Educational Administration: Theory and Practice
2024, 30(5), 3589-3603
ISSN: 2148-2403
https://kuey.net/

Research Article

Video Morphing Attack Detection Using Convolutional Neural Networks On Deep Fake Detection Algorithm

Boovaneswari .S[1*], Sri Nihil .RS[2], Madhavan .R[3], Daniyel .A[4]

[1*,2,3,4]Department of Computer Science and Engineering, Manakula Vinayagar Institute of Technology, Puducherry, India.
[*]Email:boovana2825@gmail.com,Email:srinihil010@gmail.com, Email:madhavan3102002@gmail.com, mail:danidaniyel6@gmail.com

Citation: Boovaneswari S, et.al, (2024), Video Morphing Attack Detection Using Convolutional Neural Networks On Deep Fake Detection Algorithm, *Educational Administration: Theory and Practice*, *30*(5), 3589-3603
Doi: 10.53555/kuey.v30i5.3495

ARTICLE INFO

ABSTRACT

A method called deepfake produces fake video and films with artificial or substituted faces. Deepfakes are turning into a worrying societal phenomenon because they may be used maliciously to spread harmful information, fabricate electronic convincing proof, make fake political news, and even participate in online harassment and fraud. Regarding the fight against the ubiquitous threat provided by deepfake videos, our suggested technique, named "Video Morphing Attack Detection Using CNN," is a stronghold. We provide our system the capacity to model temporal relationships across sequences and extract complex characteristics from video frames by integrating the ResNeXt-50 and LSTM neural network designs, respectively. Our method effectively detects abnormalities suggestive of video morphing assaults via forward propagation and SoftMax activation. We guarantee strong detection performance while maintaining the veracity and integrity of visual information by utilizing dynamic face localization and specific assessment measures. Our system provides a comprehensive solution to reduce the negative impacts of deep fake manipulation in visual media, including vital functionalities for picture conversion, prediction creation, and data pretreatment. Convolutional neural network architecture ResNeXt-50 is a member of the ResNeXt family, which was first presented as an advancement of the ResNet design. To overcome certain constraints of conventional convolutional neural networks (CNNs), ResNeXt aims to achieve state-of-the-art performance in image classification tasks while managing intricate visual input. The "50" in ResNeXt-50 stands for the network's depth or the total number of layers it has. ResNeXt-50 comprises 50 layers in total, comprising fully connected, pooling, and convolutional layers.

Keywords: Deep Fake detection, Morphing Detection, CNN, generation, detection, fake videos, neural network, Machine Learning, DFFMD, Deep Learning, ResneXT-50.

I. INTRODUCTION

The development of computer-generated editing software in recent years has made it simpler than ever to create and alter audiovisual assets [1]. Misinformation spreading has become much more possible, particularly with the emergence of since Deepfake. Deepfake is a deep learning technique. can manipulate already-existing films, fabricate new ones, or even synthesis the voice of someone speaking. It is therefore a risky instrument. Ali Kashif Bashir was the assistant editor in charge of organizing the manuscript's evaluation and granting publication approval.in applications that propagate misleading or hazardous information and fake news [2]. Therefore, the topic of identifying Deepfakes with machine learning approaches has been the scientific community as of 2023. Research has addressed the task from many perspectives, such as focusing or examining faces on particular areas, such as the eyes and lip motions, to produce.Fresh deep-learning frameworks. These days, a lot of Deepfake tools are abundant in learning resources, free, and open-source.
These tools generate a new video by replacing the target's face in relation to the source individuals. The objective's face was added to the input action [3]. Real and fake items from these handy services are quite tough for people to tell apart [4]. The importance of deep fake face detection is growing these days since deepfake

technology's subset of artificial intelligence. AI is becoming more and more prevalent. Deepfake Intelligence is a kind of artificial media that makes use of practical techniques for using machine learning to create edited videos that is really believable and lifelike. It produces information that appears authentic but was produced using methods for deep learning such as computer vison, convolutional neural networks and natural language processing [5].

The current methods for identifying deepfake videos have an average accuracy of 90%. However, since facemasks were introduced. This has encouraged criminals to modify surveillance footage to disguise their illicit activity and utilize face Morphing as a means of evading the law. The Deep Fake Face Mask Dataset (DFFMD) is suggested in this work. It is built on a unique Inception-Alex Net and includes batch normalization, feature-based residual connections, and preprocessing phases [6]. Unlike the ResNet-50 techniques, these improve the detection accuracy of deepfake films when facemasks are included.

The Deepfake algorithms enable video morphing assaults, which are a serious risk to the integrity of multimedia information. The demand for strong and efficient detection techniques is growing as these assaults become more complex. Modern methods are not always sufficient to detect deepfake films, which emphasizes the need for more sophisticated methods [7]. By using convolutional neural networks (CNNs) to improve the detection of video morphing assaults in deepfake material, this study seeks to close this gap The main hurdles are figuring out the complex visual signals that indicate morphing, tailoring the CNN architecture to this particular job, and making sure the model is robust against different morphing strategies. Through an examination of these facets, this research Endeavor seeks to advance the creation of dependable methods for identifying deepfake movies and protecting the legitimacy of multimedia files in an increasingly digital and altered environment [8].

A dataset comprising a vast number of human face films with labels indicating whether the faces were produced by facial editing techniques is a crucial component of the challenge. Every video the dataset is produced by signing contracts with compensated performers, and the dataset will be openly accessible to the group for creating, evaluating, and analysing methods for identifying recordings where the faces have been altered. Developers that want to take on this challenge will need to consent to the terms of service and seek access to the DFDC [9].

ResNeXt-50 can efficiently learn high-level features from raw picture data since it has been pretrained on extensive image datasets like ImageNet. Through the use of transfer learning, which involves refining the pre-trained model on particular datasets or tasks, ResNeXt-50 may be used to a variety of image-related tasks, including object identification, picture segmentation, and facial recognition [10].

II. CONTRIBUTION

A. Advanced Deep Learning Architecture

We present a novel convolutional neural network (CNN) based deep learning architecture for detecting deep fakes in movies. The ResNeXt-50 model, a state-of-the-art CNN architecture known for its remarkable photo recognition skills, is used in our design. We enhance our system's ability to efficiently extract high-level features from video frames, leading to a more accurate detection of altered data, by building on the fundamental architecture of ResNeXt-50 [11].

B. Temporal Analysis with LSTM

In contrast to conventional deep fake detection techniques, which examine each frame separately, our method uses Long Short-Term Memory (LSTM) networks for temporal analysis. Our model can detect minor changes and inconsistencies that are suggestive of deep fake manipulation more robustly because the LSTM module enables our model to gather contextual information and temporal connections over numerous frames.

C. Dynamic Face Localization

We introduce a dynamic face localization technique that adaptively identifies and extracts facial regions from video frames [12]. By integrating the face recognition library with our system, we efficiently detect and isolate facial regions, focusing the analysis on the most relevant areas for deep fake detection while minimizing computational overhead.

D. Evaluation Metrics for Video-Based Detection

Acknowledging the particular difficulties associated with video-based deep fake detection, we suggest and utilize assessment measures specifically designed to measure our system's effectiveness. A thorough assessment of our system's performance in identifying deep fakes in videos is provided by these measures, which take temporal factors including frame-level accuracy, video-level precision, and temporal consistency into consideration.

III. RELATED WORD

The worries about Deepfake technology are discussed by NORAH M. ALNAIM [1] ZAYNAB M. ALMUTAIRI. Problematic worries were particularly heightened by the widespread use of face masks during the global health crisis. The fact that masks obscure face characteristics makes advanced identification techniques essential. In order to address this, the research suggests creating the Deepfake Face Mask Dataset (DFFMD) and assesses a number of deep learning models, such as transfer learning models like Inception-ResNet-v2 and VGG19 and convolutional neural networks (CNNs). To reduce training difficulties in deep structures, Inception-ResNet-v2, which has 164 layers and residual connections, integrates the ResNet and Inception designs. With its 19 layers, VGG19 uses 3x3 convolution filters to reduce complexity in terms of structure and parameters. Moreover, a custom CNN model is created with Keras that includes convolutional layers, max-pooling, dropout to minimize overfitting, and softmax activation for classification.

Using a CNN-based method for deepfake face identification entails utilizing neural network capabilities to identify altered media. The algorithm has been trained to discriminate between authentic and falsified material. CNNs are a well-liked option for deepfake detection because they are especially ideal overall jobs involving video processing. The convolutional layers receive input data in the form of images, which are composed of pixels in a matrix [13].

A Deepfake video recognition model analysing a video clip. The model receives a video as input, splits it up into separate frames, and uses a sequence of convolutional layers to analyse each frame and extract properties such as textures, edges, and forms. Normalization layers are positioned after every convolutional layer in order to stabilize the learning process. Then, an activation layer adds non-linearity, which is essential for classifying videos. Next, a pooling layer summarizes information from nearby pixels to lower the dimensionality of the data. Capable of managing sequential data such as movies, an LSTM layer examines the characteristics that have been extracted and records the temporal correlations among the frames. Lastly, two thick layers estimate whether the video is real or phony by combining the learnt information [14].

CNN-based method for deepfake face identification, highlighting the ability of neural networks to discern between authentic and altered information, especially pictures and videos by Alben Richards MJ [15]. The model starts by importing the input picture and preprocessing it to eliminate undesirable noise and standardize size and format. Then, features that are crucial for differentiating between actual and synthetic faces are retrieved, such as texture, shape, and colour. For training and validation, the model uses Flickr's 140k dataset, which contains 50K actual faces and 50K deep fakes. Five convolutional blocks and one classifier block make up the architecture [16]. Pooling, activation, and dropout layers come after the thirteen convolution layers.

Four convolutional layers with progressively more kernels—from Conv 6 with 24 kernels to Conv 9 with 128—are included in the architecture developed by Zhiqing Guo [17]. Conv 9 with an 11x11 kernel promotes cross-channel interaction prior to forwarding to the classification module. Each convolution layer employs a modest stride of 1 to extract detailed information. MaxPool, ReLU, and BN are the layers for dimension reduction, non-linearity, and regularization that come after convolution. MaxPool layers save important data in a sliding window, ReLU improves non-linearity, and BN helps output regularization. Interestingly, manipulation traces in MaxPool layers minimize the dimensionality of feature maps with a stride of two and a constant kernel size of 3x3. Effective feature extraction, non-linearity, and regularization are ensured by this thorough design, maximizing.

Realistic Deepfake videos are produced using a GAN-based technique called face swapping or identity swapping. The face swap technique replaces a subject's never-before- compared the face in the origin videos with the saw face in the objective video. The most common usage for it is to add well-known actors to different movie segments. GANs and conventional CV methods, such as Face Swap (an application for face swapping), can be used to synthesis face swaps these methods is implemented by Wodajo and Atnafu.[18]. Darius proposed a CNN model called ResNeT network to automatically detect hyper-realistic false movies created with Deepfake and Face to Face. The scientists used two topologies of networks (Meso-4 and MesoInception-4) that focus on the mesoscopic features of a picture. In order to capitalize on the picture transform inconsistencies—Yuzu as well as Sawai proposed a CNN architecture that addresses issues—like movement, cutting, and scaling that come up during the production of Deepfakes. Their technique centres on nonlinear face-warping effects as a method of differentiating between real and fake movies [10].

Video morphing detection method is proposed by Liming Jiang [19]. Two notable benchmarks for face forgery detection are Face Forensics Benchmark and Celeb-DF. The former introduces six image-level forgery detection baselines but lacks exploration of various perturbation types and their combinations. Celeb-DF, on the other hand, offers a benchmark with seven methods, but the assumption that the test set mirrors the training set's distribution introduces biases and limits practicality for real-world scenarios [20]. To address these shortcomings, a new benchmark is proposed, featuring a challenging hidden test set with manipulated videos, aiming to simulate diverse real-world distributions. This benchmark analyses various perturbations for a more comprehensive evaluation and primarily focuses on video-level forgery detection baselines. Additionally, to tackle issues of low visual quality and face style mismatches, a Deep Fake Variational Auto-

Encoder (DF-VAE) framework is introduced, which emphasizes generality, scalability, and temporal continuity in generating high-quality videos through three key components: a structure extraction module, a disentangled module, and a fusion module.

To differentiate authentic films from deepfake ones using optical flow fields. Seeming motion in a video is captured by optical flow, which is calculated between consecutive frames this proposed by Irene Amerini and Leonardo Galteri [21]. There is a theory that deepfakes, especially with regard to facial motions, differ in motion from those that were taken in real life. Extracting forward flow between frames is done using PWC-Net, a CNN model for optical flow. We then feed this flow into Flow-CNN, a semi-trainable CNN that uses pre-trained backbones such as VGG16 or ResNet50. Due to dataset size limitations, transfer learning is used, with certain network components pre-trained and the remainder refined using deepfake data. After layers are educated, they are frozen for the fine-tuning process.10,000 face animation films from ten distinct activities are included in the Deep Fake MNIST+ dataset, a noteworthy addition, along with 10,000 actual human face videos from separate datasets by G. Luo [22].

Interestingly, these animated movies continue to provide a challenge to more modern detectors by successfully parodying the liveness detection methods currently available in the market. The dataset was produced using Siarohin's framework, which combines motion features and local affine transformations from driving videos to produce high-quality animations while maintaining the original image's identity. The actions included in the dataset include blinking, yawing, nodding, and emotional expressions like surprise and embarrassment [24]. Videos that met the VoxCeleb1 dataset format were cropped after being captured using the front-facing Pro cameras. Each action in the dataset has 1,000 spoof videos in order to guarantee difficult samples. This was achieved by filtering movies that potentially evade detection using two liveness detection APIs. This study emphasizes the challenge of successfully identifying Deep Fake MNIST+ films, highlighting the need for better detection models to thwart such assaults. It does this by analysing the efficacy of current detectors trained on datasets such as FF++.

Deep learning applications for a range of jobs are examined in this article. Dr. N. Palanivel's Convolutional Neural Network (CNN) for real-time object identification is proposed in the first part. With the help of this technology, items in each frame of video would be identified. Requiring hardware acceleration and specialized CNN architectures may be necessary to achieve real-time performance, which is a significant hurdle [25]. Second, utilizing recurrent neural networks (RNNs), the paper presents a unique method for categorizing cardiac illness. Because RNNs are so good at processing sequential data, they may be used to analyse medical data such as ECG readings, which is something that traditional approaches struggle with. In order to enhance accuracy over a single model, the authors suggest integrating several RNN classifiers [26].

IV. PROPOSED WORK

The study used a number of different technologies, including information gathering, datasets preparing, information dividing, constructing models, model training, validation of models, and evaluation of models.

A. Frame Extraction of video

This paper loads input video files and extracts frames using OpenCV, a popular computer vision library. The system reads video files frame by frame by utilizing OpenCV's capability, which makes preprocessing and analysis easy. Strategic frame sampling is used by the system to reduce the computing burden of processing each frame. Frequent frame sampling maximizes computing efficiency while guaranteeing thorough coverage of the video material. This sampling technique allows for efficient processing of following steps in the deep fake detection pipeline by balancing computing resources with the comprehensive inspection required for proper analysis [27].

The system incorporates the face recognition library at the first preprocessing stage in order to carry out face localization in every video frame. Before moving on to further analysis, this critical stage seeks to precisely recognize and pinpoint face areas within the video frames. The method utilizes the face recognition library's ability to identify and separate facial characteristics, enabling targeted examination of specific regions of interest [28]. This methodical technique guarantees that only frames with identifiable faces are taken into account for further examination, which is consistent with the way deep fake modifications primarily target facial features. By use of accurate face localization, the method improves the effectiveness of next processing stages, enabling more reliable and accurate identification of possible deepfake modifications.

Following the effective localization of faces inside the video frames, the system moves on to the next important step: frame extraction. This process involves taking systematic pictures and storing them for further examination of the localized faces in the frames. Every extracted frame function as a snapshot, encapsulating a discrete point in the video sequence. This allows the system to examine the temporal development of facial characteristics and identify any anomalies that would indicate the use of deep fakes [29]. The algorithm makes sure that the analysis that follows concentrates on crucial facial areas, where profound false modifications are most likely to occur, by carefully extracting frames that feature localized faces. These extracted frames provide the basis for identifying minor differences or abnormalities throughout the temporal domain, enabling the system to more accurately and consistently identify and flag possible deep fake manipulation cases. By means of methodical frame extraction, the system establishes the foundation for all-encompassing temporal analysis,

which permits the identification of minute modifications suggestive of deeply faked material present in the video feed [30].

Preprocessing procedures are used once frames are extracted to make sure the frame data is consistent and compatible with the deep fake detection model. These preprocessing steps include a range of methods designed to standardize the extracted frames' properties and format. First, resizing may be used to reduce computational cost and enable efficient processing by adjusting the frame size to a predefined resolution. Furthermore, normalization techniques are frequently used to reduce fluctuations in lighting, contrast, or colour distribution that can possibly confuse the detection process by standardizing pixel intensity values between frames.

B. Eye Blink Statistics

The following model, which was developed after experimenting with earlier models, focuses on a single facial trait that may be useful in identifying deepfake films. Online, there are surprisingly few pictures of people with their eyes closed. This makes it more challenging for programmers to produce deepfake films that faithfully mimic the blinking rate of a human. The notion that the blinking rate is a complex component of face activity, impacted by a variety of elements including emotional state, weariness, and environmental circumstances, is in line with this emphasis on eye blink statistics. The dearth of closed-eye photographs on the internet unintentionally results in a dearth of training material for deepfake algorithms that target this particular facial characteristic.

As such, it becomes more challenging for programmers to replicate real blinking dynamics, which might make it a reliable signal for distinguishing between real and fake visual material. The fact that this model was improved by testing with previous iterations emphasizes the iterative nature of deepfake detection research. The shift from earlier models to this targeted strategy points to a commitment to ongoing development and flexibility in the face of new obstacles in the dynamic field of deepfake production. The focus on eye blink data adds a crucial layer to current approaches in the wider context of deepfake identification. Ingenious strategies like this one aid in the development of more durable and dependable solutions for preserving the authenticity of visual material across a variety of online platforms and apps as the arms race between deepfake generation and detection technology continues.

C. Spectral Responses

Though these kinds of artifacts have been crucial in the battle against deepfakes, detection methods can't depend on them indefinitely. These kinds of artifacts are reasonably simple to find and correct since they are byproducts of the deconvolutional process rather than variations in the videos. These artifacts have restricted even if they work well in some situations. Their relative simplicity, which makes them rather easy to recognize and fix, is one significant drawback. Certain patterns are frequently introduced during the deconvolutional process, and these patterns can be used to identify tampering when examined in the frequency domain. But as deepfake technology advances, producers are becoming more conscious of these detection methods and may take steps to lessen the influence of spectral response-based analysis.

D. Convolution Neural Networks (CNN)

Convolutional neural networks (CNNs) are deep learning neural networks composed of pooling, convolution, completely linked, and irregular layers. Using the Keras toolkit, the suggested convolutional network in this study was constructed. It has Three distinct levels of convoluted that are REL was used as the mechanism for activation in all levels, and a layer with a maximum pooling capacity with a pooling size was used to decrease the size of the massive movie. Three different degrees of convolution are included in the design of the convolutional network that is proposed in this study. Convolutional layers are essential for identifying small-scale patterns and characteristics in a picture, which helps the network identify intricate structures.

Across all layers, the Rectified Linear Unit (ResNeXT-50) activation function was used, which improved the network's ability to simulate non-linear connections and extract complex characteristics from the input data. This study's use of CNNs highlights how successful these models are for tasks involving visual data, especially those involving deepfake detection. CNNs are successful in image-based applications because of their innate capacity to learn hierarchical representations, as well as activation functions like pooling layers and ResNeXT-50. It's vital to remember, though, that the architecture and hyperparameter selection such as the quantity of layers and their arrangement have a significant impact on how well CNNs perform overall for certain tasks. Moreover, a maximum pooling layer was added to effectively handle the size of big videos. By down sampling the incoming data, pooling layers minimize its spatial dimensions, preserving important properties while lowering computational complexity. In order to achieve a compromise between computational efficiency and feature preservation, the pooling size was carefully selected.

E. Dataset

DL models are trained using data. Because of this, the quality of their learning and the accuracy of the predictions they make depend on the careful preparation of the dataset. The faces are extracted using the Blaze Face neural face detector, MTCNN, and face recognition DL libraries. Face recognition and Blaze Face can

process a lot of videos quickly. A basis for integrating a broad range of deepfake films, including different alteration techniques and attributes, is provided by leveraging well-established deepfake datasets like Google's Deepfake Detection Dataset, Face Forensics++, and the Deepfake Detection Challenge (DFDC) dataset.
Two primary datasets are used in the training phase of a machine learning model: the training dataset, which is used to expose the model to a variety of patterns and instances, and the testing dataset, which assesses the model's learning efficiency from the training data.

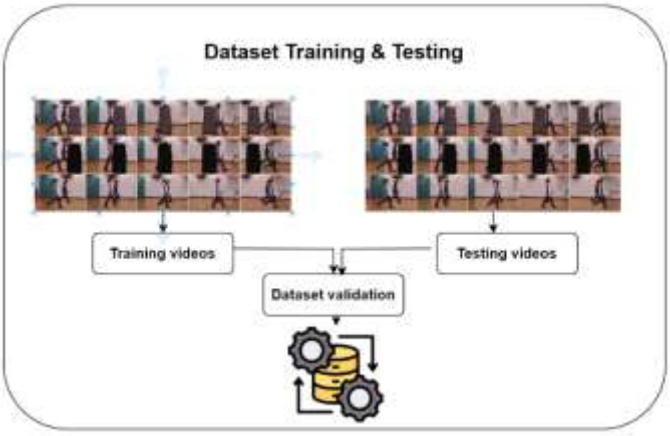

Figure 1: Dataset validation Process

Furthermore, a crucial phase known as dataset validation is performed to identify instances of overfitting, in which the model becomes too dependent on the training set and exhibits worse performance on fresh, untested data (As shown in the figure 1). Both datasets must faithfully replicate real-world data, with the training dataset being far bigger than the testing dataset for training 800 videos and then 160 videos to provide the model with a sufficient number of instances to work with. The training and testing datasets are split at random to ensure accuracy and fairness.

F. Proposed Model

The suggested model may include a number of cutting-edge methods in addition to the fundamental CNN-RNN architecture to improve its detection of video morphing threats. First, using large-scale picture datasets like as ImageNet to initialize the CNN with pre-trained weights is one way to apply transfer learning. By using this initialization, the model might possibly improve its capacity to distinguish between real and fake video footage by utilizing information gained from a variety of visual cues. Moreover, predictions from several CNN-RNN models trained with various initializations or hyperparameters may be combined using ensemble learning techniques, which lowers the possibility of overfitting and increases the detection system's overall resilience. RNN is used to train dataset with combination of the LSTM to give accurate prediction on the data validation on the process. Then the process is moved to Pre-Processing.
After the pre-processing is finished, the data is used to train a deep learning model, which is often a mix of ResNeXt-50 and LSTM (Long Short-Term Memory). The algorithm gains the ability to identify the unique patterns and traits that set authentic films apart from deepfakes as it trains. The model is now ready for testing after it has been trained. A video is first divided into its component frames, regardless of whether it is being uploaded or watched live. After that, the model carefully examines every frame, examining each one to see if it shows a genuine face or a deepfake.

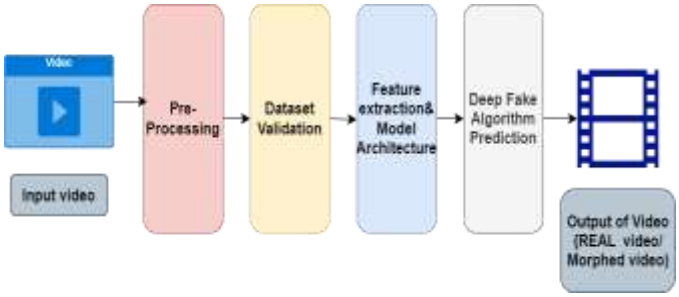

Fig 2: Process of video morphing detection Model.

To further enable the model to focus on small morphing cues while disregarding irrelevant background noise, attention mechanisms inside the RNN module may be adjusted to dynamically prioritize informative frames or areas within each frame. Furthermore, the model's resistance to complex adversarial attacks that are frequently seen in real-world settings may be increased by implementing adversarial training procedures to

supplement the training data with adversarial constructed samples. Moreover, semi-supervised or self-supervised learning approaches might be investigated to overcome the problem of insufficient labelled data for training (As shown in the figure 2). In these methods, the model learns to differentiate between authentic and fraudulent movies without the need for large labelled datasets. Synthetic video samples may be produced using methods like contrastive learning or generative adversarial networks (GANs), which can be used to increase the size of the training set and enhance the generalization capabilities of the model.
Finally, to help users understand the morphing artifacts or temporal inconsistencies that most influence the detection outcome, model interpretability techniques like saliency maps or attention visualization techniques can be used to provide insights into the CNN-RNN model's decision-making process. Through the incorporation of these sophisticated methodologies, the suggested model is capable of attaining cutting-edge results in identifying video morphing assaults, all the while preserving interpretability and resilience in practical implementation situations.
Next the Data Splitting process Feature extraction: Identifying and extracting relevant features from the data that will be used for prediction. Feature engineering is Creating new features from existing ones that may be more predictive. Model Training is the training set is used to train a machine learning model. The model learns to identify patterns and relationships in the data that can be used for prediction. There are many different machine learning algorithms that can be used for data mining. The performance of the trained model is evaluated using the testing set. This involves metrics like accuracy, precision, recall, and F1 score.
To identify Fake or Real on the final output of the pipeline may be a classification of "Fake" or "Real", depending on the specific application.
Flow-CNN each output a score through a sigmoid activation function. These scores are likely measures of how confident the network is that the input is a real face. The system then compares these scores to a threshold. If both scores are above the threshold, the input is classified as a real face and goes down the "Original" branch. If either score is below the threshold, the input is classified as a fake face and goes down the "Fake" branch. The final output of the system depends on the branch it takes. The "Original" branch may output the identity of the recognized person or a confidence score for the recognition. The "Fake" branch may simply output a label of "fake" or provide more information about why the input was classified as fake.
Deepfake algorithm finds the variation on the morphed video and then check that video by using the CNN based technique. after the verification by this algorithm the morphed video identified. If the video is normal one this algorithm gives the number of the normal one. In the existing system this method only applied for the image morphing detection but now this Paper finds out the morphing detection for the videos.

V. METHODOLOGY

A. Dataset Preprocessing

Video frames with faces can be identified using the face recognition library. Pre-trained models for precisely identifying faces in pictures or video frames are available from this collection. In order to prepare the dataset for preprocessing in a deepfake detection Paper, a variety of actual and deepfake films encompassing a range of situations and facial expressions are sourced. After that, the dataset is labelled so that each video is tagged for supervised learning with the appropriate label (fake or real). While data augmentation increases dataset variety by applying changes like rotation and flipping, data cleaning entails deleting damaged or missing video recordings. In order to ensure consistency, face identification and alignment algorithms find and standardize facial locations inside frames.
Frame sampling helps with temporal feature capture by identifying uniform-length segments from films. Pixel values are normalized and features are scaled for convergence during training using normalization and standardization. To evaluate the model, dataset splitting divides the data into test, validation, and training sets. Lastly, batches of labeled and pre-processed video sequences are rapidly loaded for training using a data loader. The dataset is guaranteed to be clear, diversified, and prepared for training a strong deepfake detection model thanks to our thorough pretreatment approach.

B. Feature Extraction

To enable precise deepfake detection, feature extraction is essential for extracting discriminative information from video frames. To maintain consistency and improve model performance, preprocessing operations like scaling and normalization are first applied to each frame. After preprocessing, high-level spatial characteristics are extracted from individual frames using convolutional neural networks (CNNs) pretrained on large-scale picture datasets like ResNeXT-50 or VGG. These CNNs use their capacity to identify patterns and objects in pictures to perform the function of feature extractors. The input frame's primary visual attributes are then captured by using the output of the CNN's final convolutional layer as a representation of the input frame.
Recurrent neural networks (RNNs) or their derivatives, such as long short-term memory (LSTM) networks, are used to evaluate sequential input and store temporal information. By processing spatial feature sequences that are taken from successive frames, these RNNs enable the model to comprehend the temporal dynamics and context of the video. Furthermore, attention processes might be included to concentrate on pertinent areas

of interest within the frames, improving the model's capacity to distinguish minute distinctions between authentic and altered material. In order to enable the model to capture both static and dynamic components of the video input, the feature extraction step combines CNNs for spatial feature extraction and RNNs for temporal modelling. This lays the groundwork for an efficient deepfake detection system.

C. Temporal Modelling with LSTM

One of the main features of this Paper's deepfake detection methodology is temporal modelling utilizing Long Short-Term Memory (LSTM) networks, which makes it possible to analyse sequential data and capture the temporal dynamics intrinsic in films. Long-term dependencies in sequential data can be preserved via recurrent neural network (RNN) architectures known as long-term switching machines (LSTMs), which are intended to solve the vanishing gradient issue that conventional RNNs face. Large-scale feature extraction sequences from successive frames of the input video are processed by LSTMs in the context of deepfake detection.

By sequentially feeding these spatial features into the LSTM network, the model learns to recognize patterns and relationships over time, effectively capturing the evolution of visual content throughout the video. This temporal modelling capability enables the model to discern subtle variations and inconsistencies introduced by deepfake manipulation techniques, such as facial reenactment or expression synthesis, which often manifest gradually across multiple frames. Furthermore, LSTMs can adaptively weigh the importance of past and present information through their gated structure, allowing the model to focus on relevant temporal cues while filtering out noise and irrelevant background changes. As a result, the LSTM-based temporal modelling enhances the deepfake detection system's robustness and accuracy by leveraging the sequential nature of video data to distinguish between authentic and manipulated content effectively.

D. Prediction on Videos

Prediction entails using the trained deep learning model to determine if a certain video segment is a deepfake or not, or whether it comprises edited or real information. Following preprocessing, feature extraction, and temporal modelling using LSTM, the trained model receives the input video data. After that, the model examines the temporal dynamics that the LSTM network recorded while processing the spatial characteristics that were taken out of each frame. The algorithm makes predictions about the video segment's authenticity based on this research. In addition to a confidence score that indicates the model's level of confidence in its forecast, these predictions usually include a binary classification label ("REAL" or "FAKE").

Greater scores indicate greater certainty. The confidence score represents the model's degree of confidence in its prediction. These visual aids can offer valuable insights into the decision-making process. All things considered, the prediction stage is essential for evaluating the veracity of video material and spotting possible deepfake manipulation cases, which helps detect and reduce misinformation and disinformation propagated by manipulated media.

E. Python Flask Web Application

The main user interface for the deepfake detection system in this Paper is the Flask web application written in Python. It has a homepage with connections to other sections of the website, a login page in case user verification is needed, and an uploader page where users may submit videos from their local systems. The program controls the file upload procedure, stores the video in the server filesystem, and starts the deepfake detection system for examination once a video is uploaded. When the analysis is finished, the program shows the prediction findings, along with a confidence score and a classification of whether the video is authentic or phony.

F. System Architecture

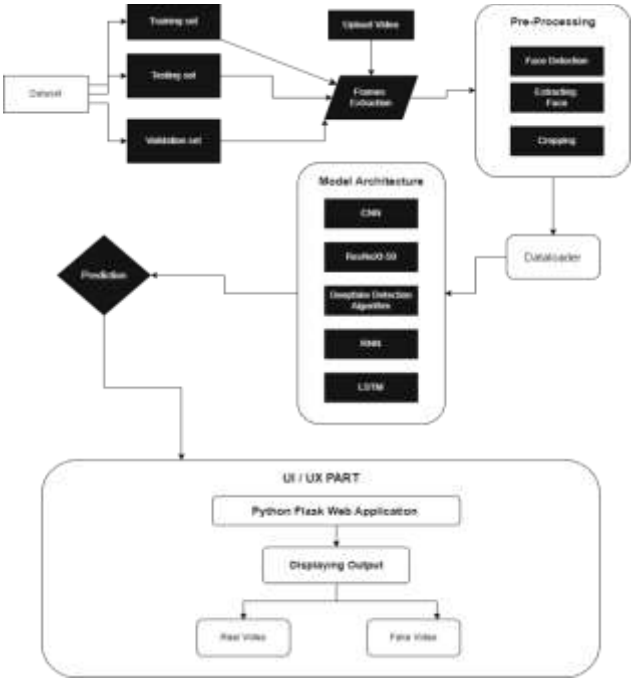

Figure 3: Proposed Video Morphing Detection System Architecture

In order to ensure an accurate description of probable circumstances and contexts where deepfakes may be encountered, the dataset is gathered by carefully compiling a varied variety of actual and fake movies from multiple sources. Training, testing, and validation sets of films are created by combining actual and fictitious footage during the dataset collecting process. From each video, frames are taken, and then cropping is done after face identification techniques are used to find human faces in these frames. Using a ResNeXt-50 CNN model, feature extraction is carried out to capture unique face features (As shown in the fig 3). To differentiate between authentic and fraudulent films, an LSTM RNN model is then utilized for feature classification. Users may upload movies and get predictions about their authenticity using the Python Flask web application, which serves as an interface between the user and the deepfake detection algorithm. This tool helps mitigate misinformation and protects against harmful usage of synthetic media by allowing users to identify potentially deepfake material.

VI. PERFORMANCE EVALUATION

Deepfake detection system may be evaluated for efficacy and dependability using a number of critical performance criteria. The Receiver Operating Characteristic (ROC) curve, Area Under the Curve (AUC), F1 score, accuracy, precision, recall, and confusion graph are all included in this set of measurements. The key component of a model's overall correctness is accuracy, whereas recall and precision provide information on false positives and false negatives, respectively. The F1 score takes into account both kinds of faults, making it a fair indicator. Error analysis is aided by the confusion graph, which offers a thorough dissection of the predictions.

Classification report				
	precision	recall	F1-score	Support
fake	0.85	0.94	0.89	10000
real	0.93	0.84	0.88	10000
accuracy	-	-	0.89	20000
Macro avg	0.89	0.89	0.89	20000
Weighted avg	0.89	0.89	0.89	20000

Table 1: Evaluation of Algorithm scores.

Table 1 shows the performance metrics of the Deep Fake Detection Algorithm Precision quantifies the percentage of affirmative forecasts that turned out to be accurate. For instance, an accuracy of 0.85 for "fake" news indicates that, of all the items the model identified as bogus, 85% were in fact fake. The percentage of real positive cases that the model properly recognized is measured by recall. For instance, a recall of 0.93 for "real" news indicates that 93% of the real news pieces were properly classified by the model out of all of them. The F1-Score is the harmonic mean of recall and accuracy. It is employed to assess the accuracy of a test on skewed or unbalanced data sets. Lower scores indicate deficiencies in either precision or recall, whereas a number of 1 denotes the ideal balance between the two.

Formula for calculating Accuracy, Precision, Recall and F1-score.

- **Accuracy:**

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

- **Precision**:

Precision=$\frac{TP}{TP+FP}$

- **Recall (Sensitivity):**

Recall = $\frac{TP}{TP+FN}$

- **F1-score:**

F1-score=$2 \times \frac{Precision \times Recall}{Precision+Recall}$

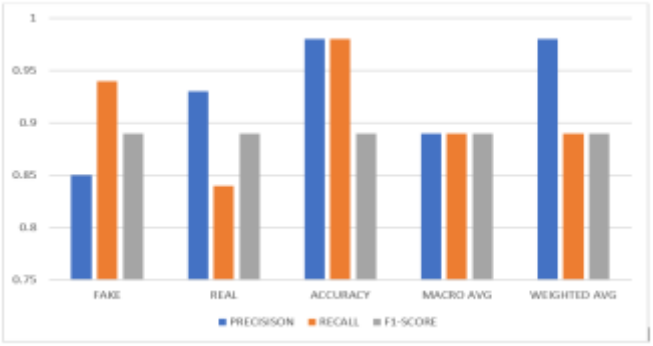

Figure.4 Accuracy, Precision, F1-score ,Macro Average for selected ResNeXT-50 Model.

The accuracy at the bottom of the table is a metric that looks at how many of the total predictions were correct (98% in this case). There are also macro and weighted averages which can be useful for comparing models trained on imbalanced datasets. Overall, the high precision, recall and F1 scores for both real and fake categories suggest that this model is performing well at classifying (As shown in the fig 4).

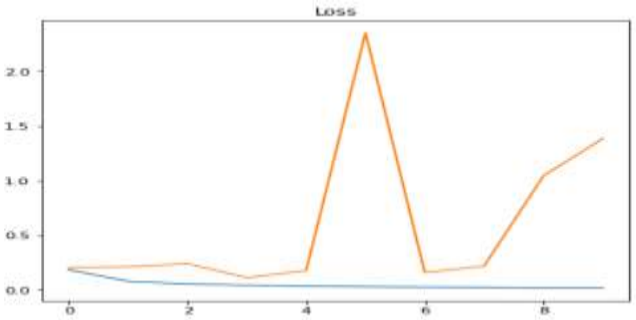

Figure 5: Loss of Data Training in ResNeXT-50 model

The difference between the model's expected output and the actual target values during training is represented by the training data loss, which is given as 2.2. A loss value of 2.2 indicates that there is an average 2.2-unit deviation between the model's predictions and the real data (As shown in fig.5). Better performance is usually shown by lower loss values, which show that the model's predictions match the real data more closely. However, there are a number of variables that might affect how loss numbers should be interpreted, including the job at

hand and the complexity of the dataset. Throughout the training phase, it's critical to keep an eye on the loss to make sure the model is learning efficiently and approaching peak performance.

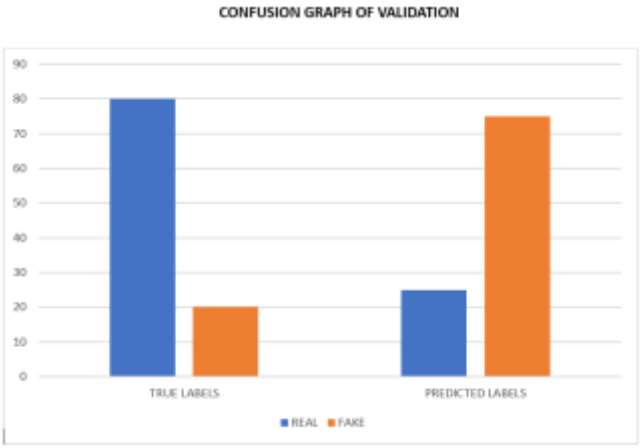

Figure.6: Deep Fake Face Detection Confusion Graph

The given confusion graph provides a thorough overview of the performance of the classification model by showing the number of accurate and inaccurate predictions for both actual and fictitious categories (As shown in the fig 6). The model predicts the labels in the columns of the graph, whereas the rows correspond to the actual labels of the data, indicating which labels are true and which are bogus. With 9392 cases successfully detected, True Positives (TP) represent the instances that are appropriately classified as real. False Positives (FP), which amount to 608, are misclassifications in which things were mistakenly identified as real. False positives with the right label are called True Negatives (TN), and they have the number 8371. False Negative (FN) on the other hand, which amount 1629, are misclassifications in which things were mistakenly categorized as fake. By using the confusion graph, it is easier to calculate other performance measures like as accuracy, recall, and F1-score, which are important to assess how well the model works to discriminate between actual and false instances.

This paper explain about the Deep Fake detection based on the face recognition library, To separate the face from the video . This separation is based on the facial expressions and eye statics .

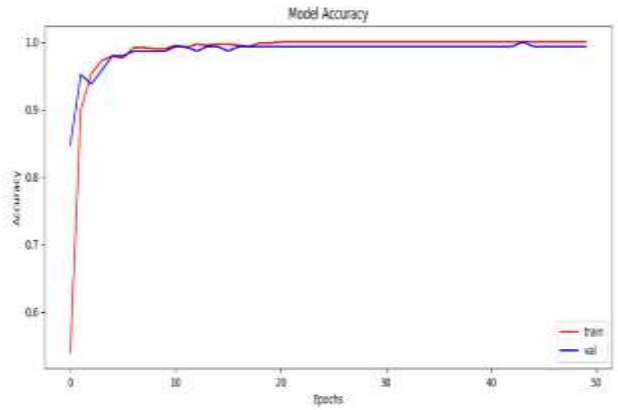

Figure.7: Model Accuracy of the machine Learning Performance

The model's accuracy throughout several epochs on the training and validation datasets is shown in the image. As the model gains knowledge from the training dataset, the accuracy on the training data is displayed by the red line, which increases steadily with each epoch. On the other hand, the accuracy on the validation data, a different dataset used to assess the model's generalization to unobserved data, is depicted by the blue line. To prevent overfitting, a situation in which the model retains the training data but finds it difficult to process fresh data, it is imperative that the model perform well on the validation set. Although it is ideal for the accuracies on the two datasets to match, overfitting might be indicated by a discrepancy between the training and validation accuracies. The x-axis indicates "Epochs," while the y-axis shows "Accuracy," with each epoch denoting shows the total accuracy of 97.8%.

VII. DISCUSSION

This Paper's topic includes assessing the deepfake detection algorithm's efficacy in identifying real and fake movies by taking into account metrics like accuracy, precision, recall, and F1-score. It also entails evaluating how the model performs in relation to various dataset preparation methods, such as data gathering, frame extraction, and face identification. It is necessary to look at how feature extraction using a ResNeXt-50 CNN model and temporal modelling with an LSTM work together to capture temporal dependencies in the video data. Overfitting, model generalization, dataset imbalance, and other issues that arise during model training, validation, and testing must be addressed. Implications for real-world application in the fight against disinformation must also be considered. Further research and development in deepfake detection and mitigation must take into account factors like computational resources, scalability, and efficiency, as well as potential improvements like enhanced deep learning architectures and dataset extension.
Our web application is built by python framework which is flask , this python framework help us to connect the webpages in the web application .

A. Application Home page UI

Figure 8. Home page UI

B. Authentication Page UI

Figure 9. Login page UI

C. Main page UI

Figure 11.Main Page UI

D. Video Morphing Detection Pages UI

FIGURE 11. VIDEO UPLOADING PAGE UI

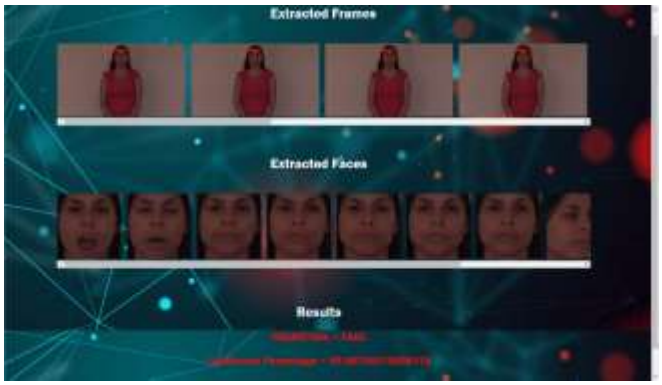

FIGURE 12. VIDEO MORPHING RESULT PAGE UI

VIII. CONCLUSION

In Conclusion, Our research effectively combined a number of cutting-edge machine learning algorithms to create a reliable deepfake detection system. The model showed good accuracy in differentiating between authentic and modified content by preprocessing a wide dataset of actual and fraudulent movies, extracting pertinent features using ResNeXt-50 CNNs, and employing LSTM RNNs for temporal modelling. The Paper yielded important insights into the model's performance and possible areas for development through a thorough review utilizing performance measures and the display of outcomes. The deepfake detection algorithm's practical value was further demonstrated by the development of a Python Flask web application, which provided users with an easy-to-use platform to evaluate the authenticity of videos.

This effort emphasizes the value of multidisciplinary approaches in solving current difficulties in media integrity and trustworthiness and adds to the expanding body of research targeted at halting the spread of

deepfake technology achieved an impressive total accuracy rate of 97.8%. In order to prevent the spread of false information and safeguard the integrity of digital material in a variety of situations, the model's efficacy and applicability may be further improved by ongoing attempts to improve it, enlarge the dataset, and incorporate real-time detection capabilities.

REFERENCES

1. DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era With Deepfake Detection Algorithms NORAH M. ALNAIM 1, (Member, IEEE), ZAYNAB M. ALMUTAIRI2 , MANAL S. A
2. Deep Fake Face Detection using Convolutional Neural Networks Alben Richards MJ, Kaaviya Varshini E, Diviya N, Kasthuri P.
3. Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain imageto-image translation," in Proc. IEEE Conf. Comput. Vis. pattern Recognin.
4. B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge (DFDC) preview dataset".
5. S. R. Ahmed, E. Sonuç, M. R. Ahmed, and A. D. Duru, "Analysis survey on deepfake detection and recognition with convolutional neural networks," in Proc. Int. Congr. Hum. -Comput. Interact., Optim. Robot. Appl. (HORA).
6. Zhiqing Guo, Gaobo Yang, Jiyou Chen, Xingming Sun (2021) "Fake face detection via adaptive manipulation traces extraction network" in Computer Vision and Image Understanding-Volume 204.
7. J. Huang, X. Wang, B. Du, P. Du, and C. Xu, "DeepFake MNIST+: A deepfake facial animation dataset," in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2021, pp. 1973–1982.
8. Pishori, B. Rollins, N. van Houten, N. Chatwani, and O. Uraimov, "Detecting deepfake videos: An analysis of three techniques," 2020, arXiv:2007.08517.
9. Y. Li and S. Lyu, "Exposing DeepFake videos by detecting face warping artifacts," 2018, arXiv:1811.00656.
10. Wodajo and S. Atnafu, "Deepfake video detection using convolutional vision transformer," 2021, arXiv:2102.11126.
11. Amerini, L. Galteri, R. Caldelli, and A. D. Bimbo, "Deepfake video detection through optical flow-based CNN," in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 1–3.
12. Y. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey," ACM Comput. Surv., vol. 54, no. 1, pp. 1–41, 2021.
13. M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: A systematic literature review," IEEE Access, vol. 10, pp.
14. 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. 3207–3216.
15. L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, "DeeperForensics1.0: A large-scale dataset for real-world face forgery detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 2889–2898.
16. Palanivel, "Design and Implementation of Real Time Object Detection using CNN" in IEEE - International Conference on System, Computation, Automation and Networking (ICSCAN2023) conducted by Manakula Vinayagar Institute of Technology, Puducherry on 17-11-2023.
17. Y. Zhang, L. Zheng, and V. L. Thing, "Automated face swapping and its detection," in 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP). IEEE, 2017, pp. 15–19.
18. Palanivel, "Novel Implementation of Heart Disease Classification Model using RNN Classification" in IEEE - International Conference on System, Computation, Automation and Networking (ICSCAN2023) conducted by Manakula Vinayagar Institute of Technology, Puducherry on 17-11-2023.
19. Singh, A. S. Saimbhi, N. Singh, and M. Mittal, "DeepFake video detection: A time-distributed approach," Social Netw. Comput. Sci., vol. 1, no. 4, pp. 1–8, Jul. 2020.
20. Palanivel, "Forest Fire and Smoke Detection using Convolution Neural Networks" in IEEE - International Conference on System, Computation, Automation and Networking (ICSCAN2023) conducted by Manakula Vinayagar Institute of Technology, Puducherry on 17-11-2023.
21. Sathishkumar, R., and M. Govindarajan. "A Comprehensive Study on Artificial Intelligence Techniques for Oral Cancer Diagnosis: Challenges and Opportunities." In 2023 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1-5. IEEE, 2023.
22. Sathishkumar, R., Govindarajan, M., Deepankumar, R. (2024). Hate Speech Detection in Social Media Using Ensemble Method in Classifiers. Mobile Radio Communications and 5G Networks. MRCN 2023. Lecture Notes in Networks and Systems, vol 915. Springer, Singapore. https://doi.org/10.1007/978-981-97-0700-3_16
23. Sathishkumar, R., Govindarajan, M., Dhivyasri, R. (2024). Detection and Classification of Neuro-Degenerative Disease via EfficientNetB7. Mobile Radio Communications and 5G Networks. MRCN 2023. Lecture Notes in Networks and Systems, vol 915. Springer, Singapore. https://doi.org/10.1007/978-981-97-0700-3_17

24. H.Khalid,S.Tariq,M.Kim,andS.S.Woo,"FakeAVCeleb:Anovelaudio video multimodal deepfake dataset," 2021, arXiv:2108.05080.
25. S. Uçan, F. M. Buçak, M. A. H. Tutuk, H. İ. Aydin, E. Semiz, and S. Bahtiyar, "Deepfake and security of video conferences," in Proc. 6th Int. Conf. Comput. Sci. Eng. (UBMK), Sep. 2021, pp. 36–41. [12] N. Graber-Mitchell, "Artificial illusions: Deepfakes as speech," Amherst College, MA, USA, Tech. Rep., 2020, vol. 14, no. 3.
26. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge (DFDC) preview dataset," 2019, arXiv:1910.08854
27. Palanivel, "Smart Parking and Parking Guidance Using ECNN Algorithm in Convolution Neural Network" in IEEE - International Conference on System, Computation, Automation and Networking (ICSCAN2023) conducted by Manakula Vinayagar Institute of Technology, Puducherry on 17-11-2023.
28. D.GüeraandE.J. Delp, "Deepfake video detection using recurrent neural networks," in Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Nov. 2018, pp. 1–6.
29. Palanivel, "Performance Evaluation of the Infected Rice Leaves Using RCNN" in IEEE - International Conference on System, Computation, Automation and Networking (ICSCAN2023) conducted by Manakula Vinayagar Institute of Technology, Puducherry on 17-11-2023.
30. C.-C. Hsu, C.-Y. Lee, and Y.-X. Zhuang, "Learning to detect fake face images in the wild," in 2018 International Symposium on Computer, Consumer and Control (IS3C). IEEE, 2018, pp. 388–391.