from human beings using signal processing techniques Yohanna [2], paving the path for a time when sounds will serve an identical function in audio, deep fake detection as in the video as well. Moreover, several automatic voice detection systems of pathological voice were developed for deep fake detection in the field of voice forensics and AI techniques.

The primary research motivations are outlined as follows. Deep audio fakes are the most common cause of deception due to advancements in synthesized speech production technologies. Consequently, identifying the difference between fraudulent and authentic sounds is getting harder. This is why it's critical to have a system that can quickly distinguish between real and fake sounds. Prior studies have been conducted on it. The current approaches require a lot of computation. Therefore, the recommended strategy combines the most effective feature extraction/ selection techniques and the Ensemble Bagging model, producing higher-quality categorization outcomes faster when compared with conventional techniques. The novelty of this research includes the features of audio files that are validated without training and are merely based on the extraction of features such as MFCC, Chroma, Spectral contrast and Centroid, and zero crossing rate.

A. Research contribution

The proposed research work suggests a technique for identifying fake information related to sound and differentiating it from authentic or fraudulent audio. The specific contributions are as follows.

1) A new technique that uses machine learning classifiers to discriminate between authentic and fraudulent audio in Kaggle metadata has been introduced.
2) To obtain the most relevant characteristics from an audio file, the best feature extraction/feature selection methodologies are used.
3) According to the experimental results, the proposed approach can precisely identify real-life false audio.
4) The suggested approach performs 30% more accurately than conventional methods.

B. Organization of the Research study

An introduction of the issue and the study context are provided in Section 1, while Section 2 describes the background of review. With the experimental setup described in Section 3, the proposed method is provided as reliable and robust. Section 4 provides a brief overview of the main results, demonstrating the effectiveness of the proposed approach. Summarize the findings and propose areas for further studies, such as investigating more complex deep learning architectures, increasing the variety of datasets, and incorporating real-time processing, as detailed in section 5.

II. RELATED WORKS

Reviewed various articles regarding deep fake video/image detection using several deep learning algorithms [3]. Surveyed 67 research study regarding deep fake detection models using machine learning and fusion models, providing guidelines for future work [4]. Used a deep-based CNN model and applied a YOLO face detector to detect and classify the frames in fake videos [5]. Utilized an unsupervised machine-learning model for deep fake video detection [6]. Moreover, Abdullah detected fake news using a deep-based sequential model [7]. Developed machine learning models for detecting deep fakes by MFCC feature, which extracts relevant information from the noisy signal [8]. Among several machine learning models, VGG-16 attained a maximum accuracy of 93% with MFCC 40 features, 67%, and 73.4% accuracy by the Support Vector Machine with MFCC 20 features and Timbre model analysis feature by [9-10] respectively. Focused on how deep fake video avoidance can be done using various deep-based recognition approaches, especially blockchain technology [11]. Employed blockchain technology for detecting fake videos in digital media [12-13]. Authentication content is efficient, enhancing the user's trust. Introduced a hybrid of Mobile Edge computing and adaptation methods for streaming videos with video streaming support [14]. Offered an approach and a general structure for tracking the beginning and development of digital media back toward its genuine source, regardless of cases where it has been replicated repeatedly by means of Ethereum smart contracts [15].

III. PROPOSEDMETHODOLOGY

The unique approach involves integrating ensemble bagging models for deep fake speech recognition to achieve high accuracy and reliability in fraud detection. This method combines advanced feature extraction techniques and thorough validation methods. Our proposed methodology framework for Fake audio recognition is depicted in Figure 1. In this work, the authors employed fake video metadata to extract audio files to conduct trials. It comprises five phases, namely data pre-processing and extraction/selection of features for deep fake audio detection. The acoustic signals transform the time to frequency region to evaluate the sound visually. The most crucial ways of extracting features are applied to extract the significant features from an audio stream. To narrow the selection process to only significant aspects, the sound stream becomes normalized, reducing its size. The audio metadata is categorized into the training phase and testing phase, in which 20% of the entire metadata is used for measuring the efficiency of the prediction, while the remaining 80% is used to train the models. The authors developed an Ensemble bagging model to identify audio metadata's authentic or fraudulent sound.

A. Description on dataset

The study utilizes metadata from the Kaggle source, which comprises video (mp4) files with a size of 471.54 GB divided into a compressed set of 10 GigaBytes per file. This video file consists of the following columns: Filename, Label (Authentic/Fraudulent), Original, and split columns.

TABLE I. DATASET-VIDEO FILE

Name of file	Label	Original	Split
Mentioned the name of the video file	Represented the findings of the audio file as authentic or fraudulent	The original videos are scheduled here; among that train data are fraudulent	Equalized to train data

TABLE II. DATASET DESCRIPTION

Kinds of data taken	Audio Dataset link
Real Audio	"/kaggle/input/deep-voice-deepfake-voice-recognition/DEMONSTRATION/DEMONSTRATION/linus-original-DEMO.mp3"
Fake Audio	"/kaggle/input/deep-voice-deepfake-voice-recognition/DEMONSTRATION/DEMONSTRATION/linus-to-musk-DEMO.mp3"

The overall file comprises training samples based on video, from which audio files are extracted. Further implementation has been done, predicting the video's probability of being either authentic or fraudulent.

B. Proposed method for Fake Audio recognition

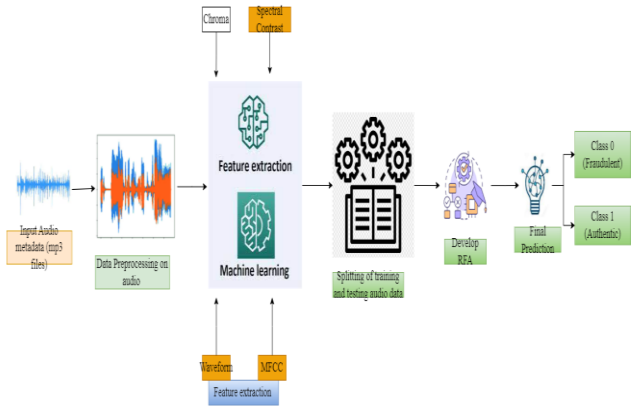

Fig. 1. Proposed Framework for Fake Audio Recognition

The audio files are loaded to extract significant features such as waveform, Mel-Frequency Cepstral Coefficients, Chroma feature, Spectral contrast, zero crossing rate, and spectral centroid. Figure 2 depicts such feature visualization of waveform.

1) Display the waveform

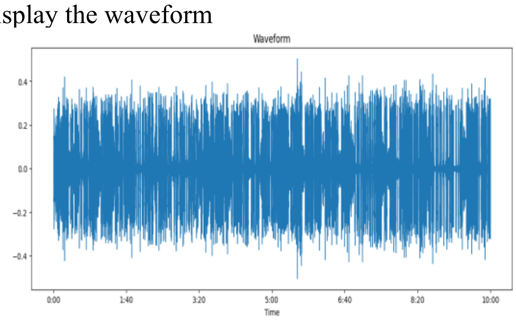

Fig. 2. Waveform representation for detecting audio signals

2) **Extracting Mel-Frequency Cepstral coefficients:** A popular feature extraction method in speech and audio processing is called MFCC. MFCCs represent the spectrum properties of sound in a fashion that can be useful for various applications based on machine learning, including analyzing music and recognizing speech, as depicted in Figure 3. MFCCs are the collection of coefficients that represent the form of an audio signal's frequency range. They are obtained by utilizing a method such as the Discrete Fourier Transform (DFT) to convert the initial audio signal

into the sound's frequency domain and then using the mel-scale to simulate how the human ear perceives a sound's frequency. Ultimately, the mel-scaled spectrum is used to compute cepstral coefficients. This can be calculated using eqn. (1).

$$mel\,(f) = 1127 \ln\left(1 + \frac{f}{700}\right) \quad (1)$$

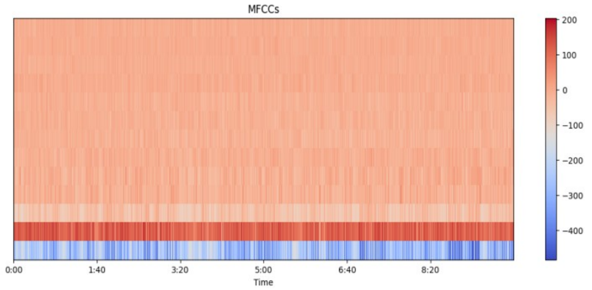

Fig. 3. Extraction of MFCC feature from Audio

C. Extracting chroma feature

This feature extraction is essential for several audio analysis and processing activities, especially those involving the comprehension and alteration of sound. It offers a brief overview of acoustic material, making it ideal for audio analysis-related musical data-retrieving activities. Figure 4 depicts its representation.

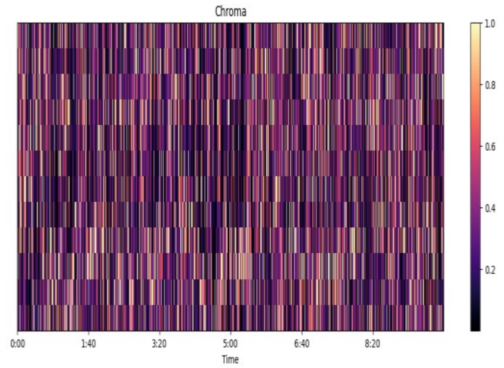

Fig. 4. Representation of chroma feature extract from audio data

1) **Extracting Spectral contrast feature:** Contrast characteristics can generally offer beneficial details underlying the layout and content of audio data for a variety of acoustic computation and analysis applications, such as sound identification, sound quality analysis, and music-related data retrieving. Figure 5 depicts the representation of spectral contrast features extracted from audio data.

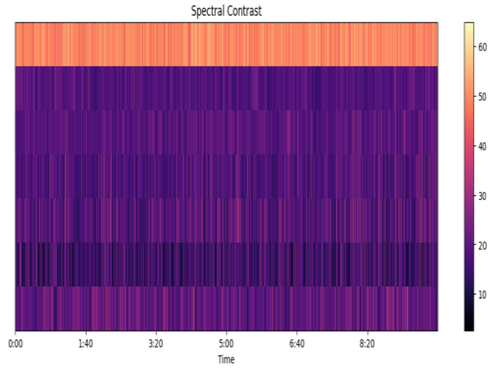

Fig. 5. Signal representation based on spectral contrast feature.

2) **Extract Zero Crossing Rate (ZCR):** The purpose of extracting the zero crossing rate feature from audio data files is to offer significant information about analyzing specific activities, including evaluation of pitches, tone of voice recognition, analysis of timbral, identification of assault, and segmenting audio files, as shown in Figure 6.

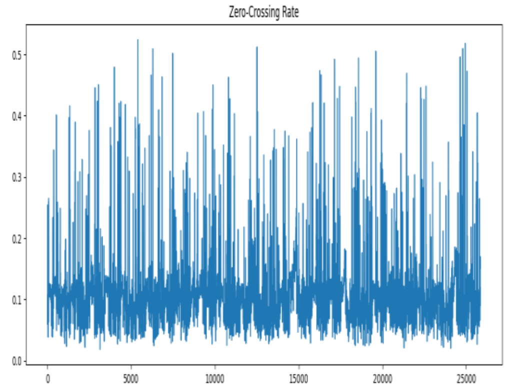

Fig. 6. Signal representation of zero crossing rate

D. Data preprocessing

Preparation processing is essential to achieve higher-quality outcomes in machine learning. The main objective of preprocessing acoustic information is to convert the input audio stream onto a structure the machine learning model can understand. For this reason, it's critical to carry out conventional processing procedures prior to the model's deployment. In this stage, duplicated acoustics are eliminated in metadata and standardize the audio characteristics. The quantity of an audio signal termed acoustic recording per second is obtained and set the sample rate of an audio signal to 44100. After Preprocessing, the real data are classified into 8 classes and 56 in fake audio files. The label encoder approach is simple and effective in converting the categorical value into a numerical value. During this phase, the data samples are labelled CLASS 0, well thought-out as fake audio, and CLASS 1, considered real audio.

E. Feature Extraction and Selection

Integrating instances representing the minority class into the training dataset over arbitrary, along with a substitute, is known as random oversampling. The editors reviewed the most popular techniques to acquire knowledge from unbalanced courses. They asserted that three key problems:

error costs, class dissipation, and accuracy are essentially to blame for the poor effectiveness of the algorithms developed by conventional machine learning techniques on unbalanced classes. Resampling techniques were proposed because it was hard to find the minority targets. A novel replication technique based on artificial conditions resulting from class-specific sub-clustering results in a proportional over-sampling of uncommon positives and an under-sampling of non-infected majorities. They claimed that their novel resampling method outperformed conventional random resampling in performance. The difference between under sampling/oversampling, which highlights majority and minority classes from audio samples, is shown in Figure 7.

Fig. 7. Differentiating undersampling and Oversampling

F. Splitting dataset into training and testing

The audio dataset is split into a Training phase of 80% and a validation phase of 20%. The audio files are trained to extract features and apply machine learning models, and finally, validation has been performed.

G. Develop machine learning techniques.

This section describes the development of a machine learning model, namely a sequential model that sequences data input/output. An ensemble bagging model has been introduced for differentiating input audio files into authentic and fraudulent.

H. Sequential Model

Sequence models are known as machine learning models that generate or receive information in patterns. Text streams, multimedia snippets, longitudinal data, historical data, and other types of data are examples of sequential data. This model helps to extract various patterns from input audio data that are effectively relevant to audio dataset classes, such as fraudulent (Class 0) and authentic (Class 1). The summary of the sequential model is illustrated in the figure, in which the layer comprises 256, 128, and 2, respectively. Moreover, layers 2 and 5 are activation layers, and layers 1 and 4 are dense layers that help extract more complex patterns from audio data via interlinked neurons. Layers 3 and 6 are dropout layer supports to avoid overfitting by randomly deactivating the neurons during the training phase. Here, the entire 71426 samples are trainable parameters using this model.

Layer (type)	Output Shape	Parameters
dense (Dense)	(None, 128)	5248
activation (Activation)	(None, 128)	0
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 256)	33024
activation_1 (Activation)	(None, 256)	0
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
activation_2 (Activation)	(None, 128)	0
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 2)	258
activation_3 (Activation)	(None, 2)	0

Total params: 71426 (279.01 KB)
Trainable params: 71426 (279.01 KB)
Non-trainable params: 0 (0.00 Byte)

H. Develop Ensemble Bagging Model

Here, the authors applied an ensemble bagging model where bootstrapped samples are produced to classify fake audio data into fraudulent and authentic. Such fake audio classification is identified by calculating the highest voting majority received by every class. Ensembling Bagging can be evaluated using eqn. (2)

$$\hat{g}_{bagg} = \hat{g}_1(X) + \hat{g}_2(X) + \cdots + \hat{g}_b(X) \qquad (2)$$

Whenever individuals exhibit overfitting (i.e.) when slight variations in training fake audio datasets result in significant variations in anticipated outcomes, bagging becomes particularly effective. Combining the individual learners of various statistical features, like mean and SD Standard Deviations, reduces the variance. For highly unpredictable models, like decision trees, it performs effectively. The technique does not affect learners if applied to minimal variability models like linear regression. The properties of the fake audio dataset determine how many base learners (trees) should be selected. While overfitting is not a result of utilizing excessive trees, it can be computationally demanding. Figure 8 demonstrates the bagging ensemble model on an audio dataset with bootstrap samples for identifying fake audio using weak learners fitted on each sample.

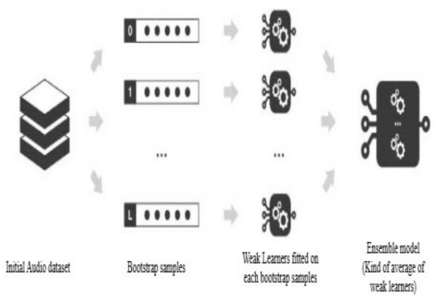

Fig. 8. Bagging Ensemble Model for Fake Audio Detection

Algorithm 1: Ensemble Bagging Model (EBM)

Input: Training data T with correct labels $\propto_1, \propto_2 \ldots. \propto_C$ where C indicates classes of samples
WLA- Weak Learning Algorithm
N- Number of iterations
Percent (P) – To create bootstrapped training data
Class 0 --→ Fraudulent
Class 1 --→ Authentic
Do n =1…. N

Take Bootstrapped replica T_n by choosing randomly P percent of T

Call WL with T_n and receive the hypothesis h_n

Add h_n to ensemble, E

End
Test case: Simple Majority Voting – Audio data samples with unlabeled Z
Ensemble estimation E = $\{h_1, \ldots. h_n\}$ on Z

$$Let\ vot_{n,m} = \begin{cases} 1 & if\ h_n takes\ the\ class\ \propto_m \\ 0, & Otherwise \end{cases}$$

Be the voting given to the class $\propto_m$ by the classifier
Achieve entire voting accepted by every class

$$V_m = \sum_{n=1}^{N} vot_{n,m} \quad m= 1, \ldots., C$$

Finally, the fake audio prediction was made based on the highest voting majority received by every class.

I. Validating audio data

The input audio metadata is validated using sequential and Ensemble Bagging Models during this phase. While validating the sequential model, only trained features, around 80% of data, are tested, whereas, in Bagging, all input audio files are tested based on feature extraction in which the features are not trained. Even though the features are untrained, the Ensembling Bagging model ensures both reliability and robustness in the prediction of fake audio signals with accuracy 100% named as authentic (Class 1) and fraudulent (Class 0).

IV. Results and Discussions

Standard performance metrics such as accuracy, precision, recall, and F1 score are employed in machine learning models to assess an algorithm's efficacy on an audio file dataset. These are a few typical predictive modeling effectiveness specifications. Such metrics are evaluated using the formulas (3), (4), (5), and (6).

$$\text{Accuracy}=\frac{\text{Amount of positively predicted with distinguished samples}}{\text{Entire voice acoustic samples}}\times 100\% \quad (3)$$

$$\text{Precision}=\frac{\text{Truly identified positive samples}}{\text{Positively predicted samples}} \quad (4)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad (5)$$

$$\text{F1-measure}=2*\frac{\text{Precision*Recall}}{\text{Precision+Recall}} \quad (6)$$

A. Training and validation Accuracy

Training and validation accuracy are tracked throughout the process to evaluate the model's performance and direct the training process. Achieving the highest precision when minimizing excessive fitting is the aim for both training and validation of audio metadata. The model's predictive accuracy can be enhanced by avoiding over-fitting and utilizing strategies like premature stopping and normalization. To achieve normalization, features may be scaled to a standard range normalization can be used. This will ensure that the data is distributed consistently and that the model performs well.

B. Training and validation loss

Both losses are evaluated to predict the model's performance in fake voice acoustic detection. Our approach recognizes the fake audio metadata, so a smaller number of losses are attained.

C. Experimental Analysis

According to categorization outcomes, the ensemble bagging model achieved an accuracy rate of 99.5% in the prediction of deep fake audio, whereas the sequential model attained 65.2% accuracy. Figure 9 illustrates the sequential model accuracy of both training and validation concerning epochs. Figure 10 represents both training and validation accuracy with respect to epochs to predict the overall performance of the layer-based ensemble bagging model.

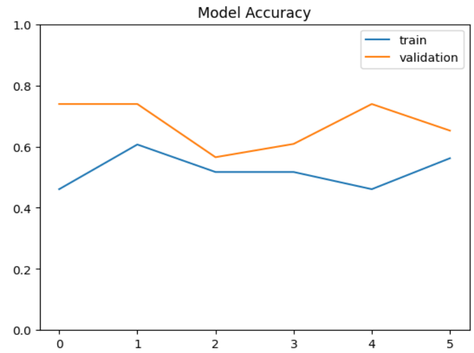

Fig. 9. Comparison of training and validation accuracy with Epochs for sequential model

The graph shows the comparison of training and validation accuracy using a sequential model to detect fake audio data samples over 20 epochs.

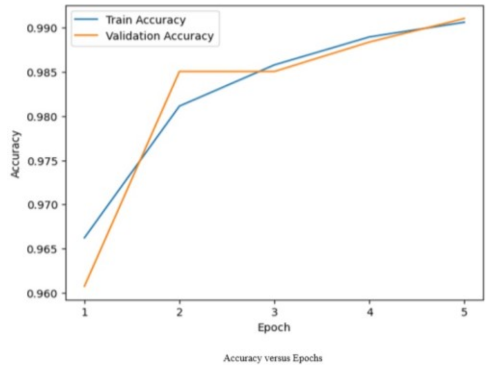

Fig. 10. Ensemble bagging model comparison on training and validation accuracy

The graph shows the comparison of training and validation accuracy with 20 epochs using an ensemble bagging model in the detection of fake audio data samples reached a maximum accuracy of 99.5%. In accordance with the outcomes stated above, Ensemble bagging performs better than the sequential model in terms of accuracy; the Ensemble bagging achieves 100% accuracy while the sequential model only reaches 65.2% and a validation loss of 0.6 milliseconds. This is because of each algorithm's unique traits and behaviors, as well as how they affect the information that is used. According to our dataset, the Ensemble bagging algorithm gives some features more weight than others. Therefore, the need for more features employed is crucial to classification accuracy.

The sequential model requires significantly more effort and time. Regarding classification, the Ensemble bagging method outperforms the sequential model when using this fake voice audio dataset. Furthermore, according to our results, the Sequential model predicts results faster than the Ensemble bagging, executing in 28 minutes compared to 11 minutes for the Ensemble bagging. The Ensemble bagging achieved 99.5% accuracy, 99% precision, 99.6% recall, and 99.9% F1-measure in predicting real or artificial sounds, resulting in improved real-time prediction. A comparison of existing approaches and proposed methods for fake audio detection using machine learning techniques is described in Table III.

TABLE III. OVERALL ASSESSMENT OF FAKE AUDIO RECOGNITION

Authors	Dataset	Methods Applied	Accuracy
Ameer et al. [1]	Fake audio metadata	VGG-16	93%
Khochare et al. [9]	Audio file	SVM with MFCC 20 features	67%
Reimao et al. [10]	Audio file from open source	SVM with timbre analysis model feature	73%
Proposed Method 1	Kaggle Audio file metadata	Sequential Model	65%
Proposed Method	Kaggle Audio file metadata	Ensemble Bagging model	99.5%

To minimize loss and obtain high accuracy in the system, it is essential to begin with a strong, varied, and balanced dataset. Thoroughness is required to extract features from audio metadata. Securely optimize hyperparameters using Ensemble Bagging Models. Successful model improvement and refinement require regularisation methods, cross-validation, suitable loss functions, and continuous performance monitoring. Performance may be enhanced by incorporating advanced ensemble approaches, optimizing model parameters by rigorous tuning, and enhancing feature extraction methods.

Maintaining accurate datasets, using strong validation techniques like cross-validation, and including error analysis to fully understand model behavior are several methods to achieve dependability. Improve the system's dependability and performance via continuous monitoring and customization. Maximize algorithm complexity, simplify data preparation, and increase efficiency. The goal of feature selection methods such as principal component analysis (PCA), recursive feature elimination (RFE), and mutual information is to reduce dimensionality while retaining crucial information to classify audio authenticity accurately.

V. CONCLUSIONS

The system created using Ensemble Bagging Models for deep fake speech acoustic detection has shown to be very capable of correctly detecting between real and fake audio recordings. The system attained high accuracy, precision, recall, and F1-measure, demonstrating its usefulness in combatting voice-based fraud via rigorous feature extraction from audio metadata and painstaking model optimization. Reliable performance across varied datasets is ensured by including sophisticated ensemble algorithms and rigorous validation methodologies, enhancing its practical use in real-world applications. More improvements and expansions may be made to the system in the future. To enhance feature representation and model performance, future research may investigate the integration of sophisticated deep learning architectures such as RNNs and CNNs. Adding real-time processing capabilities would make the system flexible and useful in real-time authentication systems and other dynamic situations where audio data is transmitted. Improving the system's resilience to novel threats requires research into physical attacks on speech recognition systems and developing strong responses. The flexibility and accessibility of the system may be enhanced by increasing the dataset to include a wider variety of languages, dialects, and speech variances. It might improve the system's usability for forensic applications and provide compliance with legal standards.

REFERENCES

[1] A. Mohammed, A. Mansour, M. Ghulam, Z. Mohammed, T.A. Mesallam, K.H. Malki, and B. Mohamed, "Automatic Speech Recognition of Pathological Voice," Indian Journal of Science and Technology, vol. 8, no. 32, pp. 1–6, 2015,

[2] A. Komathi, S.R. Kishore, A.K. Velmurugan, A.S. Begum, D. Muthukumaran, "Network load balancing and data categorization in cloud computing," Indonesian Journal of Electrical Engineering and Computer ScienceThis link is disabled, vol. 35, no. 3, pp. 1942-1951, 2024.

[3] A. Mary, and A. Edison, "Deep fake Detection using deep learning techniques: A Literature Review," International Conference on Control, Communication and Computing, pp. 1–6, 2023,

[4] M. Pravin Kumar, T. Jayaraman, M. Senthilkumar, and A. Sumaiya Begum. "Performance Investigation of Generalized Rain Pattern Absorption Attention Network for Single-Image Deraining," Journal of Circuits, Systems and Computers vol. 32, no. 13, pp. 1-12, 2023.

[5] F. Mira, "Deep Learning Technique for Recognition of Deep Fake Videos," IEEE IAS Global Conference on Emerging Technologies, GlobConET, pp. 1–4, 2023,

[6] S.B. Akbar, K. Thanupillai, and S. Sundararaj, "Combining the advantages of AlexNet convolutional deep neural network optimized with anopheles search algorithm based feature extraction and random forest classifier for COVID-19 classification, " Concurrency and Computation: Practice and Experience, vol. 34, no. 15, pp. 1-9, 2022.

[7] A. M. Ali, F. A. Ghaleb, B. A. S. Al-Rimy, F. J. Alsolami, and A. I. Khan, "Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique," Sensors, vol. 22, no. 18, 2022.

[8] S.T. Ramya, V.P.G. Pushparathi, D. Praveena, A.S. Begum, B. Kalpana, and T.P. Swamy, "Roadmap of integrated data analytics–practices, business strategies and approaches." Demystifying Graph Data Science: Graph Algorithms, Analytics Methods, Platforms, Databases, and Use Cases, vol. 48, no. 217, pp. 1-12, 2022.

[9] J. Khochare, C. Joshi, B. Yenarkar, S. Suratkar, and F. Kazi, "A deep learning framework for audio deepfake detection," Arabian Journal for Science and Engineering, vol. 47, pp. 3447-3458, 2021.

[10] S.B. Akbar, K. Thanupillai, and V. Govindarajan, "Forecasting Bitcoin price using time opinion mining and bi-directional GRU," Journal of Intelligent and Fuzzy Systems, vol. 42, no. 3 ,pp. 1825-1833, 2022.

[11] M.M. Rashid, S.H. Lee, and K.R. Kwon, "Blockchain Technology for Combating Deepfake and Protect Video/Image Integrity", Journal of Korea Multimedia Society, vol. 24, no. 8, pp. 1044-1058, 2021.

[12] A. S. Begum, and S. Poornachandra, "Curvelet based image de-noising using beta-trim shrinkage for magnetic resonance images," 2014 International Conference on Science Engineering and Management Research, pp. 1-8, 2014.

[13] M.Liu, F.R.Yu, Y.Teng, V.C.M.Leung, and M.Song, "Distributed Resource Allocation in Blockchain-Based Video Streaming Systems With Mobile Edge Computing," IEEE Transactions onWireless Communications, vol. 18, no. 1, pp. 695-708, 2019.

[14] A. Zelensky, V. Voronin, E. Semenishchev, I. Svirin and A. Alepko, “Video Content Verification Using Blockchain Technology,” IEEE International Conference on Smart Cloud, pp. 208-212, 2018.

[15] H.R. Hasan, and K. Salah, "Combating Deep fakeVideos Using Blockchain and Smart Contracts, "IEEEAccess, vol. 7, pp. 1-10, 2019.