**Mini Project Report on**



**Deepfake Detection Using Deep Learning**



**Submitted in partial fulfillment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name**  **University Roll No.**

**Aditya Semwal 2019528**

***Under the Mentorship of***

**Mr. Arnav Kotiyal**

**Assistant Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**July-2024**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“ Deepfake Detection”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Mr. Arnav Kotiyal**, **Assistant Professor** Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

**Name** **University Roll no.**

**Aditya Semwal 2019528**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | **1-2** |
| Chapter 2 | Literature Survey | **3-4** |
| Chapter 3 | Methodology | **5-10** |
| Chapter 4 | Result and Discussion | **11-12** |
| Chapter 5 | Conclusion and Future Work | **13-14** |
|  | References |  |

**Chapter 1**

**Introduction**

**1.1. Introduction**

Deepfakes are artificial media that use artificial intelligence technology to replace someone’s likeness with someone else’s in an existing image Deepfakes pose serious risks, such as disinformation, identity theft, privacy violated, although used in the creative and entertainment industry becomes.



**Fig. 1.1 Rise of the Deepfakes**

* 1. **Problem Statement**

The emergence of deepfake has created disturbing situations. These include the dissemination of misleading information, including the use of in-depth videos for propaganda or publicity, leading to social dissatisfaction and political unrest while deepfakes can be used to hack people or manipulate personal data, so Identity theft is another serious problem. Unauthorized use of human likeness violates privacy and can cause emotional pain and loss of reputation. Deepfakes are artificial media that use artificial intelligence technology to replace someone’s likeness with someone else’s in an existing image Deepfakes pose serious risks, such as disinformation, identity theft, privacy violated, although used in the creative and entertainment industry s becomes.

* 1. **Objective**

This project aims to develop a reliable and accurate deep lie detection algorithm using state-of-the-art deep learning techniques. The system must be able to recognize and recall authentic and manipulated features with high accuracy, to be effective in in-depth manufacturing and quality processes.

* 1. **Roadmap**

The design and needs assessment phase of the Deep Fake Identification Project will also include a literature review and interpretation of the project design and planning to ensure quality and diversity and we will then collect and preliminarily develop an extensive dataset of genuine fakes . Once we have tested and selected the best deep learning models, we will train and fine tune these models, using robustness testing against multiple deep learning methods and cross-validation. We will build deployment plans, user interfaces, and APIs to ensure scalability and efficiency after obtaining acceptable performance metrics. This will be followed by detailed reports and documents, such as user guides, and finally a project report. Finally, keeping an eye on new possibilities for research and development, we will regularly monitor and update the system based on usage and emerging deep threats.

**Chapter 2**

**Literature Survey**

A literature survey on deepfake detection using deep learning reveals a growing body of research addressing the challenge of detecting manipulated media. Here are some key insights and trends from recent studies:

1. **Architectures and Techniques**:
   * **Convolutional Neural Networks (CNNs)**: CNNs are commonly used due to their effectiveness in image analysis tasks. Researchers have explored various architectures, including deep and shallow networks, to optimize performance in detecting subtle visual cues indicative of deepfakes.
   * **Generative Adversarial Networks (GANs)**: GANs have been employed not only for generating deepfakes but also for detecting them. Adversarial training involves training a discriminator network to distinguish between real and fake images, leveraging the same principles used in their creation.
2. **Datasets**:
   * The availability and quality of datasets play a crucial role in training and evaluating deepfake detection models. Datasets like FaceForensics++ and Celeb-DF have been widely used, providing diverse examples of both real and manipulated videos.
3. **Feature Extraction**:
   * Studies have focused on extracting robust features from images or videos to distinguish between real and manipulated content. This includes spatial and temporal patterns, inconsistencies in facial expressions, blinking, and unnatural artifacts.
4. **Multimodal Approaches**:
   * Recent research explores combining information from multiple modalities such as audio, visual, and textual cues to improve detection accuracy. These approaches aim to exploit inconsistencies across different modalities that may be less apparent when analyzed individually.
5. **Adversarial Attacks and Defenses**:
   * As deepfake techniques evolve, researchers are investigating adversarial attacks specifically designed to evade detection systems. Countermeasures include adversarial training, where models are trained against these attacks to improve robustness.
6. **Real-Time and Scalability**:
   * Efforts are underway to develop real-time deepfake detection systems suitable for deployment in online platforms and social media networks. Scalability remains a challenge due to the computational demands of deep learning models.
7. **Ethical and Legal Implications**:
   * There is increasing awareness of the ethical and legal implications of deepfake technology. Research includes considerations of privacy, consent, and the potential for misuse, driving discussions on regulation and policy development.
8. **Benchmarking and Evaluation Metrics**:
   * Establishing standardized benchmarks and evaluation metrics is crucial for comparing the performance of different deepfake detection models. Metrics such as precision, recall, accuracy, and area under the receiver operating characteristic curve (AUC-ROC) are commonly used.

**Chapter 3**

**Methodology**

**3.1 Outline of the project**

The idea of machine learning is to let the algorithm analyze and learn by itself the best parameter from the data set to make good predictions. There are many different approaches to training a model to detect Deep Fakes but, in our case, we use TensorFlow, NumPy, Matplotlib.

**3.2 Projects Functions**

**3.2.1 Initialization:**

Import necessary libraries for deep learning, data processing, and visualization.

**import tensorflow as tf**

**import os**

**import numpy as np**

**from matplotlib import pyplot as plt**

**from tensorflow.keras.models import Sequential**

**from tensorflow.keras.layers import Conv2D,MaxPooling2D, Dense, Flatten, Dropout**

**3.2.2 Data Load and Data Visualization:**

Loading Data and Data Visualization.

**data = tf.keras.utils.image\_dataset\_from\_directory('train')**

**data\_iterator = data.as\_numpy\_iterator()**

**batch = data\_iterator.next()**

**fig, ax = plt.subplots(ncols=4, figsize=(20, 40))**

**for idx, img in enumerate(batch[0][:4]):**

**ax[idx].imshow(img.astype(int)) ax[idx].title.set\_text(batch[1][idx])**

**3.2.3 Splitting Data Set:**

Splitting Dataset into Train, Test, and Validation

**train\_size = int(len(data) \* .7)**

**val\_size = int(len(data) \* .2)**

**test\_size = int(len(data) \* .1)**

**train = data.take(train\_size)**

**val = data.skip(train\_size).take(val\_size)**

**test = data.skip(train\_size + val\_size).take(test\_size)**

* **train\_size:** The number of samples in the training set, calculated as 70% of the total dataset size.
* **val\_size:** The number of samples in the validation set, calculated as 20% of the total dataset size.
* **test\_size:** The number of samples in the test set, calculated as 10% of the total dataset size.

**3.2.4 Building model:**

Builds a convolutional neural network (CNN) model using TensorFlow and Keras for binary classification (e.g., distinguishing between real and deepfake images).

**model = Sequential()**

**model.add(Conv2D(16, (3, 3), 1, activation='relu', input\_shape=(256, 256, 3)))**

**model.add(MaxPooling2D())**

**model.add(Conv2D(32, (3, 3), 1, activation='relu'))**

**model.add(MaxPooling2D())**

**model.add(Conv2D(16, (3, 3), 1, activation='relu'))**

**model.add(MaxPooling2D())**

**model.add(Flatten())**

**model.add(Dense(256, activation='relu'))**

**model.add(Dense(1, activation='sigmoid'))**

**model.compile('adam', loss=tf.losses.BinaryCrossentropy(), metrics=['accuracy'])**

**model.summary()**

**3.2.5 Training and Evaluate model performance:**

**pre = Precision()**

**re = Recall()**

**acc = BinaryAccuracy()**

**for batch in test.as\_numpy\_iterator():**

**X, y = batch yhat = model.predict(X)**

**pre.update\_state(y, yhat)**

**re.update\_state(y, yhat)**

**acc.update\_state(y, yhat)**

**print(f'Precision: {pre.result().numpy() \* 100}, Recall: {re.result().numpy() \* 100}, Accuracy: {acc.result().numpy() \* 100}')**

**3.2.6 Testing the Model on a New Image:**

**import cv2**

**img = cv2.imread('test/modideep.jpeg')**

**plt.imshow(cv2.cvtColor(img, cv2.COLOR\_BGR2RGB))**

**plt.show()**

**resize = tf.image.resize(img, (256, 256))**

**plt.imshow(resize.numpy().astype(int))**

**plt.show()**

**yhat = model.predict(np.expand\_dims(resize / 255, 0))**

**if yhat > 0.5:**

**print('Predicted class is original')**

**else:**

**print('Predicted class is deepfake')**

* Loads an image, resizes it, and makes a prediction using the trained model.
* Displays the image and the resized version.
* Predicts whether the image is original or a deepfake and prints the result.

**3.2.7 Saving and Loading the Model:**

**from tensorflow.keras.models import load\_model**

**model.save(os.path.join('models', 'deepfakemodel.h5'))**

**new\_model = load\_model(os.path.join('models', 'deepfakemodel.h5'))**

**yhatnew = new\_model.predict(np.expand\_dims(resize / 255, 0))**

**if yhatnew > 0.5:**

**print('Predicted class is original')**

**else:**

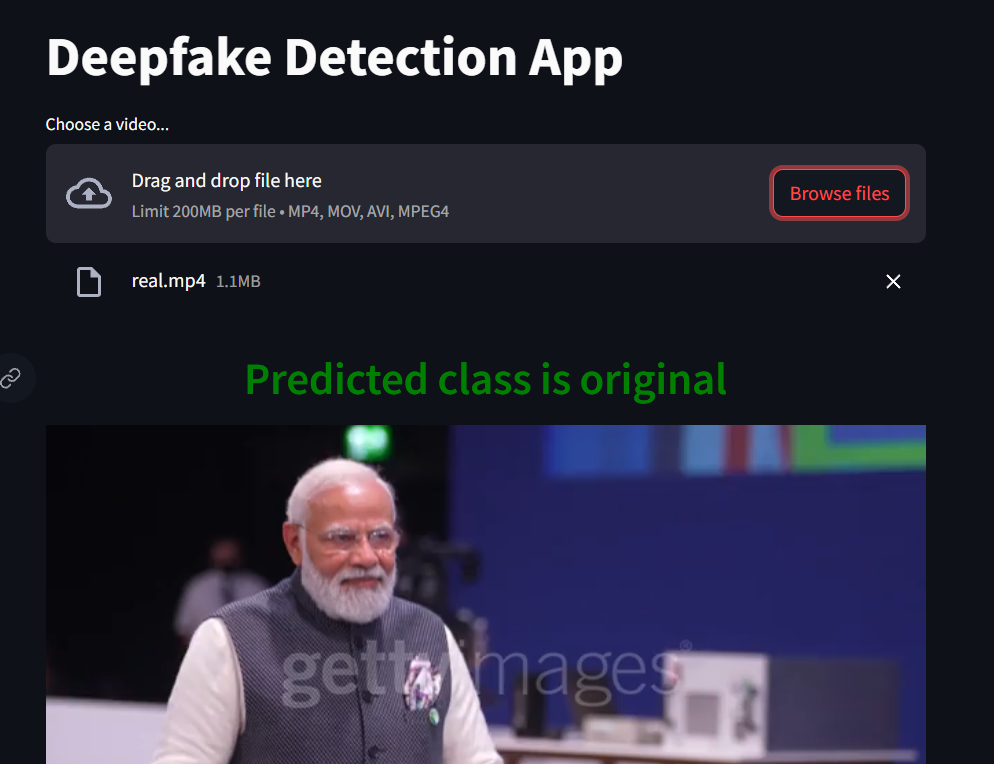
**print('Predicted class is deepfake')**

* Saves the trained model to a file.
* Loads the model from the file and makes a prediction to verify that the saved model works correctly.

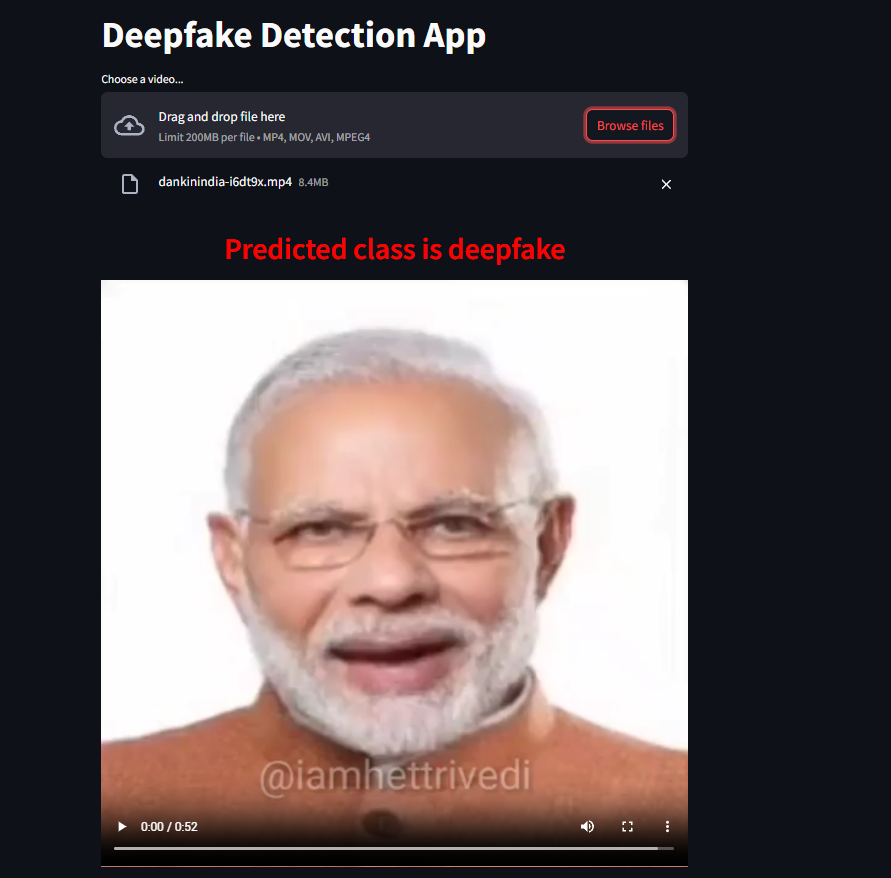
**3.2.8 Final Output:**

Streamlit is an open-source Python library that allows data scientists and developers to create web applications for data visualization and interactive data exploration with minimal effort. It is designed to turn data scripts into shareable web applications quickly and easily.

With Streamlit, you can transform Python scripts that produce charts, plots, dataframes, and other visualizations into web apps with just a few lines of code. It eliminates the need for HTML, CSS, or JavaScript, making it accessible to those with little to no web development experience



**Fig. 3.1 Deep Fake detection using Deep Learning**

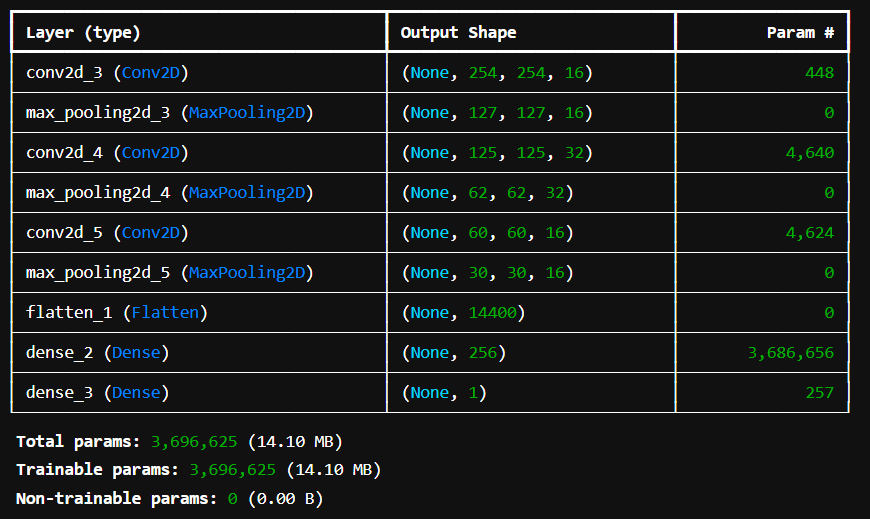
****

**Fig. 3.2 Deep Fake detection using Deep Learning**

**Chapter 4**

**Result and Discussion**

**4.1 CNN model Summary:**

****

**Fig. 4.1 CNN model Summary:**

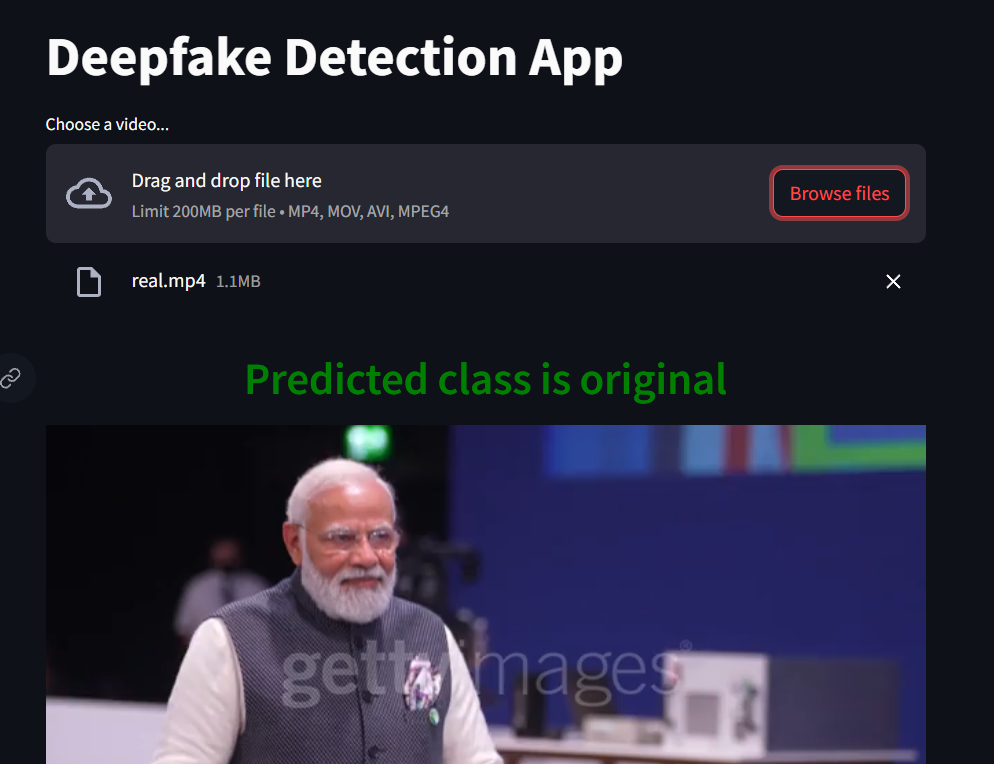
**4.2 Accuracy Graph:**

**A graph with orange and blue lines

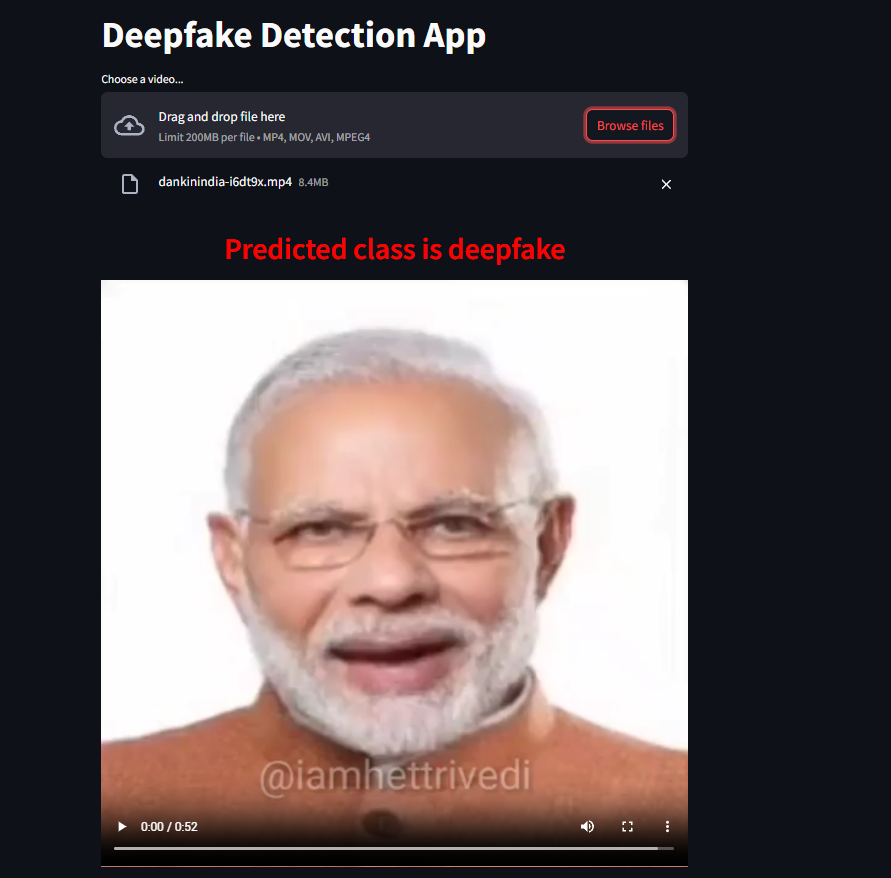
Description automatically generated**

**Fig. 4.1 Accuracy Graph**

**4.2.3 Final Output:**



**Fig. 3.1 Deep Fake detection using Deep Learning**

****

**Fig. 3.2 Deep Fake detection using Deep Learning**

**Chapter 5**

**Conclusion and Future Work**

**5.1 Conclusion:**

In this study, we developed a convolutional neural network (CNN) for high-depth fake image detection. To ensure that the research decisions were reliable, the sample was divided into training, validation, and tests. Model performance can be effectively monitored thanks to Tensor Board logging and visualization capabilities, and final analysis with untested data revealed that model reliability can meet This effort to address the growing problem of deep fake detection contributes to the preservation of trust in digital media through the apparent potential for deep learning strategies.

**5.2 Future scope**

Deep learning-based deepfake detection has interesting directions to pursue in the future. First, the ability of the model to detect deeper features can be enhanced by adding more complex structures, such as perspective-based networks or transducers, and extending the dataset to include deepfake a wide variety of methods and sources have been added. Algorithms can be built and embedded in online content, social media networks, video conferencing applications and more, those that use text, audio and video data to create comprehensive and accurate and deep lie detection systems love -Different options can be explored.

**5.3 Future Work**

To further improve the performance and applicability of the model, future research in this area could focus on critical areas. The use of more complex neural network structures such as transformers or anti-generational networks (GANs) is one way to increase the generalizability of the model even further by incorporating higher and more diverse deepfake patterns into the dataset ho that can help reveal the more subtle parts on the deeper level. With the help of real-time deepfake detection and integration of this system with applications such as social media platforms, video conferencing software, the fight against misinformation and identity theft can be facilitated Finally, deep visual more recognition by examining multiple channels -A complete should be provided with a channel that includes audio and video data

**References**

* Y. Patel et al., **"Deepfake Generation and Detection: Case Study and Challenges,"** in IEEE Access, vol. 11, pp. 143296-143323, 2023

**Doi: 10.1109/ACCESS.2023.3342107**.

<https://ieeexplore.ieee.org/abstract/document/10354308>

* D. Güera and E. J. Delp**, "Deepfake Video Detection Using Recurrent Neural Networks,"** 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6,

Doi: 10.1109/AVSS.2018.8639163.

<https://ieeexplore.ieee.org/abstract/document/8639163>

* H. A. Khalil and S. A. Maged, **"Deepfakes Creation and Detection Using Deep Learning,"** *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, Cairo, Egypt, 2021, pp. 1-4,

Doi: 10.1109/MIUCC52538.2021.9447642.

<https://ieeexplore.ieee.org/abstract/document/9447642>

* M. S. Rana, B. Murali and A. H. Sung, **"Deepfake Detection Using Machine Learning Algorithms,"** *2021 10th International Congress on Advanced Applied Informatics (IIAI-AAI)*, Niigata, Japan, 2021, pp. 458-463,

Doi: 10.1109/IIAI-AAI53430.2021.00079.

<https://ieeexplore.ieee.org/abstract/document/9790940>

* A. Malik, M. Kuribayashi, S. M. Abdullahi and A. N. Khan, **"DeepFake Detection for Human Face Images and Videos: A Survey,"** in *IEEE Access*, vol. 10, pp. 18757-18775, 2022,

doi: 10.1109/ACCESS.2022.3151186.

<https://ieeexplore.ieee.org/abstract/document/9712265>