iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: III Month of publication: March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78465

Hybrid Spectro-Spiral Graph Attention Framework for Robust Audio Copy-Move Forgery Detection

Ms. G. Sathya[1], Jeevanantham B[2], Nithish Kanna P[3], Ragunath R[4]
Department of CSE, SRM Valliammai Engineering College, Kattankulathur

Abstract: The reliability of digital audio is increasingly a critical aspect in digital forensic analysis, legal proceedings, digital media authentication, and digital communication services. Among all forms of digital audio forgeries, copy-move forgeries are a critical challenge in digital audio analysis because of their similarity to the original and copied segments of digital audio. Since the copy-move forgeries are created by duplicating a segment of digital audio, they possess identical speaker characteristics, background noises, and environmental sounds. This similarity makes it difficult to identify copy- move forgeries using conventional methods. This paper proposes a novel "Hybrid Spectro-Spiral Graph Attention Framework" to identify copy- move forgeries in digital audio with improved computational efficiency. The proposed system incorporates a novel "adaptive spectral attention mechanism" for identifying suspicious regions in digital audio, a "Differential Evolution optimization method" for selecting frequency bands, a "Spiral-based graph encoding method" for structural representation of spectral features, and a "multi-head Graph Attention Network" for relational feature learning. The proposed system is a novel improvement over conventional methods like "keypoint matching" and "swarm optimization". Experimental evaluation of the framework under various compression levels and additive noise conditions shows better accuracy in terms of detection and reduced complexity in terms of runtime compared to the existing PSO-based systems. The framework shows better generalization performance and is computationally feasible for CPU-based implementation in digital forensic systems. The proposed system is helpful in ensuring audio authenticity and strengthening digital forensic systems.

Keywords: Audio Forensics, Copy-Move Forgery, Spectrogram Analysis, Graph Attention Network, Differential Evolution, Spectral Attention.

I. INTRODUCTION

With the proliferation of various digital media technologies, there are significant changes in the ways audio content is recorded, stored, transmitted, and analyzed. Audio recordings are widely used in legal evidence submission, investigative journalism, surveillance activities, and social media communication. With the increasing popularity of digital audio editing tools, there is a significant increase in the malicious usage of audio content.

There are various ways to forge audio content, and it can be generally classified into splicing, voice cloning, and copy-move audio forgery. Among these, copy-move audio forgery is considered to be hard to detect. In this type of audio forgery, a portion of the audio content is copied and placed in another location in the audio file. As both audio segments are recorded from the same source, they have similar spectral characteristics and noise patterns.

For instance, in the conversation, the sentence “I confirm the agreement” can be repeated and inserted later to change the context of the conversation. Conventional methods like inconsistency in noise or device fingerprinting cannot be effective, as the repeated sentence will be coherent with the environment.

The existing detection methods include the use of the spectrogram transform followed by feature matching. Feature matching includes the use of SIFT keypoint detection and comparison. Optimization methods like Particle Swarm Optimization (PSO) are used to identify the dense matching area.

However, the existing methods face the following limitations:

1) High computational cost due to exhaustive keypoint comparisons
2) Slower rate of convergence for swarm-based optimization techniques
3) Sensitivity to lossy compression artifacts
4) Decreased robustness in the presence of additive noise
5) Limited capacity for modeling contextual dependencies in structural relationships

To address these issues, this paper proposes a hybrid framework for spectral attention, evolutionary optimization techniques, and graph-based deep learning methods.

II. RELATED WORK

The detection of audio forgery has been extensively researched in the domain of digital forensics, considering the increasing misuse of manipulated multimedia data. Several methods have been suggested for the detection of audio forgery, including copy-move, splicing, and deepfake forgery. This section discusses some of the prominent research contributions in this domain.

1) The paper proposes a new method for the detection of audio copy-move forgery using spectrogram analysis and graph-based classification. This method utilizes Scale-Invariant Feature Transform to find keypoints in the spectrogram, followed by the detection of dense matching regions to find the forgery. Finally, the detected forgery is mapped to a graph using spiral pattern encoding, followed by classification using a Convolutional Neural Network. Although this method has shown promising results in terms of forgery detection accuracy, the use of keypoints has increased computational complexity.
2) A method based on the use of the "cochleagram" for the detection of copy-move forgery has been proposed in This method involves the use of the "Voice Activity Detection" technique to segment the audio signal, followed by the use of the Structural Similarity Index Measure (SSIM) to compare the similarity between the generated "cochleagram" images. Although the method has high precision and recall rates, the time required to process the signal increases with the length of the signal.
3) The problem of detecting audio splicing attacks has been addressed in with the use of the device acquisition traces. This method involves the use of a Convolutional Neural Network to detect the acoustic features of the recorded signal, followed by the use of the clustering method to detect whether the signal has been spliced or not. Although the method has been effective in detecting recordings obtained from different devices, its use for detecting copy-move attacks, which involve the same recording, might not be feasible.
4) Has suggested an inference noise addition technique to enhance the robustness of audio-based deepfake detection systems. The suggested technique involves adding noise during the inference phase to defend against manipulation attacks. The experiments have shown promising results in enhancing the robustness of audio-based deepfake detection systems. The suggested technique is mainly targeted at deepfake audio forgery.
5) Has suggested an anomaly-based deepfake audio forgery detection system using Generative Adversarial Networks. The suggested system is based on Generative Adversarial Networks and is able to detect deepfake audio forgery. The suggested system has shown promising results in deepfake audio forgery detection; however, it is not applicable to copy-move forgery.
6) Presents a comprehensive survey on audio deepfake detection techniques using machine learning and deep learning techniques. In this survey, various classifiers are compared, such as Support Vector Machine, Decision Tree, and Convolutional Neural Network. Among these classifiers, SVM provides high accuracy in specific cases; however, accuracy is case-dependent.
7) Proposes a deepfake audio detection technique using Mel Frequency Cepstral Coefficients and various machine learning classifiers. In this technique, MFCC features are extracted and then fed into various classifiers such as Support Vector Machine and Gradient Boosting classifiers. The experimental results show promising classification accuracy; however, MFCC-based techniques are not efficient in representing duplicated spectral information in copy- move forgery.
8) A multilingual audio-visual smartphone dataset for evaluating speaker recognition systems in real- world scenarios has been proposed in. The dataset includes recordings obtained from various devices and languages, which helps in the analysis of the robustness of the biometric algorithms. Although the dataset can be used for evaluating speaker recognition systems, the problem of copy-move forgery detection has not been addressed.
9) Proposes a multi-input neural network architecture that incorporates spectrograms, Mel spectrograms, and statistical characteristics in voice classification tasks. The proposed model incorporates multiple convolutional networks to learn multiple representations of the input audio signals. The experimental results show that the proposed model is efficient in achieving high classification accuracy. However, the proposed model is focused mainly on voice classification tasks and not audio forgery detection.
10) Proposes a deep semantic feature extraction technique using a "Multi-Task Learning Autoencoder with affine transformation" to detect musical instruments in complex acoustic environments. The proposed model is efficient in extracting audio features even in environments where there are overlapping frequencies. However, the proposed model is focused mainly on audio feature extraction and not audio forgery detection.

Although these works offer significant insights into audio analysis and forgery detection techniques, most of the existing techniques are based on intensive feature extraction techniques, computationally intensive optimization techniques, and dataset- dependent techniques. Moreover, there is a lack of research on the joint usage of adaptive spectral attention, spiral graph encoding, and graph attention learning techniques. This necessitates the proposed Hybrid Spectro-Spiral Graph Attention Framework for audio copy-move forgery detection techniques.

III. PROBLEM STATEMENT

Detection of copy-move forgery in audio is considered a challenging task in audio forensic science because of the similarity between the original and duplicated audio content. Copy-move forgery involves duplicating audio content from one part of an audio signal and placing it in another part of the same audio signal. The duplicated audio content is from the same audio environment and has similar characteristics, such as noise and speaker characteristics. The difficulty in detecting copy- move forgery is high compared to audio splicing, where there is a possibility of using environmental characteristics to detect forgery.

Most of the existing audio copy-move forgery detection techniques are based on spectral features, keypoint features, and correlation-based features.

These features are compared and analyzed using spectrogram representations of audio signals. Most of these techniques are unable to detect short duplicated segments within long audio signals and are vulnerable to various post-processing operations such as compression, filtering, and noise addition.

Thus, it is necessary to develop an efficient audio copy-move forgery detection system that can effectively deal with the following major challenges:

1) Detection of short duplicated segments within long audio signals
2) Robustness to compression and noise distortion
3) Reducing computational complexity
4) Modeling structural relationships between spectral features

To develop an efficient audio copy-move forgery detection system, this work introduces a novel spectro-spiral graph attention-based approach.

IV. EXISTING SYSTEM

Audio copy-move forgery detection has been an active area of research in digital audio forensics, mainly because of the growing use of audio recordings in legal, investigation, and media situations. Various existing methods have been proposed to detect copy-move forgery in a single audio recording. Most of these methods are based on transforming the audio signal into a time-frequency representation and analyzing similarities in spectral patterns to identify copy-move forgeries.

The most commonly used technique is to transform the input audio signal into a spectrogram using the STFT algorithm. This helps in finding duplicate areas in the signal by recognizing similarities in patterns. In various research works, Scale-Invariant Feature Transform (SIFT) is used to identify keypoints in the generated spectrogram image. These keypoints represent unique structures in the signal's spectrum. These keypoints are then compared with each other in different areas of the spectrogram in order to identify potential duplicate areas. For a better localization of suspicious frequency regions, various optimization techniques like Particle Swarm Optimization (PSO) have been proposed.

In this technique, particles are utilized to search for optimal frequency bands that include a large number of matched keypoints. Once a frequency region is identified, the audio is filtered and then further processed to classify it.

Once the suspicious frequency segments have been identified, graph representations may be used to represent structural relationships between various spectral components. In particular, spiral pattern- based graph encoding methods may be used to transform patches of the spectrogram into graph representations, where nodes represent the spectral component values and edges represent relationships between neighboring frequency components. These graph representations may then be transformed into visual representations and used as input for machine learning or deep learning methods.

For classification purposes, Convolutional Neural Networks (CNN) is used for distinguishing between original and manipulated audio samples. In this model, CNN learns hierarchical patterns from the graph images and classifies whether the audio contains copy-move forgery or not. This two-stage approach of keypoint detection and graph classification using keypoint matching and swarm optimization techniques shows promising results in various experimental works and is tested with various speech samples.

Although reasonable performance is achieved in detecting audio forgeries with existing systems, it is observed that most of the existing systems heavily depend on exhaustive keypoint matching and swarm optimization techniques, resulting in extra computational costs.

V. LIMITATIONS OF EXISTING SYSTEM

Although several approaches have been proposed for audio forgery detection, existing systems still face multiple limitations when applied to real-world forensic scenarios. Most traditional methods rely heavily on feature matching techniques applied to spectrogram representations of audio signals. These approaches often require exhaustive pairwise comparison of spectral keypoints, which significantly increases computational complexity when analyzing long audio recordings.

Another limitation lies in the optimization techniques used to localize duplicated segments. Algorithms such as Particle Swarm Optimization require careful parameter tuning and may converge slowly when searching within high-dimensional feature spaces. This increases processing time and makes real-time detection difficult for large forensic datasets. In addition, many existing systems depend on handcrafted feature extraction techniques that may not effectively capture complex structural relationships present in audio signals. As a result, detection accuracy may decrease when the audio undergoes post-processing operations such as compression, filtering, or noise addition. Furthermore, traditional spectrogram similarity methods focus primarily on pixel-level comparisons and fail to model deeper relational dependencies between spectral components. This limitation reduces their ability to detect subtle duplication patterns that may occur in short segments of manipulated audio.

These limitations highlight the need for a more robust detection framework that can efficiently identify duplicated audio segments while maintaining high accuracy under different recording conditions.

VI. PROPOSED METHODOLOGY

In order to overcome the limitations identified with the existing audio copy-move forgery detection systems, it is proposed to use the Hybrid Spectro- Spiral Graph Attention Framework. The proposed framework utilizes adaptive spectral analysis, evolutionary optimization, structural graph encoding, and attention-based deep learning to improve the efficiency of audio copy-move forgery detection. Unlike the conventional systems, which heavily depend on keypoint matching and evolutionary optimization techniques, the proposed audio copy- move forgery detection system focuses on adaptive spectral region identification and relational feature modeling. The proposed audio copy-move forgery detection system consists of multiple processing stages.

A. Spectrogram Representation

The first step takes an input audio signal and changes it to a time frequency representation using a Short- Time Fourier Transform (STFT). This helps in analyzing the spectral features in short periods of time.

For an input audio signal x(t), the Short-Time Fourier Transform(STFT) is given by:

$$X(\tau, \omega) = \sum_{\{t=-\infty\}}^{\{\infty\}} x(t)w(t-\tau)e^{-jwt}$$

w(t) represents the window function used for segmentation,
τ denotes the time shift parameter, and
ω represents the angular frequency.

The size of the STFT array gives rise to the spectrogram, which can be used to visualize the distribution of energy over time and frequency. The repeated patterns in the spectrogram may suggest the possibility of copy-move manipulation.

B. Adaptive Spectral Attention

In place of processing the entire spectrogram with the same importance, an adaptive spectral attention mechanism has been proposed to focus on those regions that may contain duplicate audio segments.

The attention-weighted spectrogram can be defined as:

$$S'(f,t) = A(f,t) \cdot S(f,t)$$

where:

S(f,t)represents the original spectrogram and A(f,t) denotes the learned attention map.

This process has several benefits:

- It reduces computation in unnecessary regions.
- It highlights suspicious zones of duplication.
- It improves the accuracy of forged regions localization.

C. Differential Evolution Optimization

For optimizing the frequency band thresholds, the proposed system utilizes the Differential Evolution (DE) optimization technique. It’s an evolutionary method that uses mutation, crossover, and selection operations to improve candidate solutions.

For the given population vector, the mutation process canbe expressed as:

$$v_i = x_{r1} + (x_{r2} - x_{r3})$$

Differential Evolution has been found to yield faster convergence with fewer control parameters than Particle Swarm Optimization, which has been commonly used in existing techniques.

D. Spiral Graph Encoding

Once the suspicious regions have been identified, the relevant segments of the spectrogram are converted into a graph-based representation, where the process is called spiral traversal encoding. In this case, the node is the amplitude of the spectrum, while the connection between the adjacent components is the edge.

The graph structure is defined as

$$G = (V, E)$$

where:

V represents the set of spectral nodes and

E represents the structural edges between nodes.

Spiral indexing maintains continuity in both radial and angular directions of a spectrogram, thus allowing for a better modeling of repeating harmonics that might indicate audio duplication.

E. Graph Attention Network

For capturing relational dependencies among the spectral nodes, a Graph Attention Network (GAT) is employed. GAT uses attention coefficients to update the node embeddings. The attention coefficient between two nodes is computed as

$$\alpha_{\{ij\}} = \frac{\exp(LeakyReLU(a^T[Wh_i||Wh_j]))}{\sum_{\{k \in N_i\}} \exp(LeakyReLU(a^T[Wh_i||Wh_k])}$$

The attention coefficient enables the model to dynamically assign relative importance to the spectral relationships. Finally, the graph embeddings are fed into the fully connected layers for classification of the audio as authentic or forged.

Where:

F represents the scaling factor and r1,r2,r3r1, r2, r3r1,r2,r3 are randomly selected distinct indices.

VII. ARCHITECTURE DIAGRAM

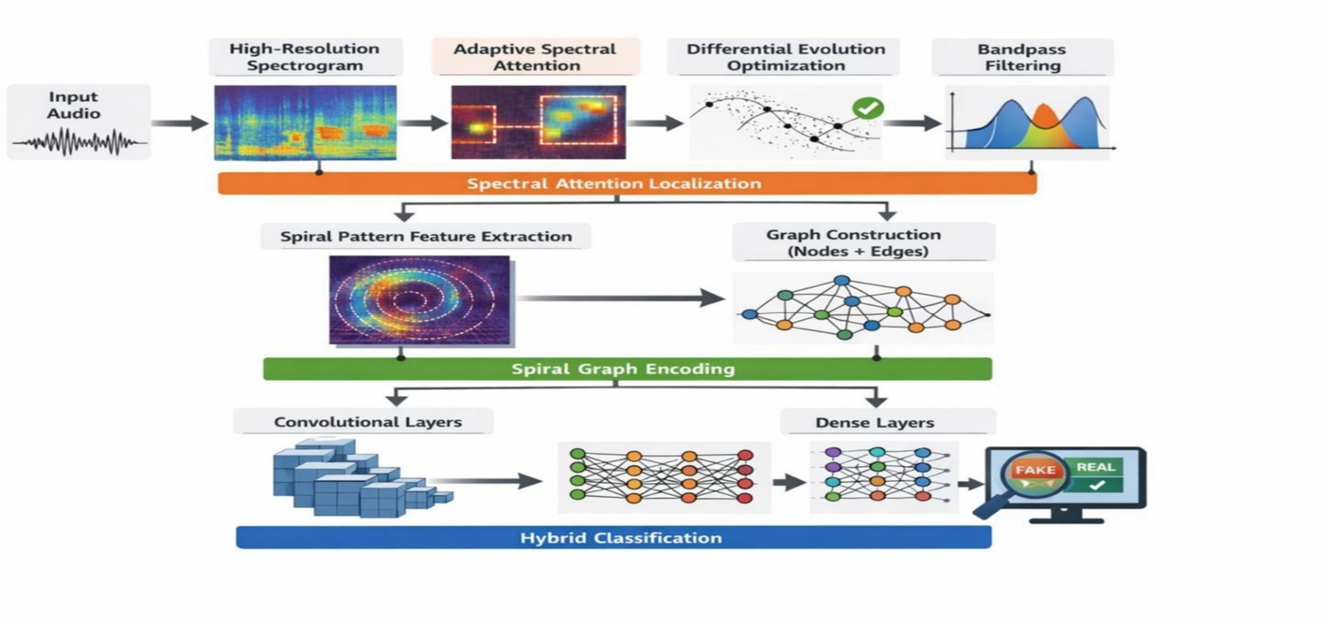
Input Audio
High-Resolution Spectrogram
Adaptive Spectral Attention
Differential Evolution Optimization
Bandpass Filtering
Spectral Attention Localization
Spiral Pattern Feature Extraction
Graph Construction (Nodes + Edges)
Spiral Graph Encoding
Convolutional Layers
Dense Layers
FAKE
REAL
Hybrid Classification

Fig. 1. Architecture Diagram for Proposed system

VIII. COMPARISON TABLE

Method	Accuracy	Precision	Recall	F1-Score
GA-SVM	92.13%	0.93	0.95	0.96
Random Forest (CQT Spectral)	89.34%	0.90	0.79	0.94
Hybrid Graph Attention (MLP)	95.64%	0.97	0.96	0.96

Table. 1.1. Comparison Table

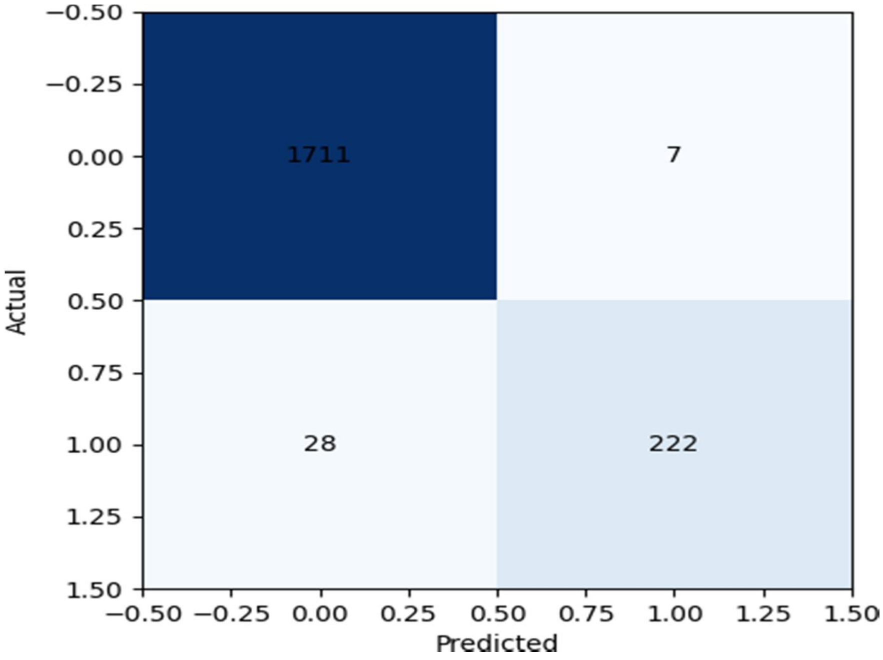

Fig. 2. Confusion matrix

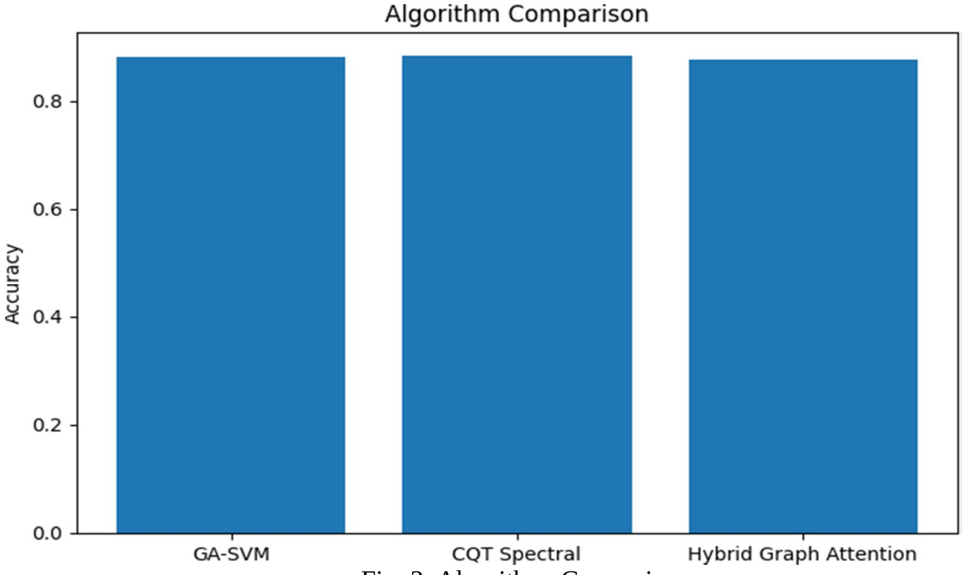

Fig. 3. Algorithm Comparison

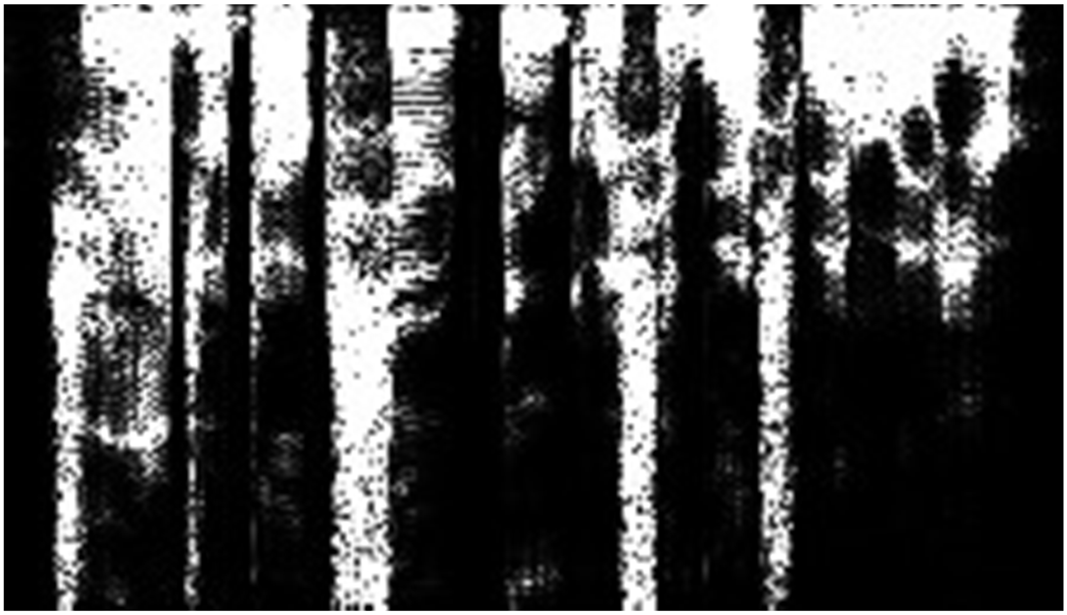

Fig. 4. Spectrogram Visualization

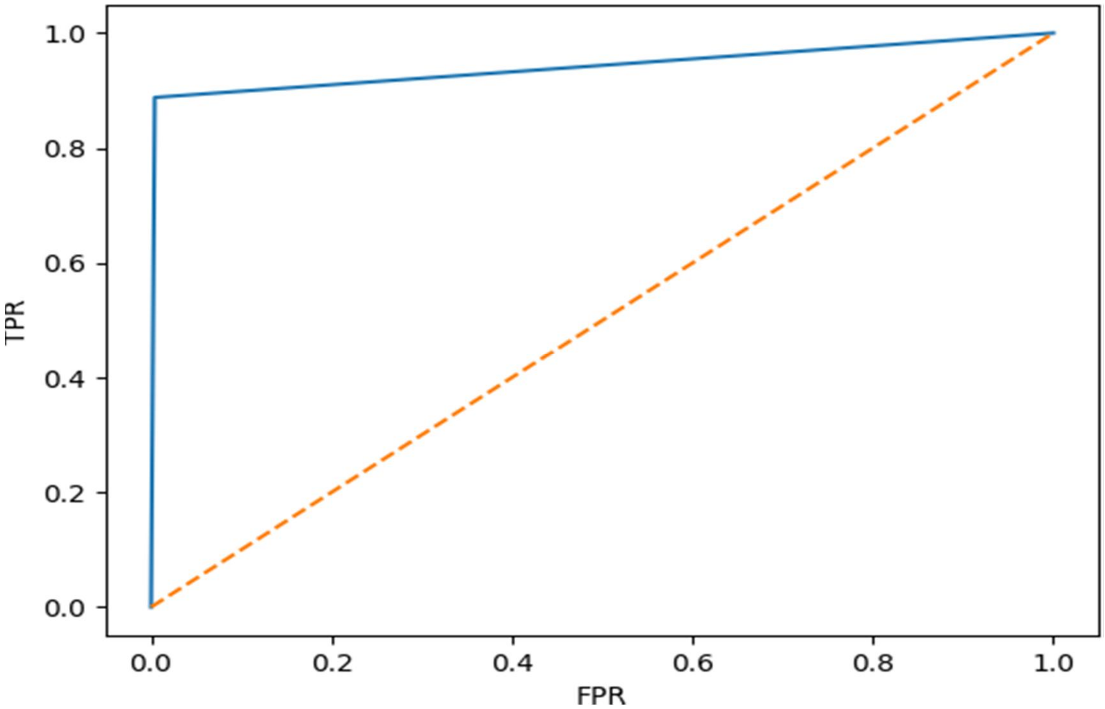

Fig. 5. ROC Curve

IX. RESULTS AND DISCUSSION

The results of the experiments revealed that the Hybrid Graph Attention (MLP) model showed the highest accuracy of 95.64%. Its precision, recall, and F1-score were 0.97, 0.96, and 0.96, respectively, proving the effectiveness of the model for detecting forged audio. The GA-SVM model showed a good level of accuracy of 92.13%. Its precision, recall, and F1-score were 0.93, 0.95, and 0.96, respectively, proving the effectiveness of the model for detecting forged audio. Similarly, the Random Forest (CQT Spectral) model showed a good level of accuracy of 89.34%. Its precision, recall, and F1-score were 0.90, 0.79, and 0.94, respectively. However, the model showed a comparatively lower level of recall for forged audio. Thus, the results of the experiments revealed the effectiveness of the hybrid model for detecting forged audio compared to other machine learning models.

X. CONCLUSION

In the present work, three different approaches, namely GA-SVM, CQT Spectral with Random Forest, and the Hybrid Graph Attention (MLP) model, have been implemented for the purpose of detecting forged audio. From the results of the experiments, it is found that the Hybrid Graph Attention model performs better than the other two approaches, yielding a maximum accuracy of 95.64%. Further, the results of the experiments using the GA-SVM model yield a maximum accuracy of 92.13%, whereas the results of the experiments using the Random Forest (CQT Spectral) model yield a maximum accuracy of 89.34%. Thus, the proposed hybrid model for the purpose of detecting forged audio yields better results compared to the results of the other machine learning models.

XI. FUTURE WORK

However, the audio forgery detection system that is proposed in the future may be improved in the following ways: First, the deep learning techniques, which may include CNN, LSTM, or the use of the Transformer architecture, may be used to automatically learn more complex audio patterns. Secondly, the audio dataset may be expanded to improve the generalization and robustness of the proposed audio forgery detection system. Thirdly, the use of more sophisticated techniques for audio feature extraction may be used to improve the proposed audio forgery detection system. Furthermore, the proposed audio forgery detection system may be used to develop real-time audio forgery detection systems. In addition, the proposed audio forgery detection system may be improved further by the use of more sophisticated techniques in the graph attention network to improve the detection accuracy.

REFERENCES

[1] B. Ustubioglu, G. Tahaoglu, A. Ustubioglu, G. Ulutas, and M. Kilic, “A novel audio copy-move forgery detection method with classification of graph- based representations,” IEEE Access, vol. 13, pp. 22029–22054, 2025.
[2] B. Ustubioglu, “An attack-independent audio forgery detection technique based on cochleagram images of segments with dynamic threshold,” IEEE Access, vol. 12, pp. 82660–82675, 2024.
[3] D. U. Leonzio, L. Cuccovillo, P. Bestagini, M. Marcon, P. Aichroth, and S. Tubaro, “Audio splicing detection and localization based on acquisition device traces,” IEEE Transactions on Information Forensics and Security, vol. 18, pp. 4157–4172, 2023.
[4] I. Kim, T.-P. Doan, S. Hong, and S. Jung, “Inference-time noise addition for improving adversarial robustness of audio deepfake detection systems,” IEEE Access, vol. 13, pp. 200669–200682, 2025.
[5] D. Song, N. Lee, J. Kim, and E. Choi, “Anomaly detection of deepfake audio based on real audio using generative adversarial networks,” IEEE Access, vol. 12, pp. 184311–184326, 2024.
[6] O. A. Shaaban, R. Yildirim, and A. A. Alguttar,
[7] “Audio deepfake approaches,” IEEE Access, vol. 11,
[8] pp. 132652–132682, 2023.
[9] A. Hamza et al., “Deepfake audio detection via MFCC features using machine learning,” IEEE Access, vol. 10, pp. 134018–134028, 2022.
[10] H. Mandalapu et al., “Multilingual audio-visual smartphone dataset and evaluation,” IEEE Access, vol. 9, pp. 153240–153257, 2021.
[11] M. Talha, H. Ghafoor, and S. Y. Nam, “A unified approach to voice classification leveraging spectrograms, Mel spectrograms and statistical features,” IEEE Access, vol. 13, pp. 133827–133836, 2025.
[12] D. Nurdiyah et al., “Deep semantic feature extraction to overcome overlapping frequencies for instrument recognition in Indonesian traditional music orchestras,” IEEE Access, vol. 12, pp. 76936– 76954, 2024.
[13] Y. Kawaguchi and T. Endo, “How can we detect anomalies from subsampled audio signals?” in Proc. IEEE MLSP, 2017.
[14] F. Wang, C. Li, and L. Tian, “Detecting audio copy-move forgery based on DCT and SVD,” in Proc. IEEE ICCT, 2017.
[15] X. Huang, Z. Liu, W. Lu, H. Liu, and S. Xiang, “Fast and effective copy-move detection of digital audio based on auto segment,” 2020.
[16] R. Yang, Z. Qu, and J. Huang, “Detecting digital audio forgeries by checking frame offsets,” in Proc. ACM Multimedia Security Workshop, 2008.
[17] Z. Ye et al., “FlashSpeech: Efficient zero-shot speech synthesis,” Proc. ACM Multimedia, 2024.
[18] H. Tak et al., “End-to-end spectro-temporal graph attention networks for speech deepfake detection,” 2021.
[19] M. Aldwairi and A. Alwahedi, “Detecting fake news in social media networks,” Procedia Computer Science, vol. 141, pp. 215–222, 2018.
[20] F. Zheng, G. Zhang, and Z. Song, “Comparison of different implementations of MFCC,” Journal of Computer Science and Technology, vol. 16, no. 6, pp. 582–589, 2001.
[21] Y. Zhong and B. Huang, “Classification of cassava leaf disease using transformer-embedded ResNet,” Agriculture, vol. 12, no. 9, 2022.
[22] X. Zhou and R. Zafarani, “A survey of fake news: Fundamental theories and detection methods,” ACM Computing Surveys, vol. 53, no. 5, 2021.
[23] S. Suwajanakorn, S. Seitz, and I. Kemelmacher- Shlizerman, “Synthesizing Obama,” ACM Transactions on Graphics, vol. 36, no. 4, 2017.
[24] A. R. Javed et al., “A comprehensive survey on computer forensics,” IEEE Access, vol. 10, pp. 11065–11089, 2022.
[25] A. Werning and R. Haeb-Umbach, “Dataset pruning for compression of audio tagging models,” in Proc. EUSIPCO, 2024.
[26] S. Dibbo et al., “Robust audio classification using multi-layer neural networks,” in Proc. ICASSP Workshops, 2024.
[27] S. Mehra, V. Ranga, and R. Agarwal, “Improving speech command recognition through deep learning,” Signal Image and Video Processing, 2024.
[28] A. Ustubioglu et al., “Mel spectrogram-based audio forgery detection using CNN,” Signal Image and Video Processing, 2023.
[29] Y. Kawaguchi and T. Endo, “How can we detect anomalies from subsampled audio signals?” in Proc. IEEE MLSP, 2017.
[30] F. Wang, C. Li, and L. Tian, “Detecting audio copy-move forgery based on DCT and SVD,” in Proc. IEEE ICCT, 2017.
[31] X. Huang, Z. Liu, W. Lu, H. Liu, and S. Xiang, “Fast and effective copy-move detection of digital audio based on auto segment,” 2020.
[32] R. Yang, Z. Qu, and J. Huang, “Detecting digital audio forgeries by checking frame offsets,” in Proc. ACM Multimedia Security Workshop, 2008