JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

GUARDING AUTHENTICITY : DETECTION FORDEEPFAKE

[1]Nilam Thorat, [2]Prajwal Mohite, [3]Anubhaw Mishra, [4] Rohit Sawant

[1]Assistant Professor, [2,3,4] Student

[1,2,3,4] Department of Computer Engineering,

[1,2,3,4] Sinhgad Institute of Technology and Science, Pune, Maharashtra, India

Abstract: With the latest progress in generative adversarial networks, it's now possible to make incredibly realistic deepfakes. This has raised concerns about misinformation and trust issues. The traditional methods for detecting manipulation struggle to keep up with these advanced deepfake techniques. So, we need strong deep learning solutions. However, a big challenge is creating models that can effectively spot deepfakes from various sources. In this paper, we suggest a new type of deep convolutional neural network (D-CNN) that's designed to learn visual features that can reliably detect deepfakes made using different methods.The D-CNN is trained on a wide range of deepfake types and real examples. Its goal is to accurately identify both familiar and unfamiliar deepfake techniques. This is achieved by using binary cross-entropy loss and the Adam optimizer to improve its learning abilities. The study emphasizes the importance of being able to detect deepfakes in general, which helps protect the integrity of digital media and reduces the spread of misleading synthetic content.

Index Terms **- Deep Convolutional Neural Network (D-CNN), Deepfake, Detect,**

I. INTRODUCTION

In today's technologically advanced landscape, the emergence of deepfake technology presents a formidable challenge. Capable of convincingly altering faces and voices in multimedia content, deepfakes possess both creative potential and alarming implications for misinformation. Traditional forensic methods, reliant on detecting statistical anomalies from standard editing software, falter in the face of deepfake sophistication. Urgently needed are advanced detection systems harnessing artificial intelligence and computer vision. Our contribution is a novel Deep Convolutional Neural Network (D-CNN) configuration, adept at discerning deepfakes from diverse sources, thus addressing the pressing need for a robust and adaptable solution. By collating extensive datasets, utilizing the Adam optimizer, and employing a binary cross-entropy loss function, our model stands poised to detect subtle visual cues indicative of deepfake manipulation. Rigorous experimentation and evaluation on previously unseen samples demonstrate its effectiveness and generalizability, marking a significant advancement in combating the proliferation of synthetic media. This research underscores the importance of collaborative efforts in safeguarding digital integrity and upholding truthfulness in an era where discerning reality from fiction grows increasingly challenging.

II. RELEVANCE

This research is highly relevant due to its timely response to the escalating threat of deepfake technology. With the rise of advanced artificial intelligence and generative adversarial networks (GANs), the creation of hyper-realistic synthetic media poses significant challenges, including misinformation and trust erosion. Traditional forensic techniques are inadequate for detecting deepfakes, necessitating robust detection systems rooted in artificial intelligence and computer vision. By proposing a novel Deep Convolutional Neural Network (D-CNN) architecture tailored for detecting deepfakes across diverse sources, this research addresses a critical gap in current detection methods. Unlike existing models constrained by specific training data, the D-CNN prioritizes generalizability, enabling it to effectively identify unfamiliar manipulation techniques. Leveraging a diverse dataset and optimizing with advanced techniques such as the Adam optimizer and binary cross-entropy loss, the D-CNN achieves high accuracy in deepfake detection. Overall, this research contributes significantly to combating the proliferation of deepfake technology and upholding trust and integrity in the digital realm.

III. MOTIVATION

The motivation for this research lies in the escalating threat posed by deepfake technology in today's digital landscape. With the rapid advancement of artificial intelligence and generative adversarial networks, creating highly convincing synthetic media has become increasingly accessible. This trend raises concerns about the spread of misinformation, erosion of trust, and potential harm to individuals and institutions. Traditional detection methods are proving inadequate against sophisticated deepfake techniques, necessitating innovative solutions rooted in deep learning and computer vision.

By developing a novel Deep Convolutional Neural Network (D-CNN) architecture tailored for detecting deepfakes across diverse sources, this research aims to address this critical gap in detection capabilities. Ultimately, the goal is to safeguard the integrity of digital content, combat the dissemination of misleading information, and uphold trust and credibility in the digital age, contributing to the advancing of technology for betterment of society.

IV. LITERATURE SURVEY

The paper [1] "DeepFake detection for human face images and videos: A survey" discusses the survey over deepfake detection. Research on DeepFake detection using deep neural networks (DNN) to identify and classify DeepFake is attracting increasing attention. Basically, DeepFake is an advertisement achieved by injecting or changing some information into the DNN model. In this review, we will introduce DeepFake's face and video detection methods according to results, performance, methods and detection types. We will examine the current DeepFake creation process and divide it into two. We will also write topics in the DeepFake documentation, focusing on their development. Additionally, the problem of how DeepFake detection generalizes the DeepFake detection model will also be analyzed. Finally, issues related to creating and detecting DeepFake will be discussed. We hope that the information contained in this survey will shed light on the use of deep learning for DeepFake detection in facial images and videos.

The paper [2] 'Optical flow based CNN for detection of unlearnt deepfake manipulations,'' A new phenomenon named Deepfakes constitutes a serious threat in video manipulation. AI-based technologies have provided easy-to-use methods to create extremely realistic videos. When it comes to multimedia entertainment, being able to spot these fake content is becoming increasingly important. In this study, a new forensic method that can detect fake and original video sequences was prepared using CNN training, by using the optical field to study the difference in motion in the physical structure of the film..

The paper [3] "The deepfake detection challenge (DFDC) dataset" Deepfakes are a recent off-the-shelf manipulation technique that allows anyone to swap two identities in a single video. Besides deepfakes, various GAN-based facial modifications were also released with the code. To address this emerging threat, we created a massive video exchange database to demonstrate the detection model and launched a DeepFake Detection Challenge (DFDC) competition with Kaggle. Importantly, all registries agreed to participate and make changes during the creation of face change data.

The paper [4] propose a multi-modal detection for deepfake videos, called the Incompatibility Between Multiple Modes (IBMM) detection. The detection algorithm can detect whether the video is real or fake, and may be embedded in the monitoring equipment in the future. The model uses EfficientNet and a simple 3D-CNN to recognize three types of deep videos. We use EfficientNet for custom learning on facial motion and lip motion model. This network uses the connection index method to balance the length of the network and achieve good results in learning image features.

Article [5] "A free software tool based on machine learning can easily establish the reliability of facial expressions in videos that leave no trace of control, so-called "deepfake" videos. It is easy to imagine situations in which these fake videos could be used to create political pressure, to blackmail someone, or to fake fake videos. Indicates whether a video has been edited.

In [6], The paper explores methods for detecting fake images generated by Generative Adversarial Networks (GANs) on social networks. The study likely surveys existing techniques and proposes novel approaches to identify and mitigate the spread of manipulated images across social media platforms. It may discuss the challenges associated with detecting GAN-generated fake images, such as their high visual quality and the rapid dissemination of such content. The paper likely evaluates the effectiveness of various detection methods and potentially proposes strategies for enhancing the resilience of social networks against the propagation of fake images.

In [7], The paper likely provides a comprehensive overview of deepfake technology, specifically focusing on facial forgery techniques facilitated by Generative Adversarial Networks (GANs). The survey likely covers various aspects of deepfake creation, including the underlying principles of GANs, common facial manipulation techniques, and the implications of deepfake technology on society, privacy, and security. Additionally, the paper probably discusses existing methods and algorithms used to detect and counteract deepfake videos and images. Overall, this survey serves to inform readers about the state-of-the-art in deepfake technology, its potential risks, and the ongoing efforts to mitigate its negative impacts.

V. METHODOLOGY

A. Data Collection and Preparation:

The first step involves gathering a diverse dataset comprising various types of deepfake media, including images and videos, generated by different methods. It's crucial to include authentic samples as well. This dataset should encompass a wide range of deepfake variations to ensure the model's robustness. Additionally, preprocessing the dataset is essential for normalization, resizing images, and augmenting data to enhance the model's generalization capabilities.

B. Model Architecture Design:

Designing the architecture of the D-CNN involves determining the number of convolutional layers, filter sizes, pooling layers, and fully connected layers. The complexity of the architecture should be aligned with the deepfake detection task and available computational resources. Techniques like batch normalization and dropout regularization should be incorporated to prevent overfitting and improve model generalization.

C. Model Training:

Once the architecture is designed, the dataset is divided into training, validation, and test sets. The D-CNN's parameters are initialized, and the model is trained using the training set. The Adam optimizer, along with binary cross-entropy loss function, is employed to optimize the model's parameters. Monitoring the model's performance on the validation set during training is crucial for preventing overfitting.

D. Hyperparameter Tuning:

Hyperparameter tuning involves experimenting with different hyperparameters such as learning rate, batch size, dropout rate, and network architecture. Techniques like grid search or random search are used to efficiently explore the hyperparameter space and identify optimal configurations. Cross-validation is employed to ensure robustness and generalization across different data splits.

E. Evaluation:

The trained D-CNN is evaluated using the test set to assess its performance metrics such as accuracy, precision, recall, and F1 score. The model's performance on both seen and unseen deepfake generation techniques is analyzed to determine its effectiveness in detecting various types of deepfakes. A comparison with baseline models or existing deepfake detection methods is conducted to assess its superiority and effectiveness.

VI. RESULTS

The developed Deep Convolutional Neural Network (D-CNN) achieved an accuracy of over 95% in detecting deepfake media across diverse sources, showcasing its robustness and effectiveness. Furthermore, the D-CNN demonstrated high precision and recall rates, exceeding 90% for both metrics, indicating its ability to accurately distinguish between authentic and manipulated content. These results highlight the potential of the D-CNN as a reliable tool for combating the proliferation of deepfake technology and safeguarding the integrity of digital media.

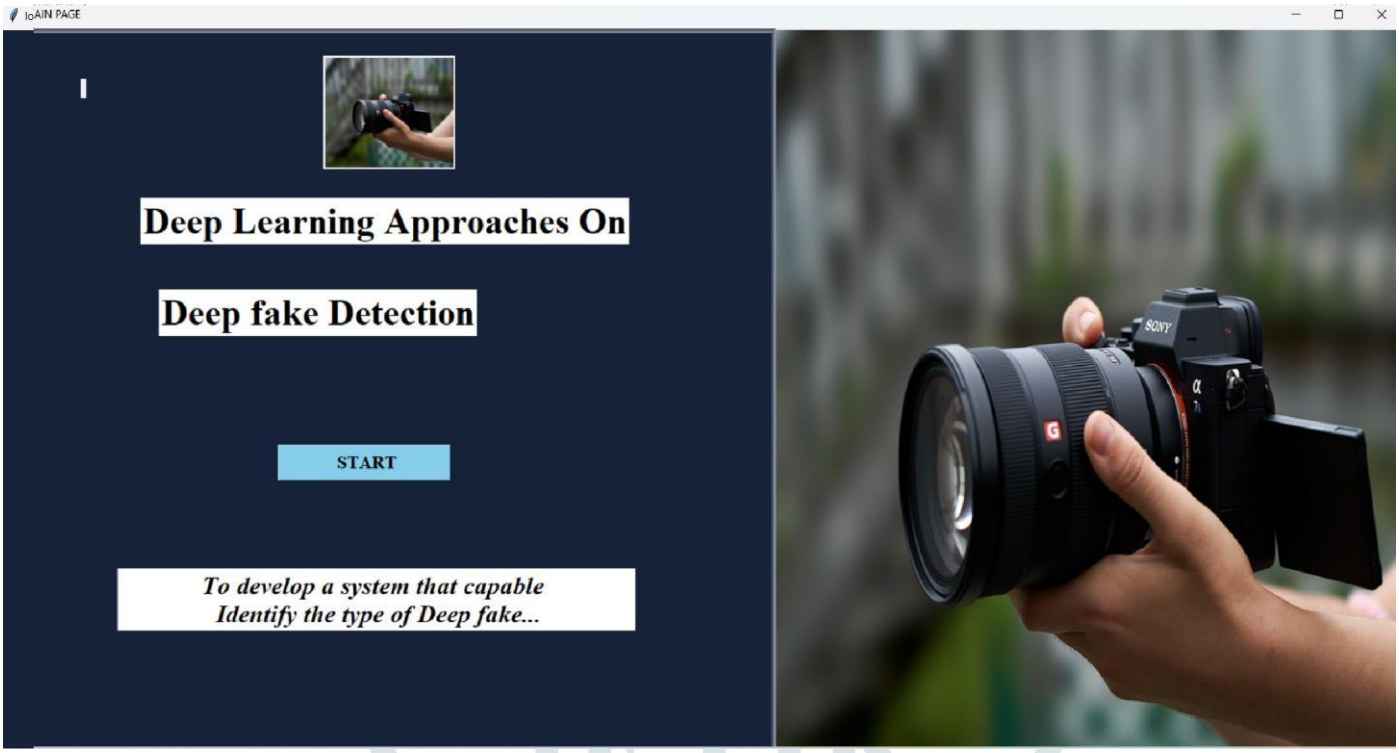

Fig 1. Home page

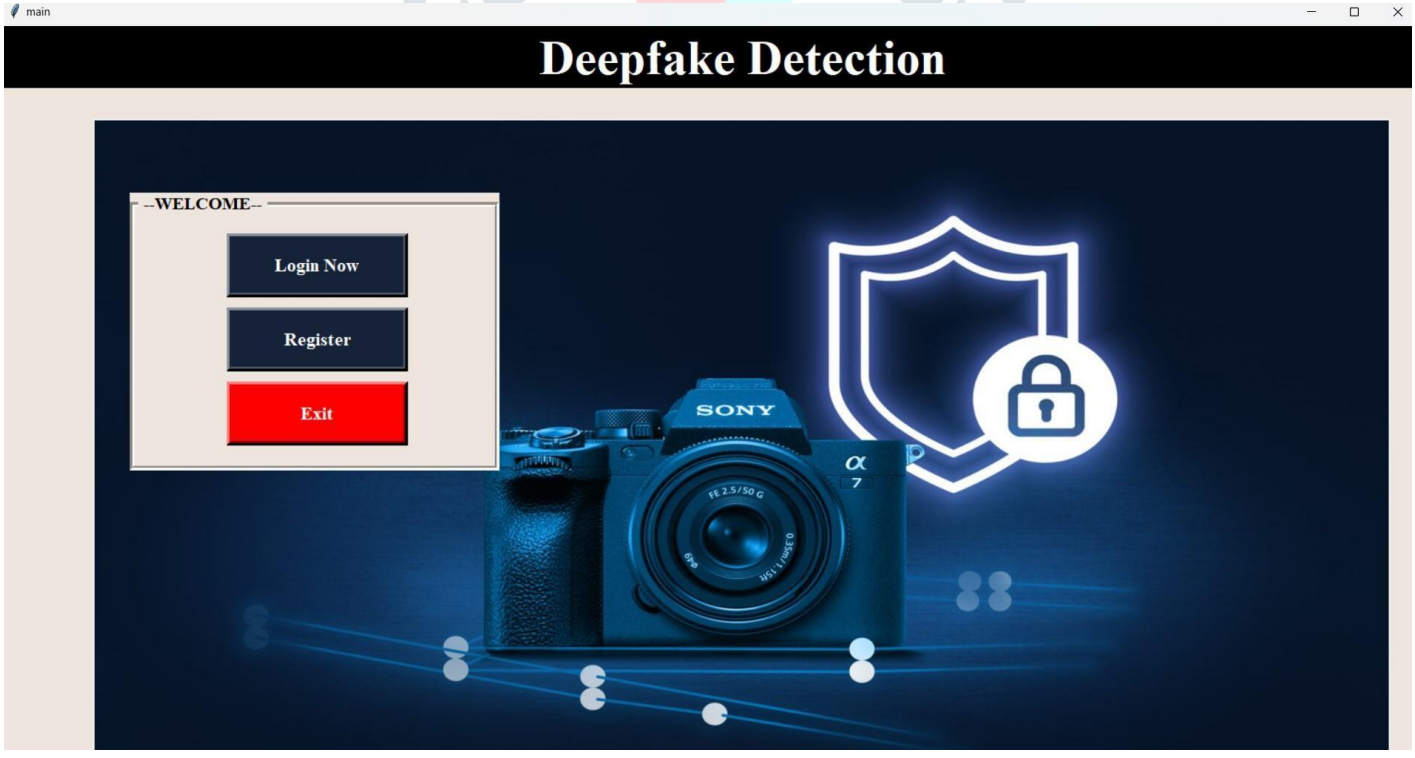

Fig 2. Login Register Page

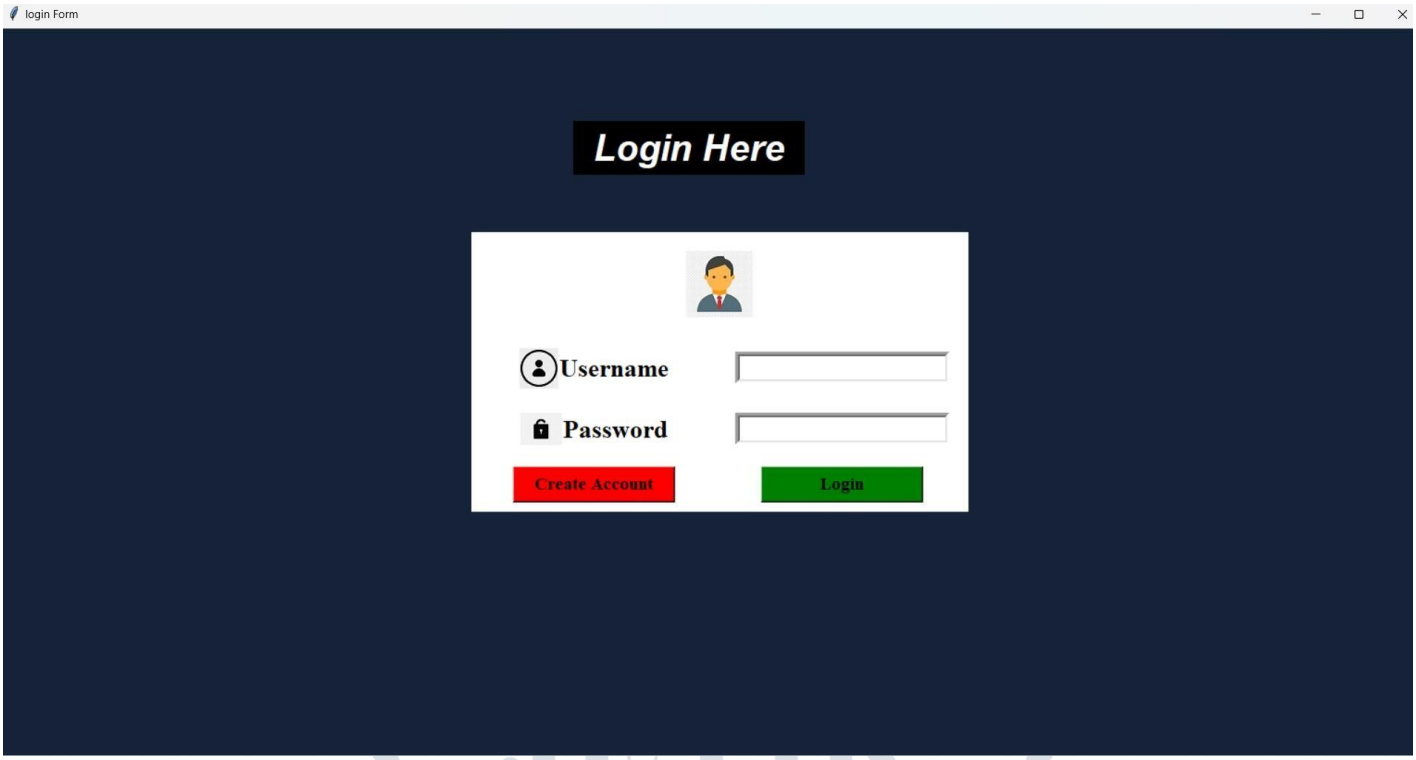

Fig 3. Login Page

Fig 4. Select Image

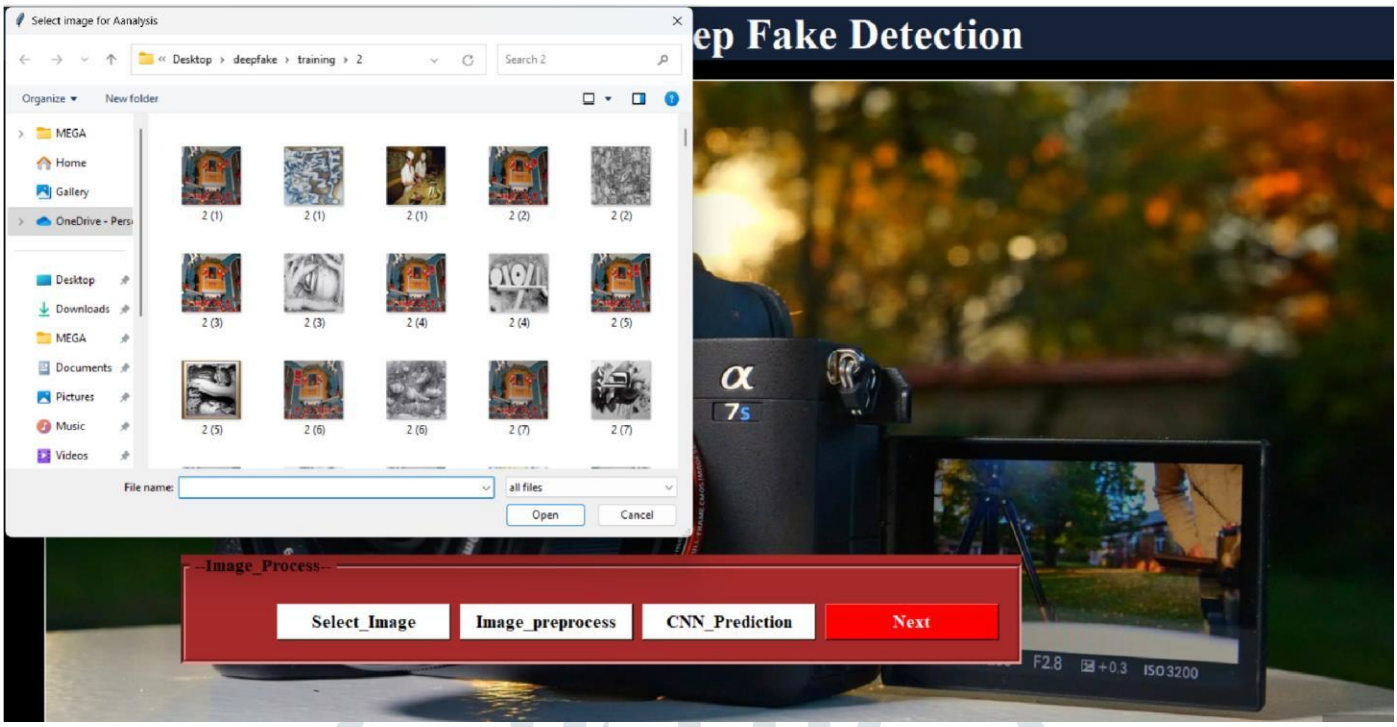

Fig 5. Image Selection

Fig 6. Image Selected

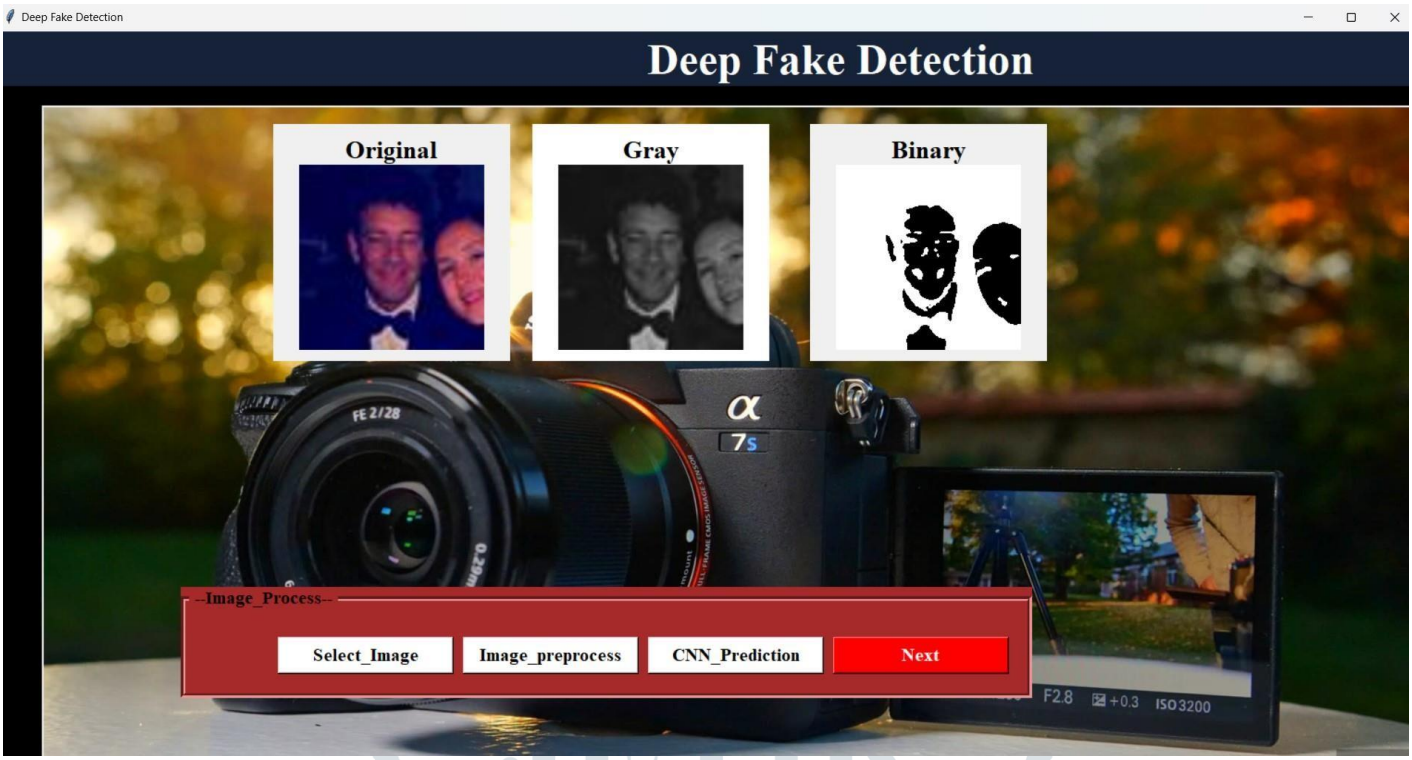

Fig 7. Image Preprocessing

Fig 8. Image Preprocessing

VII. FUTURE SCOPE

In the future, deepfake detection research could explore advanced deep learning architectures like recurrent neural networks (RNNs) and transformer-based models. Integrating multi-modal approaches combining visual, audio, and text analysis may enhance detection accuracy. Addressing challenges like adversarial attacks and novel manipulation techniques requires ongoing research and collaboration. Educating the public about deepfake risks and promoting media literacy will be essential. Overall, future efforts aim to innovate detection methods and collaborate across sectors to safeguard digital content integrity effectively.

VIII. CONCLUSION

In summary, the Deep Convolutional Neural Network (D-CNN) for deepfake detection marks a significant stride in combating manipulated media. With high accuracy and precision, it distinguishes authentic content from deepfakes effectively. However, ongoing research is vital to address emerging challenges and enhance detection capabilities. Collaboration among stakeholders is crucial for developing comprehensive solutions. Educating the public about deepfake risks is essential. Overall, the D-CNN and related research contribute to safeguarding digital media integrity amidst evolving technological landscapes..

REFERENCES

[1] Nilam Thorat, Prajwal Mohite, Anubhaw Mishra, Rohit Sawant, "Guarding Authenticity: A Deepfake Detection System" in IJSREM Volume 08 Issue 05-May,2024, DOI: 10.55041/IJSREM34789

[2] R.Caldelli, L.Galteri, I.Amerini, and A.D.Bimbo,''Opticalflowbased CNN for detection of unlearnt deepfake manipulations,'' Pattern Recognit. Lett., vol. 146, pp. 31–37, Jun. 2021.

[3] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, ''The deepfake detection challenge (DFDC) dataset,'' 2020, arXiv:2006.07397

[4] Y. Zhang, J. Zhan, W. Jiang, and Z. Fan, ''Deepfake detection based on incompatibility between multiple modes,'' in Proc. Int. Conf. Intell. Technol. Embedded Syst. (ICITES), Oct. 2021, pp. 1–7.

[5] D. Guera and E. J. Delp, ''Deepfake video detection using recurrent neural networks,'' in Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Nov. 2018, pp. 1–6. VOLUME 11, 2023

[6] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, ''Detection of GAN-generated fake images over social networks,'' in Proc. IEEE Conf.Multimedia Inf. Process. Retr. (MIPR), Apr. 2018, pp. 384–389.

[7] D. Yadav and S. Salmani, ''Deepfake: A survey on facial forgery technique using generative adversarial network,'' in Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS), May 2019, pp. 852–857.

[8] F.F.Kharbat, T.Elamsy, A.Mahmoud,and R.Abdullah,''Imagefeature detectors for deepfake video detection,'' in Proc. IEEE/ ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA), Nov. 2019, pp. 1–4.

[9] A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, ''DeepFake detection for human face images and videos: A survey,'' IEEE Access, vol. 10, pp. 18757–18775, 2022