International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Unified Deepfake Detection System for Image, Video and Audio Media

Prof. Neelakantappa T T[1], Ms. Ashwini J[2], Ms. Ashwini H[3], Ms. Lavanya K[4], Ms. Sushmitha D[5]

[1]Assistant Professor, Computer Science And Engineering, Sjm Institute Of Technology Chitradurga
[2,3,4,5]Student, Computer Science And Engineering, Sjm Institute Of Technology Chitradurga

ABSTRACT

The rise of deepfake technology poses significant threats to the authenticity of content on social media. This research introduces Sach-AI, a pioneering framework for detecting various deepfakes in video, audio, and image data. Leveraging the power of deep neural networks, Sach-AI utilizes Eulerian Video Magnification combined with the ResNext architecture for enhanced detection. For video deepfakes, Long Short-Term Memory (LSTM) networks are integrated to improve classification tasks. This combination allows Sach-AI to effectively address the evolving, multimodal nature of deepfakes. The framework has been rigorously evaluated using diverse datasets, such as Celeb-DF and FaceForensics++, demonstrating its robustness and accuracy. Sach-AI achieved 97.76% accuracy in video deepfake detection, surpassing Intel's FakeCatcher, 99.13% accuracy in audio deepfake detection, and 93.64% accuracy in image deepfake detection. These results underscore Sach-AI's reliability in safeguarding digital media integrity against deceptive synthetic content in an era increasingly dominated by artificial technologies.

INDEX TERMS Deepfake, multimedia, audio-video analysis, temporal analysis, media authenticity.

INTRODUCTION

In recent times, the most rapid evolution and progress by taking the potential of artificial intelligence technologies has brought forth the best innovative solutions in every field along with new challenges [1], [2]. One of these challenges that has gained enough importance and prominence is the proliferative expansion of superimposed content or ''deep- fake'' content across various digital and multimedia-based platforms [3]. The Deepfakes are based mainly on synthetic media, which are also generated by different AI algorithms based on GAN (Generative Adversarial Networks) [4], [5]. These had the strong potential for manipulation of different visual, auditory, and textual information, ultimately posing remarkable danger in the case of originality and gaining The associate editor coordinating the review of this manuscript and approving it for publication was Zijian Zhang . the trustworthiness of virtual content [6]. Sach-AI, state- of-the-art research, which stands as a strong pioneering framework, was designed and made to address this type of sophisticated nature of different deepfake contents, which spanned itself over video [7], [8], audio [9] and image [10] domains. This comprehensive research framework is aimed at the integration of all the deepfake content detection in a single place and the enhancement of the robustness and accuracy of deepfake detection mechanisms by leveraging different efficient deep neural networks. The methodology of this framework would involve the integration of the Eulerian Video Magnification

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

and ResNext for the working of feature extraction in the video domain. Next, after the feature extraction, Long-Short-Term Memory or LSTM does the job of classification. In the audio analysis, CNN played a very important role while DenseNet121 pre-trained was fine-tuned on its dataset for image processing. **The work done in** 48683

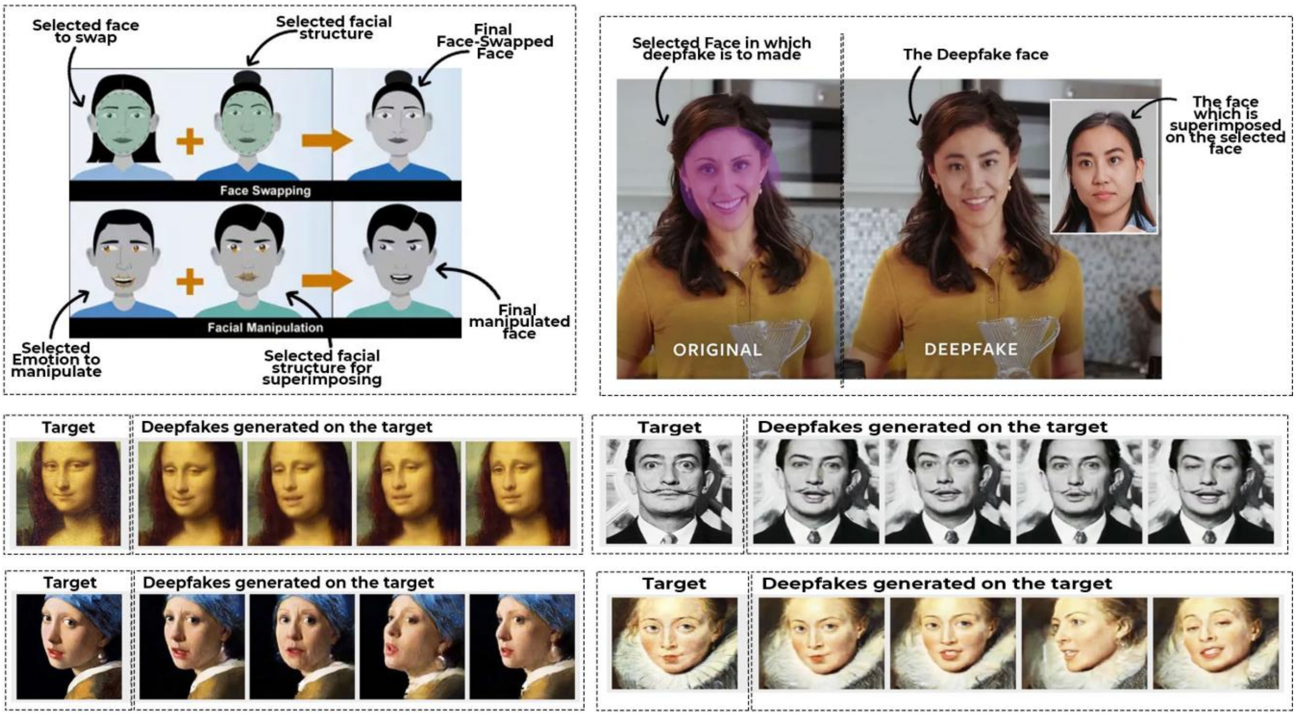


FIGURE 1. A visual representation of the real and superimposed image (deepfake) created. The original face, selected facial structure, and final manipulated face are displayed. The process involves superimposing a selected face onto a target face, often with the manipulation of facial emotions.

this research is novel in its kind as a multimedia-based framework combining all three forms of media is first proposed.

The **novelty of this work** highlighted through this research framework are

Improved accuracy achieved with advanced neural network technology in video deepfake detection, which **surpassed the accuracy achieved by Intel's FakeCatcher. [11]**

The novel approach in audio deepfake detection with **the Mel spectrogram approach**, which is unique of its kind and provides improved robustness.

Fine-tuned pre-trained model on a custom dataset for image deepfake detection and **surpassed the existing architectures**, which faced difficulties in advanced deepfake generation algorithms.

Able to beat the deepfakes created with the most **recent FaceSwap algorithms** from the perspective of consumers.

The different face-swapping techniques used as part of the feature extraction, such as realistic lip-syncing and adaptive generation methods, ultimately contribute to the increasing complexity of detecting different manipulated videos, as shown in Fig 1. The video content, or the audio- visual content, has become a primary medium of focus for the propagation of deepfake, and this framework Sach-AI uses its own

Eulerian Video Magnification. This technique uses its effective method, which enhances the subtle colour and motion variations within a video. This, in turn, allowed for more improved feature extraction, and made it more adaptable for the identification of different anomalies which were introduced by deepfake generation algorithms. The fusion of ResNext with this made this a very powerful convolutional neural network (CNN) and, with further refined feature representation, enhanced the working capability of the framework in the discerning of the intricate patterns inside any manipulated video content and sequences. This groundbreaking proposed framework of Sach-AI promised the whole revolutionisation of deepfake detection, with all the cutting-edge and encompassing solutions to combat fake content across all formats of multimedia. Advanced neural network architectures like Eulerian Video Magnification, ResNext, LSTM, and CNN, along with DenseNet121, were integrated into this framework, and this system stood as the main frontline defender in this digital age. The risks posed by the deepfake content in the digital landscape and the system would ensure the safest online experience for all types of end-user consumers, giving protec- tion from misinformation and manipulation. The full working methodology summarised in a video with working example of testing is shown in the unlisted Youtube video and Google Drive. The manuscript is organised as follows: Section II contains the literature review of the works done in this field, and Section III contains the information of the dataset used for the development of framework. Section IV contains the methodology and algorithms discussed, and Section V shows

TABLE 1. Literature review of different existing image/video deepfake detection methods.

Types of Approach used	Year	Techniques/Algorithms Used	Methodology used	Limitations and Challenges faced
Signal-level features-based detection [12]	2017	GoogleNet, Patch Triplet Stream	Two streams which included GoogleNet for Facial Classification along with Patch Triplet Stream for local Noise residual	The proposed methodology performed well but failed on advanced deepfake generative technology
Physiological attributes-based detection [13]	2018	Eye Blink Pattern Analysis	This methodology detected the deepfake videos using the eye blink pattern	The assumption of this is that there was a time difference between the original and deepfake videos. The current advanced deepfake videos were vigorous in detecting visual features
Data-driven models [14]	2018	InceptionNet, DenseNet, XceptionNet	This manuscript performed a comparative study over different pre-trained models including InceptionNet, DenseNet, and XceptionNet. The last one performed best.	The process lacked generalizability
Data-driven models [15]	2018	CNN (Five-layer architecture)	This proposed a total of five-layer CNN architecture, which was called a deep forgery discriminator.	This total process missed generalizability
	2018	Inception Module	The CNN architecture utilized the Inception module as the backbone of the architecture.	The process worked well with compressed videos, but the models outperformed it in every dataset present
	2019	CNN with Motion Analysis	This method tried to exploit different discrepancies in motion across frames at f(t) and f(t+1). This process also used CNN as an algorithm classification.	The other algorithms outperformed this method
Physiological attributes-based detection [18]	2019	Virtual Artifacts (Eye Color Differences, Shadows)	Here, virtual artifacts were used, such as differences in the eye colors, and various disproportionate shadows in the detection of deepfakes.	Visual features were not reliable with the advanced deepfakes
Signal-level features-based detection [19]	2020	Scale-Invariant Feature Transform (SIFT)	Here, a popular feature extraction method known as Scale-invariant Feature Transform (SIFT). This detects the keypoint pixels and extracts their features.	The method was good but with more advanced deepfakes, the local features were becoming difficult to detect
Gaze-Guided Spatial Inconsistency Learning [20]	2023	GazeForensics (Gaze-Guided Learning)	GazeForensics employs gaze-guided spatial inconsistency learning to detect deepfakes by focusing on the eye movements in videos	Difficulty in detecting deepfakes with minimal eye movements; dependency on high-quality eye tracking data
Analysis of Deepfake Detection vs. Image Quality Metrics [21]	2024	Image Quality Metric Correlation Analysis	Examined the correlation between detection performance and image quality, highlighting metrics that influence detection accuracy	Lack of robustness against different types of deepfakes; difficulty in balancing image quality and detection performance

Data-driven models [16]

the proposed system design for the framework. Section VI shows the results obtained along with the test results on different multimedia, and finally Section VII concludes the findings of the paper with Section VIII showing the future scope of the framework.

LITERATURE SURVEY

The deepfake technology, which is mainly driven by the different advancements in the field of artificial intelligence and machine learning, has always raised major and significant concerns regarding the manipulation of digital audio-visual content. The literature review is aimed here at the exploration of different important aspects related to deepfake technology along with its challenges and its ongoing efforts in the detection and combating of deepfake content. Table 1 shows the literature review of the existing architectures in the image/video deepfake detection methods. Westerlund et al.
in [22] in 2019 provided a thorough and elaborated review of the various emergence of deepfake technologies. The emphasis is on the fast expansion and its excessive effect on society. The author also focused on the demand for several cautious measures to address the various ethical and security use cases of deepfake technology. Yu et al. [23] in 2021, conducted an elaborate survey on deepfake video detection. They talk about the different approaches and algorithms as well as their issues in the identification of distorted videos. The authors also focus on the importance of their substructure and its strong detection mechanism for opposing the changing behavior of several deepfake technologies. Dolhansky et al.

[24] 2020 showed a novel dataset named as Deepfake Detection Challenge (DFDC) dataset, which came out as a very precious asset for researchers and developers for setting a point of reference and showing a remarkable enhance- ment in the deepfake detection algorithms. Additionally, Li et al. [19] also put forward their dataset titled ‘‘Celeb-DF’’, a wide-ranging, challenging dataset purposely designed for executing forensics in the deepfake technology. Zhao et al. [25] put forward a multi-attentional deepfake detection method, which influences the strength of recently developed attention mechanisms in the amplification of accuracy in the detection procedures. Rana et al. [26] presented a structured and methodical assessment of deepfake detection and eventually gave an outline of its major findings and speedy growth and development in the area. The assessment represents an understanding of the current predicament of the research foundation with the identification of capable fields for future investigations. John et al.
[27] conducted a detailed comparative analysis of the different deepfake detection methods while including a discussion on several semi-supervised GAN architectures for improving the procedure of detection. This study also contributed to the understanding of the several strengths and limitations of various approaches in the detection process. Garg et al. [28] presented a study on the deepfake generation and detection in an exploratory manner, ultimately shedding light on the challenges which were associated with the creation and identification of manipulated content. Khder et al. [29] explored the fusion of artificial intelligence into the world of deepfake creation and detection, with deep discussions on the implications it could have on the AI-driven advancements in this field. Zhang et al. [30] proposed a heterogeneous feature that was based on the ensemble learning method. From this, a combination of the diversified features could also help in the enhancement of accuracy in the detection of deepfakes. Jia et al. in [31] mainly focused on the attribution of different models which was based on the face-swapping videos generated by deepfakes. The identification of specific models in this field of deepfake detection was also significant, as pointed out. Trabelsi et al.
[32] discussed the improvements in the detection of deepfakes by the mixture of the top solutions

available. The collaborative approaches in this framework could also help in the enhancement of deepfake detection along with gaining abilities for further robust detection. Yadav et al. also proposed a more robust and generalized method of detection [33]. This was very important for aiming for more robust detection methods which could be fit in various scenarios. Malik et al. [34] also explored the different detection of deepfakes in the video domain which consisted of driving videos in animated mode. The detection of different manipulated content within these contexts, like that of the driving videos, was also important. Verma et al. [35] leveraged the potential of pre-trained Inception-ResNetV2 for deepfake detection. The various advanced pre-trained neural network architectures could also contribute to the overall improvement in the field of accuracy in deepfake detection. Jaleel et al. [36] proposed a novel method, which was based on facial behavior analysis using different types of Generative Adversarial Networks (GAN). The integration of the analysis of facial behavior along with GAN differentiation would also certainly help in the enhancement of different capabilities of the detection. Alnaim et al. [37] presented their dataset called DFFMD, which was a comprehensive dataset mainly used for deepfake face masks along with various algorithms for detection. The addressing of various challenges, which were also introduced by face masking, was very crucial, as stated in this paper. Pipin et al. [38] proposed a novel method that mainly leveraged the potential of spatiotemporal convolutional networks for the detection of deep fakes. The proper utilization of spatiotemporal features, which were embedded along with the photo response, improved the accuracy of detection. Liu et al. in [39] investigated the various types of irregular identities in the deepfake videos. Hsu et al. [40] suggested pairwise learning, which increased the accuracy of the method after comparing the pairs of images and analysing the discrepancies. Kohli et al. [41] presented a frequency-based Convolutional Neural Network (CNN) for detecting facial forgeries, which means recognising the irregularities of the facial image in terms of the frequency domain. Qiao et al. [42] proposed a novel method for deepfake video detection that is fully unsupervised and built upon ECL. Using a self-supervised approach, the model separates the real from the fake videos to eliminate the need for labeled data, making the model more generalizable in the actual world setting. To address these problems, Chen and Hsu [43] present a decoy mechanism to guard against deepfake manipulations and adversarial attacks. The current approach adds defensive decoys to the model to distract the endangering entity while not compromising the model's primary function and thus improves its functionality against various risks. Shao, R. et al. [44] respond to the afore-said challenge of detecting and grounding multi-modal media manipulation through the proposal of a broader framework beyond detection. The approach uses visual and textual signals to detect manipulations and has a detailed description of each change, which contributes to a better understanding of the result of the recognition. The due exploration of different methodologies, and datasets would underscore its commitment in order to mitigate the different ethical and security implications associated with the deepfake technology.

A. CASE STUDY ON INTEL'S "FAKECATCHER "

In a study regarding deepfake detection conducted by Intel [11] in 2022, they developed a real-time deepfake detector in the category of audio-visual content or video content. They developed their framework named Fake- Catcher, which was a technology that could detect fake videos with an accuracy rate of 96%. This ran on the basis of hardware and software by Intel on a web-based

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Boost by Intel. The accuracy demanded by Intel was nearly about 96%, but no such significant information was available on the algorithms they used or the proper methodology they followed.

While the solution provided by Intel represented a significant stride in this field of deepfake detection, this framework Sach-AI stood out with the integration of a multi-modal approach with all three types of content. The integration of Eulerian Video Magnification with ResNext for feature extraction would be a unique algorithm developed and this would give the improved accuracy than any previous works done. The future scope of this framework also includes the integration of blockchain for enhancement of security and attribution, along with its geographical origin tracking of the content produced. The proposed blockchain algorithm would go beyond the detection and allow the identification of the main culprit of deepfake creators and their source of origin.

DATASET

The dataset used for the development of the framework consisted of diverse videos from different categories. For the dataset in the video domain, the framework was built on two datasets, and the best was taken into consideration. The two datasets include FaceForensics [45] and Celeb-DF [19].

The FaceForensics++ dataset (https://niessnerlab.org/ projects/roessler2018faceforensics.html) is mainly a foren- sics dataset that consists of a total of 1000 original video frames and sequences. The videos are collected in such a way that the videos have been manipulated with the four face manipulation methods in automation mode: Deepfakes, FaceSwap, Face2Face and NeuralTextures. The videos in the dataset were sourced from the social platform YouTube, and every video consists of a trackable, mostly frontal face without any occlusions, which ultimately enabled the automated tampering methods for the generation of realistic deepfakes.

The second dataset used here was Celeb-DF [19] (https:// cse.buffalo.edu/ siweilyu/celeb-deepfakeforensics.html), which evolved due to the problem with existing datasets of low visual quality and poor resemblance with the videos circulated over the World Wide Web. This dataset was a new and large scale challenging dataset, which consisted of 5630 high-quality Deepfake videos of the celebrities generated using the process of improved synthesis process. The comprehensive evaluation of deepfake detection methods along with the datasets was also conducted for the demonstration of the uplifted level of the different challenges posed by Celeb-DF.

The dataset used for this study consisted of all 70k REAL faces from the Flickr dataset collected by Nvidia, as well as 70k fake faces sampled from the 1 Million FAKE faces (generated by StyleGAN) that was provided by Bojan.

The dataset used for audio deepfake detection was ASVspoof 2019. This is the Third Automatic Speaker Verification Spoofing and Countermeasures and the goal is to continue the research on countermeasure techniques related to ASV. This data is of logical and physical access and it is based on the VCTK corpus, was obtained with the spoofing attacks as text-to-speech, voice conversion and replay attack. It assists progress through a standard with a fixed test set, comparative assessments, and a novel, ASV-centered measure for measuring countermeasures on behalf of emerging spoofing techniques (t-DCF) [46].

METHODOLOGY AND ALGORITHMS DISCUSSED

Here, this research aims to give a long-vision proposal for the smooth working of this pioneering framework, which is mainly designed for the detection of deepfake content across multiple diversified domains, including video, audio and image. The methodology kept here is simple, which leverages the potential of different advanced techniques with a combination of different state-of-the-art algorithms. For the domain in the video, the approach used in the framework utilized the Eulerian video magnification technique for the enhancement of subtle temporal variations, which was fused with the ResNet architecture for the extraction of different features for the robustness of the model. The temporal features were extracted from the videos and then passed on to and fed into the pre-defined Long-Short Term Memory (LSTM) networks for undergoing classification. From the audible perspective, the Convolutional Neural Networks were used for feature extraction. This ultimately allowed the effective and accurate characterization and differentiations of different audio patterns identified in the deepfake content. In the visual content, the DenseNet121 pre-trained model, along with its versatile architecture, was employed for the extraction of discriminative features from the data. The real-time integration of all these diversified modalities would ultimately facilitate the process with a more holistic and multi-dimensional approach toward deepfake detection. The application of LSTM, CNN, and other dense neural networks with their respective domains ensured the true ability of the model to discern intricate patterns and certain anomalies that are indicative of deepfake manipulation. The true collaboration between the different algorithms and maintaining the synergy enhanced the overall robustness and accuracy of Sach-AI. It made it a comprehensive and more effective framework in fighting against deepfake content. This proposed methodology would showcase the different significances of leveraging the state-of-the-art architectures specifically designed for each modality and ultimately contributing to the collaborative fight of the nation against deepfake threats.

Sach-AI is developed as a cutting-edge framework for deepfake detection across different types of content, includ- ing audio-visual and individuals. The algorithms applied here form the backbone of the architecture of the system. The video analysis, which mainly focused on the Eulerian Video Magnification and ResNext for the feature extraction, formed the crucial part of Sach-AI.

A. VIDEO DEEPFAKE DETECTION

1) EULERIAN VIDEO MAGNIFICATION(EVM)

The process of Eulerian Video Magnification is a very effective signal-processing technique, which mainly focuses on the amplification of temporal color variations in video sequences. This was originally proposed by Wadhwa et al.

[47] in 2013. The main backend idea behind this is the accurate application of spatial decomposition and temporal filtering in order to enhance the subtle color changes that might be indicative of different deepfake manipulations.

EVM Spatial Decomposition

$$I_c(x, y, t) = I(x, y, t) - \bar{I}_c(x, y, t). \qquad (1)$$

This formula in Equation 1 represented the spatial decomposition step in Eulerian Video Magnification. Here, I_c denoted the color channel (Red, Green, Blue) at different pixel coordinates (x,y) and time t. I(x,y,t) is the original pixel value, and I_c(x,y,t) is the spatially filtered version of the color channel. The result I_c(x,y,t) represented the amplified color variations that are visible after spatial decomposition.

EVM Temporal Filtering

$\partial I_c(x, y, t)$

Algorithm 1 Video Deepfake Detection

Input: Video Sequence $V = \{I_1, I_2, \ldots, I_n\}$, where I_i is the i th frame of the video. Parameters for EVM, ResNext, and LSTM networks.
Output: Prediction of whether the video contains deepfake manipulation or not.
Initialization:
each frame I_i
Apply Spatial decomposition using Eq. [1] to obtain $I_{c,i}$.
Apply temporal filtering to each $I_{c,i}$ using Eq. [2] to obtain $\Delta I_{c,i}$.
each $\Delta I_{c,i}$
6: Extract features using the ResNext architecture as in Eq.
[3] to obtain feature maps $F = \{F_1, F_2, \ldots, F_n\}$.
Initialize LSTM cell states C_0 and hidden states h_0. each time step t from 1 to n
Compute the forget gate f_t using Eq. [4].
Compute the input gate i_t using Eq. [5].
Compute the candidate cell state $\tilde{C}t$ using Eq. [6].
Update the cell state C_t using Eq. [7].
Compute the output gate o_t using Eq. [8]. 13: Update the hidden state h_t using Eq. [9]. 14: Obtain the final hidden state h_n.
Feed the final hidden state h_n into a fully connected layer for binary classification (real vs deepfake).
End.
$\Delta I_c(x, y, t) =$
*(2)
Here, in Equation 2, $I_c(x,y,t)$ represented the temporal derivative of the color channel I_c with respect to time. This helped in the measurement of the rate of change of color at a given pixel over time. The temporal filtering was essential for the highlighting of the various subtle temporal variations, which made them more discernible for different subsequent analyses.

ResNext for Feature Extraction

ResNext is a renowned deep neural network architecture that extended the ResNet architecture with the introduction of a unique cardinality parameter. This enhanced the capability of the architecture to capture more diverse and fine-grained features from the input data. In this framework, Sach-AI, ResNext was employed for the feature extraction from the magnified video frames obtained through the Eulerian Video Magnification.

$$FeatureMap : F = \sigma(B(F_{in} + W_{i,j}xin) \quad (3)$$

Here, in Equation 3, F_{in} is the input feature map, $W_{i,j}$ represented the weights, x_{in} is the input data, and σ denoted the activation functions. The feature map F aims to capture the high-level representations of all the video frames, which is crucial for subsequent classification. The algorithm developed for the proposed framework for video deepfake detection is discussed below as Algorithm 1.

LSTM(Long-Short-Term-Memory)

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (5)$$

$$\tilde{C}t = tanh(W_C.[h_{t-1}, x_t] + b_c) \quad (6)$$

$$C_t = f_t.C_{t-1} + i_t.\tilde{C}t \quad (7)$$

$$\sigma_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t.tanh(C_t) \quad (9)$$

Here, all the equations from 4 - 9 represented the updating process within the LSTM cell. In Equation 4, The forget gate f_t helps in the determination of how much the previous cell state $C_{t\ 1}$ is to retain. This is computed by the application of the sigmoid activation function to the weighted sum of the previous hidden state h_{t-1} along with the current input x_t. In Equation 5, the input gate i_t determines and controls the amount of new information $\tilde{C}t$ to store in the cell state. This is computed similarly to the forget gate, keeping in mind the weighted sum of the previous hidden state h_{t-1} and the current input x_t. In Equation 6, the candidate cell state denoted by $\tilde{C}t$ shows the new information that is to be added to the cell state. It is computed by the application of the hyperbolic tangent (tanh) activation function to the weighted sum of the previous hidden state h_{t-1} along with the current input x_t. In Equation 7, the cell state C_t is updated here by the combination of the retained information from the previous state C_{t-1} (controlled by the forget gate) along with some

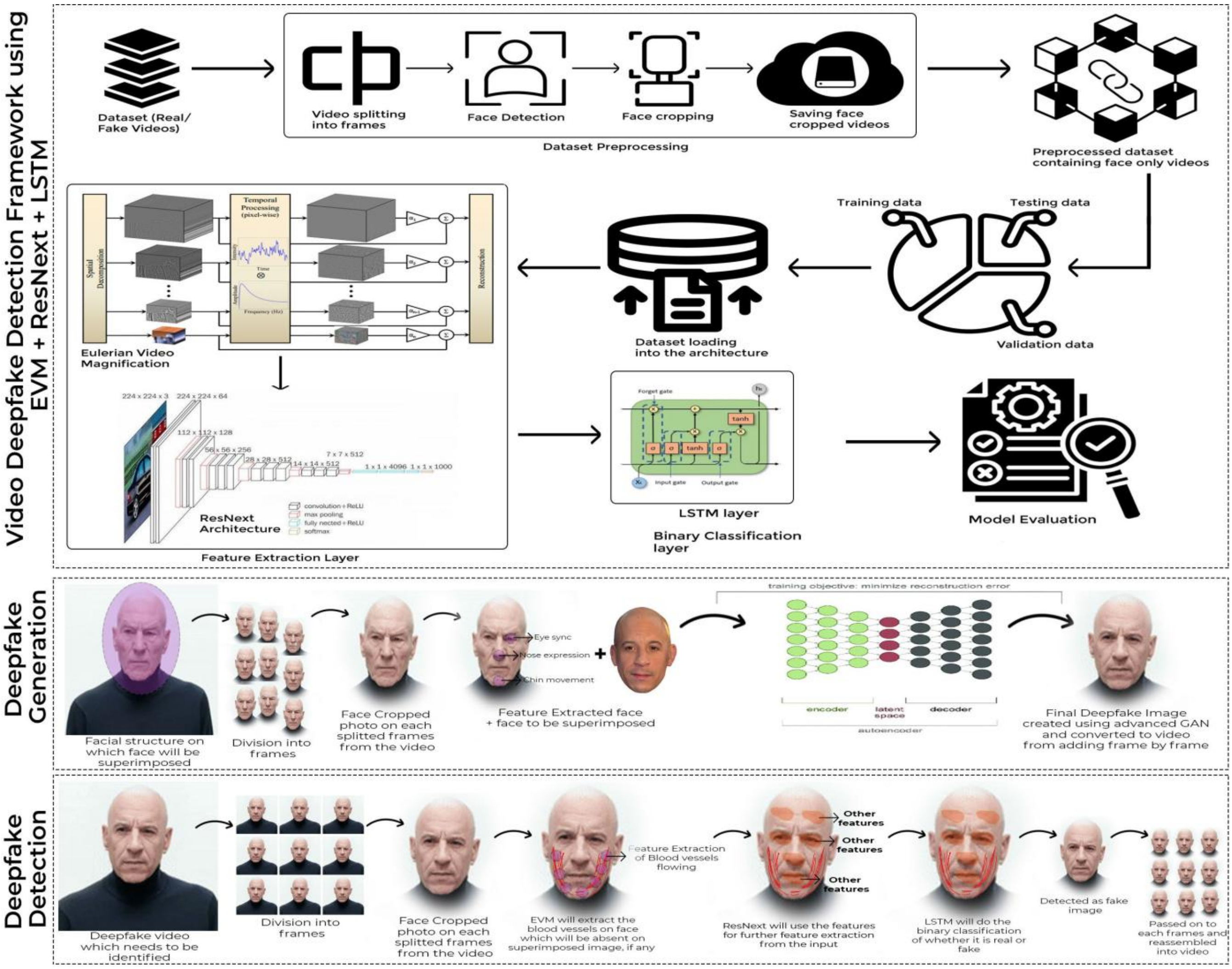


FIGURE 2. Working Diagram of the Video Deepfake Detection framework developed. The full methodology followed for the video deepfake detection framework using advanced neural network is well designed and described with this figure. Secondly, deepfake generation method is also shown using advanced GAN (Generative Adversarial Networks). Thirdly, the method used for deepfake generation is also shown with proper visual representation of the image at every step of the process.new information $\tilde{C}_t$ (controlled by the input gate). Equation 8 shows the output gate o_t , which is determined by the next hidden state h_t based on the current input x_t along with the previous hidden states h_{t-1}. This is also computed using the sigmoid activation function. Lastly, Equation 9 shows the new hidden state h_t is computed by the application of the hyperbolic tangent (tanh) activation function to the updated cell state C_t and multiplying it by the output gate o_t .
Fig. 2 describes the working diagram of the video deepfake detection.

B. *AUDIO DEEPFAKE DETECTION*

1) MEL SPECTROGRAM

A Mel spectrogram is a representation of the short-term power spectrum of a specific sound signal that has been converted from its original form into the Mel frequency scale. This type of spectrogram is very useful, particularly in audio signal processing, speech recognition tasks, and many similar applications. Fig. 3 is created by the application of a Mel Filterbank to the spectrogram, where the scale mimicking the human perceptual sensitivity in connection to different frequencies is used for designing the overlapping of triangular filters. These filters help in capturing the energy within specific ranges of frequency, and the resulting values within the filter were subjected to logarithmic compression. Next, stacking all the filtered values for each time frame makes the Mel spectrogram, where each column represents the energy distribution across Mel frequency bins. The representation is crucial for capturing the perceptually relevant and significant information in the generated audio signals, which will further help in the extraction of features from Mel spectrograms for accurate audio classification.

2) **CONVOLUTIONAL NEURAL NETWORKS**

The Convolutional Neural Network is a deep learning algorithm specifically designed for the processing of

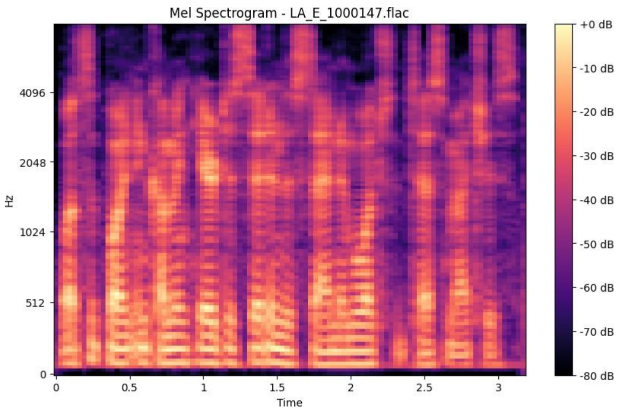


FIGURE 3. Sample MeL Spectrogram plot of the test set on audio deepfake detection.

grid-like data, such as images and spectrograms. In the context of this framework developed, the CNN algorithm was employed for the accurate extraction of features, specifically local, from the Mel spectrogram. The convolutional layer applied various filters across different parts of the Mel spectrogram generated for the detection of patterns and features. The output of the convolutional layer here is calculated using the Equation 10.

Algorithm 2 Audio Deepfake Detection

nput: Audio Sequence A $a_1, a_2, \ldots, a_n$, where a_i is the i th segment of the audio. Parameters for Mel Spectrogram, CNN, and other network layers.

Output: Prediction of whether the audio contains deepfake manipulation or not.

Initialization: each audio segment a_i

Convert a_i to Mel Spectrogram using the Mel Filterbank.

Apply logarithmic compression to the Mel Spectrogram.

nitialize the CNN with convolutional layers and apply it to the Mel Spectrograms to extract feature maps.

each

feature map

Apply MaxPooling to downsample the spatial dimen- sions.

Apply Batch Normalization to normalize the activations.

Apply ReLU activation function element-wise.

Apply Dropout for regularization during the training process.

Apply Global Average Pooling to obtain a fixed-size representation.

Feed the pooled representation into a Dense layer with a sigmoid activation function.

Obtain the final output as a probability score indicating the likelihood of the input being a deepfake.
End.

Output $(i, j) =$
$F-1$ $F-1$

$m=0$ $n=0$*Input*

$i+m$
$j+n$
$\times$ *Filter* $(m, n) + $ *Bias*
(10)

which helps in the prevention of division by zero, γ is the scaling parameter, and β is the shifting parameter. Here, in Equation 10, the Output (i,j) is the output at position (i,j). Input (i m, j n) is the input at position (Im,j n).+Filter+(m,n) is the filter at position (m,n). Bias stands as the bias term, and F is the filter size.+ +

Max Pooling Layer

After the convolutional layer, the MaxPooling layer plays

Dropout Layer

This layer actually performs the technique for regular- ization, which is used to prevent the overfitting situation. During the training of the model, various random neurons get deactivated with a certain probability. The output to this layer is calculated with the Equation 13.
its role as it helps in downsampling and in the reduction of the spatial dimensions while retaining the most important information. This is achieved by the improved value within

Dropout(x)
$\underline{Mask * x}$

$1 -$
(13)
every region of the feature map. The operation could be represented as in Equation 11.
$MaxPooling(i, j) = max_{m,n} Input(ixs + m, jxs + n)$ (11)
Here, in Equation 11, MaxPooling(i,j) is the output of the MaxPooling operation at position (i,j). s is the stride of the MaxPooling operation.

Batch Normalization

This process is applied for the normalizing of the

(13)
every region of the feature map. The operation could be represented as in Equation 11.
$MaxPooling(i,j) = max_{m,n} Input(ixs + m, jxs + n)$ (11)
Here, in Equation 11, MaxPooling(i,j) is the output of the MaxPooling operation at position (i,j). s is the stride of the MaxPooling operation.
Batch Normalization
This process is applied for the normalizing of the
Here, in Equation 13, a Mask is a binary mask with values
only 0 or 1. This indicates whether a particular neuron is activated or it is deactivated. p is the rate of dropout.
Global Average Pooling Layer
The global average pooling layer aggregates the feature maps globally in order to obtain a fixed size of the representation. It then calculates the average value of each feature map, which would result in the one-dimensional vector. This could be expressed as in Equation 14.
activations or activation functions within a layer, ultimately
preventing the internal covariate shift and stabilizing the

$GAP(x) = \frac{1}{\quad} \sum\sum x(i,j)$ (14)
training process. The normalized output BN(x) for the input x is calculated as in Equation 12.

$H * W$

$i=1 \; j=1$
$BN(x) = \frac{x-\mu}{\quad}$
$+$

$* \gamma + \beta$ (12)

Here, in Equation 14, H and W are the height and width of
the feature maps.
Dense Layer
Here, in Equation 12, μ denoted the mean of the batch, σ is the standard deviation of the full batch, ϵ is a small constant
The final layer in the audio deepfake detection architecture is the Dense layer, which performs the task of classification

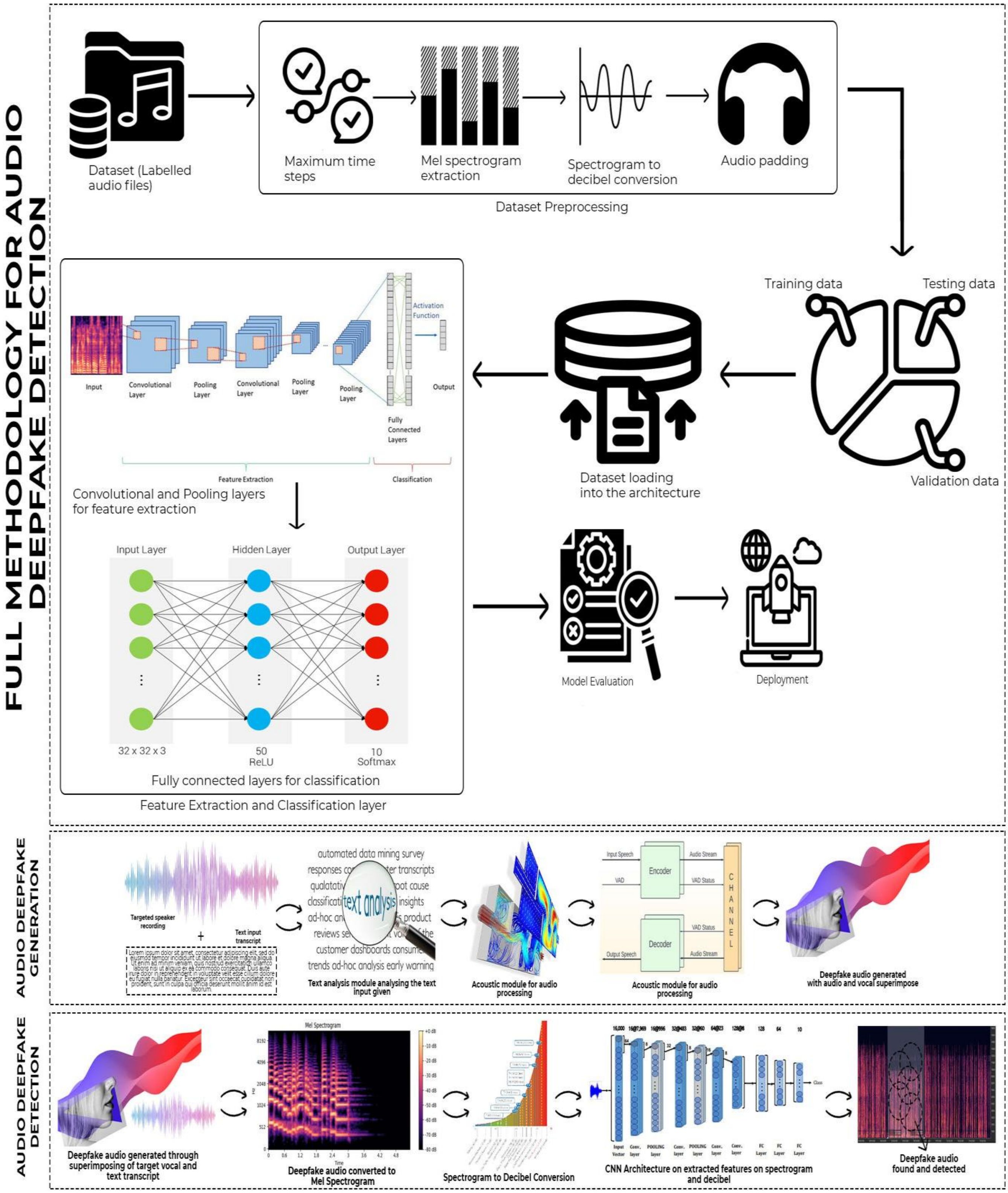


FIGURE 4. Working Diagram of the Audio Deepfake Detection framework developed. The full methodology followed for the audio deepfake detection framework using convolutional neural network using MeL Spectrogram approach is well designed and described with this figure. The steps for deepfake audio generation and proposed way of detection through Mel Spectrograms is also well described of every step.

using the sigmoid activation function. The output will be a probability score, which would indicate the likelihood of the input is deepfake or not. The equation for this computation of this layer is shown in Equation 15.

$$Dense(x) = \sigma(W.x + b) \quad (15)$$

Here, in Equation 15, σ is the sigmoid activation function, W is the weight matrix, and b is the bias vector. The workflow for the audio deepfake detection is clearly shown in Fig. 4, along with the procedural approach in Algorithm 2.

C. IMAGE DEEPFAKE DETECTION

1) DENSENET ARCHITECTURE

Dense blocks, which are known as the DenseNet network, is a densely connected convolutional network, which is a deep neural network architecture typically characterised by its strongly connected connection patterns. Unlike CNN (convolutional neural network), which is based on a skip connection between each layer, DenseNet connects

hierarchically, meaning that every layer is connected to every other layer. This given structure also ensures the effective sharing of resources within the network, as well as passes gradient flow and parameter economy between units. Densely packed blocks form architecture, with every block consisting of multiple layers squashing. Every segment of the block from each layer will pass on the feature path maps from the previous layers and also use its feature path maps, which, in turn, will be used by all the next conforming layers. This event was followed by the emergence of a highly codified and entangled spatial structure.

Dense Connectivity

This process is applied for the normalizing of the activations or activation functions within a layer, ultimately preventing the internal covariate shift and stabilizing the training process. The normalized output BN(x) for the input x is calculated as in Equation 16.

$$H_l = H_0 \oplus H_1 \oplus \ldots \oplus H_{l-1} \quad (16)$$

hierarchically, meaning that every layer is connected to every other layer. This given structure also ensures the effective sharing of resources within the network, as well as passes gradient flow and parameter economy between units. Densely packed blocks form architecture, with every block consisting of multiple layers squashing. Every segment of the block from each layer will pass on the feature path maps from the previous layers and also use its feature path maps, which, in turn, will be used by all the next conforming layers. This event was followed by the emergence of a highly codified and entangled spatial structure.

Dense Connectivity

This process is applied for the normalizing of the activations or activation functions within a layer, ultimately preventing the internal covariate shift and stabilizing the training process. The normalized output BN(x) for the input x is calculated as in Equation 16.

$$H_l = H_0 \oplus H_1 \oplus \ldots \oplus H_{l-1} \quad (16)$$

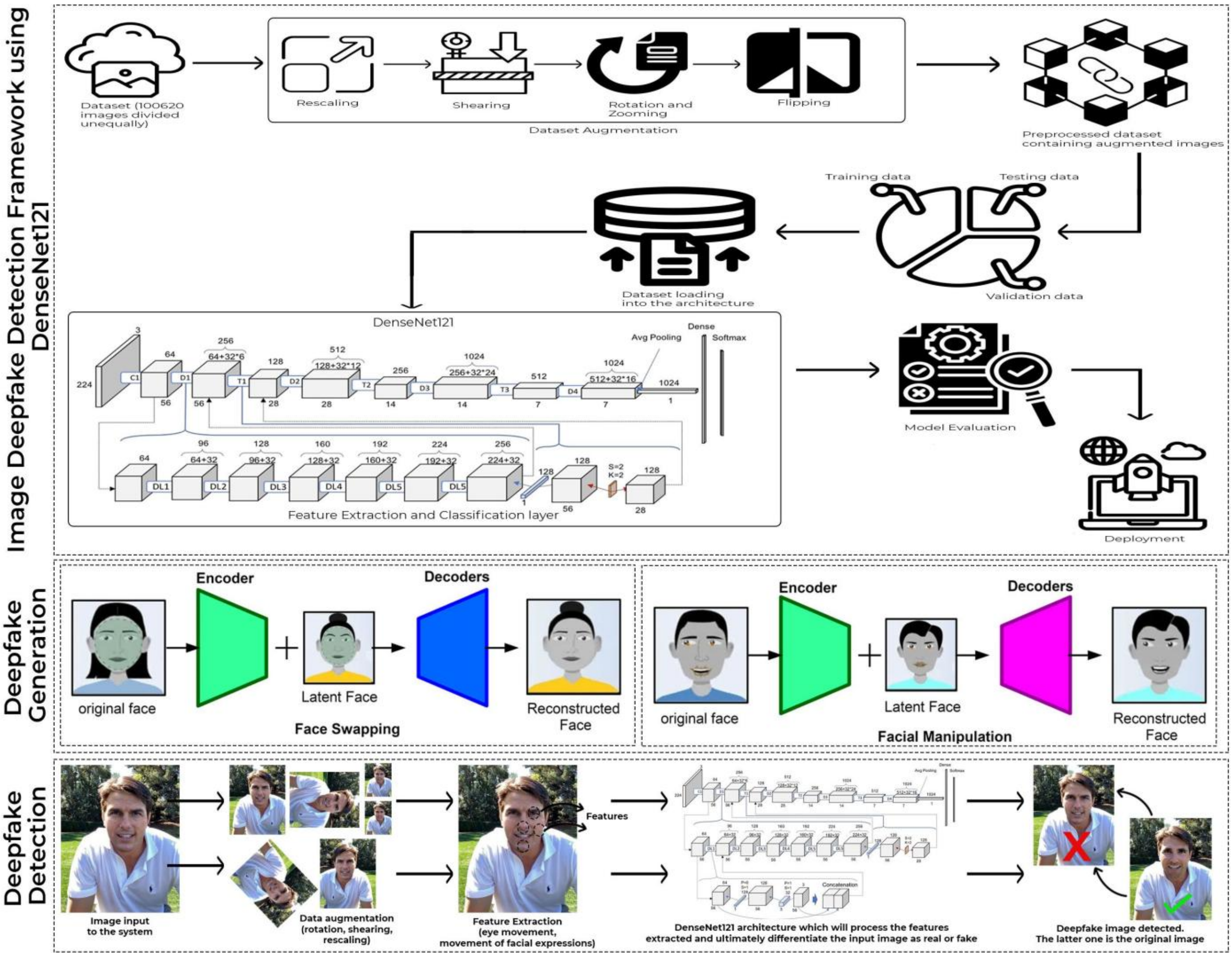


FIGURE 5. Working Diagram of the Image Deepfake Detection framework developed. The full methodology followed for the audio deepfake detection framework using a pre-trained model and fine tuning is well-designed and described in this figure. The image deepfake generation is also shown using Face Swap and Face Manipulation methods. The proposed methodology also shows how these methods can be detected with the proposed methodology.

Let H_l represent the output of the l-th layer. The feature maps, which are concatenated, were denoted as $[H_0, H_1, \ldots, H_{l\ 1}]$, and this will ultimately serve as the input for the l-th layer, as shown in Equation 16. The output of the n-th layer was given by Equation 17, where F represented the composite function (convolution, batch normalization, along activation function).

$$H_l = F_l([H_0, H_1, \ldots, H_{l-1}]) \qquad (17)$$

The growth rate(k) falls under the hyperparameters, which helps determine the number of feature maps produced by each layer within a dense block. This influenced the capacity of the network along with its computational requirements. DenseNet121, the model used in the framework for image deepfake detection, was initially trained on a large dataset for general image classification tasks. The pretrained weights for this model capture the generic features which are helpful for various visual recognition tasks. For making the pre-trained

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

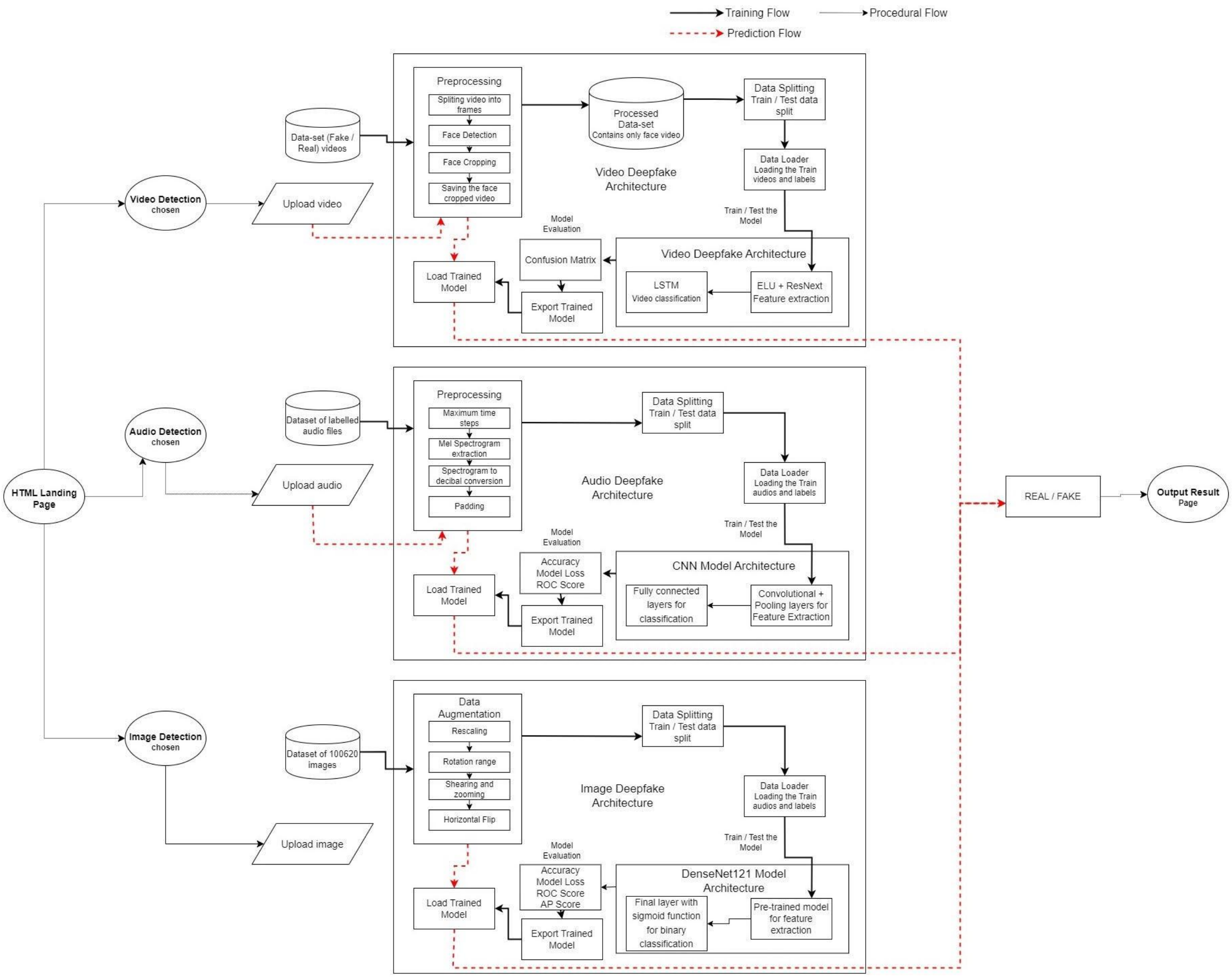
model adapt to the deepfake detection, this was further trained on its dataset of 100620 images related to deepfake scenarios. The original classification layer was replaced with the new layer using fine-tuning, which made the model best suitable for binary classification between fake and genuine.

Loss Function

The model training process involved minimising the loss function, such as binary cross-entropy, for optimising the network parameter for accurate deepfake detection.

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)] \qquad (18)$$

Here in Equation 18, N is the number of samples used, y_i is the ground-truth label, and $\hat{y}_i$ is the pre- dicted probability of being fake. The roadmap methodology for image deepfake detection with its working diagram in Fig. 5.



I. PROPOSED SYSTEM DESIGN

Sach-AI presented the innovative framework for the most comprehensive deepfake detection across multiple platforms and content types, which include video, audio, and image, ultimately leveraging the potential of advanced deep neural networks. The proposed system, as shown in Fig. 6, utilized the Eulerian Video Magnification, which was fused with the ResNext architecture to increase the robustness of the feature extraction process in video content types. This combination allowed for more enhanced analysis of various temporal patterns and subtle inconsistencies, making the detection of deepfake more accurate and effective. In the audio content, convolutional neural networks were employed for the extraction of discriminative features, which enabled the model to identify different anomalies in the acoustic digital signatures of genuine and manipulated or ‘‘fake’’ content. In the field of visual deepfake detection, DenseNet121 architecture was employed for more effective feature extraction along with the analysis. The diversity
in full neural network architecture catered to the different unique characteristics and challenges posed by each of the modalities. The classification task was handled using long short-term memory networks for the audio-visual part, which ensured the ability of the model to capture temporal dependencies and different patterns essential in the differ- entiation of deepfake from authentic videos. This approach enhanced the overall detection accuracy and robustness of Sach-AI, which made it the most formidable solution against the evolving landscape for deepfake detection across all media types. Through this fusion of all the state-of-the-art neural network architectures, this framework stood as the most versatile and effective framework in combatting the proliferation of different synthetic media.

II. RESULTS AND DISCUSSIONS

The proposed system was evaluated on various aspects and parameters such as security, transfer learning capabilities, response time, accuracy, and robustness. Table 2 highlights

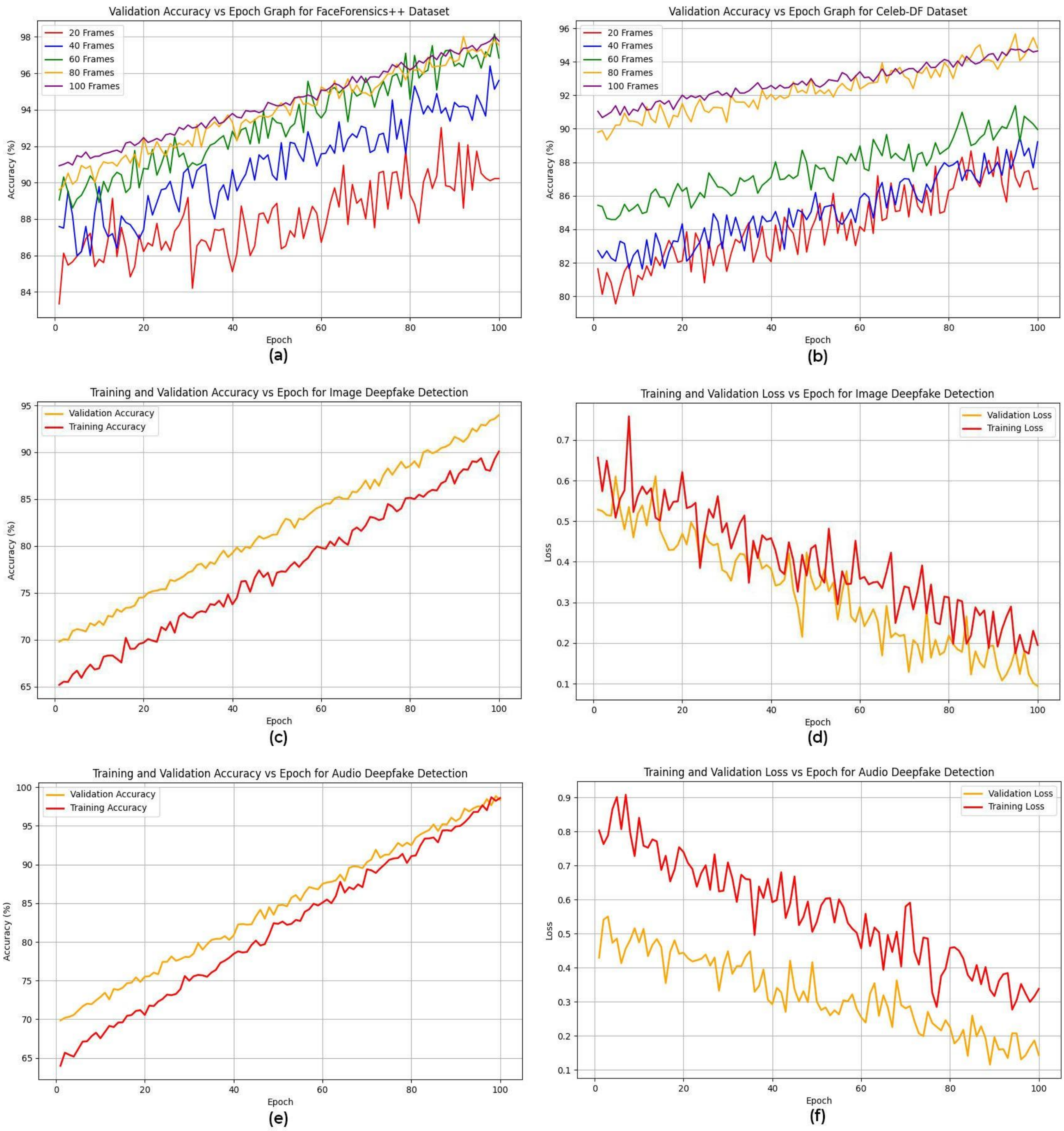


FIGURE 7. Output graphs of the proposed system developed for multimedia-based deepfake detection. **(a)** represents the validation accuracy graph tested on the FaceForensics++ dataset for different frames of 20,40,60,80,100 on the videos. Note that the purple line here represented the **highest accuracy** obtained by splitting the video into 100 frames and then applying the deepfake detection method, as compared to Intel's FakeCatcher [11]. **(b)** shows the Validation Accuracy graph tested on the Celeb-DF dataset for different frames of 20,40,60,80,100 on the videos. Here, the yellow line showed the highest accuracy on 80 frames though it was less than that of the earlier. **(c)** represents Training (Red) accuracy and Validation (Orange) accuracy for the proposed methodology for Image Deepfake Detection. The higher validation accuracy here through the yellow line showed

that the model is robust enough. **(d)** shows Training (Red) loss and Validation (Orange) loss for the proposed methodology for Image Deepfake Detection. **(e)** represents Training (Red) accuracy and Validation (Orange) accuracy for the proposed methodology for Audio Deepfake Detection. **(f)** shows Training (Red) loss and Validation (Orange) loss for the proposed methodology for Audio Deepfake Detection.

how various methods, from CNNs to GANs and Auto- Encoders, perform across these metrics, offering insights into the effectiveness and limitations of each approach in detecting
deepfakes. Table 3 shows the detailed comparative study and performance evaluation with existing architectures in the multimedia-based deepfake detection methods.

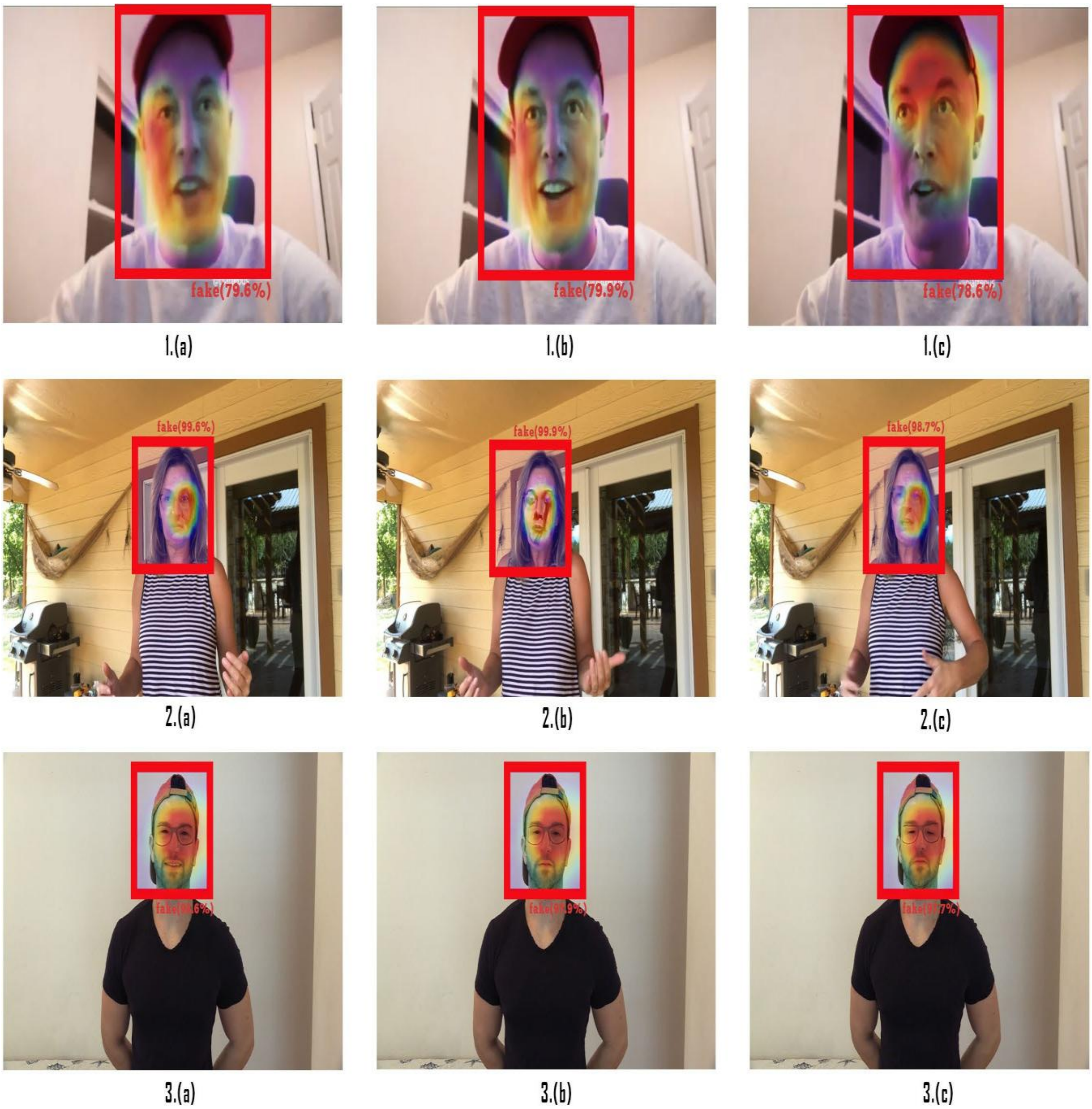


FIGURE 8. Results achieved for **video deepfake detection** through **proposed algorithm and methodology framework** on the test dataset as well as publicly available video from Youtube. Here, **1.(a), 1.(b), 1.(c)** are three different frames of a video taken from Youtube. The link to the video can be accessed at https://www.youtube.com/watch?v=XuKUkyPegBE. **2.(a), 2.(b), 2.(c)** along with **3.(a), 3.(b) and 3.(c)** was taken from DFDC dataset [24]. The **red box** indicated the face detected was superimposed and fake. The **confidence score** of the fake was also shown with the detection.

A. *RESULTS ON VIDEO DEEPFAKE DETECTION*

Here, the outcomes of the proposed system are shown where the framework, with its novel approach of deepfake detection across different modalities, is being employed. The part of video detection, where the Eulerian Video Magnification is combined with the ResNext architecture for the efficient feature extraction in the videos, while the LSTM layer performs the job of classification.

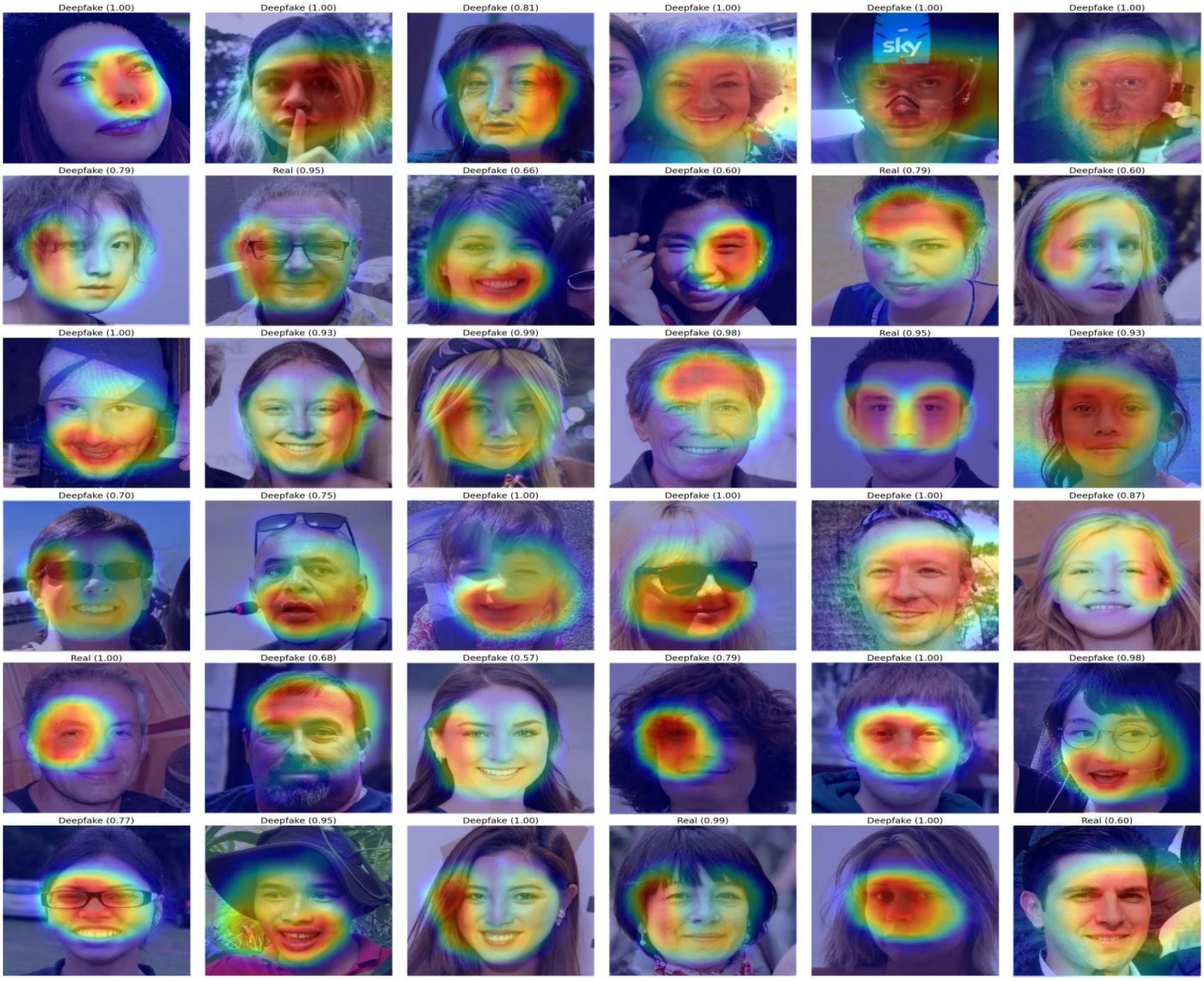
Here, the Sach-AI framework demonstrated its robust performance in detecting deepfakes within the limit of audio-visual content. This was evaluated on two benchmark datasets: FaceForensics [14] and Celeb-DF [19]. This experimentation was very tough and done with varying frame counts during the training phase to assess the impact on the accuracy of the detection. The initial testing, on 20 frames, achieved a remarkable accuracy of 90.95% in the detection of deepfakes. Next, in the increased frame of 40, the detection accuracy rose to 95.23%, indicating the framework's efficiency in capturing nuanced features in audio-visual content. In the increased number of frames of 60, the accuracy increased to that of 97.49%. Next, experimentation was done for 80 frames, where the accuracy slightly increased to 97.73%, and finally, for 100 frames, the model achieved the highest accuracy of 97.76%. This showcased the robust performance of the model in detecting

deepfakes in videos on the FaceForensics dataset as in Fig. 7.(a). Next, the model was trained on the Celeb-DF dataset and it again gave promising results. Similarly to the previous training on the FaceForensics dataset where the increase in the number of frames led to more improved accuracy, this framework Sach-AI achieved a remarkable accuracy of 94.66% on the Celeb-DF [19] dataset also, as shown in Fig. 7.(b). This was experimented on 100 frames, which indicated the effectiveness of detecting framewise on the diverse range of video datasets.

The results demonstrated the true effectiveness of the framework in the detection of deepfakes across diversified and uniquely different datasets. The utilization and applica- tion of Eulerian Video Magnification and ResNext for the feature extraction, and ultimately integrated with LSTM for the classification, contributed to the robustness of the model. The increased amount of accuracy with a larger number of frames also suggested the true significance of temporal information in the detection of deepfakes. The realism and variety features of the challenging FaceForensics dataset used to evaluate participants' performances were widely commonly agreed upon. This place has an advantage concerning very distinct signals of deepfakes and arrived at an accuracy of 97.76%. Moreover, its flexibility is proven in Celeb-DF [19] task results, which showed that the model

deepfakes in videos on the FaceForensics dataset as in Fig. 7.(a). Next, the model was trained on the Celeb-DF dataset and it again gave promising results. Similarly to the previous training on the FaceForensics dataset where the increase in the number of frames led to more improved accuracy, this framework Sach-AI achieved a remarkable accuracy of 94.66% on the Celeb-DF [19] dataset also, as shown in Fig. 7.(b). This was experimented on 100 frames, which indicated the effectiveness of detecting framewise on the diverse range of video datasets.

The results demonstrated the true effectiveness of the framework in the detection of deepfakes across diversified and uniquely different datasets. The utilization and applica- tion of Eulerian Video Magnification and ResNext for the feature extraction, and ultimately integrated with LSTM for the classification, contributed to the robustness of the model. The increased amount of accuracy with a larger number of frames also suggested the true significance of temporal information in the detection of deepfakes. The realism and variety features of the challenging FaceForensics dataset used to evaluate participants' performances were widely commonly agreed upon. This place has an advantage concerning very distinct signals of deepfakes and arrived at an accuracy of 97.76%. Moreover, its flexibility is



proven in Celeb-DF [19] task results, which showed that the model

FIGURE 9. Results achieved through the test dataset for Image Deepfake Detection. The proposed system was able to detect the deepfakes with Grad-CAM heatmap applied to the results image. On the above result, the predicted class is above the images with its confidence score.

could generalize to another different set of video datasets. The output of the video deepfake detection has been shown in Fig. 8. The methodology has been tested on publicly available data from Youtube as well as test dataset from the DFDC dataset [24].

B. RESULTS ON IMAGE DEEPFAKE DETECTION

The dataset used for this study consisted of all 70k REAL faces from the Flickr dataset collected by Nvidia,
as well as 70k fake faces sampled from the 1 Million FAKE faces (generated by StyleGAN) that was provided by Bojan.
In the visual domain, the model has been trained on a set of custom images, of exactly 1,42,041 images– which had been carefully selected and labelled into different segments of real and fake. The model was able to reach an excellent accuracy measure of 93.64% because it had the capability of detecting real shown in Fig. 7.(c).

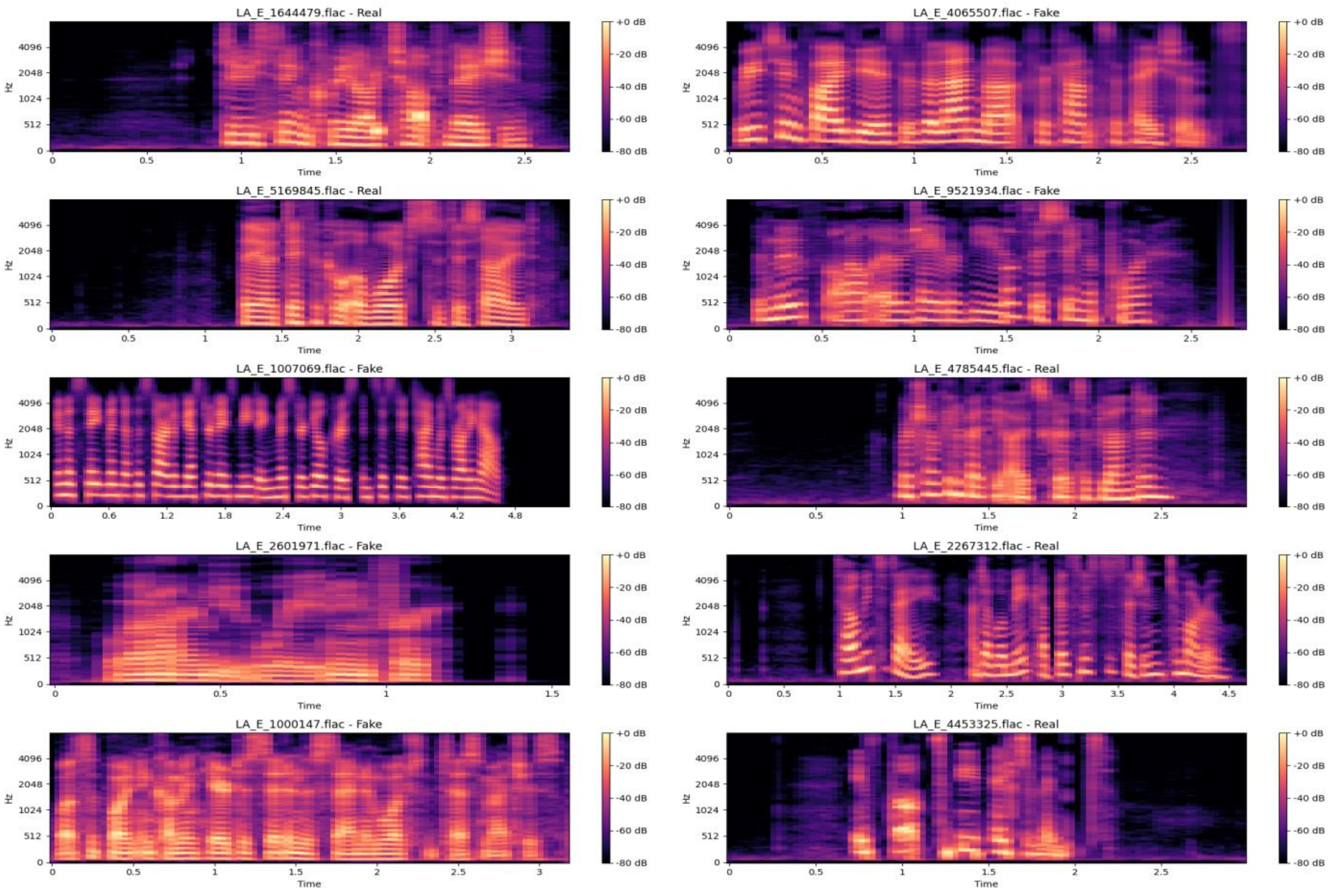


FIGURE 10. Results achieved through the test dataset for Audio Deepfake Detection. The proposed system was able to detect the deepfakes through Mel Spectrogram approach applied to the test (unseen) audio. On the above result, the predicted class (real or fake) is above the Mel Spectorgram images generated.

The training and validation loss was also reduced till the last of epochs, as in 7.(d). The ROC AUC score, which measured the model's ability to distinguish between different classes, was found to be 0.992. This also indicated the higher level of differentiation power in the model's predictions. The average precision score, considered for the precision- recall trade-off, was found to be 0.991. This was further attested to the robustness of the mode in the capturing of some actual positive instances with the minimization of false positives. The higher accuracy and robustness of the model in visual deepfake detection presented promising applications and utilisations in various fields, including media forensics and moderation within content. However, along with further investigation and validation across different video and audio modalities, it is essential to establish the comprehensive effectiveness and accuracy of the framework. The results obtained from the study along with Grad-CAM technique
as well as 70k fake faces sampled from the 1Million FAKE faces (generated by StyleGAN) that was provided by Bojan.
In the visual domain, the model has been trained on a set of custom images, of exactly 1,42,041 images– which had been carefully selected and labelled into different segments of real and fake. The model was able to reach an excellent accuracy measure of 93.64% because it had the capability of detecting real and fake images, as shown in Fig. 7.(c).

TABLE 2. Considered evaluation metrics for performance evaluation of the proposed system along with other architectures.

Type	Authors	Methodology	Security	Transfer Learning	Response Time	Accuracy	Delay	Complexity	Area under the Curve (AUC)	Robustness
Deepfake Image	I-J Yu et al. (2020) [48]	CNN	NR	✓	NR	✓	NR	NR	✓	✓
	Guo et al. (2021) [49]	CNN	NR	NR	✓	✓	NR	NR	✓	NR
	Muthukumar et al. (2024) [50]	DenseNet121	✓	✓	NR	✓	NR	✓	✓	✓
	Hyunjoon Kim et. al. (2024) [21]	Performance analysis based on image quality	✓	NR	NR	✓	NR	✓	✓	✓
Deepfake Video		GAN	NR	NR	NR	✓	NR	NR	✓	NR
		CNN	NR	NR	NR	✓	NR	NR	✓	NR
		Convolutional Neural Networks	✓	NR	NR	✓	NR	NR	✓	✓
		Gaze-guided Spatial Inconsistency Learning	✓	✓	NR	✓	NR	✓	✓	✓
Deepfake Audio		RNN + LSTM	NR	NR	NR	NR	NR	NR	✓	NR
		CNN + RNN	NR	✓	NR	✓	NR	NR	✓	NR
		LSTM	NR	NR	NR	✓	NR	✓	✓	NR
Deepfake Multimedia		CNN	NR	✓	NR	✓	NR	NR	NR	NR
		LSTM + CNN + Blockchain	✓	NR	NR	✓	NR	NR	NR	NR
		LBP + CNN	NR	NR	NR	NR	NR	NR	✓	NR
		Auto-Encoders	NR	✓	NR	✓	NR	NR	NR	✓

thresholds in decision-making. The Precision-Recall curve is also a crucial metric in evaluating classification models. Here, this provided insights into the ability of the model to correctly identify the positive instances with the minimisation of false positives. The calibration curve is also important for properly understanding predicted probabilities' reliability. This helped in the comparison of predicted probabilities against the original outcomes, which allowed for proper assessment of the confidence estimates of the model. The exceptional accuracy in the audio deepfake detection of 99.13% underscored the robustness and effectiveness of this multi-modal innovative idea. The integration of CNN in the analysis of audio showcased the ability of the model the capturing complex, intricate patterns within the audio signals. The real-time combination of advanced feature extraction along with the powerful neural network architectures ulti- mately contributed to the high performance of the model. The Mel spectrograms illustrated the capability of the model

Jung et al. (2020) [51]
Z. Zhao et al. (2021) [25]
Hussain, Z.F. et.al. (2023) [52]
He, Q. et. al.(2023) [20]
Mittal et al. (2020) [53]
Wijethunga et al. (2020) [54]
Borrelli et al. (2021) [55]
Nasar et al. (2020) [56]
Chan et al. (2020) [57]
Khalil et al. (2021) [58]
Kong et al. (2021) [59]

Note: NR shows that the corresponding metrics are not reported for these architectures.

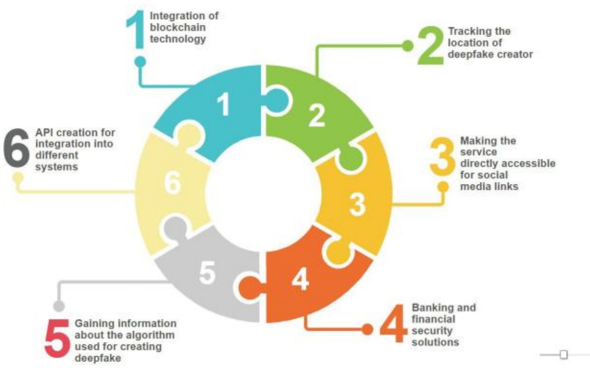


FIGURE 11. Future applications and scope of the proposed system.

TABLE 3. Performance evaluation and detailed comparative study on different parameters required for deepfake detection systems and models.

Domain	Author	Methodology	Accuracy	Recall	Precision	AUC Score
Deepfake Video Detection	Li et al. (2018) [13]	Detecting eye blinking in AI-generated fake videos	0.920	0.915	0.925	0.940
	Amerini et al. (2019) [17]	Optical flow-based CNN for deepfake detection	0.935	0.920	0.940	0.950
	Li et al. (2020) [19]	Celeb-DF dataset for DeepFake forensics	0.960	0.950	0.970	0.980
	Zhao et al. (2021) [25]	Multi-attentional deepfake detection	0.965	0.950	**0.975**	0.980
	Hussain, Z.F. et.al.(2023) [52]	Convolutional Neural Networks based Jaya algorithm	0.972	**0.958**	0.964	0.980
	He, Q. et. al.(2023) [20]	Gaze-guided Spatial Inconsistency Learning	0.965	0.942	0.955	0.970
	Proposed Algorithm	EVM + ResNext + LSTM	**0.977** (↑)	0.920	0.923	**0.988** (↑)
Deepfake Image Detection	Marra et al. (2018) [14]	GAN-generated fake images detection	0.890	0.875	0.900	0.900
	Matern et al. (2019) [?]	Visual artifact exploitation for deepfake detection	0.910	0.900	0.920	0.920
	Zhang et al. (2020) [60]	Counterfeit feature extraction using deep learning	0.875	0.850	0.880	0.880
	Guo et al. (2021) [49]	Adaptive manipulation traces extraction network	0.890	0.865	0.905	0.890
	Muthukumar, A. et al.(2024) [50]	Automated adaptive gaining sharing knowledge algo-rithm with DenseNet121	0.938	0.921	0.932	0.950
	Hyunjoon Kim et. al. (2024) [21]	Correlation between deepfake detection performance and image quality metrics	**0.947**	**0.935**	0.940	0.960
	Proposed Method	Fine-Tuning of Pre-Trained model on custom dataset	0.936	0.920	**0.991** (↑)	**0.992** (↑)
Deepfake Audio Detection	Wijethunga et al. (2020) [54]	CNN + RNN for audio deepfake detection	0.940	**0.930**	**0.950**	**0.960**
	Borrelli et al. (2021) [55]	LSTM for audio deepfake detection	0.935	0.920	0.940	0.950
	Proposed Algorithm	CNN on MeL Spectrograms	**0.991** (↑)	0.870	0.870	0.900

Note: Bold indicates the improved achieved in its category for different evaluation metrics.

CONCLUSION

This study introduced the Sach-AI framework, which proved to be a pioneering framework mainly designed for the detection of deepfakes across different modalities, including video, audio, and image. Leveraging the potential of the state-of-the-art technologies employed in the Eulerian Video Magnification technique which was coupled with ResNext for the accurate feature extraction in the video data. The LSTM network was also integrated for classification for its learning capabilities in sequential data. Convolutional Neural Networks along with Mel spectrograms were also used for extraction of discriminative features, while in image-based deepfake detection, the DenseNet121 was employed. The results were also promising, which demonstrated the efficacy of the framework in the accurate detection of deepfakes. The framework exhibited promising performance metrics, which also showcased its potential for real-world applications in the mitigation of adverse impacts of deepfake technology. The threat of manipulated media, as it continues to grow, this framework will serve as an invaluable tool in the proper addressing of challenges associated with the detection of deepfakes.

III. FUTURE SCOPE

This framework can address deepfake problems and has considered several future research directions as shown in Fig. 11. A possible addition could be the use of blockchain in order to add an extra layer of security to anti-money laundering processes and also to bring more transparency. Blockchain would assist in the creation of a decentralized and immutable database for the verification of multimedia content from sources. Future developments can also enhance Sach A.I.'s ability to detect deepfake authors using blockchain attributions. Further, methods of location detection could also assist in identifying the source of the fake content, thus assisting policing in arresting the individuals involved.

REFERENCES

[1] V. Maniyal and V. Kumar, ''Unveiling the deepfake dilemma: Framework, classification, and future trajectories,'' *IT Prof.*, vol. 26, no. 2, pp. 32–38, Mar. 2024, doi: 10.1109/mitp.2024.3369948.

[2] K. Sharma, S. Taneja, and S. Monga, *DeepFake Technologies: Concepts, Methods, and Applications*. Harrisburg, CA, USA: IGI Global, Jul. 2024,
pp. 59–65, doi: 10.4018/979-8-3693-5298-4.ch004.

[3] S. H. Al-Khazraji, H. H. Saleh, A. I. Khalid, and I. A. Mishkhal, ''Impact of deepfake technology on social media: Detection, misinformation and societal implications,'' *Eurasia Proc. Sci. Technol. Eng. Math.*, vol. 23,
pp. 429–441, Oct. 2023, doi: 10.55549/epstem.1371792.

[4] M. T. Soto-Sanfiel and S. Saha, *Predicting Deepfake Enjoyment: A Machine Learning Perspective*. Cham, Switzerland: Springer, 2024,
pp. 384–402, doi: 10.1007/978-3-031-61281-7_28.

[5] L. Whittaker, T. C. Kietzmann, J. Kietzmann, and A. Dabirian, '''All around me are synthetic faces': The mad world of AI- generated media,'' *IT Prof.*, vol. 22, no. 5, pp. 90–99, Sep. 2020, doi: 10.1109/MITP.2020.2985492.

[6] E. Lundberg and P. Mozelius, ''The potential effects of deepfakes on news media and entertainment,'' *AI Soc.*, Oct. 2024, doi: 10.1007/s00146-024- 02072-1.
[7] N. Rathoure, R. K. Pateriya, N. Bharot, and P. Verma, ''Combating deep- fakes: A comprehensive multilayer deepfake video detection framework,'' *Multimedia Tools Appl.*, vol. 83, no. 38, pp. 85619–85636, Aug. 2024, doi: 10.1007/s11042-024-20012-5.
[8] P. P. Sudharsana, R. R. Rajalaxmi, A. Devak, R. Gokul, R. Annamalai, and P. G. Brindha, ''Enhancing deepfake detection: An ensemble deep learning approach for efficient attribute manipulation identification,'' in *Proc. Int. Conf. Cognit. Robot. Intell. Syst. (ICC-ROBINS)*, Apr. 2024, pp. 352–359, doi: 10.1109/ICC-ROBINS60238.2024.10533994.
[9] A. Hamza, A. R. R. Javed, F. Iqbal, N. Kryvinska, A. S. Almadhor, Z. Jalil, and R. Borghol, ''Deepfake audio detection via MFCC features using machine learning,'' *IEEE Access*, vol. 10, pp. 134018–134028, 2022, doi: 10.1109/ACCESS.2022.3231480.
[10] Z. N. Ashani, ''Comparative analysis of deepfake image detection method using VGG16, VGG19 and ResNet50,'' *J. Adv. Res. Appl. Sci. Eng. Tech- nol.*, vol. 47, no. 1, pp. 16–28, Jun. 2024, doi: 10.37934/araset.47.1.1628.
[11] *Intel Deepfake Detector FakeCatcher*. Accessed: Mar. 8, 2025. [Online]. Available: https://www.intel.com/content/www/us/en/newsroom /news/intel-introduces-real-time-deepfake-detector.html#gs.jwhbjd
[12] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, ''Two-stream neural networks for tampered face detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1831–1839, doi: 10.1109/CVPRW.2017.229.
[13] Y. Li, M.-C. Chang, and S. Lyu, ''In ictu oculi: Exposing AI created fake videos by detecting eye blinking,'' in *Proc. IEEE Int. Workshop Inf. Foren- sics Secur. (WIFS)*, Dec. 2018, pp. 1–7, doi: 10.1109/WIFS.2018.8630787.
[14] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, ''Detection of GAN-generated fake images over social networks,'' in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, pp. 384–389, doi: 10.1109/MIPR.2018.00084.
[15] C.-C. Hsu, C.-Y. Lee, and Y.-X. Zhuang, ''Learning to detect fake face images in the wild,'' in *Proc. Int. Symp. Comput., Consum. Control (IS3C)*, Dec. 2018, pp. 388–391, doi: 10.1109/IS3C.2018.00104.
[16] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, ''MesoNet: A compact facial video forgery detection network,'' in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2018, pp. 1–7, doi: 10.1109/WIFS.2018.8630761.
[17] I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, ''Deepfake video detection through optical flow based CNN,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1205–1207, doi: 10.1109/ICCVW.2019.00152.
[18] F. Matern, C. Riess, and M. Stamminger, ''Exploiting visual artifacts to expose deepfakes and face manipulations,'' in *Proc. IEEE Winter Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2019, pp. 83–92, doi: 10.1109/WACVW.2019.00020.
[19] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, ''Celeb-DF: A large-scale challenging dataset for DeepFake forensics,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3204–3213, doi: 10.1109/CVPR42600.2020.00327.

[20] Q. He, C. Peng, D. Liu, N. Wang, and X. Gao, ''GazeForensics: DeepFake detection via gaze-guided spatial inconsistency learning,'' *Neural Netw.*, vol. 180, Dec. 2024, Art. no. 106636, doi: 10.1016/j.neunet.2024.106636.
[21] H. Kim, J. Lee, L. H. Park, and T. Kwon, ''On the correlation between deepfake detection performance and image quality metrics,'' in *Proc. 3rd ACM Workshop Secur. Implications Deepfakes Cheapfakes*, Jul. 2024,
pp. 14–19, doi: 10.1145/3660354.3660358.
[22] M. Westerlund, ''The emergence of deepfake technology: A review,'' *Technol. Innov. Manage. Rev.*, vol. 9, no. 11, pp. 39–52, Jan. 2019, doi: 10.22215/timreview/1282.
[23] P. Yu, Z. Xia, J. Fei, and Y. Lu, ''A survey on deepfake video detection,'' *IET Biometrics*, vol. 10, no. 6, pp. 607–624, Apr. 2021, doi: 10.1049/bme2.12031.
[24] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and
C. C. Ferrer, ''The DeepFake detection challenge (DFDC) dataset,'' 2020,
arXiv:2006.07397.
[25] H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen, and N. Yu, ''Multi-attentional deepfake detection,'' in *Proc. IEEE/CVF Conf. Com- put. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2185–2194, doi: 10.1109/CVPR46437.2021.00222.
[26] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, ''Deepfake detection: A systematic literature review,'' *IEEE Access*, vol. 10,
pp. 25494–25513, 2022, doi: 10.1109/ACCESS.2022.3154404.
[27] J. John and B. V. Sherif, ''Comparative analysis on different deepfake detection methods and semi supervised GAN architecture for deepfake detection,'' in *Proc. 6th Int. Conf. I-SMAC*, Nov. 2022, pp. 516–521, doi: 10.1109/I-SMAC55078.2022.9987265.
[28] D. Garg and R. Gill, ''Deepfake generation and detection—An exploratory study,'' in *Proc. 10th IEEE Uttar Pradesh Sect. Int. Conf. Electr., Electron. Comput. Eng. (UPCON)*, Dec. 2023, pp. 888–893, doi: 10.1109/UPCON59197.2023.10434896.
[29] M. A. Khder, S. Shorman, D. T. Aldoseri, and M. M. Saeed, ''Artificial intelligence into multimedia deepfakes creation and detection,'' in *Proc. Int. Conf. IT Innov. Knowl. Discovery (ITIKD)*, Mar. 2023, pp. 1–5, doi: 10.1109/ITIKD56332.2023.10099744.
[30] J. Zhang, K. Cheng, G. Sovernigo, and X. Lin, ''A heteroge- neous feature ensemble learning based deepfake detection method,'' in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 2084–2089, doi: 10.1109/ICC45855.2022.9838630.
[31] P. Yu, Z. Xia, J. Fei, and Y. Lu, ''A survey on deepfake video detection,'' *IET Biometrics*, vol. 10, no. 6, pp. 607–624, Apr. 2021, doi: 10.1049/bme2.12031.
[32] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and
C. C. Ferrer, ''The DeepFake detection challenge (DFDC) dataset,'' 2020,
arXiv:2006.07397.
[33] H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen, and N. Yu, ''Multi-attentional deepfake detection,'' in *Proc. IEEE/CVF Conf. Com- put. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2185–2194, doi: 10.1109/CVPR46437.2021.00222.
[34] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, ''Deepfake detection: A systematic literature review,'' *IEEE Access*, vol. 10,pp. 25494–25513, 2022, doi: 10.1109/ACCESS.2022.3154404.
[35] J. John and B. V. Sherif, ''Comparative analysis on different deepfake detection methods and semi

supervised GAN architecture for deepfake detection,'' in *Proc. 6th Int. Conf. I-SMAC*, Nov. 2022, pp. 516–521, doi: 10.1109/I-SMAC55078.2022.9987265.
[36] P. Yu, Z. Xia, J. Fei, and Y. Lu, ''A survey on deepfake video detection,'' *IET Biometrics*, vol. 10, no. 6, pp. 607–624, Apr. 2021, doi: 10.1049/bme2.12031.
[37] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and
C. C. Ferrer, ''The DeepFake detection challenge (DFDC) dataset,'' 2020,
arXiv:2006.07397.
[38] H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen, and N. Yu, ''Multi-attentional deepfake detection,'' in *Proc. IEEE/CVF Conf. Com- put. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2185–2194, doi: 10.1109/CVPR46437.2021.00222.
[39] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, ''Deepfake detection: A systematic literature review,'' *IEEE Access*, vol. 10,
pp. 25494–25513, 2022, doi: 10.1109/ACCESS.2022.3154404.
[40] J. John and B. V. Sherif, ''Comparative analysis on different deepfake detection methods and semi supervised GAN architecture for deepfake detection,'' in *Proc. 6th Int. Conf. I-SMAC*, Nov. 2022, pp. 516–521, doi: 10.1109/I-SMAC55078.2022.9987265.
[41] D. Garg and R. Gill, ''Deepfake generation and detection—An exploratory study,'' in *Proc. 10th IEEE Uttar Pradesh Sect. Int. Conf. Electr., Electron. Comput. Eng. (UPCON)*, Dec. 2023, pp. 888–893, doi: 10.1109/UPCON59197.2023.10434896.
[42] M. A. Khder, S. Shorman, D. T. Aldoseri, and M. M. Saeed, ''Artificial intelligence into multimedia deepfakes creation and detection,'' in *Proc. Int. Conf. IT Innov. Knowl. Discovery (ITIKD)*, Mar. 2023, pp. 1–5, doi: 10.1109/ITIKD56332.2023.10099744.
[43] J. Zhang, K. Cheng, G. Sovernigo, and X. Lin, ''A heteroge- neous feature ensemble learning based deepfake detection method,'' in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 2084–2089, doi: 10.1109/ICC45855.2022.9838630.
[44] S. Jia, X. Li, and S. Lyu, ''Model attribution of face-swap deepfake videos,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022,
pp. 2356–2360, doi: 10.1109/ICIP46576.2022.9897972.
[45] P. Yu, Z. Xia, J. Fei, and Y. Lu, ''A survey on deepfake video detection,'' *IET Biometrics*, vol. 10, no. 6, pp. 607–624, Apr. 2021, doi: 10.1049/bme2.12031.
[46] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and
C. C. Ferrer, ''The DeepFake detection challenge (DFDC) dataset,'' 2020,
arXiv:2006.07397.
[47] H. Zhao, T. Wei, W. Zhou, W. Zhang, D. Chen, and N. Yu, ''Multi-attentional deepfake detection,'' in *Proc. IEEE/CVF Conf. Com- put. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2185–2194, doi: 10.1109/CVPR46437.2021.00222.
[48] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, ''Deepfake detection: A systematic literature review,'' *IEEE Access*, vol. 10,
pp. 25494–25513, 2022, doi: 10.1109/ACCESS.2022.3154404.
[49] J. John and B. V. Sherif, ''Comparative analysis on different deepfake detection methods and semi supervised GAN architecture for deepfake detection,'' in *Proc. 6th Int. Conf. I-SMAC*, Nov. 2022, pp. 516–521, doi: 10.1109/I-SMAC55078.2022.9987265.
[50] D. Garg and R. Gill, ''Deepfake generation and detection—An exploratory study,'' in *Proc. 10th IEEE*

Uttar Pradesh Sect. Int. Conf. Electr., Electron. Comput. Eng. (UPCON), Dec. 2023, pp. 888–893, doi: 10.1109/UPCON59197.2023.10434896.

[51] M. A. Khder, S. Shorman, D. T. Aldoseri, and M. M. Saeed, ''Artificial intelligence into multimedia deepfakes creation and detection,'' in *Proc. Int. Conf. IT Innov. Knowl. Discovery (ITIKD)*, Mar. 2023, pp. 1–5, doi: 10.1109/ITIKD56332.2023.10099744.

[52] J. Zhang, K. Cheng, G. Sovernigo, and X. Lin, ''A heteroge- neous feature ensemble learning based deepfake detection method,'' in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 2084–2089, doi: 10.1109/ICC45855.2022.9838630.

[53] S. Jia, X. Li, and S. Lyu, ''Model attribution of face-swap deepfake videos,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2022, pp. 2356–2360, doi: 10.1109/ICIP46576.2022.9897972.

[54] A. Trabelsi, M. M. Pic, and J.-L. Dugelay, ''Improving deepfake detection by mixing top solutions of the DFDC,'' in *Proc. 30th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2022, pp. 643–647, doi: 10.23919/EUSIPCO55093.2022.9909905.

[55] S. Yadav, S. Bommareddy, and D. K. Vishwakarma, ''Robust and generalized deepfake detection,'' in *Proc. 13th Int. Conf. Comput. Com- mun. Netw. Technol. (ICCCNT)*, Oct. 2022, pp. 1–6, doi: 10.1109/ICC- CNT54827.2022.9984553.

[56] Y. S. Malik, N. Sabahat, and M. O. Moazzam, ''Image animations on driving videos with deepfakes and detecting deepfakes generated animations,'' in *Proc. IEEE 23rd Int. Multitopic Conf. (INMIC)*, Nov. 2020, pp. 1–6, doi: 10.1109/INMIC50486.2020. 9318064.

[57] A. Verma, D. Gupta, and M. K. Srivastava, ''Deepfake detection using inception-ResnetV2,'' in *Proc. 1st Int. Conf. Adv. Comput. Future Commun. Technol. (ICACFCT)*, Dec. 2021, pp. 39–41, doi: 10.1109/ICACFCT53978.2021.9837351.

[58] Q. Jaleel and I. H. Ali, ''Facial behavior analysis-based deepfake video detection using GAN discriminator,'' in *Proc. Int. Conf. Data Sci. Intell. Comput. (ICDSIC)*, Nov. 2022, pp. 36–40, doi: 10.1109/ICD- SIC56987.2022.10075660.

[59] N. M. Alnaim, Z. M. Almutairi, M. S. Alsuwat, H. H. Alalawi, A. Alshobaili, and F. S. Alenezi, ''DFFMD: A deepfake face mask dataset for infectious disease era with deepfake detection algorithms,'' *IEEE Access*, vol. 11, pp. 16711–16722, 2023, doi: 10.1109/ACCESS.2023.3246661.

[60] S. J. Pipin, R. Purba, and M. F. Pasha, ''Deepfake video detection using spa- tiotemporal convolutional network and photo response non uniformity,'' in *Proc. IEEE Int. Conf. Comput. Sci. Inf. Technol. (ICOSNIKOM)*, Oct. 2022, pp. 1–6, doi: 10.1109/ICOSNIKOM56551.2022.10034890.

[61] H. Liu, P. Bestagini, L. Huang, W. Zhou, S. Tubaro, W. Zhang, and N. Yu, ''It wasn't me: Irregular identity in deepfake videos,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2023, pp. 2770–2774, doi: 10.1109/icip49359.2023.10222654.

[62] C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, ''Deep fake image detection based on pairwise learning,'' *Appl. Sci.*, vol. 10, no. 1, p. 370, Jan. 2020, doi: 10.3390/app10010370.

[63] A. Kohli and A. Gupta, ''Detecting DeepFake, FaceSwap and Face2Face facial forgeries using frequency CNN,'' *Multimedia Tools Appl.*, vol. 80, no. 12, pp. 18461–18478, Feb. 2021, doi: 10.1007/s11042-020-10420-8.

[64] T. Qiao, S. Xie, Y. Chen, F. Retraint, and X. Luo, ''Fully unsupervised deepfake video detection via enhanced contrastive learning,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 7, pp. 4654–4668, Jul. 2024, doi: 10.1109/TPAMI.2024.3356814.
[65] G.-L. Chen and C.-C. Hsu, ''Jointly defending deepfake manipulation and adversarial attack using decoy mechanism,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 9922–9931, Aug. 2023, doi: 10.1109/TPAMI.2023.3253390.
[66] R. Shao, T. Wu, J. Wu, L. Nie, and Z. Liu, ''Detecting and grounding multi-modal media manipulation and beyond,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, pp. 5556–5574, Aug. 2024, doi: 10.1109/TPAMI.2024.3367749.
[67] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, ''FaceForensics ++: Learning to detect manipulated facial images,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1–11, doi: 10.1109/ICCV.2019.00009.
[68] J. Yamagishi, M. Todisco, Sahidullah, H. Delgado, X. Wang, N. Evans, T. Kinnunen, K. A. Lee, V. Vestman, and A. Nautsch. (2019). *Asvspoof 2019: The 3rd Automatic Speaker Verification Spoofing and Counter- measures Challenge Database*. Univ. Edinburgh. [Online]. Available: https://datashare.ed.ac.uk/handle/10283/3336
[69] N. Wadhwa, H.-Y. Wu, A. Davis, M. Rubinstein, E. Shih, G. J. Mysore, J. G. Chen, O. Buyukozturk, J. V. Guttag, W. T. Freeman, and F. Durand, ''Eulerian video magnification and analysis,'' *Commun. ACM*, vol. 60, no. 1, pp. 87–95, Dec. 2016, doi: 10.1145/3015573.
[70] I.-J. Yu, S.-H. Nam, W. Ahn, M.-J. Kwon, and H.-K. Lee, ''Manipulation classification for JPEG images using multi-domain features,'' *IEEE Access*, vol. 8, pp. 210837–210854, 2020, doi: 10.1109/ACCESS.2020.3037735.
[71] Z. Guo, G. Yang, J. Chen, and X. Sun, ''Fake face detection via adaptive manipulation traces extraction network,'' *Comput. Vis. Image Under- stand.*, vol. 204, Mar. 2021, Art. no. 103170, doi: 10.1016/j.cviu.2021. 103170.
[72] A. Muthukumar, M. T. Raj, R. Ramalakshmi, A. Meena, and P. Kaleeswari, ''Fake and propaganda images detection using automated adaptive gaining sharing knowledge algorithm with DenseNet121,'' *J. Ambient Intell. Humanized Comput.*, vol. 15, no. 9, pp. 3519–3531, Jul. 2024, doi: 10.1007/s12652-024-04829-4.
[73] T. Jung, S. Kim, and K. Kim, ''DeepVision: Deepfakes detection using human eye blinking pattern,'' *IEEE Access*, vol. 8, pp. 83144–83154, 2020, doi: 10.1109/ACCESS.2020.2988660.
[74] Z. F. Hussain and H. R. Ibraheem, ''Novel convolutional neural networks based Jaya algorithm approach for accurate deepfake video detection,'' *Mesopotamian J. CyberSecur.*, vol. 2023, pp. 35–39, Feb. 2023, doi: 10.58496/mjcs/2023/007.
[75] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, ''Emotions don't lie: An audio-visual deepfake detection method using affective cues,'' in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 2823–2832, doi: 10.1145/3394171.3413570.
[76] R. L. M. A. P. C. Wijethunga, D. M. K. Matheesha, A. A. Noman, K. H. V. T. A. De Silva, M. Tissera, and L. Rupasinghe, ''Deepfake audio detection: A deep learning based solution for group conversations,'' in *Proc. 2nd Int. Conf. Advancements Comput. (ICAC)*, vol. 1,

Dec. 2020,
pp. 192–197, doi: 10.1109/ICAC51239.2020.9357161.

[77] C. Borrelli, P. Bestagini, F. Antonacci, A. Sarti, and S. Tubaro, ''Synthetic speech detection through short-term and long-term prediction traces,'' *EURASIP J. Inf. Secur.*, vol. 2021, no. 1, pp. 1–14, Apr. 2021, doi: 10.1186/s13635-021-00116-3.

[78] B. F. Nasar, T. Sajini, and E. R. Lason, ''Deepfake detection in media files–audios, images and videos,'' in *Proc. IEEE Recent Adv. Intell. Comput. Syst. (RAICS)*, Dec. 2020, pp. 74–79, doi: 10.1109/RAICS51191.2020.9332516.

[79] C. C. K. Chan, V. Kumar, S. Delaney, and M. Gochoo, ''Combating deepfakes: Multi-LSTM and blockchain as proof of authenticity for digital media,'' in *Proc. IEEE/ITU Int. Conf. Artif. Intell. Good (AI4G)*, Sep. 2020,
pp. 55–62, doi: 10.1109/AI4G50087.2020.9311067.

[80] S. S. Khalil, S. M. Youssef, and S. N. Saleh, ''ICaps-Dfake: An integrated capsule-based model for deepfake image and video detection,'' *Future Internet*, vol. 13, no. 4, p. 93, Apr. 2021, doi: 10.3390/fi13040093.

[81] C. Kong, B. Chen, W. Yang, H. Li, P. Chen, and S. Wang, ''Appearance matters, so does audio: Revealing the hidden face via cross-modality transfer,'' *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1,
pp. 423–436, Jan. 2022, doi: 10.1109/TCSVT.2021.3057457.

[82] W. Zhang, C. Zhao, and Y. Li, ''A novel counterfeit feature extraction technique for exposing face-swap images based on deep learning and error level analysis,'' *Entropy*, vol. 22, no. 2, p. 249, Feb. 2020, doi: 10.3390/e22020249.

...