

***School of Computer Science Engineering and Information Systems***

**Department of Computer Applications**

**Winter Semester 2023-2024**

**SET CONFERENCE**

**First Review (28th February 2024)**

**Reg. No: 23MCA0304, 20MCA0310, 20MCA0314,**

**Student Name: Sayantan Bhattacharyya, Milind Chakraborty, Nitin Sharma**

**Guide Name: Prof. DHARMENDRA SINGH RAJPUT**

# Paper Title: Deepfake Detection Using Deep Learning Techniques

**Problem Description:**

The widespread creation and distribution of deepfake images and videos, facilitated by advancements in smartphone technology and social media platforms, pose a significant threat to individual and national security. These fabricated visuals, often indistinguishable from genuine footage, have the potential to manipulate public opinion, spread misinformation, and undermine trust in legitimate information sources. Deepfakes can be weaponized to humiliate and endanger individuals by creating damaging and defamatory content, potentially leading to social and legal repercussions. Moreover, they can be used to fuel radicalization and terrorism by manipulating religious and ideological sentiments, potentially inciting violence and promoting extremist agendas. In the context of national security, deepfakes can be employed to create disinformation campaigns, manipulate foreign policy decisions, and sow discord within societies.

Despite the growing threat, current methods for detecting deepfakes are often limited in their accuracy and effectiveness. This research aims to investigate and develop robust techniques for identifying and mitigating the risks associated with deepfake videos. By exploring the latest advancements in vision transformers and inception net technology, this study seeks to establish a highly accurate and reliable method for differentiating authentic videos from deepfakes.

This research addresses a critical need for innovative solutions in the face of the ever-evolving landscape of deepfakes. By developing a reliable deepfake detection method, we can safeguard individuals, strengthen national security, and foster a more trustworthy information ecosystem.

**Literature Survey:**

[1] D. Güera and E. J. Delp in their paper used CNN and LSTM for frame feature extraction and temporal sequence analysis, with a shallow network comprising two fully-connected layers and one dropout layer. The dataset used contains 600 deepfake videos from multiple video-hosting websites and HOHA dataset, with an accuracy of 97.1% with 80 frames.

[2] This paper by X.Chang et al. proposed a type of VGG network based on noise and image augmentation (NA-VGG) by adding an SRM filter layer and an image augmentation layer in front of the VGG16 network. Celeb-DF dataset is used for training evaluation. The image is extracted from the Deepfake video. The model achieved an accuracy of 85.7 %.

[3] Huaxiao Mo et al. have proposed an architecture that converts RGB images into residuals and then passes them through three-layer groups consisting of a convolutional layer, LReLu activation function, and a max pooling layer. The resulting output is then fed into two fully-connected layers, and the SoftMax layer is used to generate the final output. The dataset used for this architecture is prepared from the CELEBAHQ dataset.

[4] This study uses optical flow to differentiate between a deepfake and a genuine picture. A CNN pre-trained with VGG-16/ResNet50 is fed optical flows, followed by a sigmoid activation to determine if a frame is false or real. The FaceForensics++ dataset gives an accuracy of 81.61% with VGG16 and 75.46% with ResNet50.

[5] The proposed CFFN consists of three dense units with a transition layer of 0.5 & a growth rate of 24. A convolution layer with 128 channels and 3x3 kernel size is concatenated to the output layer of the last dense unit. To obtain the discriminative feature representation, a dense layer is inserted last. The dataset used in the experiments was extracted from CelebA, with 10,177 of identities and 202,599 aligned face images. This method has a recall value of 0.900.

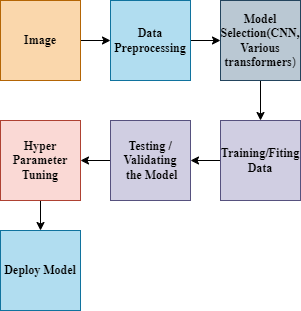
[6] The authors of this work Hasin Shahed Shad et al. employed the CNN basic architecture and pre-trained the model using several DenseNet and ResNet iterations. Data was supplied into the model and the matching output was produced. The Flickr dataset had 70,000 genuine faces and one million phony faces, and the images were downsized to 256 pixels and combined. The architecture achieved an accuracy of 81.6% with ResNet50, which is the highest.

[7] The study utilized the InceptionNet Convolutional Neural Network (CNN) algorithm to detect deep fakes. Different types of transitions in real images were used for the test, with parameters including the number of key points in the images, comparison rate and performance time required for each algorithm. The results showed that the algorithm had an overall accuracy of 93% for the DFDC dataset.

**Proposed work:**   
The proposed conceptual framework for deepfake detection starts with the collection of a diverse dataset containing both authentic and manipulated images from reliable sources. Following data gathering, an intricate preprocessing phase unfolds, involving tasks such as image resizing and normalization, along with facial landmarks extraction to capture essential facial features. Through tokenization and embedding techniques, images are formatted to suit convolutional neural network (CNN) and pretrained transformer architectures. The dataset is intelligently divided into training and testing subsets, ensuring a balanced representation of genuine and manipulated samples in both categories. A hybrid model is crafted, integrating CNNs and various pretrained transformers to effectively capture spatial and temporal dependencies within the images. Trained to discern subtle visual cues indicative of manipulated content, the model undergoes rigorous evaluation using the testing dataset, with metrics like accuracy, precision, recall, and F1 score computed for comprehensive assessment. Comparative analysis spans multiple CNN architectures and pretrained transformer models, shedding light on the strengths and weaknesses of each approach. This exhaustive evaluation aids in the selection of the most robust and accurate deepfake detection model, poised for effective deployment in real-world scenarios and enhancing the ongoing efforts to combat the challenges posed by manipulated images.

**Detailed design:**

**Architecture Diagram**



# Methodology and Module Description:

1. **Data Collection:**

We plan to gather a diverse dataset comprising both authentic and manipulated images from reputable sources, ensuring a broad representation of potential manipulation techniques.

1. **Data Preprocessing:**

We plan to resize and normalize images to a standardized format suitable for deep learning models. Facial landmarks extraction will be conducted to capture crucial facial features. We'll prepare images for processing by a convolutional neural network (CNN) and pretrained transformer architectures.

1. **Dataset Splitting:**

We intend to intelligently partition the dataset into training and testing subsets, maintaining a balance between authentic and manipulated samples in both categories.

1. **Model Architecture Design:**

Our plan is to craft a hybrid model that integrates both CNNs and various pretrained transformers to effectively capture spatial and temporal dependencies within the images.

1. **Training:**

We will train the model using the training dataset to enable it to discern subtle visual cues indicative of manipulated content.

1. **Evaluation Metrics:**

We'll assess the model's performance using the testing dataset. Metrics such as accuracy will be computed to comprehensively evaluate the model's efficacy in detecting manipulated images.

1. **Hyperparameter Tuning:**

We will perform hyperparameter tuning to optimize the model's configuration, exploring different combinations to enhance performance

**1.**

**2.**

**3**

# Signature (Students) Signature (Guide)

# 