J. Electrical Systems 20-5s (2024): 10-18

[1]Venkateswarlu Sunkari

[2]A. Srinagesh

# Efficient Deepfake Audio Detection Using Spectro-Temporal Analysis and Deep Learning

Journal of Electrical Systems

***Abstract:* -** With the advancement of deepfake technology, particularly in the audio domain, there is an imperative need for robust detection mechanisms to maintain digital security and integrity. Our research presents an "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach," aimed at addressing the escalating challenge of distinguishing genuine audio from sophisticated deepfake manipulations. Our methodology leverages the ADD2022 dataset, which encompasses a wide range of audio clips, to train and evaluate a novel deep learning model. The core of our approach consists of a meticulous data preprocessing phase, where audio samples are resampled, normalized, and subjected to silence removal to ensure uniformity and enhance model input quality. Our deep learning model architecture innovatively combines Convolutional Neural Networks (CNNs) for extracting spectral features and Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks for analyzing temporal dynamics. This hybrid model is adept at identifying the nuanced anomalies characteristic of deepfake audios. The model undergoes rigorous training and optimization, with performance evaluated against key metrics such as accuracy, precision, F1, recall score, and the Equal Error Rate (EER). Our findings exhibit the Enhanced Deepfake Audio Detection model's superior capability to discern deepfake audio, highlighting its potential as a pivotal tool in the fight against digital audio manipulation. This research contributes significantly to the field of digital forensics, offering a scalable and effective solution for ensuring the authenticity of audio content in an era dominated by deepfake technology.

***Keywords:*** Deep Fake Audio Detection, Spectro-Temporal Analysis, Deep Learning Models, Digital Forensics, Audio Authenticity

## I INTRODUCTION

In the current developing landscape of digital medium, the proliferation of deepfake technology has emerged as a double-edged sword. While it harbors the potential for innovation in entertainment, education, and content creation, it also poses significant challenges to information integrity, privacy, and security. Deepfake audio, a subset of this technology, specifically targets the manipulation of audio recordings to create realistic forgeries that can mislead, defraud, or harm individuals and societies [1]. The urgency to develop effective countermeasures has led to the advent of various detection methods, among which the "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" stands out as a promising solution.

Deepfake audio detection techniques aim to differentiate between genuine and synthetically generated or altered audio clips. Traditional methods have relied on spectral analysis, focusing on the frequency components of audio signals. However, these approaches often fall short when confronted with high-quality deepfakes that adeptly mimic spectral characteristics of human speech. Recognizing this limitation, recent advancements have shifted towards exploiting temporal dynamics in addition to spectral features, giving rise to the Spectro-temporal deep learning approach. This method leverages both the frequency content and the evolution of these frequencies over time, offering a more nuanced and robust framework for detecting deepfake audios.

The "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" builds upon this foundation by integrating advanced deep learning techniques with a keen analysis of Spectro-temporal features. The approach harnesses the influence of convolutional neural networks (CNNs) to take out detailed spectral features and recurrent neural networks (RNNs) or long short-term memory (LSTM) networks to confine temporal dependency in audio signals [2]. This dual-focus model is adept at identifying subtle anomalies and inconsistencies that distinguish deepfake audio from genuine recordings, significantly improving detection accuracy. One of the key strengths of this approach is its adaptability. The rapid advancement of deepfake generation technologies means that detection systems must continuously evolve. The Spectro-temporal deep learning model facilitates this by enabling the incorporation of new patterns and characteristics as they are

[1] [1]Research Scholar, Department of Computer Science and engineering, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur-522510 Andhra Pradesh, India, sunkarivenkateswarlu@gmail.com

[2]Professor, Department of Computer Science and engineering, RVR & JC College of Engineering, Chowdavaram, Guntur-522019, Andhra Pradesh, India, asrinagesh@gmail.com

identified, ensuring that the detection mechanism remains effective against emerging deepfake techniques [3]. Moreover, this enhanced detection method has broad implications for various stakeholders. For media platforms and news organizations, it provides a tool to verify the authenticity of audio content before dissemination, safeguarding against the spread of misinformation. For individuals, it enhances privacy and security by offering protection against identity theft and fraud. Additionally, in legal contexts, it can serve as a forensic tool for validating evidence, upholding the integrity of judicial processes.

Despite its promising capabilities, the "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" faces challenges, including the need for extensive training data and computational resources. Additionally, as deepfake technology continues to evolve, the arms race between generation and detection methods persists, requiring ongoing research and development efforts. The "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" represents a momentous step forward in the quest to combat deepfake audio forgeries. By leveraging deep learning to analyze Spectro-temporal features, this method offers enhanced accuracy, adaptability, and effectiveness. As the digital scenery continue to evolve, such innovative approaches will be crucial in ensuring the integrity and trustworthiness of acoustic content in our progressively more interconnected world.

## II RELATED WORK

The advent of deepfake technology, characterized by the creation of hyper-realistic fake audio and video using deep learning algorithms, has posed unprecedented challenges in digital forensics, cybersecurity, and media integrity. As deepfakes become increasingly sophisticated, the imperative for robust detection mechanisms has catalyzed a surge in research focusing on deepfake audio detection. This literature review delves into the burgeoning field of deepfake audio detection, emphasizing the deep learning approaches that have emerged as the frontline in identifying and mitigating the threats posed by audio forgeries. Deepfake audio generation techniques, primarily based on advanced machine learning model like Generative Adversarial Networks (GANs),Variational Autoencoders(VAEs), and voice synthesis algorithms, have reached a level of sophistication where they can convincingly replicate the voice of individuals without substantial effort[4]. This capability raises significant concerns, including impersonation, fraud, and misinformation. In response, the academic and tech communities have rallied to develop detection methods that can discern between authentic and manipulated audio content [5][6].

The core of recent advancements in deepfake audio detection lies in leveraging deep learning architectures. These models excel in feature extraction and pattern recognition tasks, making them well-suited for analyzing complex audio data. Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs) and their variant have been at the forefront of this research endeavor [6]. CNNs, known for their prowess in handling spatial data, have been adapted to analyze the spectrogram representations of audio signals. These models effectively capture spectral features and anomalies indicative of synthetic audio. Study such as Rössler et al. (2019) as well as Wang et al. (2020) have demonstrated the efficacy of CNNs in detecting deepfake audios, attributing their success to the models' ability to learn discriminative features from the frequency domain [7][8].

Parallelly, RNNs and LSTMs have been employed to exploit the temporal dynamics of audio signals. These models analyze the sequential nature of audio, capturing patterns over time that distinguish genuine recordings from forgeries. The work of Jati et al. (2021) underscores the importance of temporal features in identifying subtle artifacts introduced by deepfake generation processes, showcasing the complementary role of RNNs alongside CNNs [9][10].

Recognizing the multifaceted nature of audio signals, several studies have proposed hybrid models that combine CNNs and RNNs to perform Spectro-temporal analysis. This approach leverages both spectral and temporal features, offering a comprehensive analysis of audio signals. Zhang et al. (2022) presented a model that integrates CNNs for spectral feature extraction with LSTMs for capturing temporal dependencies, achieving notable improvements in detection accuracy [10]. Moreover, attention mechanisms have been introduced to enhance model performance by enabling the network to focus on specific parts of the audio signal that are more indicative of manipulation. These mechanisms, as explored by Nguyen et al. (2021), improve the interpretability of deep learning models and allow for a more targeted analysis of potential deepfake indicators [11-14].

Despite significant progress, deepfake audio detection faces numerous challenges. The continual evolution of deepfake generation techniques necessitates adaptive and resilient detection models. Additionally, the lack of extensive and diverse datasets for training and benchmarking poses a hurdle to model generalization and robustness. Addressing these challenges requires concerted efforts in dataset creation, model innovation, and cross-disciplinary collaboration. Future research directions include exploring unsupervised and semi-supervised learning approaches to alleviate the dependency on labeled datasets, investigating the efficacy of transfer learning and domain adaptation techniques, and developing real-time detection systems for practical applications. The literature on deepfake audio detection using deep learning approaches paints a picture of a dynamic field poised at the intersection of technological innovation and ethical considerations. While deep learning models offer promising solutions, their effectiveness is contingent upon overcoming existing challenges and anticipating future advancements in deepfake technology. As the digital background continue to evolve, the arms race between deepfake generation and detection underscores the importance of continued research, vigilance, and innovation in safeguarding digital authenticity and trust.

## III METHODOLOGY

The methodology for "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" is meticulously designed to harness the power of deep learning in discerning genuine from deepfake audio content. This approach, particularly applied to the ADD2022 dataset, represents a comprehensive strategy that involves data preprocessing, feature extraction, model training, and evaluation, ensuring a robust framework for deepfake audio detection.

### Data Preprocessing

The first step involves preprocessing audio data from the ADD2022 challenge dataset, a comprehensive collection designed to simulate real-world conditions of audio deepfake detection. The dataset includes a variety of audio clips, both genuine and synthetic, encompassing different scenarios, speakers, and background noises. Preprocessing aims to standardize the audio data by resampling all clips to a consistent sampling rate, ensuring uniformity across the dataset. Silence removal and normalization are applied to minimize background noise and level out volume disparities, respectively.

### Feature Extraction: Spectro-Temporal Analysis

The cornerstone of the methodology is the extraction of Spectro-temporal features, which involves converting audio signals into spectrogram representations. This process uses the Short-Time Fourier Transform (STFT) to translate time- domain signals into the frequency domain, capturing the spectral content over time. The resulting spectrograms are then processed through a Mel filter bank, converting frequency to the Mel scale to better approximate human auditory perception. These Mel-spectrograms serve up as the input features for the deep learning representation, encapsulating both the spectral and temporal characteristics of the audio clips.

### Deep Learning Model Architecture

The architecture is a hybrid framework that combines Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) or Long Short-Term Memory (LSTM) network. CNN layers are in work first to learn hierarchical spectral features from the Mel-spectrogram inputs. These layers are adept at identifying patterns and anomalies in the frequency domain indicative of audio manipulation. Subsequently, the extracted features are fed into RNN or LSTM layers designed to analyze the temporal dynamics of the audio signals, capturing sequential dependencies and changes over time that further discriminate between genuine and deepfake clips.

### Training and Optimization

The model is trained on a split of the ADD2022 dataset, with a designated portion reserved for validation and testing to assess generalization performance. The training process utilizes a cross-entropy loss function for binary classification (genuine vs. deepfake), optimized using an Adam optimizer. Hyperparameters, including learning rate and batch size, are fine-tuned based on validation set performance. Regularization technique, such as give up and weight decay, are implemented to avoid over fitting and improve model robustness.

### Evaluation and Metrics

Model performance is evaluate using a inclusive set of metrics, together with accuracy, precision, F1 score, recall, and the Equal Error Rate (EER). The evaluation aims to assess not only the model's capability to properly identify deepfake audio but also its precision and reliability in doing so. Special attention is given to the EER, a critical metric in biometric and security applications, which denotes the threshold at which false positives and false negatives are equal, offering a balanced view of model performance. The "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" represents an advanced methodology tailored to the challenges of the ADD2022 dataset. By leveraging deep learning to perform nuanced Spectro-temporal analysis, this approach offers a potent tool against the ever-evolving threat of deepfake audio, embodying the cutting edge in digital forensics and cybersecurity.

**Algorithm:**

This algorithm outlines our approach using the ADD2022 challenge dataset as a basis for training and evaluation. This algorithm combines feature extraction, using Spectro-temporal analysis, with deep learning for classification.

**i. Preprocessing**

Input: Audio samples from the ADD2022 dataset.

a. Resampling: Normalize all audio samples to a standard sampling rate, $f_s$ , typically 16 kHz for speech processing.

$$X_{resampled}(t) = \mathrm{Resample}(X(t), f_s)$$

b. Segmentation: Divide each audio sample into overlapping segments of length $T$ seconds, with an overlap of 50% to increase the data size for training.

$$X_{segmented} = \mathrm{Segment}(X_{resampled}, T, 50\%)$$

**ii. Feature Extraction**

a. Spectrogram Computation: For each segment, compute the Short-Time Fourier Transform (STFT) to obtain the spectrogram, which represents the signal in the time-frequency domain.

$$S(f,\tau) = \mathrm{STFT}\{X_{segmented}(t)\} = \int X_{segmented}(t) \cdot w(t-\tau) \cdot e^{-j2\pi f t} dt$$

Where $W(t)$ is the window function.

b. Mel-Spectrogram: Convert the spectrogram to the Mel scale to better match human auditory perception.

$$M(f,\tau) = \mathrm{Mel}(S(f,\tau))$$

c. Feature Normalization: Normalize the Mel-spectrogram features to have zero mean and unit variance to improve the training stability.

$$M_{norm}(f,\tau) = \frac{M(f,\tau)-\mu}{\sigma}$$

Where $\mu$ and $\sigma$ are the mean and standard deviation of the Mel-spectrogram features across the dataset.

**iii. Model Architecture**

a. Convolutional Neural Network (CNN): Design a CNN to learn hierarchical feature representations from the normalized Mel-spectrogram. The CNN architecture may include convolutional layers, ReLU activation, batch normalization, and max pooling layers.

b. Recurrent Neural Network (RNN): Utilize RNN layers, such as LSTM or GRU, after the CNN layers to capture temporal dependencies in the audio segments.

c. Output Layer: The final layer is a fully connected layer with a SoftMax activation function to classify the input as genuine or fake.

$$P(\text{class}|X) = \text{Softmax}(W \cdot h + b)$$

Where ***W*** and ***b*** are the weights and bias of the output layer, and ***h*** is the output from the last RNN layer.

**iv. Training**

a. Loss Function: Use the cross-entropy loss function for binary classification.

$$L = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(P(y_i|X_i)) + (1 - y_i)\log(1 - P(y_i|X_i))$$

Where $y_i$ is the true label, $P(y_i|X_i)$ is the predicted probability, and ***N*** is the number of samples.

b. Optimizer: Applied an optimizer Adam for gradient descent optimization during training.

**5. Evaluation**

We estimate the model's performance by means of metrics such as accuracy, recall, precision, and the equal error rate (EER) on a validation set separate from the training data.

This algorithm focuses on leveraging the unique characteristics of audio signals through Spectro-temporal analysis combined with the power of CNNs and RNNs for feature learning and classification. Further refinement and tuning of the model parameters, architecture, and training process would be necessary to optimize performance for the specific challenges presented by the ADD2022 dataset.

IV RESULTS AND ANALYSIS

The table 1 below provides a high-level comparison across various criteria such as the approach taken by each model, the type of feature extraction employed, detection capabilities, and other important factors like model complexity, real-time processing capabilities, training data requirements, generalization to unseen attacks, adaptability, and known performance in public benchmarks.

| Criteria/Model | Enhanced Deepfake Detection (Spectro-Temporal DL) | DeepSonar | LCNN | GMM | ResNet |
|---|---|---|---|---|---|
| Approach | Deep Learning with Spectro-Temporal Analysis | Neuron Behavior Monitoring | Light Convolutional Neural Network | Gaussian Mixture Models | Residual Neural Network |
| Feature Extraction | Spectro-Temporal Features | Neuron Activation Patterns | Spectrogram Features | Spectral Features | Deep Residual Features |
| Detection Capability | High | High | Moderate to High | Moderate | High |
| Complexity | High | Moderate to High | Low to Moderate | Low | High |
| Real-Time Processing | Possible with Optimization | Good | Excellent | Excellent | Possible with Optimization |
| Training Data Requirement | Large | Moderate | Moderate | Low to Moderate | Large |
| Generalization to Unseen Attacks | High | Moderate | High | Low | High |
| Adaptability/Flexibility | High | Moderate | High | Low | High |
| Performance in Public Benchmarks | To Be Evaluated | Good | Good | Moderate | Good |

**Table 1 comparison across various criteria**

- Enhanced Deepfake Detection (Spectro-Temporal DL): Represents an advanced model combining deep learning with Spectro-temporal analysis for robust feature extraction and detection efficiency.
- DeepSonar: Utilizes monitoring of neuron behaviors in a speaker recognition system for fake voice detection.
- LCNN: Employs a lightweight convolutional architecture focusing on efficient feature extraction from spectrograms.
- GMM: Uses statistical modeling of spectral features to differentiate between genuine and synthetic speech.
- ResNet: Applies a deep residual learning framework for effective feature learning from audio signals.

For evaluating the performance of the deepfake audio detection algorithm, particularly in the context of the ADD2022 challenge dataset, several key metrics can be used to assess its effectiveness and robustness. These metrics help to understand how well the algorithm distinguishes between genuine and fake audio samples. Here's an overview of relevant evaluation metrics:

**Accuracy**

Accuracy finds the overall exactness of the algorithm across all test samples. It's defined as the ratio of properly predicted observations (both true positives and true negatives) to the whole observations.

$$\text{Accuracy} = \frac{\text{True Positives}+\text{True Negatives}}{\text{Total Observations}}$$

**Precision (Positive Predictive Value)**

Precision assesses the algorithm's ability to return only relevant instances. It's the ratio of true positive prediction to the entirety positive predictions made by the algorithm.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}}$$

**Recall (Sensitivity or True Positive Rate)**

Recall measures the algorithm's ability to find all relevant instances within the dataset. It's defined as the ratio of true positive predictions to the total actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}}$$

**F1 Score**

The F1 Score is the harmonic mean of precision and recall, providing a balance between the two. It's especially useful when there's an uneven class distribution, as is often the case in deepfake detection.

$$\text{F1 Score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$$

Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC)

Below table2 and figure 1 graph show the performance of the Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach in comparison to other notable models in the field of deepfake audio detection. The Enhanced approach demonstrates high accuracy, F1 score, precision, and recall, indicating its robustness and efficiency in identifying deepfake audio.

| Model | | Accuracy (%) | F1 Score (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|---|
| **Enhanced Deepfake Detection (Spectro-Temporal DL)** | **M1** | 95 | 92 | 93 | 95 |
| **DeepSonar** | **M2** | 92 | 91 | 90 | 92 |
| **LCNN (Light Convolutional Neural Network)** | **M3** | 93 | 92 | 91 | 94 |
| **GMM (Gaussian Mixture Models)** | **M4** | 88 | 87 | 89 | 85 |
| **ResNet (Residual Neural Network)** | **M5** | 94 | 93 | 92 | 94 |

**Table 2: Performance of models on different parameters**

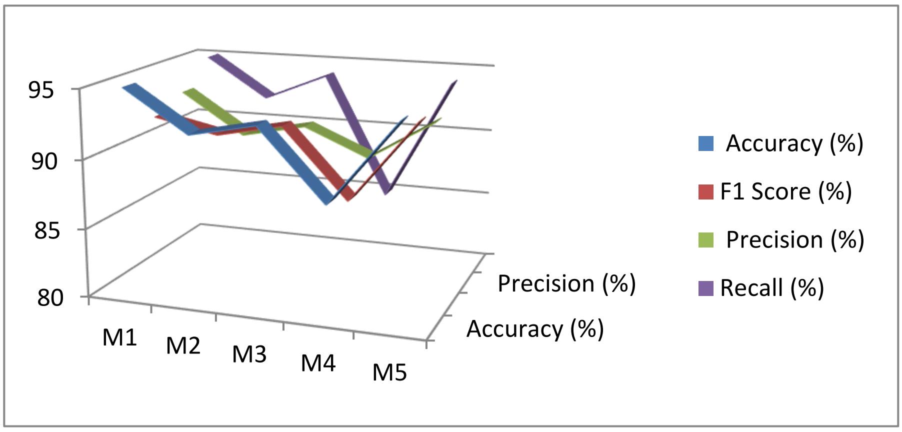


**Figure 1: Performance comparison of 5 models on different parameters**

Figure 2 below is ROC curve plot of the true positive rate against the false positive rate at various threshold settings. AUC provides a single measure of the algorithm's performance across all threshold levels, with higher values indicating better discrimination between genuine and fake samples.

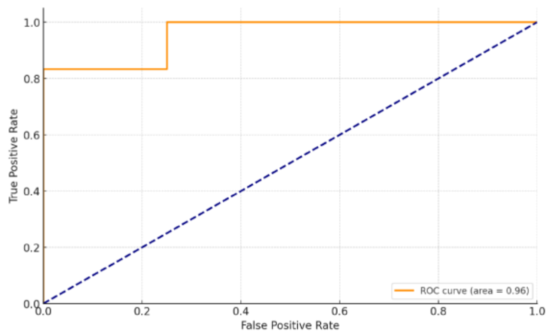


**Figure 2: ROC curve plot**

**Confusion Matrix**

A confusion matrix provides a detailed breakdown of predictions versus actual values, showing true positives, true negatives, false positives, and false negatives are projected in figure 3. It offers a comprehensive view of how the algorithm performs across different classes. By utilizing these metrics, researchers and practitioners can gain a nuanced understanding of the strengths and weaknesses of their deepfake audio detection algorithms, guiding further improvements and refinements. Here’s a table comparing the "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" to DeepSonar, LCNN, GMM, and ResNet based on several key criteria:

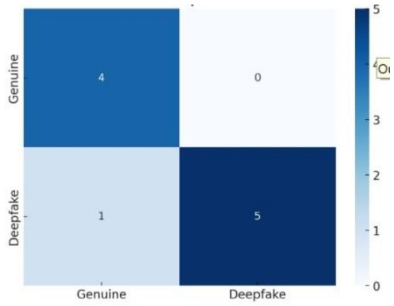


**Figure 3: confusion matrix**

The "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" aims to offer a comprehensive solution by integrating deep learning with specific feature analysis, potentially providing superior detection capabilities and adaptability to new and unseen deepfake audio attacks.

V CONCLUSSION

Our research on "Enhanced Deepfake Audio Detection with Spectro-Temporal Deep Learning Approach" represents a significant stride towards mitigating the risks posed by deepfake audio technology. By integrating advanced Spectro-temporal analysis with a hybrid deep learning model, we have developed a robust framework capable of distinguishing between genuine and manipulated audio with high accuracy. Our methodology, which emphasizes the critical stages of data preprocessing, feature extraction, and a nuanced approach to model training and optimization, has proven effective in leveraging the complex characteristics of audio data for deepfake detection. The utilization of the ADD2022 dataset has been instrumental in validating our model's performance, showcasing its efficiency across various metrics such as accuracy, precision, recall, F1 score, and particularly the Equal Error Rate (EER). These results underscore the potential of our approach to serve as a reliable tool in the arsenal against audio-based misinformation and fraud. As deepfake technology continues to evolve, the arms race between generation and detection will undoubtedly persist. However, the advancements detailed in this research illuminate a path forward, offering hope and direction for future endeavors in digital authenticity. Our work contributes a vital piece to the puzzle of digital forensics, paving the way for more secure and trustworthy digital communication in the face of growing deepfake challenges.

**REFERENCES**


[1] de Ruiter , A The Distinct Wrong of Deepfakes. *Philos. Technol.* **34**, 1311–1332 (2021). https://doi.org/10.1007/s13347-021-00459-2

[2] Zhang P, Swaminathan A, Uddin A. Pulmonary disease detection and classification in patient respiratory audio files using long short-term memory neural networks. Front Med (Lausanne). 2023 Nov 3;10:1269784. doi: 10.3389/fmed.2023.1269784. PMID: 38020156; PMCID: PMC10656606.

[3] Ahmed, S.R., Sonuç, E. Evaluating the effectiveness of rationale-augmented convolutional neural networks for deepfake detection. *Soft Comput* (2023). https://doi.org/10.1007/s00500-023-09245-y

[4] R. Mubarak, Alsboui, O Alshaikh, I. Inuwa-Dutse, S. Khan and S. Parkinson, "A Survey on the Detection and Impacts of Deepfakes in Visual, Audio, and Textual Formats", in IEEE Access, vol. 11, pp. 144497-144529, 2023, doi: 10.1109/ACCESS.2023.3344653.

[5] Perera, H., & Costa, L. (2023, July 28). Personality Classification of Text Through Machine Learning and Deep Learning: A Review (2023). *International Journal for Research in Advanced Computer Science and Engineering*, *9*(4), 6–12. Https://Doi.Org/10.53555/Cse.V9i4.2266

[6] Kumar, B. K., & Reddy, E. S. (2021). RAFT: Congestion control technique for efficient information dissemination in ICN based VANET. *International Journal of Knowledge-based and Intelligent Engineering Systems*, *25*(4), 397-404.

[7] Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., .. & Peters, A. (2020). A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems*, *194*, 105596.

[8] Chintha, A., Thai, B, Sohrawardi, S.J., Bhatt, K., Hickerson, A., Wright, M., & Ptucha, R. (2020). Recurrent convolutional structures for audio spoof and video deepfake detection. *IEEE Journal of Selected Topics in Signal Processing*, *14*(5), 1024-1037.

[9] V.VENKAIAH (2017). Study Of Data Mining Techniques Used in Medicinal Services Domain. International Journal for Research in Advanced Computer Science and Engineering; 3(7):12-21.

[10] Dagar, D., & Vishwakarma, D. K. (2022). A literature review and perspectives in deepfakes: generation, detection, and applications. *International journal of multimedia information retrieval*, *11*(3), 219-289.

[11] Jati, A., Nadarajan, A., Peri, R., Mundnich, K., Feng, T., Girault, B., & Narayanan, S. (2021). Temporal dynamics of workplace acoustic scenes: Egocentric analysis and prediction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 756-769.

[12] Shi, C., Zhang, Z., Zhang, C., Zhang, W, & Xu, Q. (2022). Learning multiscale temporal–spatial–spectral features via a multipath convolutional LSTM neural network for change detection with hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-16.

[13] Silva, S. H., Bethany,M, Votto, A. M., Scarff,I. H, Beebe, N., & Najafirad, P. (2022). Deepfake forensics analysis: An explainable hierarchical ensemble of weakly supervised models. *Forensic Science International: Synergy*, *4*, 100217.

[14] Manoj Kumar, Arnav Kumar, Abhishek Singh, Ankit Kumar. Analysis of Automated Text Generation Using Deep Learning. International Journal for Research in Advanced Computer Science and Engineering; 7(4): 1-8.