DECODING DEEPFAKES: A DEEP LEARNING APPROACH TO UNVEILING SYNTHETIC MEDIA

A Major Project – II Report
Submitted in Partial Fulfilment of Requirements for the
Award of the Degree of

MASTER OF TECHNOLOGY

in
Information Systems
by

Sarthak Kulkarni
(Roll No. 2K22/ISY/16)

Under the Supervision of
Prof. Dinesh Kumar Vishwakarma & Dr. Virender Ranga

Department of Information Technology

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi – 110042, India

May, 2024

DECODING DEEPFAKES: A DEEP LEARNING APPROACH TO UNVEILING SYNTHETIC MEDIA

A Major Project – II Report
Submitted in Partial Fulfilment of Requirements for the
Award of the Degree of

MASTER OF TECHNOLOGY

in
Information Systems
by

Sarthak Kulkarni
(Roll No. 2K22/ISY/16)

Under the Supervision of
Prof. Dinesh Kumar Vishwakarma & Dr. Virender Ranga

Department of Information Technology

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi – 110042, India

May, 2024

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I Sarthak Kulkarni hereby certify that the work which is being presented in the thesis entitled "**Decoding Deepfakes: A Deep Learning Approach to Unveiling Synthetic Media**" in partial fulfilment of the requirements for the award of the Degree of Master of Technology, submitted in the Department of Information Technology, Delhi Technological University is an authentic record of my own work carried out during the period from 2022 to 2024 under the supervision of Prof. Dinesh Kumar Vishwakarma & Dr. Virender Ranga.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

Place: Delhi **SARTHAK KULKARNI**

Date : 30/05/2024 **2K22/ISY/16**

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that **Sarthak Kulkarni** (2K22/ISY/16) has carried out their search work presented in this thesis entitled “**Decoding Deepfakes: A Deep Learning Approach to Unveiling Synthetic Media**” for the award of **Master of Technology** from Department of Information Technology, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Signature	Signature
Prof. Dinesh Kumar Vishwakarma	Dr. Virender Ranga
SUPERVISOR	**CO-SUPERVISOR**
Head Of Department	Associate Professor
Department Of Information Technology	Department Of Information Technology
Delhi Technological University	Delhi Technological University

Place : New Delhi

Date : 30/05/2024

ACKNOWLEDGEMENTS

I am thankful to my supervisor Prof. Dinesh Kumar Vishwakarma , Head Of Department of Information Technology and Dr. Virender Ranga, Associate Professor at Delhi Technological University, Delhi and all the department's faculty members. They all assisted me whenever I needed any help. This work would not be possible without their direction and assistance. I am also grateful to the IT department for providing the various resources required to complete this work.

SARTHAK KULKARNI

2K22/ISY/16

ABSTRACT

Identifying deepfakes is essential to combat misinformation, protect personal and national security, and prevent financial fraud and reputational harm. The increasing sophistication of deepfake technology heightens these threats, requiring advanced detection methods. Convolutional Neural Networks (CNNs) are good models for a classification task within deep learning. "The State of Deepfakes" from Deeptrace registered over 14,000 deep fake videos, the majority of which are not pornographic, marking a high potential to cause political and social disruption; as AI continues to advance deep fakes grow far more realistic—and more accessible to make. It is simply a working mechanism that is constant on detection and delays involving the technological, legal, and educational components. Underlining that the threat is newly emerging and changing all the time, the report puts stress on the fact that such future risks as fraud and misinformation, among others, arise from such types of malicious activities as deepfakes, which call for comprehensive action.

Alternatively, can the development of the Vision Transformer (ViTs) help inspire turning the image processing problem into an approach where an image is regarded as a string of patches, in other words, self-attention to global relations? Whereas the current architecture is entirely based on Convolutional Neural Networks, it uses a sequence of local receptive fields and a building-up process hierarchically. This is believed to provide better ViTs with the ability to learn relationships at long ranges and information from contexts much more effectively, giving better performances for a more enormous scope of image identification. They are more robust, flexible, and scalable than CNNs and generally have a much more comprehensive set of visual inputs..

The new ViT model will apply deepfakes determination. The model's performance will be judged on performance metrics: accuracy, precision, recall, and F1-score, thereby making it more feasible. Through these metrics, we can determine how good the ViT model can be in differentiating an actual image from a deepfake.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

CNN	Convolutional Neural Network
ViT	Vision Transformer
DFDC	Deepfake Detection Challange
AI	Artificial Intelligence
VAE	Variational Autoencoders
NLP	Natural Language Processing
ReLU	Rectified Linear Unit
SOTA	State-of-the-Art
GAN	Generative Adversarial Network

CHAPTER 1

INTRODUCTION

In the days of generative synthetic media and AI, a chilling and remarkable epiphany ensues: the belief taken for granted—that audio and video can be given and received in good faith as a representation of reality—no longer holds. Indeed, lots of different advanced tools make it fast and easy to produce fake information. Lots of positive and, on the other hand, quite probably negative impacts lie in the power of such content production. This exposes a crying need to work out this problem in the growth of synthetic media and generative AI. Deepfakes are completely false media created by replacing the appearance of an existing person, either from an image or video, with another person's, making it look genuine. More specifically, these manipulations yield plausible changes that are indiscernible from authentic materials due to the use of cutting-edge machine learning algorithms, in particular Generative Adversarial Networks (GANs).

1.1 Types of Deepfakes

Currently Face Manipulation is being done in two ways :

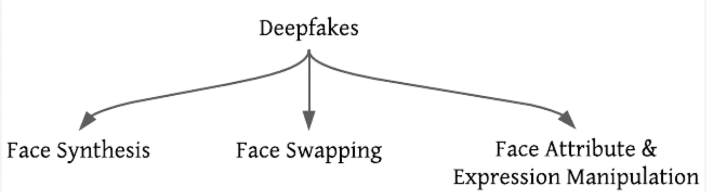


Fig 1.1: Types of Deepfakes

1.1.1 Face Synthesis

Face synthesis yields new face creations—works of models as complex as GANs. So, this way, using the most advanced deep learning algorithms, one can generate pictures or videos in which the face looks similar to a natural human face. That draws on and creates much in the process a very subtle dance between the generator network of a GAN that is responsible for producing synthetic faces and the discriminator network that evaluates how authentic fake images look. Such adversarial training helps a GAN iteratively refine the created synthetic faces with an exceptional level of fidelity, much like humans.

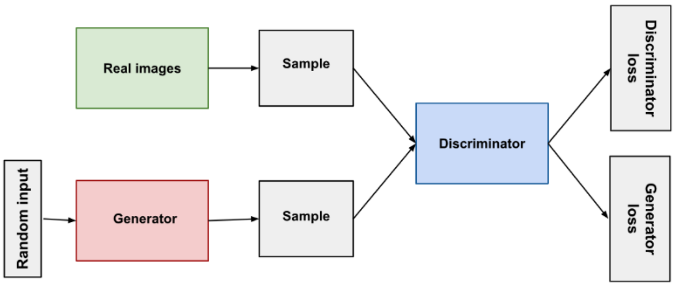


Fig 1.2: GAN Architechture[6]

These synthetic faces can be put to use in a wide array of purposes devised by researchers within the deep-fake technology framework. For instance, they can underlie the basis of populating the development of entirely fictitious online personas to execute everything from fraud to criminal activities. Synthetic faces could also be used as representations in manipulated media to create disinformation or threaten someone. It is used for the same technology but can be promising for various applications to protect individuals, e.g., protecting anonymity in sensitive cases like witness protection or privacy-preserving scientific studies[1]. Face synthesis is the process of creating new faces that do not even exist but look realistic using generative adversarial networks. It is trendy for the same, and currently in use with StyleGAN, for exercising control over multiple facial properties, namely pose, identity, freckles, hair, among other things. This technology has been applied in the design of synthetic personas and the improvement of virtual environments[3].

1.1.2 Face Swapping

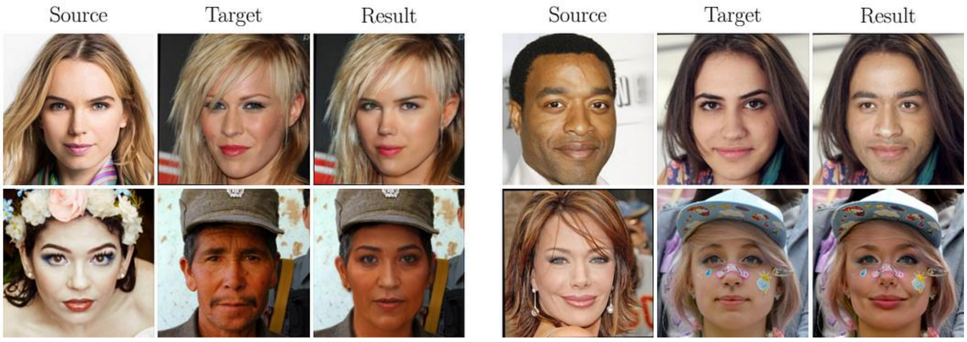


Fig 1.3: Types of Deepfakes[5]

This is called face swapping, a technique in deepfaking where a face in a photo or video is exchanged with another person's face. The system for face detection and alignment before the blending is photo-realistically capable of incorporating the target by the features of the source face using deep learning algorithms. Overlays could be simple to quite complex, involving even the correction of lighting, pose, and expression. Besides entertainment and art applications, there's also a more practical issue: privacy concerns and information dissemination based on its highly convincing fakes.

1.1.3 Face Attribute and Expression Manipulation

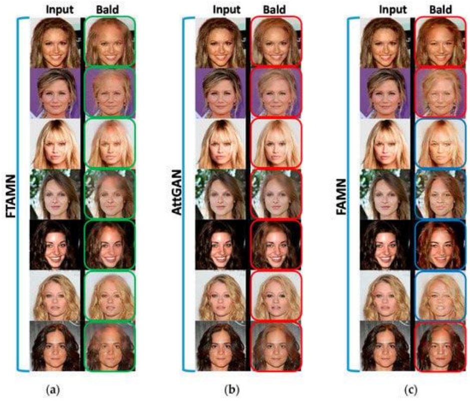


Fig 1.3: Face Expression Manipulation[7]

Face attribute and expression modification is the process of modifying particular facial traits or expressions in photos or videos by applying sophisticated deep learning techniques. This method produces incredibly lifelike alterations by changing features including age, gender, hair colour, and facial expressions. For this, Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) are frequently utilised.

1.2 Why is Deepfake identification critically important?

In just a year and a half, the prevalence of deep fake videos has skyrocketed. In June 2019, IBM detected a mere 3,000 of these manipulated videos. By January 2020, that number had skyrocketed to 100,000, and as of March 2020, there are now over one million circulating on the internet. Surprisingly, according to research by Deeptrace, this trend began in December 2018 with 15,000 deep fake videos, which then soared to 558,000 by June 2019, and has now reached a staggering one million in just over a year. Unfortunately, the vast majority of these fraudulent videos are created for malicious purposes, particularly non-consensual pornography. In fact, a whopping 96% of the deep fake videos studied by Deeptrace featured women in this exploitative and violating manner[2].

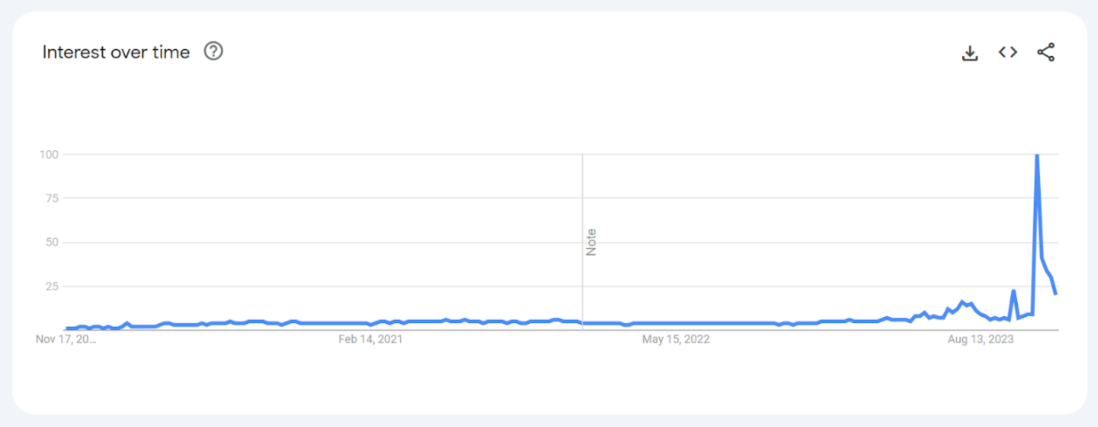


Fig 1.4: Popularity of Deepfakes according to Google Trends

1.2.1 Case Study of threats of Deepfakes

1. The destabilizing political impact of Deepfakes

   In late 2018, there was considerable speculation regarding the well-being of Gabonese President Ali Bongo, who had been absent from public life for several months. To dispel these rumors, the government released a video featuring Bongo delivering a traditional New Year's address. However, the president's unusual appearance in the video prompted many on social media, including Gabonese politician Bruno Moubamba, to assert that it was a deepfake. This fueled suspicions that the government was concealing Bongo's poor health or possible demise. About a week after the video's release, amidst escalating unrest, members of Gabon's military initiated an attempted coup against the government. In a video announcing the coup, the military cited the video's peculiarities as evidence of something amiss with the President[2].

2. Nancy Pelosi: Manipulated voice audio

   A recent high profile shallow fake involved a manipulated video of US Speaker of the House and Democrat Congresswoman Nancy Pelosi.11 In the video that was shared on May 23rd 2019, Pelosi’s speech had been slowed down, making it sound like she was slurring her words. The edited version of the video went viral on social media and was retweeted by the official Twitter account of US President Trump, receiving over 6.3m views as of July 31st 2019.12 On a popular Facebook page, the video received over 2.2m views in the 48 hours following its initial upload, with commenters calling Pelosi “drunk” and a “babbling mess”[2].

Fig 1.5: a) Ali Bongo Deepfake, b) Nancy Pelosi: Manipulated voice audio[2]

1.3 Convolutional Neural Network & Transforms for Deepfake Detection

One such image-based task in which CNN architectures have outperformed all others up to this date is deepfake detection. The CNN models are the archetypical base necessary for next-generation deepfake detectors since it has more salient characteristics in extracting the features hierarchically from the data. CNNs have very discriminating powers and can detect slight differences between false and true images because they spot different styles of patterns and anomalies to varying levels of detail through the stacks of convolutional filters. Hierarchical processing of the data enables the CNNs to pick up the fine details and context, which belongs to the needs of the hour of successful deepfake detection. Despite possessing great image-related features, CNN has some weaknesses. The primary task of CNNs is performing work on local features; they rely on the convolutional filters and receptive fields, whereas the latter sometimes causes the former to overlook global context and long-range dependencies provided in the images. In addition, CNNs naturally face problems at high resolutions and enormous datasets, which means they require much more processing power. This will not only help extract the features of the image but also, as the network grows deeper, this will cause a more loss of spatial information. Finally, large volumes of labeled data are another limitation for other applications because CNNs usually require them to train for high-accuracy performance.

Vision Transformers (ViTs) address some specific limitations of Convolutional Neural Networks (CNNs). They ignore local features more than Convolutional Neural Networks do. That said, they come with some self-attention mechanisms to attend and develop dependencies globally on an image to take into account all the contexts. This makes it easier for ViTs to understand complex patterns and long-range relationships. They process images as patches and then transform them into a word sequence in a phrase. This way, adaptability to varying resolutions in the image could take place without dramatically increasing the computational complexity and keeping spatial information accordingly. The ViTs, however, have a lot less inductive bias relative to CNNs, making them general for more varied visual data, with fewer heuristics to invent on their own. The performance of ViTs relative to other methods looks nice on the datasets with relatively little data labeled yet pre-trained on giant data collections and optimized for some target task. This makes ViT generalize successfully for all sorts of visual tasks thanks to pretraining, capturing the possibility of large-scale learning. Therefore, ViTs follow as an interesting alternative to CNNs, filling some of these gaps in a very specific manner for those tasks demanding holistic understanding of context that spans the entire image.

CHAPTER 2

LITERATURE REVIEW

The word "deepfake" dates back to 2017 coined by a reddit user and spread on the internet after that very quickly. Since then, much research and development has been made on developing highly potent models for effective deepfake detection.

Deepfakes have indeed been created using GANs. In their paper, "Deep Fakes using Generative Adversarial Networks (GANs)," Smith et al. went deep into the artificial media cooked up by GANs. They used the term to connote that one is making a highly realistic but fake image, video, or audio using GANs, which is often indistinguishable from natural media. Apparently, in 2017, for the first time, when someone intended to produce a highly realistic but fake image, video, or audio that would be indistinguishable from the actual media, they used the term on a Reddit user. A sharp review of GAN architecture is provided by Smith et al., which involves a generator and the neural network in the composition of a noisy discriminator that, in turn, works against each other in a one-against-another manner up to the moment when the generator yields the ability to create such a content that the discriminator cannot detect. The present paper discusses a vast number of applications of deepfakes, from creative and funny to the malicious use of disinformation campaigns and invasions of privacy[3].

In the influential paper "Attention Is All You Need," by Vaswani et al. introduced a novel neural network architecture: the Transformer. In this architecture, self-attention mechanisms circumvent recurrence and convolutions entirely. This is an attention-based model that solves the long-range dependency problem, and it is characterized by much better parallelizability and short path length. Between dependencies, it trains at the beginning of time for very efficient training and success. In other words, the Transformer model operates under the mechanism of self-attention that somehow allows it to calculate some kind of representation of every single word in a sentence, given all other words in that sentence, this time weighted by its attention score. In this way, long-range dependencies can be learned along with more complex dependencies within data with current dependencies that were impossible with any of the earlier-proposed architectures. Vaswani et al. argues that a model trained on these tasks can reach the state-of-the-art performance on neural machine translation, actually outperforming all other models on the WMT-2014 English-to-German and English-to-French datasets. The Transformer also generalizes very well to many other NLP tasks because of its scalability and capability to model long-range dependencies. These generalizations, together, make the paper a real milestone in neural architectures that consequently set the grounds for a great deal of successive work both in natural language processing and way beyond. [9].

This paper presents to readers "Efficient Estimation of Word Representations in Vector Space" by Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Two classes of new model architecture—continuous bag of words and continuous skip-gram—are introduced. Huge data sets can use methods for computing word representations in a constant space of a neural network. Such models learn to predict context words given a specific target word with Skip-gram or predict the target word from its context words using CBOW, all by modeling semantical and syntactic regularities in their embedding space. Despite the results, the models showed near-equivalent performance to some other benchmarks in the field of word similarity estimation, and they tend to be computationally effective since high-quality word vectors are obtained in less than one day. The contributions transcend the effectiveness of the models also to economize computational cost, allowing for the use of larger datasets and more iterations of projects by researchers and practitioners. In addition, it evaluates the learned word embeddings, thus being in a better position to capture more of the semantics and syntactic-based word relationships. Overall, "Efficient Estimation of Word Representations in Vector Space" can be a forward step in the field of NLP research, and the conjectures it makes over novel solutions that may pump much power into some NLP applications are hopeful [16].

The paper "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale" by Dosovitskiy et al. presents a Vision Transformer model for image recognition. In the paper, the authors propose a novel model architecture for large-scale image recognition, borrowed from the classic applications of transformer architectures in processing natural language right into the domain of image recognition. On the other hand, the approach followed so far breaks the images into fixed-size patches that consist of 16×16 pixels, and the patches are considered the token sequences, which are like words in the text. That will be able to provide the ViT model with an ability to capture large-scale dependencies and global contexts within the image when the earlier conventional models of CNN used to focus mainly on local features. The main strength of all ViT models is that they are pre-trained with huge datasets, but there is the generalization ability innately. Thus, one acquires state-of-the-art performances over a comprehensive benchmark. It matches, but in most cases, it even outperforms what was traditionally provided by ViT, giving new state-of-the-art results on a set of essential tasks. This leads to a more efficient and scalable Transformer that can do away with most task-special modifications, with the plausibility of high resolution on complex visual input tasks. Notably, large-scale pretraining empowers the ViT model dramatically to enhance performance on tasks related to image recognition: accuracy and robustness. Hence, what it does is it places Transformers, to a great extent, as one solid and flexible option for CNNs, opening new possibilities in enhanced computer vision and image processing areas. The performance of the ViT model demonstrates the breakthrough value of the transformer-based architecture for changing the issue of image recognition at a larger scale. In addition, it does place a careful framework for the upcoming development of the field of visual data analysis[11].

The paper titled "Fused Swish-ReLU Efficient-Net Model for Deepfakes Detection" by Ilyas et al. presents a novel approach to detecting deepfake videos. The authors propose a fused activation function, combining Swish and ReLU, to enhance the efficiency of the Efficient-Net model for deepfake detection. The study was presented at the 9th International Conference on Automation, Robotics and Applications (ICARA) in 2023. By integrating Swish and ReLU activation functions, the proposed model aims to improve the detection accuracy of deepfake videos. The paper provides insights into the design and implementation of the fused activation function and evaluates its performance using experimental results. This research contributes to the ongoing efforts to develop more effective and efficient methods for detecting deepfake content, addressing concerns related to misinformation and digital manipulation [12].

In the paper "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Mingxing Tan et al. tackle a dramatically increasing computational cost that is associated with scaling out the performance improvement of models by a new methodology in scaling convolutional neural networks. The paper further argues that, in the going trend, the model should be bigger and more complex, hence a justified equilibrium between model complexity and computational efficiency. In this paper, we carefully investigate the scaling of network design and present a straightforward but highly effective compound scaling method applied to the scaling of depth, width, and resolution of our baseline network. Our approach should transfer to other scales of models, which is critical for demonstrating better SOTA performance while simultaneously being an order of magnitude smaller and faster than large models in efficiency. With our resulting EfficientNet, we reach state-of-the-art accuracy on 3 out of 5 highly competitive datasets while being an order of magnitude smaller and faster on these tasks. In other words, the compound scaling technique adopted by this paper completely reformed current neural architecture design practices and should spur the next wave of more efficient and scalable architectures. Therefore, "EfficientNet" has already become a seminal paper and presents an attractive alternative against the prevalent practice in the deep learning community [13].

In the paper "Video Face Manipulation Detection Through Ensemble of CNNs," these ensembles of Convolutional Neural Networks (CNNs) are once more implemented for video facial manipulation detection. When authors try to establish the regions of the face that are manipulated in the videos, it places them in very delicate conditions since the processes are further complicated by the different changes of lighting, various facial expressions, and light occlusion. A key novelty in this work is the ensemble strategy: it amalgamates the outputs of several CNNs, each taking a different view of the task. The good thing about these nets is that they are very efficient in discovering visual cues linked to the manipulated facial areas: irregular textures, temporal inconsistencies, and geometric distortion. The ensemble of FCNN can also help the hybrid model to efficaciously deal with the detection performance in many manipulation techniques over varied video contexts. The experimental characterization and evaluation under the proposed approach are realized in a comparative benchmark setting that presents results on its ability to detect different types of facial manipulations through the Deep-Face, FaceSwap, and Face2Face manipulations. Notably, the CNN ensemble outperforms single models and provides competitive results against state-of-the-art solutions. Therefore, this paper hands to the community a quite effectively armed general framework for detecting video face tampering so that the strength of CNN representation regarding cooperative techniques in ensemble learning could be exploited in the fight against the rise of manipulated content across online platforms and media [14].

In the paper "FaceForensics++: Learning to Detect Manipulated Facial Images," Rössler et al. develop an in-depth framework that has been claimed to be accurate in detecting manipulated facial images. This is a critical issue, somewhat like the question of whether there is a way to distinguish between genuine facial content and manipulated faces, the latter of which is on the rise very quickly through technology development in the digital era. They propose a novel deep-learning method focusing on analyzing and classifying intricacies in a facial image. The first important feature behind their methodology is the diversified dataset: FaceForensics++, which encompasses vast manipulative techniques such as FaceSwap, DeepFake, and neural texture synthesis. Most importantly, it becomes the backbone for training the proposed model, since it is explicitly included that the model will be able to distinguish complex features and patterns related to both natural and manipulated facial images. The authors further show an effective method of theirs through rigid experiments and evaluation in the detection of manipulated facial content over various techniques. Inferring the results from our study with the above framework, accuracies, and robustness are found to be high compared to the one in hand, being the earlier approaches in the specific problem domain of detection of altered facial images. The importance of this study is, therefore, in the development not only as an efficient detection mechanism but also in the broad application for digital forensics and media integrity. This therefore, places the proposed framework in line with the continued efforts being made towards arresting the situation of mass misinformation and deceptive practices in the digital media. The paper puts an extra emphasis on the need for continuous advancements and research in digital forensics following the developing landscape of feats in digital manipulation. New mechanisms in place for creating this manipulated content have put forth the need to develop some of the complex detection mechanisms to counter these threats. This proposed framework will take one giant step in that direction and will be an appreciated resource for the community of researchers, practitioners, and policymakers in the fight against digital manipulation and disinformation [15].

CHAPTER 3

PROBLEM DEFINITION AND OBJECTIVE

3.1 Problem Statement

It is just amazing how deepfake technology has been growing, and that has been witnessed within the last few years. The spread of deepfakes is dangerous in itself, as it means creating very convincing, yet completely unreal, videos or audio recordings through action or relay, which may be harmful to persons within the media. There is, hence, a need to come up with some strong countermeasures to deal with and reduce the risks associated with deepfakes: this might include better and more advanced detection tools, tighter regulations, and public awareness about the dangers posted by them. Without such a comprehensive approach, there is no hope for us to have the possibility of maintaining the integrity and privacy of digital media against this rapidly growing threat. The deep-learning algorithms can be used to stop the growth of deepfakes. Moreover, with advanced pattern recognition that combines machine learning, the pinpointed amount of touch-up manipulations is done with copperplate accuracy. The most recent actions in mitigating the threats that deepfakes present in ensuring that the information made or modified continues to have the authenticity upheld and, more importantly, individuals' safety looked after from malicious forgeries are indeed deep learning technologies. Therefore, the constant upgrading and implementation of such algorithms is, in turn, a significant issue in combating profound fake technology misuse.

3.2 Objective

The aim of this research is to develop a comprehensive solution for identifying all types of deepfakes. While Convolutional Neural Networks (CNNs) have proven useful in this area, this study focuses on employing Vision Transformers due to their ability to leverage global context and knowledge. The Vision Transformer approach is expected to enhance the accuracy and reliability of deepfake detection by capturing more extensive patterns and relationships within the data. By applying this advanced method, the research seeks to improve upon existing techniques and provide a more robust defense against the growing threat of deepfake technology.

CHAPTER 4

METHODOLOGY

In this chapter, we will look over the dataset used and the methodology employed in our research.

4.1 Dataset Used

Fig 4.1: Examples from OpenForensics dataset [16]

The dataset used in this work is the OpenForensics benchmark dataset for forgery detection. The present work adopted this challenging, large-scale dataset to detect and segment a forged face from over two faces in diversified natural scenes. It was developed to focus on accelerating and promoting research aimed at preventing DeepFake. This contrasts sharply with the more traditional deepfake detection tasks focusing on spotting a single face in far more straightforward and less realistic contexts. This therefore, sets a more complex platform for the development and high-level testing of algorithms that can work robustly in the wild and solve consequently the problems of today, in which current datasets serve the purpose in repetitive backgrounds and under controlled conditions. Hence, OpenForensics would expand openly at the research level concerning multifacial forgery detection and segmentation, being a source to be able to achieve cutting-edge technology in response to the deepening threat of deepfakes[16].

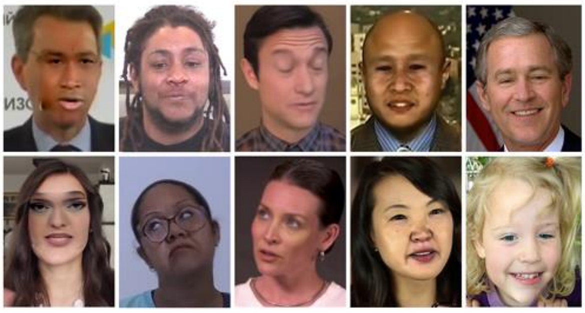


Fig 4.2: Visual artifacts of forged faces in datasets. From left to right, FaceForensics++ , DFDC , DeeperForensics, Celeb-DF , and our OpenForensics. Faces generated in our dataset have the highest resolution and best quality[16].

4.1.1 Features of Dataset

- Total Images: 115,325 images.
- Total Faces: 334,136 faces, comprising 160,676 real and 173,660 fake faces.

The OpenForensics dataset is comprehensively distributed into various subsets for effective training, validation, and testing of deepfake detection models. The training subset comprises 44,122 images containing 151,364 faces, with 85,392 being real and 65,972 forged. The validation subset includes 7,308 images with a total of 15,352 faces, out of which 4,786 are real and 10,566 are forged. For test development, there are 18,895 images with 49,750 faces, consisting of 21,071 real and 28,670 fake faces. Additionally, the test challenge subset contains 45,000 images with 117,670 faces, including 49,218 real and 68,452 fake faces. This detailed distribution ensures a robust and comprehensive evaluation framework for developing and assessing multi-face forgery detection and segmentation methods.

4.2 Data PreProcessing

The dataset is loaded from the directory specified by users with images labeled "Real" or "Fake." Inside, the technique of Random Over-sampling is applied, which duplicates instances of the minority class randomly to balance the dataset such that both classes will have equal numbers of samples. Considering the given dataset was broken up into training and test sets in the ratio of 60:40, respectively, ensuring a model was developed on one part and tested on another part. It, therefore, gives quite great explanations of how the model will perform.

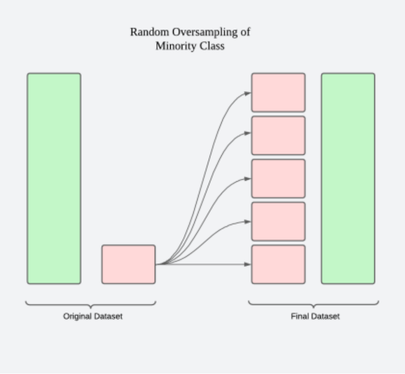
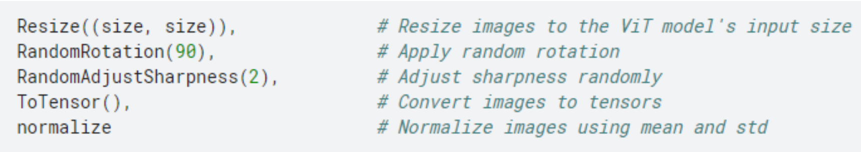


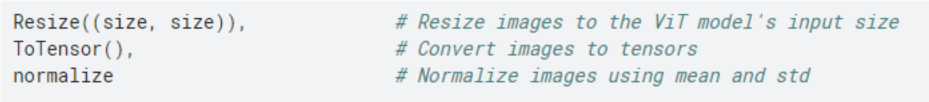
Fig 4.3: Random Oversampling

Once we are done with oversampling we perform various transformations on the images, including resizing, rotating, adjusting sharpness, and normalizing.



```
Resize((size, size)),               # Resize images to the ViT model's input size
RandomRotation(90),                 # Apply random rotation
RandomAdjustSharpness(2),           # Adjust sharpness randomly
ToTensor(),                         # Convert images to tensors
normalize                           # Normalize images using mean and std
```

Fig 4.4: Transformation on Dataset



```
Resize((size, size)),           # Resize images to the ViT model's input size
ToTensor(),                     # Convert images to tensors
normalize                       # Normalize images using mean and std
```

Fig 4.5: Resizing & converting image to tensor

These steps prepare the images for the ViT model, ensuring they are correctly formatted and scaled for optimal performance.

4.3 Vision Transformer Architecture

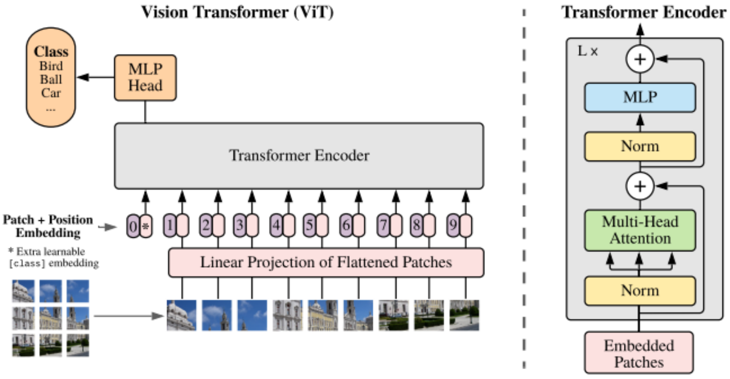


Fig 4.6: Vision Transformer Architecture

The Vision Transformer (ViT) architecture adapts the Transformer model, originally designed for natural language processing, to process image data effectively. Here is a detailed explanation of its architecture:

1. Patch extraction and embedding within patch

   This partitions the input image into fixed-size patches, for instance, 16 by 16 pixels, and then flattens the patches into a 1D vector. This is equivalent to changing each patch to a sequence of pixel values. These are linearly projected to high dimensions, hence the patch embeddings back.

2. Positional Embeddings

   Positional embeddings carry information on where patches lie in an image. Learnable vectors encode information on the positions of those patches in the overall sequence, such that patches in different places will thus be distinguished from each other in the original image.

3. Classification Token

   This is done by using sequence patch embeddings, after which a unique token CLS is prepended at the beginning of the sequence to represent the state of the Transformer as a summary of the full-image classification task after the Transformer layers.

4. Transformer Encoder Layers

   The sequence of embeddings, including the [CLS] token, is passed through multiple Transformer encoder layers. Each encoder layer consists of:

   - Multi-Head Self-Attention (MHSA): It enables the model to focus on the different parts of input sequences in a diversified way, consequently completing the capture of dependencies between very far-away patches. Self-attention is computed by each head independently over the input. This is followed by linear transform operations on the outputs of the concatenated heads.
   - Layer normalization :It is done during preprocessing before performing the layers using attention and feed-forward networks to maintain stability and speed up training.
   - FFN: A feed-forward neural network with two fully connected layers, applying subsequent processing for every embedding independently at each position. It is most often used as a structure of just two simple linear transformations with a ReLU activation in between.

- Residual connections: Skip connections around MHSA and FFN to keep gradients and enable training deep networks.

5. Output Layer

   After passing through the Transformer layers, the output corresponding to the [CLS] token is used for classification. This output is fed into a feed-forward network, often just a single linear layer, to produce the final classification logits.

4.4 Proposed Workflow

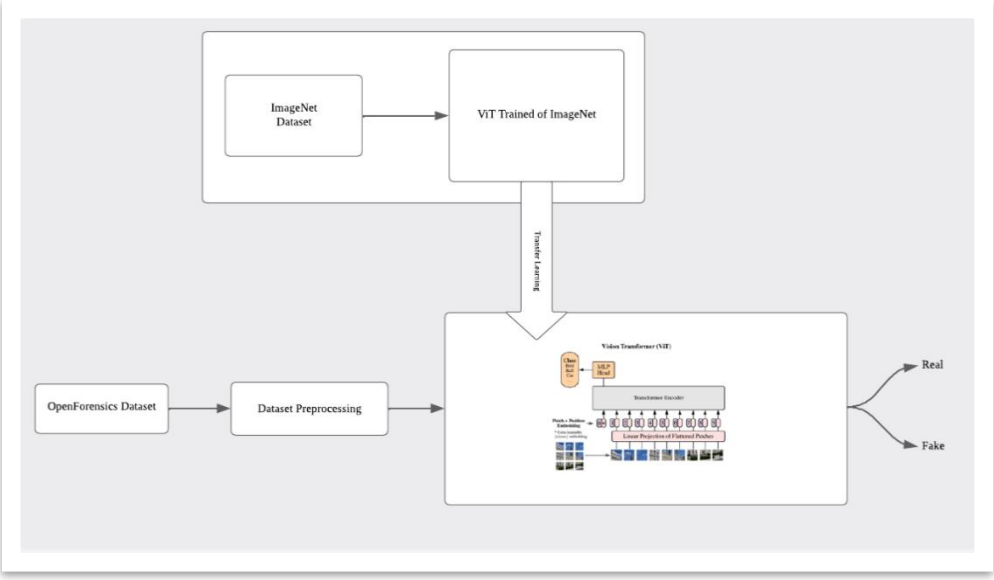


Fig 4.7: Proposed Architecture

The model workflow for detecting deepfakes utilizing the OpenForensics dataset is illustrated in the accompanying diagram. This workflow involves several key steps to effectively identify forged faces in images.

4.4.1 Dataset Acquisition and Pre-processing:

- OpenForensics Dataset: The first step involves obtaining the dataset from OpenForensics for multi-face forgery detection and segmentation. It is a mixed dataset with real and forged faces of different backgrounds in diverse natural scenes.

- Dataset Pre-processing: The dataset was also preprocessed beforehand, before the training of the model: image preprocessing included standardization, resizing to a joint resolution, and augmentation to add variability. This, in turn, will ensure that the data is in a form that will be valuable in training the model.

4.4.2 Pre-Trained Model Utilization:

- ImageNet Dataset and ViT Model: Concurrently, Vision Transformer (ViT) model that has been pre-trained on the ImageNet dataset is used. The ImageNet dataset is collection of large-scale of labeled images.Once the ViT is Pre-trained on Image-Net it can extract meaningful features from images.
- Transfer Learning: The pre-trained ViT model is then fine-tuned using the processed OpenForensics dataset. Transfer learning allows the model to leverage the knowledge it gained from the ImageNet dataset.This knowledge is than used to apply it to the specific task of deepfake detection. Fine-tuning process adjusts the model's parameters to improve its ability to distinguish between real and deepfake faces based on the new dataset.

4.4.3 Model Training and Evaluation:

- Vision Transformer (ViT): The Vision Transformer model takes the preprocessed images as input. The model divides each image into smaller patches and processes these patches through its transformer encoder to capture both local and global features.
- Classification: After processing the patches, the ViT model classifies the images as either real or fake. This classification is based on the features learned during the fine-tuning process.

4.4.4 Output:

The final output of the model is a classification label indicating whether the face in the image is real or fake. The accuracy and robustness of the model in identifying forged faces are enhanced through the use of pre-trained models and domain-specific fine-tuning.

CHAPTER 5

RESULT & DISCUSSION

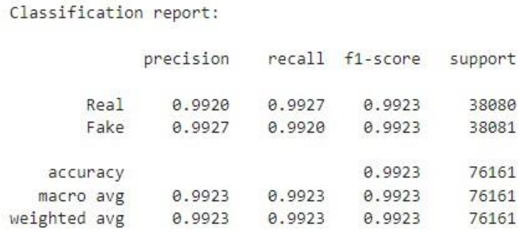
The ViT model is trained on the Openforensics dataset, and its performance is evaluated using a range of metrics, including F1 score, area under the curve (AUC), receiver operating characteristic (ROC) curve analysis, accuracy, recall, among others. These metrics provide a comprehensive assessment of the model's ability to classify images accurately. Additionally, the model undergoes further testing where it is presented with images labeled as real or fake, and its ability to predict the likelihood of an image being real or fake is examined. This testing involves analyzing the model's output probabilities, providing insights into its confidence levels and decision-making processes. By assessing the model's performance across multiple parameters and conducting real-world testing scenarios, a thorough evaluation of its capabilities in image classification and authenticity detection is achieved.

The model is trained for 4 EPOCHS ,however the performance and effectiveness for the model good, mainly because of transfer learning and a new innovative architecture of the Vision Transformer (ViT). The used transfer learning is undertaken such that, baselined with pre-training on massive datasets of the ImageNet kind, the ViT model employs the associated knowledge as the first bootstrapping step over the OpenForensics dataset. Help it learn well and generalize to complex patterns and features with a lesser number of epochs. Other than this, the intrinsic features of the ViT for extracting global context and its dependencies over long-range help in the outstanding performance of the model.

Epoch	Training Loss	Validation Loss	Accuracy
1	0.060300	0.023107	0.992726
2	0.049700	0.023382	0.992713
3	0.060300	0.023107	0.992726
4	0.049700	0.023382	0.992713

Table 1 : Accuracy in Epochs

Below is the classification report for the model



```
Classification report:

              precision    recall  f1-score   support

        Real     0.9920    0.9927    0.9923     38080
        Fake     0.9927    0.9920    0.9923     38081

    accuracy                         0.9923     76161
   macro avg     0.9923    0.9923    0.9923     76161
weighted avg     0.9923    0.9923    0.9923     76161
```

Fig 5.1 : Classification Report

Vision Transformer, over the Deepfake dataset. It obtained a precision of 0.9920 for the class "Real" and 0.9927 for the class "Fake," proving how accurately it determines the correct instances of both classes of interest. The recall scores were relatively balanced, at 0.9927 for class "Real" and 0.9920 for class "Fake." Hence, "the model generalizes well with virtually all real instances of both classes." The sound quality for F1 of both classes, at 0.9923, represents a delicate balance between precision and recall in the model setup. It means that the model's performance is high, amounting to 0.9923 in general measurements, thereby distinguishing between the real and fake states of the data. Such macro average and weighted average metrics give 0.9923, indicating the fact of relative equality in the performance of the model for all classes despite their support levels. Such a result provides a high reliability and robustness of the ViT model in solving this task of deepfake detection within these datasets.

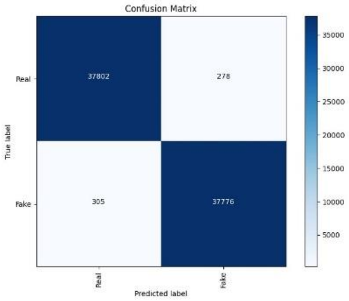


Fig 5.2 : Confusion Matrix

The confusion matrix shows that the ViT model accurately classifies the majority of real and fake instances, with 37802 true positives for real and 37776 for fake, and very few misclassifications: 278 real instances as fake and 305 fake instances as real. This indicates high precision and recall, confirming the model's robustness in detecting deepfakes.

CHAPTER 6

CONCLUSION AND FUTURE WORK

Since the accuracy of the ViT model is very high and stands at 99.23%, the effectiveness of ViT for deepfake detection is highly considered. The model has a very high precision, recall, and F1-score concerning both classes, real and fake, respectively, which proves its strength and reliability in the detection of deepfakes. This is further confirmed by the confusion matrix, with very minimal misclassifications.

For future work, it would be beneficial to undertake several initiatives to further enhance the effectiveness and robustness of the ViT model in deepfake detection:

1. **Test the ViT model on larger and more diverse datasets**: Evaluating the model on a broader range of datasets with varying types of deepfakes will help ensure its generalizability. This involves using datasets that include different manipulation techniques, varying quality, and diverse content to verify that the model performs consistently well across all scenarios. This step is crucial to confirm that the model is not overfitting to a specific dataset and can adapt to new, unseen data effectively.
2. **Investigate the model's performance in real-time deepfake detection scenarios**: Implementing the ViT model in real-time applications is critical for practical use cases, such as live video streams or social media monitoring. This involves optimizing the model for speed and efficiency so it can process data quickly without significant delays, ensuring that deepfakes are detected and flagged immediately as they appear.
3. **Explore the integration of ViT with other models or techniques**: Combining the ViT model with other approaches, such as traditional machine learning algorithms, convolutional neural networks (CNNs), or anomaly detection methods, can enhance its robustness. This hybrid approach can help in identifying deepfakes more accurately and can be particularly useful in

defending against adversarial attacks designed to deceive the detector by introducing subtle manipulations that might bypass a single model.

4. **Conduct a comparative study with other state-of-the-art models**: Performing a thorough comparison with other leading deepfake detection models will help identify the strengths and weaknesses of the ViT model. This study can highlight areas where ViT excels and where it needs improvement, guiding future development efforts. Benchmarking against a variety of models can provide a comprehensive understanding of the current state of deepfake detection technologies.
5. **Develop methods to explain the model's decisions**: Enhancing the transparency of the model's decision-making process is essential for building trust and understanding. This involves creating techniques to provide clear, interpretable explanations for why the model classified a particular video as real or fake. Such methods might include visualizing which parts of the video were most influential in the model's decision, or using explainable AI frameworks to break down the model's internal workings. Transparency is particularly important for gaining user confidence and ensuring that the model's predictions can be trusted and verified.

These initiatives will help ensure that the ViT model remains at the forefront of deepfake detection technology, offering reliable, efficient, and explainable performance in diverse and practical applications.

REFERENCES

[1] Karras, Tero, Samuli Laine, and Timo Aila. "A Style-Based Generator Architecture for Generative Adversarial Networks." ArXiv, (2018).

[2] Deeptrace. "The State of Deepfakes." https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

[3] Shen, Tianxiang, et al. "Deep Fakes using Generative Adversarial Networks (GAN)." University of California, San Diego, La Jolla, USA.

[4] Karras, Tero, Samuli Laine, and Timo Aila. "A Style-Based Generator Architecture for Generative Adversarial Networks." ArXiv, (2018).

[5] John Negoita , " Deepfake AI Face Swap. What is it? How does it work?"

[6] Pokroy, Artem A., and Alexey D. Egorov. "EfficientNets for deepfake detection: Comparison of pretrained models." In 2021 IEEE conference of russian young researchers in electrical and electronic engineering (ElConRus), pp. 598-600. IEEE, 2021.

[7] https://developers.google.com/machine-learning/gan/gan_structure

[8] Kohli, Aditi, and Abhinav Gupta. "Detecting deepfake, faceswap and face2face facial forgeries using frequency cnn." Multimedia Tools and Applications 80, no. 12 (2021): 18461-18478.

[9] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. "Attention Is All You Need." ArXiv, (2017). Accessed May 30, 2024. /abs/1706.03762.

[10] Chen, Hanting, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. "Pre-Trained Image Processing Transformer." ArXiv, (2020). Accessed May 30, 2024. /abs/2012.00364.

[11] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani et al. "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale." ArXiv, (2020). Accessed May 30, 2024. /abs/2010.11929.

[12] Ilyas, Hafsa, Ali Javed, Muteb Mohammad Aljasem, and Mustafa Alhababi. "Fused swish-relu efficient-net model for deepfakes detection." In 2023 9th International Conference on Automation, Robotics and Applications (ICARA), pp. 368-372. IEEE, 2023.

[13] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ArXiv. /abs/1905.11946

[14] Bonettini, Nicolò, Edoardo D. Cannas, Sara Mandelli, Luca Bondi, Paolo Bestagini, and Stefano Tubaro. "Video Face Manipulation Detection Through Ensemble of CNNs." ArXiv, (2020). /abs/2004.07676.

[15] Rössler, Andreas, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. "FaceForensics++: Learning to Detect Manipulated Facial Images." ArXiv, (2019). /abs/1901.08971.

[16] Le, Trung-Nghia, Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. "Openforensics: Large-scale challenging dataset for multi-face forgery detection and segmentation in-the-wild." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10117-10127. 2021.

[17] Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient Estimation of Word Representations in Vector Space." ArXiv, (2013). Accessed May 30, 2024. /abs/1301.3781.

[18] A. M. B. a. A. C. Bajaj, "Recent trends in internet of medical things: a review," *Advances in Machine Learning and Computational Intelligence,* pp. 645-656, 2021.

[19] Salvi, Davide, Honggu Liu, Sara Mandelli, Paolo Bestagini, Wenbo Zhou, Weiming Zhang, and Stefano Tubaro. "A robust approach to multimodal deepfake detection." Journal of Imaging 9, no. 6 (2023): 122.

PUBLICATIONS

1. Sarthak Kulkarni, Dinesh Kumar Vishwakarma, Virender Ranga, “AI vs. AI: Deep Learning Approaches for Detecting Deepfake Content” accepted in 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N2024)
2. Sarthak Kulkarni, Dinesh Kumar Vishwakarma, Virender Ranga, “Multi-Head Self-Attention Mechanism for Enhanced Deepfake Detection” accepted in 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N2024)

Gmail

Sarthak Kulkarni <sarthak1906k@gmail.com>

Acceptance Notification 1st IEEE ICAC2N-2024 & Registration: Paper ID 891

1 message

Microsoft CMT <email@msr-cmt.org> 25 June 2024 at 00:22
Reply-To: "Dr. Vishnu Sharma" <vishnu.sharma@its.edu.in>
To: Sarthak Kulkarni <sarthak1906k@gmail.com>

Dear Sarthak Kulkarni,
Delhi Technological University

Greetings from ICAC2N-2024 ...!!!

Congratulations....!!!!!

On behalf of the ICAC2N-2024 organising Committee, we are delighted to inform you that the submission of “Paper ID- 891 “ titled “ AI vs. AI: Deep Learning Approaches for Detecting Deepfake Content “ has been accepted for presentation and further publication with IEEE at the ICAC2N- 24. All accepted papers will be submitted for inclusion into IEEE Xplore subject to meeting IEEE Xplore’s scope and quality requirements.

For early registration benefit please pay your fee and complete your registration by clicking on the following Link: https://forms.gle/tLuA3qnVbR1zp3XT9 by 30 June 2024.

Registration fee details are available @ https://icac2n.in/register.

You can pay the registration fee by the UPI. (UPI id - icac2n@ybl) or follow the link below for QR code:
https://drive.google.com/file/d/1OIz9DB5CrxQM_cUnOjy2sc6rSDCDSOFn/view?usp=sharing

You are directed to ensure incorporating following points in your paper while completing your registration:

Comments:
The topic chosen "AI vs. AI: Deep Learning Approaches for Detecting Deepfake Content" is interesting and relevant.
Paper is well written and technically sound.
Problem identified is important and critical.
Accepted

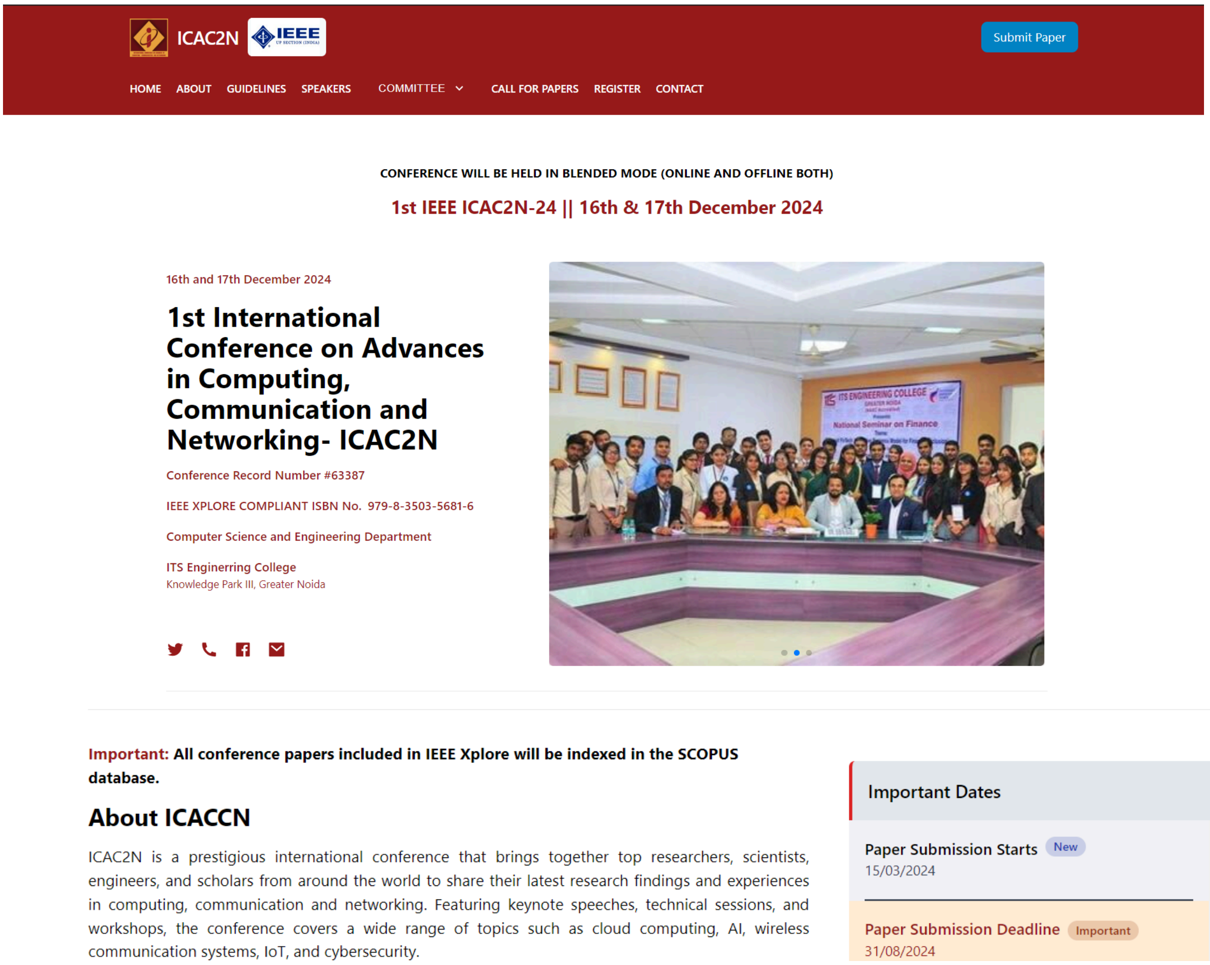
Note:
1. All figures and equations in the paper must be clear.
2. Final camera ready copy must be strictly in IEEE format available on conference website.
3. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE.
4. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 15% (checked through Turnitin).
5. Change in paper title, name of authors or affiliation of authors will not be allowed after registration of papers.
6. Violation of any of the above point may lead to rejection of your paper at any stage of publication.
7. Registration fee once paid will be non refundable.

If you have any query regarding registration process or face any problem in making online payment please call at 8168268768, write us at icac2n.ieee@gmail.com.

Regards:
Organizing committee
ICAC2N – 2024

*Steps for Completing Registration:
1. Make Payment
2. Fill Registration Form (Screenshot of Payment made by you to be uploaded in this)
3. You will receive Registration Confirmation Email within Next 3 working Days.

**If you need a fee receipt for reimbursement purpose, please send a request email on icac2n.ieee@gmail.com.



CONFERENCE WILL BE HELD IN BLENDED MODE (ONLINE AND OFFLINE BOTH)

1st IEEE ICAC2N-24 || 16th & 17th December 2024

16th and 17th December 2024

1st International Conference on Advances in Computing, Communication and Networking- ICAC2N

Conference Record Number #63387

IEEE XPLORE COMPLIANT ISBN No. 979-8-3503-5681-6

Computer Science and Engineering Department

ITS Enginerring College
Knowledge Park III, Greater Noida

Important: All conference papers included in IEEE Xplore will be indexed in the SCOPUS database.

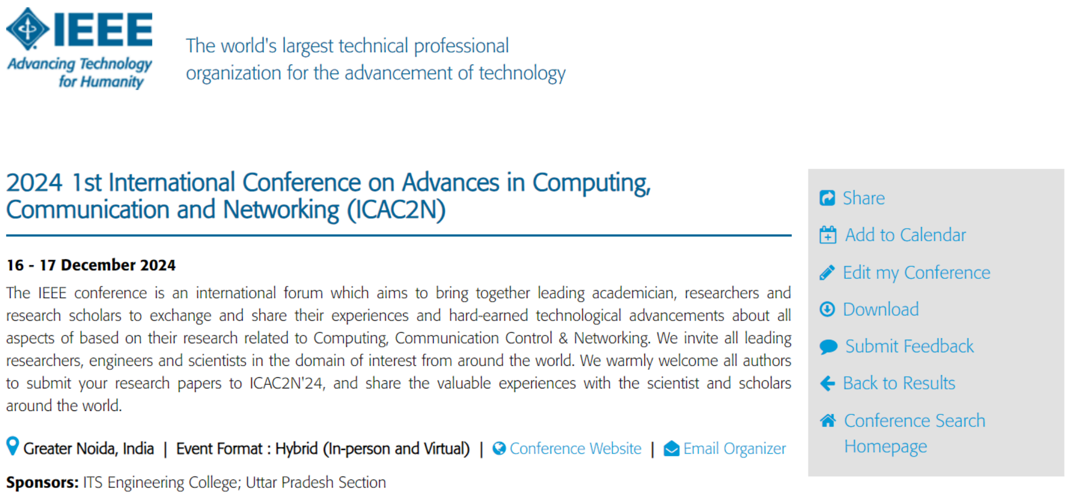
About ICACCN

ICAC2N is a prestigious international conference that brings together top researchers, scientists, engineers, and scholars from around the world to share their latest research findings and experiences in computing, communication and networking. Featuring keynote speeches, technical sessions, and workshops, the conference covers a wide range of topics such as cloud computing, AI, wireless communication systems, IoT, and cybersecurity.

Important Dates

Paper Submission Starts New
15/03/2024

Paper Submission Deadline Important
31/08/2024


IEEE
Advancing Technology for Humanity
The world's largest technical professional organization for the advancement of technology
Share
Add to Calendar
Edit my Conference
Download
Submit Feedback
Back to Results
Conference Search Homepage

2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)

16 - 17 December 2024

The IEEE conference is an international forum which aims to bring together leading academician, researchers and research scholars to exchange and share their experiences and hard-earned technological advancements about all aspects of based on their research related to Computing, Communication Control & Networking. We invite all leading researchers, engineers and scientists in the domain of interest from around the world. We warmly welcome all authors to submit your research papers to ICAC2N'24, and share the valuable experiences with the scientist and scholars around the world.

Greater Noida, India | Event Format : Hybrid (In-person and Virtual) | Conference Website | Email Organizer

Sponsors: ITS Engineering College; Uttar Pradesh Section

IEEE

INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING COMMUNICATION AND NETWORKING

iTS ENGINEERING COLLEGE
GREATER NOIDA | SINCE 2006

1st International Conference on Advances in Computing, Communication and Networking (ICAC2N-2024)

Conference Record Number # 63387

IEEE XPLORE COMPLIANT ISBN No # 979-8-3503-5681-6

Remuneration Acknowledgement Receipt ICAC2N-2024:

Received a sum of **Rs. 7000/-** f**rom Sarthak Kulkarni** of **Delhi Technological University (DTU)** as registration fee for **Paper ID – 891 "AI vs. AI: Deep Learning Approaches for Detecting Deepfake Content"** in 1st IEEE International Conference, ICAC2N-2024, organized by Department of CSE, ITS Engineering College, Gr. Noida, India on 16th-17th Dec., 2024.

★ ITS Engineering College ★ 46 Knowledge Park-III Greater Noida

Prof. (Dr.) Vishnu Sharma

Convener & Conference Organizing Chair

1st IEEE ICAC2N-2024

Sarthak Kulkarni <sarthak1906k@gmail.com>

Acceptance Notification 1st IEEE ICAC2N-2024 & Registration: Paper ID 841

1 message

Microsoft CMT <email@msr-cmt.org> 8 June 2024 at 15:09
Reply-To: "Dr. Vishnu Sharma" <vishnu.sharma@its.edu.in>
To: Sarthak Kulkarni <sarthak1906k@gmail.com>

Dear Sarthak Kulkarni,
Delhi Technological University

Greetings from ICAC2N-2024 ...!!!

Congratulations....!!!!!

On behalf of the ICAC2N-2024 organising Committee, we are delighted to inform you that the submission of “Paper ID- 841 “ titled “ Multi-Head Self-Attention Mechanism for Enhanced Deepfake Detection “ has been accepted for presentation and further publication with IEEE at the ICAC2N- 24. All accepted papers will be submitted for inclusion into IEEE Xplore subject to meeting IEEE Xplore’s scope and quality requirements.

For early registration benefit please pay your fee and complete your registration by clicking on the following Link: https://forms.gle/tLuA3qnVbR1zp3XT9 by 15 June 2024.

Registration fee details are available @ https://icac2n.in/register.

You can pay the registration fee by the UPI. (UPI id - icac2n@ybl) or follow the link below for QR code:
https://drive.google.com/file/d/1OIz9DB5CrxQM_cUnOjy2sc6rSDCDSOFn/view?usp=sharing

You are directed to ensure incorporating following points in your paper while completing your registration:

Comments:
The theme and title of the article "Multi-Head Self-Attention Mechanism for Enhanced Deepfake Detection" is suitable and appropriate for publication.
Paper is well written and technically sound.
Overall flow of paper is very good.
Result are well explained
English language is satisfactory.
Accepted

Note:
1. All figures and equations in the paper must be clear.
2. Final camera ready copy must be strictly in IEEE format available on conference website.
3. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE.
4. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 15% (checked through Turnitin).
5. Change in paper title, name of authors or affiliation of authors will not be allowed after registration of papers.
6. Violation of any of the above point may lead to rejection of your paper at any stage of publication.
7. Registration fee once paid will be non refundable.

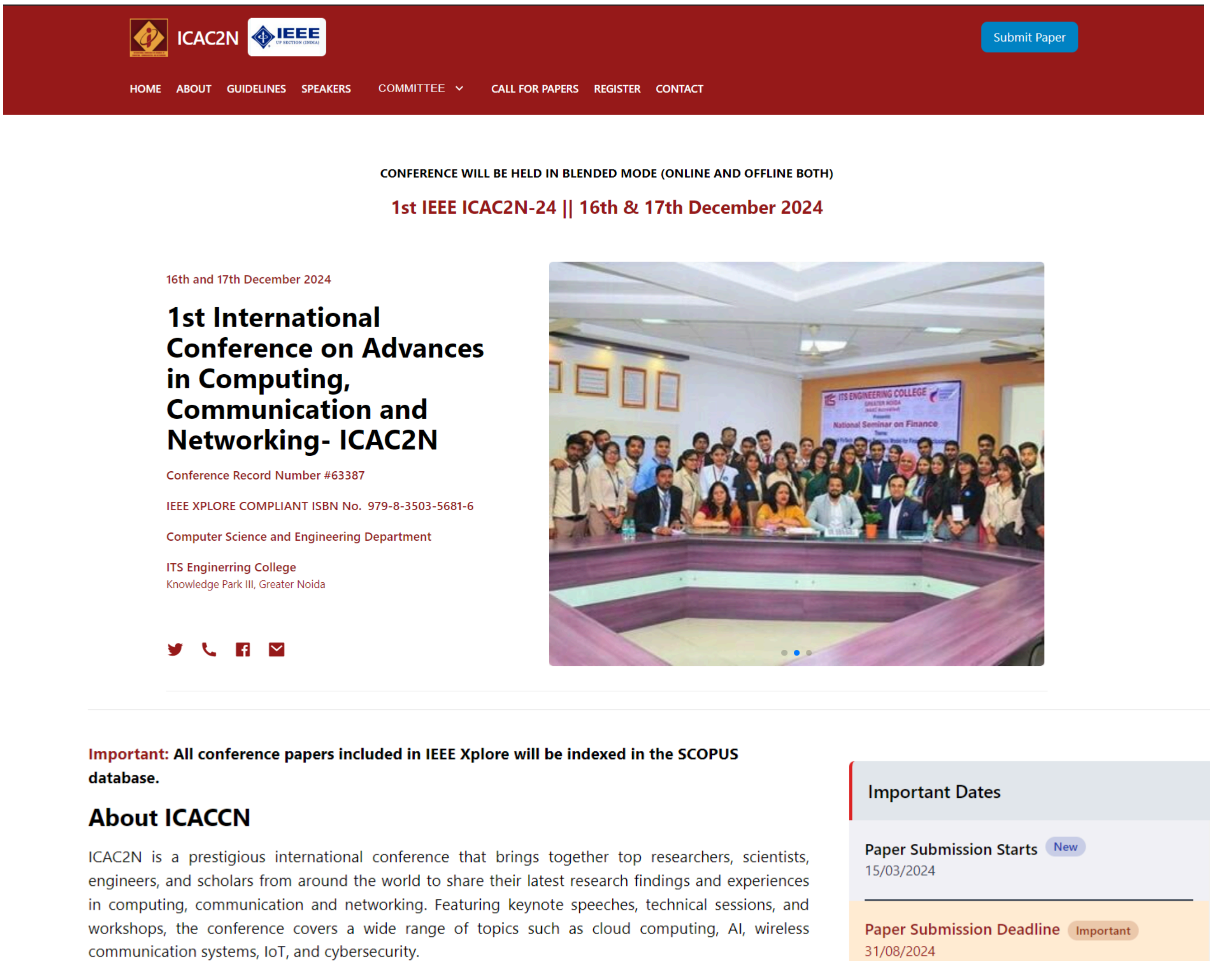
If you have any query regarding registration process or face any problem in making online payment please call at 8168268768, write us at icac2n.ieee@gmail.com.

Regards:
Organizing committee
ICAC2N – 2024

*Steps for Completing Registration:
1. Make Payment
2. Fill Registration Form (Screenshot of Payment made by you to be uploaded in this)
3. You will receive Registration Confirmation Email within Next 3 working Days.

**If you need a fee receipt for reimbursement purpose, please send a request email on icac2n.ieee@gmail.com.

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.



CONFERENCE WILL BE HELD IN BLENDED MODE (ONLINE AND OFFLINE BOTH)

1st IEEE ICAC2N-24 || 16th & 17th December 2024

16th and 17th December 2024

1st International Conference on Advances in Computing, Communication and Networking- ICAC2N

Conference Record Number #63387

IEEE XPLORE COMPLIANT ISBN No. 979-8-3503-5681-6

Computer Science and Engineering Department

ITS Enginerring College
Knowledge Park III, Greater Noida

Important: All conference papers included in IEEE Xplore will be indexed in the SCOPUS database.

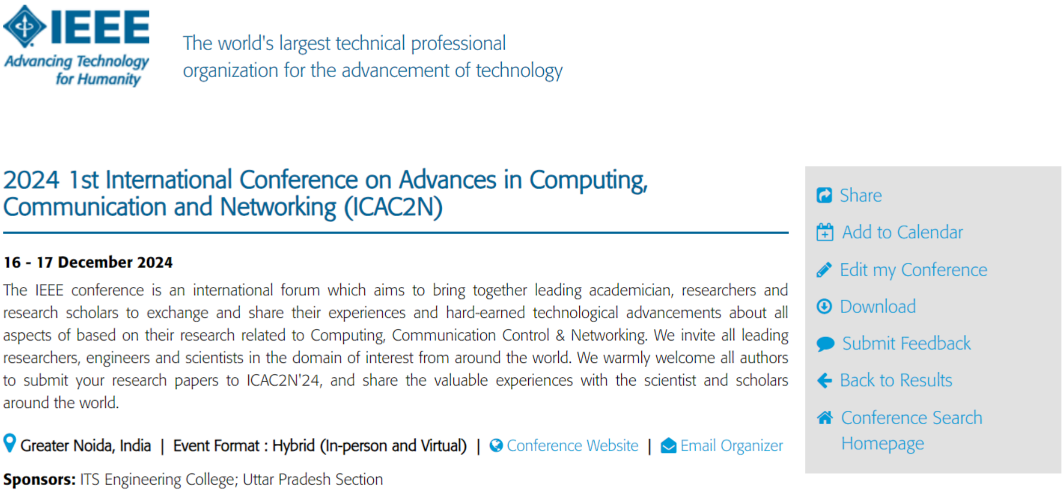
About ICACCN

ICAC2N is a prestigious international conference that brings together top researchers, scientists, engineers, and scholars from around the world to share their latest research findings and experiences in computing, communication and networking. Featuring keynote speeches, technical sessions, and workshops, the conference covers a wide range of topics such as cloud computing, AI, wireless communication systems, IoT, and cybersecurity.

Important Dates

Paper Submission Starts New
15/03/2024

Paper Submission Deadline Important
31/08/2024



The world's largest technical professional
organization for the advancement of technology

2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)

16 - 17 December 2024

The IEEE conference is an international forum which aims to bring together leading academician, researchers and research scholars to exchange and share their experiences and hard-earned technological advancements about all aspects of based on their research related to Computing, Communication Control & Networking. We invite all leading researchers, engineers and scientists in the domain of interest from around the world. We warmly welcome all authors to submit your research papers to ICAC2N'24, and share the valuable experiences with the scientist and scholars around the world.

Greater Noida, India | Event Format : Hybrid (In-person and Virtual) | Conference Website | Email Organizer

Sponsors: ITS Engineering College; Uttar Pradesh Section

IEEE

INTERNATIONAL CONFERENCE ON ADVANCES IN COMPUTING COMMUNICATION AND NETWORKING

ITS ENGINEERING COLLEGE
GREATER NOIDA | SINCE 2006

1st International Conference on Advances in Computing, Communication and Networking (ICAC2N-2024)

Conference Record Number # 63387

IEEE XPLORE COMPLIANT ISBN No # 979-8-3503-5681-6

Remuneration Acknowledgement Receipt ICAC2N-2024:

Received a sum of **Rs. 7000/-** from **Sarthak Kulkarni** of **Delhi Technological University (DTU)** as registration fee for **Paper ID – 841 " Multi-Head Self-Attention Mechanism for Enhanced Deepfake Detection"** in 1st IEEE International Conference, ICAC2N-2024, organized by Department of CSE, ITS Engineering College, Gr. Noida, India on 16th-17th Dec., 2024.

ITS Engineering College * 46 Knowledge Park-III Greater Noida *

Prof. (Dr.) Vishnu Sharma

Convener & Conference Organizing Chair

1st IEEE ICAC2N-2024

ANNEXURE-IV

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis DECODING DEEPFAKES: A DEEP LEARNING APPROACH TO UNVEILING SYNTHETIC MEDIA Total Pages 28 Name of the Scholar Sarthak Kulkarni Supervisor (s)

(1) Prof Dinesh Kumar Vishwakarma
(2) Dr.Virender Ranga

Department INFORMATION TECHNOLOGY, DELHI TECHNOLOGICAL UNIVERSITY

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 11 % , Total Word Count: 6774

Date: 31/05/2024

Candidate's Signature	**Signature of Supervisor**	**Signature of Co-Supervisor**
Sarthak Kulkarni	Prof Dinesh Kumar Vishwakarma	Dr.Virender Ranga

PAPER NAME

Decoding Deepfakes A Deep Learning Approach to Unveiling Synthetic Media.pdf

WORD COUNT **6774 Words**	CHARACTER COUNT **38207 Characters**
PAGE COUNT **28 Pages**	FILE SIZE **683.3KB**
SUBMISSION DATE **May 31, 2024 8:50 AM GMT+5:30**	REPORT DATE **May 31, 2024 8:50 AM GMT+5:30**

● 11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 7% Internet database
- Crossref database
- 8% Submitted Works database
- 3% Publications database
- Crossref Posted Content database

● Excluded from Similarity Report

- Bibliographic material
- Cited material
- Quoted material
- Small Matches (Less then 8 words)

● 11% Overall Similarity

Top sources found in the following databases:

- 7% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 8% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	storage.googleapis.com Internet	2%
2	Roehampton University on 2024-05-14 Submitted works	<1%
3	Middlesex University on 2023-03-27 Submitted works	<1%
4	web.archive.org Internet	<1%
5	Universidade do Porto on 2023-02-23 Submitted works	<1%
6	University of Hertfordshire on 2024-04-29 Submitted works	<1%
7	huggingface.co Internet	<1%
8	Southampton Solent University on 2024-01-08 Submitted works	<1%

9	link.springer.com Internet	<1%
10	frontiersin.org Internet	<1%
11	jcheminf.biomedcentral.com Internet	<1%
12	Bilkent University on 2023-12-28 Submitted works	<1%
13	De Montfort University on 2024-05-10 Submitted works	<1%
14	Liverpool John Moores University on 2023-03-08 Submitted works	<1%
15	University at Buffalo on 2023-10-13 Submitted works	<1%
16	Khaja Mannanuddin, V.R. Vimal, Angalkuditi Srinivas, S.D. Uma Mages... Crossref	<1%
17	valser.org Internet	<1%
18	Qian, Yang. "Enhancing Automatic Emotion Recognition for Clinical Ap... Publication	<1%
19	Saima Waseem, Syed R. Abu-Bakar, Bilal Ashfaq Ahmed, Zaid Omar, T... Crossref	<1%
20	Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, ... Crossref	<1%

21	University of North Texas on 2023-05-10 Submitted works	<1%
22	samit.khpi.edu.ua Internet	<1%
23	conference2go.com Internet	<1%
24	Columbia University on 2023-11-11 Submitted works	<1%
25	deepai.org Internet	<1%
26	github.com Internet	<1%
27	scholar.archive.org Internet	<1%
28	aaai.org Internet	<1%
29	mdpi.com Internet	<1%
30	slideshare.net Internet	<1%
31	Asare, Bernard. "“AWAM” – A Dual-Pathway Deepfake Discriminator fo... Publication	<1%
32	Chao Fan, Litao Yang, Hao Lin, Yingying Qiu. "DFE-Net: Dual-branch fea... Crossref	<1%

33	Hooper, Sarah McIlwaine. "Label-Efficient Machine Learning for Medic... Publication	<1%
34	Katholieke Universiteit Leuven on 2024-05-13 Submitted works	<1%
35	Lee, Ho Hin. "Exploring Explainable Optimization in Medical Segmentat... Publication	<1%
36	Muhammad Salihin Saealal, Mohd Zamri Ibrahim, Mohd Ibrahim Shapi... Crossref	<1%
37	University College London on 2023-08-18 Submitted works	<1%
38	april.zju.edu.cn Internet	<1%
39	dspace.bracu.ac.bd Internet	<1%
40	norma.ncirl.ie Internet	<1%
41	par.nsf.gov Internet	<1%
42	thuvienso.ktkt.edu.vn:8080 Internet	<1%
43	tuprints.ulb.tu-darmstadt.de Internet	<1%
44	politesi.polimi.it Internet	<1%

45	q-group.org Internet	<1%
46	springerprofessional.de Internet	<1%
47	yorkspace.library.yorku.ca Internet	<1%