**Machine Learning Classification and Feature Selection for Efficient Fake Face**

**Synthesized Video Identification**

A project report submitted to **Indian Institute of Engineering Science and Technology**

in partial fulfillment for the award of the degree of

### Bachelor of Technology in

**Information Technology**

### By

**Arijit Dalui (Enrollment Number- 510819100)**

**Soumik Mukhopadhyay (Enrollment Number- 510819102)**

**Ritaban Bhattacharya (Enrollment Number- 510819101)**

**Under the supervision of Prof. Ruchira Naskar**



### Department of Information Technology

**Indian Institute of Engineering Science and Technology, Shibpur May, 2023**

### DEPARTMENT OF INFORMATION TECHNOLOGY

**INDIAN INSTITUTE OF ENGINEERING SCIENCE AND TECHNOLOGY, SHIBPUR, HOWRAH - 711103, INDIA.**

**CERTIFICATE**

This is to certify that the project report entitled **“Machine Learning Classification and Feature Selection for Efficient Fake Face Synthesized Video Identification”** submitted by Arijit Dalui (Roll No. 510819100), Soumik Mukhopadhyay (Roll No. 510819102), Ritaban Bhattacharya (Roll No. 510819100) to Indian Institute of Engineering Science and Technology towards partial fulfillment of requirements for the award of the degree of Bachelor of Technology in Information Technology is a record of Bonafede work carried out by them under my supervision. This dissertation, in my opinion, is worthy of consideration for the purpose for which it is submitted and it fulfills the requirements of the regulations of this Institute. The results incorporated in this dissertation are original and have not been submitted to any University or Institute for the award of any Degree or Diploma.

Dr. Ruchira Naskar Professor

Department of Information Technology

Indian Institute of Engineering Science and Technology, Shibpur Howrah 711103, West Bengal, India (May 2023)

Dr Prasun Ghosal Head of Department

Department of Information Technology

Indian Institute of Engineering Science and Technology, Shibpur

**ACKNOWLEDGEMENT**

We are grateful that we had the opportunity to experience an academic journey at IIEST, Shibpur that lasted four years. These four years have been an eye-opening experience for us and we consider ourselves fortunate to have learnt life-changing lessons which shall always be useful for us.

We are indebted to Professor Ruchira Naskar of the Department of Information Technology, IIEST, Shibpur for guiding us through the minute aspects of research in the domain of surface generation and plane detection. We were able to explore some interesting problems and implement certain niche concepts under his supervision.

We are also grateful to all the faculties of the Department of Information Technology, IIEST, Shibpur. The knowledge imparted through the courses as part of the B. Tech programme will always be useful in preparing us for future technological challenges.

We are also grateful to Arnab Kumar Das, PhD scholar of the Department of Information Technology, IIEST, Shibpur, for his valuable guidance and precious inputs that were crucial to the development of the project in its entirety.

Finally, we would like to thank all our friends and fellow students for enriching our experience in IIEST, Shibpur.

**Arijit Dalui (Examination Roll No.): 510819100**

**Ritaban Bhattacharya (Examination Roll No.): 510819101**

**Soumik Mukhopadhyay (Examination Roll No.): 510819102**

# Table of contents

# *Abstract*

# *Introduction*

# Abstract

Recent developments in deep learning have enabled media synthesis and manipulation to reach previously unheard-of degrees of realism. The widespread use of deepfake technology to produce fake media has the potential to have a large negative influence on the reliability of multimedia data, including videos, photos, and audio recordings. The frequency domain spectrum of deepfake videos is examined in this paper, and a unique method is suggested that uses high-frequency Discrete Cosine Transform (DCT) coefficients that are retrieved from videos as a recognizable fingerprint. We create a robust model that can identify between genuine videos and their deepfakes by examining the variance of differences between subsequent video frames' DCT coefficients. We extend our work by not only using a binary classification model to classify real and fake videos but also, we go beyond classification in our research and look into the sources of bogus videos. We offer a thorough study of deepfake media by integrating our feature-based classification model with a look into the origin of the fake videos. Our comprehension of deepfake techniques and their possible impact on multimedia integrity is improved by this all-encompassing approach. We ran tests on the publicly accessible dataset Face Forensics++ to determine how well our suggested model worked. Amazingly, our model produced excellent results, with a multilevel classification accuracy of 98.56% and a binary classification accuracy of 98.91%, respectively.

**Introduction**

In 2017, the Deepfake phenomenon emerged notably as a consequence of artificial intelligence (AI) methods for creating synthetic media and their distribution on the internet. These methods included tweaking (or generating) audiovisual content using ad-hoc machine learning generative models, such as the Generative Adversarial Network (GAN) [1]. On television as well as internet, we found videos and images of high-profile individuals that could initially seem real, but could really be the output of an AI process that produces incredibly accurate fake media. Such fraudulent tapes are prone to be misused in personal defamation lawsuits, crimes involving child pornography, and deceiving court proceedings and the general public. The horrific consequences of deepfakes were witnessed when a Reddit user of the same name posted doctored porn clips on the site [2]. In the clips, renowned individuals like Gal Gadot, Taylor Swift, Scarlett Johansson, and others had their visages switched with those of porn performers. Recent reports claim that deepfake films were employed as a political tool during elections. In order to stop the dissemination of false information in the run-up to the 2020 US elections, Facebook banned Deepfake and its synthetic videos [3]. The usage of deepfake video in a news story as opposed to actual footage demonstrated how this affects the behavior of the general public, according to Shin et al. [4]. The abusive, misleading, and fraudulent use of this technology has increased hazards rather than possibilities for stakeholders. We can invariably see the necessity for an immediate and trustworthy solution to confront the Deepfake technology, considering that anybody could be its next target.

In this paper, an explainable method of Deepfake detection, a DCT-FADE (DCT Frequency-based Authentication for Deepfake Evaluation) technique is proposed based on the analysis of Discrete Cosine Transform (DCT) [5] video coefficients. It was demonstrated through experiments using Deepfake frames of human faces that the spectral frequencies

contain an effective generative process signature. The main contributions of this research are the followings:

* A novel approach for detecting fake faces is developed using the Discrete Cosine Transform, which produces excellent generalization results.
* This “explainable” technique allows us to identify deepfake-generated anomalous frequencies that significantly deviate from the frequencies of an authentic image.

In addition to that we present a machine-learning classification model for the blind

detection of Deepfake videos. For this purpose, we propose a multimodal detection

technique by combining an efficient set of prefabricated Histogram of

Oriented Gradients (HOG) based features, and a set of features automatically

learned by Convolutional Neural Networks (CNNs). Our contributions in this

paper may be summarized as follows:

* Development of a multimodal Deep-Fake detection scheme combining both deep learning and traditional computer vision techniques, hence optimizing performance efficiency.
* We propose a multi-stage classification approach, for identifying a Deepfake video, followed by identifying its source network using the above set of features.