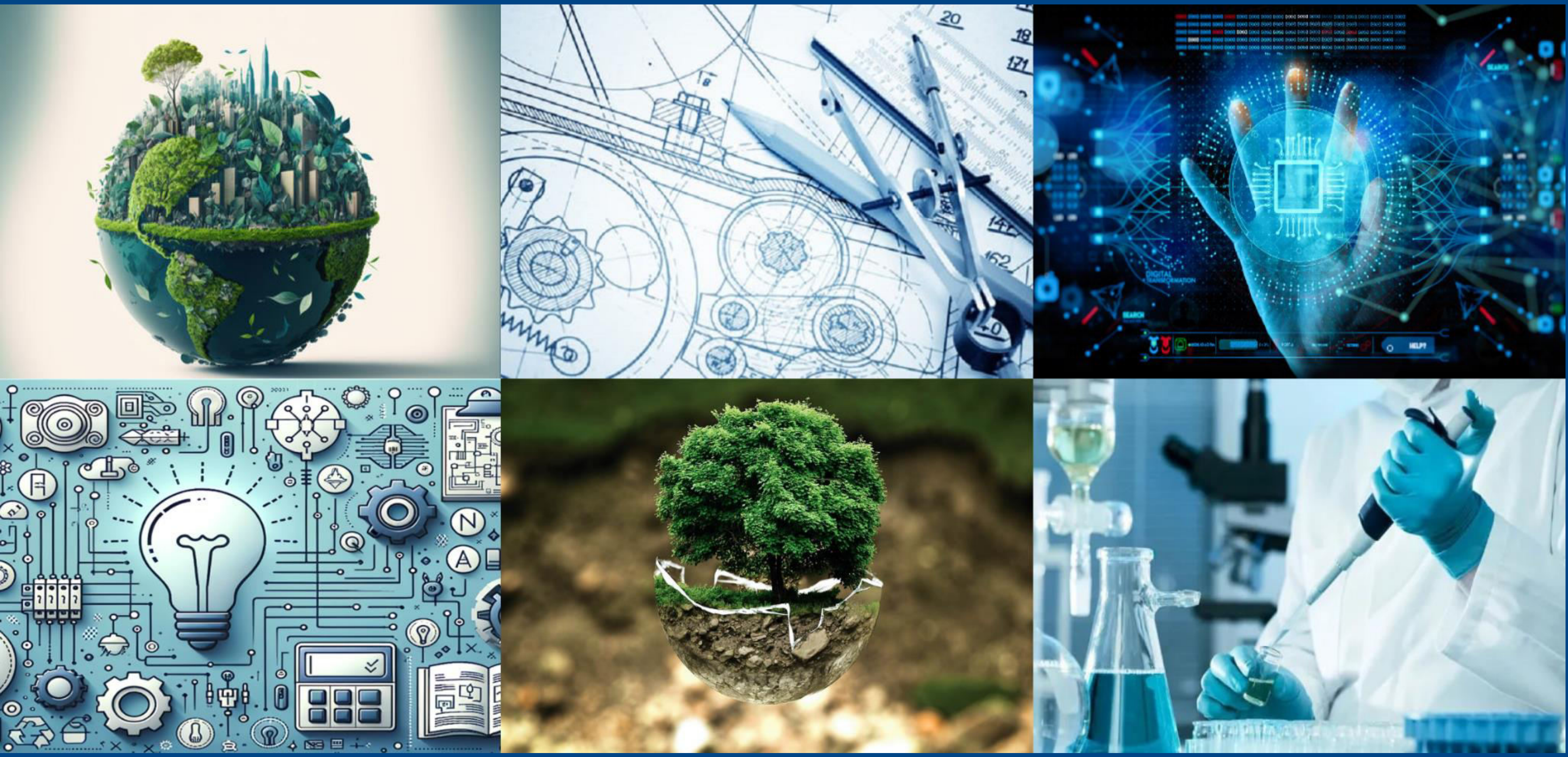
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 4, April 2025

🌐 www.ijmrset.com ✉ ijmrset@gmail.com ☎ +91 6381 907 438 ☎ +91 63819 07438

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804100

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Deeplearning Approaches for Robust Deepfake Detection and Localisation

K.Spandana[1], M.Sivalingareddy[2], R.SyamKumar[3], B.Balaji[4]

Professor, Department of ECE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, A.P, India[1]

Undergraduate Students, Department of ECE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, A.P, India[2-4]

ABSTRACT: In today's digital era, the proliferation of deepfake technology has raised significant concerns about the authenticity and integrity of multimedia content. Deepfake involves creation of highly realistic fake images and videos and misuse them for spreading fake news, defaming individuals, and possess a significant threat to the integrity of digital content.

To combat this challenge, our project "Deep learning approaches for robust deepfake detection and localisation" aims to address this critical issue by developing a robust system for identification and localization of deepfake content by introducing 'Densenet121'. Based on Densenet121 , we proposed a framework with the Multiscale Adapter, which can capture short- and long-range forgery contexts for efficient fine- tuning. For better identification 'Grad-Cam' module is proposed. The accurate identification of tampered regions holds the utmost importance for identifying the intentions of the offenders which is performed using Binary mask. This proposed framework seamlessly integrates forgery detection and localization. KEYWORDS: Deep Learning, Densenet121,Grad-Cam,Binary mask,Detection.

I. INTRODUCTION

In today's digital era, the proliferation of deepfake technology has raised significant concerns about the authenticity and integrity of multimedia content. Deepfake involves creation of highly realistic fake images and videos and misuse them for spreading fake news, defaming individuals, and poses a significant threat to the integrity of digital content.

To combat this challenge, our project "Deep learning approaches for robust deepfake detection and localisation" aims to address this critical issue by developing a robust system for identification and localisation of deepfake content by introducing 'Densenet121'. Based on Densenet121, we proposed a framework with the Multiscale Adapter, which can capture short- and longrange forgery contexts for efficient fine-tuning.

For better identification, the 'Grad-Cam' module is proposed. The accurate identification of tampered regions holds the utmost importance for identifying the intentions of the offenders, which is performed using a Binary mask. This proposed framework seamlessly integrates forgery detection and localization. Additionally, our approach focuses on ensuring scalability and real-time detection, which makes it suitable for large-scale applications like social media platforms, news agencies, and other digital spaces where deepfake content is a growing concern.al for efficient information retrieval and understanding.

Deep Learning models can produce high-quality summaries. Recent advancements in deep learning, including the development of self-attention mechanisms and transformer models, have enabled the creation of highly effective abstractive summarization models.

II. MODEL IMPLEMENTATION

We employ three different deep learning models for robust deepfake detection and localization:

1. Densenet121 Model: A convolutional neural network (CNN) model that excels in image classification tasks. We adapt Densenet121 for deepfake detection by incorporating it into our framework to capture both local and global features. By using a dense connectivity pattern, Densenet121 promotes efficient information flow, which is critical for detecting subtle tampering in deepfake images and videos. Fine-tuned for deepfake detection tasks, it can identify fake content by learning from large datasets of authentic and manipulated media.

Advantages:

- High accuracy in detecting deepfake content.
- Efficient feature extraction with reduced overfitting
- Integration in GUI:
- Users can select Densenet121 model for deepfake detection.
- Detection results displayed along with the probability score of authenticity.

2. Multiscale Adapter with Densenet121: This model integrates a Multiscale Adapter with the Densenet121 architecture, which allows the system to detect and localize deepfake content by capturing both short- and long-range contextual information. The adapter helps in efficient fine-tuning of the model by enabling it to focus on forgery patterns across multiple scales, improving the localization of tampered regions.

Advantages:

3. Improved localization of tampered regions.
4. Captures both global and local manipulation patterns efficiently.
5. Integration in GUI:
6. Option to enable Multiscale Adapter for enhanced localization.
7. Displays identified tampered regions and confidence levels.

3. Grad-Cam Module: Grad-Cam is used for visualizing and interpreting the regions of an image or video that contribute most to the deepfake classification decision. By providing heatmaps that highlight manipulated areas, this model aids in both detection and localization, making it easier for users to understand the system's decision-making process and to identify the specific regions of the content that are tampered.

Advantages:

- Provides visual explain ability of model predictions.
- Identifies and localizes tampered regions clearly with heatmaps.
- Integration in GUI:
- Users can toggle Grad-Cam for visualizing tampered regions.
- Heatmaps displayed along with a clear summary of localized areas.

Workflow:

Image/Video Pre-processing ➜ Feature Extraction with Densenet121 ➜ Multiscale Adapter for Contextual Analysis ➜ GradCam for Visualization ➜ Detection and Localization Results.

Integration in GUI:

- Users can choose from Densenet121, Multiscale Adapter, or Grad-Cam for different detection tasks.
- Summary displayed with tampered region localization and model accuracy.

III. SYSTEM ARCHITECTURE

1.1 Architecture of the proposed model

A cutting-edge architecture called DenseNet121 has demonstrated exceptional performance in image classification challenges. DenseNet, which stands for "Densely Connected Convolutional Networks," gets its name from the feedforward connections it makes between each layer and every other layer. Two salient characteristics of DenseNet distinguish it from other CNN architectures. Its dense block structure is the first feature; every layer is feed forwardly coupled to every other layer. Secondly, it makes use of bottleneck layers, which assist in lowering the number of parameters without lowering the total amount of characteristics the network learns. Every convolutional layer of a conventional feed-forward convolutional neural network (CNN), with the exception of the first, which receives input, gets the output of the preceding convolutional layer and generates an output feature map, which is then forwarded to the following convolutional layer. As a result, there are 'L' direct connections for 'L' layers, one for each layer and the layer after that. The 'vanishing gradient' issue, however, appears as the CNN gets deeper—that is, as the number of layers increases. This implies that certain information may vanish or get lost when the path for information from the input to the output layers lengthens, which lowers the network's capacity for efficient training. By altering the conventional CNN architecture and streamlining the layer-to-layer connectivity structure, DenseNets alleviate this issue. Densely Connected Convolutional Network is the name given to an architecture in which every layer is directly connected to every other layer: DenseNet. There are L(L+1)/2 direct connections for 'L' layers.

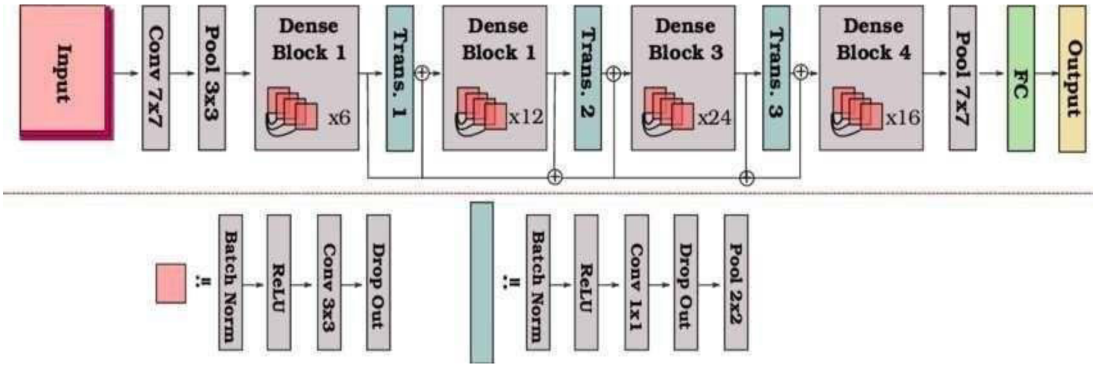


Fig 1: DenseNet121 Architecture

1.2 Working of the model

The Deepfake detection and localization system using DenseNet121 and binary masking operates through a sequence of steps to accurately classify and localize the deepfakes. Firstly, the system is given a dataset consisting of a combination of real and fake images obtained from publicly available repositories like Kaggle. These images undergo preprocessing step which includes image resizing, noise removal, and pixel value normalization thereby standardizing the data. Then, the processed dataset is fed to DenseNet121 model which learns features from the input images. As the images pass through the DenseNet121 layers, low level features like edges, lines, and corners are identified in early layers and high-level features like shapes, textures and local object parts are more focused in the deeper layers. The training phase unfolds iteratively, with batches of preprocessed images being fed into the model. Throughout this iterative process, the model fine-tunes its parameters to minimize the disparity between predicted and actual labels, thereby refining its ability to distinguish between authentic and manipulated imagery. Upon completion of training, the efficacy of the trained model is scrutinized through validation and testing using a distinct dataset. A set of evaluation metrics, including accuracy, precision, recall, and F1 score, is computed to gauge the model's performance in deepfake detection and localization. ‘140k Real and Fake Faces’ is the dataset used for training the DenseNet121 model which contains images of various sizes so in the preprocessing step all the images are resizes to 224*224 pixels. The localization is carried out using Binary masking. If the image is identified as fake image, then the image is first converted to gray scale and then a specific threshold is selected to convert that gray scale image to binary image.

Regions highlighted in white within the binary image signify areas where tampering has occurred, facilitating precise localization of deepfake alterations.

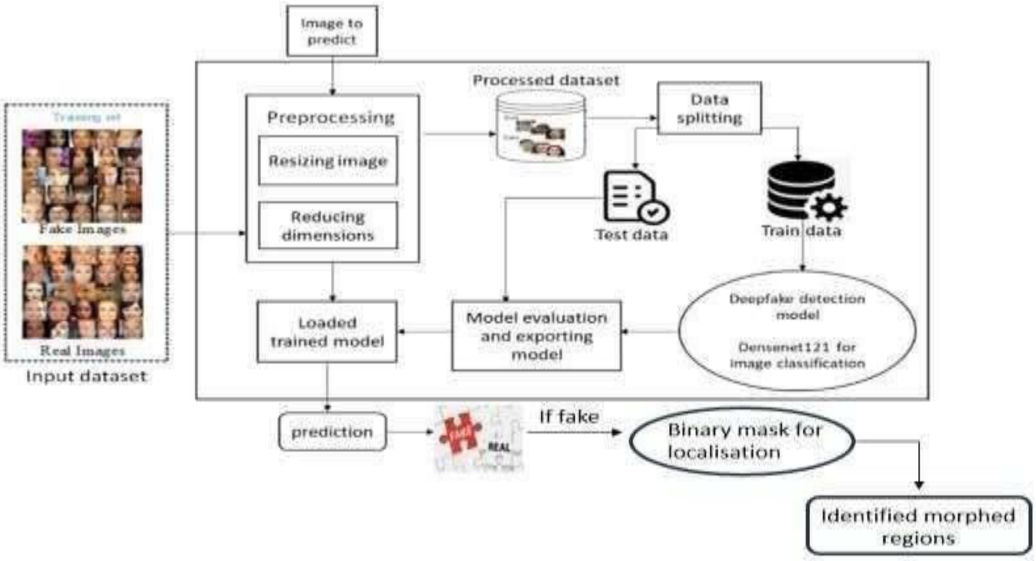
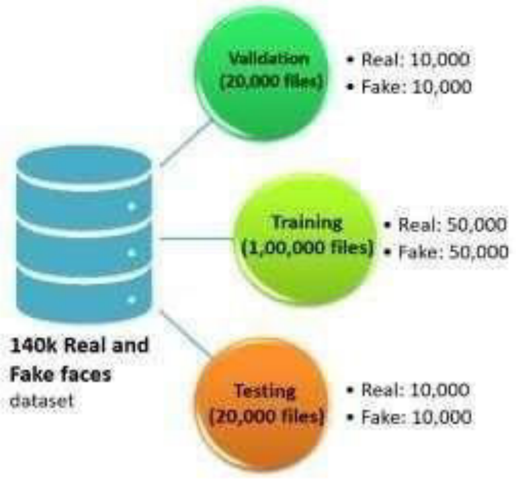


Fig 2: Process flow of proposed system

1.3 Dataset Analysis

The dataset utilized in this project, named '140k Real and Fake Faces,' comprises 70,000 authentic facial images sourced from the Flickr dataset curated by Nvidia, alongside 70,000 fabricated facial images sampled from the extensive 1 million FAKE faces corpus produced by StyleGAN. This amalgamation ensures a diverse representation of both real-world and synthetic facial features. The dataset is meticulously partitioned into training, testing, and validation subsets to facilitate robust model development and evaluation. Each subset maintains a balanced distribution of real and fake faces, with an equal 50% split between the two categories, fostering fair learning conditions.

Within this framework, the 140,000 images are meticulously allocated, with 20,000 images dedicated to validation, evenly distributed between 10,000 real and 10,000 fake faces. The training set encompasses 100,000 images, consisting of 50,000 real faces and an equivalent number of fake faces, ensuring parity in class representation during model training. Lastly, the testing set comprises 20,000 images, meticulously divided into 10,000 real and 10,000 fake faces, serving as an independent benchmark to assess model performance on unseen data. This comprehensive dataset structure, meticulously organized and balanced, underpins the efficacy and reliability of the ensuing model for discerning real from fake facial images.

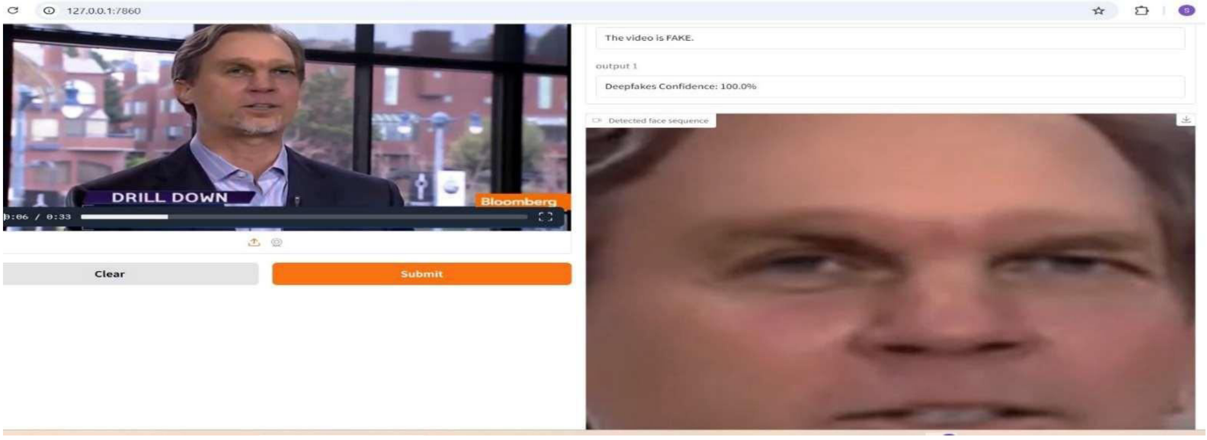


IV. RESULTS AND PERFORMANCE ANALYSIS

1.1 Training data

The training process entails the iterative refinement of a neural network model designed to distinguish between real and fake faces. Initially, the dataset of 140,000 images is prepared, involving standard preprocessing steps and partitioning into training, validation, and testing subsets. The model architecture, likely comprising convolutional and fully connected layers, is then defined using Keras, and the model is compiled with specific optimization settings and evaluation metrics. Training unfolds over multiple epochs, with each epoch involving the sequential processing of batches of training data through the network. The model updates its parameters based on computed loss, aiming to minimize the difference between predicted and actual labels. Simultaneously, validation datasets are used to monitor generalization and detect overfitting. Progress logs track metrics such as loss and accuracy for both training and validation datasets, with a focus on improving validation loss over epochs. Upon completion of training, the final model is evaluated on a separate testing dataset to assess its generalization performance. The process continues until a stopping criterion is met, such as convergence or a predefined number of epochs. Overall, this training process exemplifies the iterative refinement of a neural network model to effectively classify images as real or fake, demonstrating the robustness and adaptability of deep learning techniques in image recognition tasks.

1.2 Model prediction results

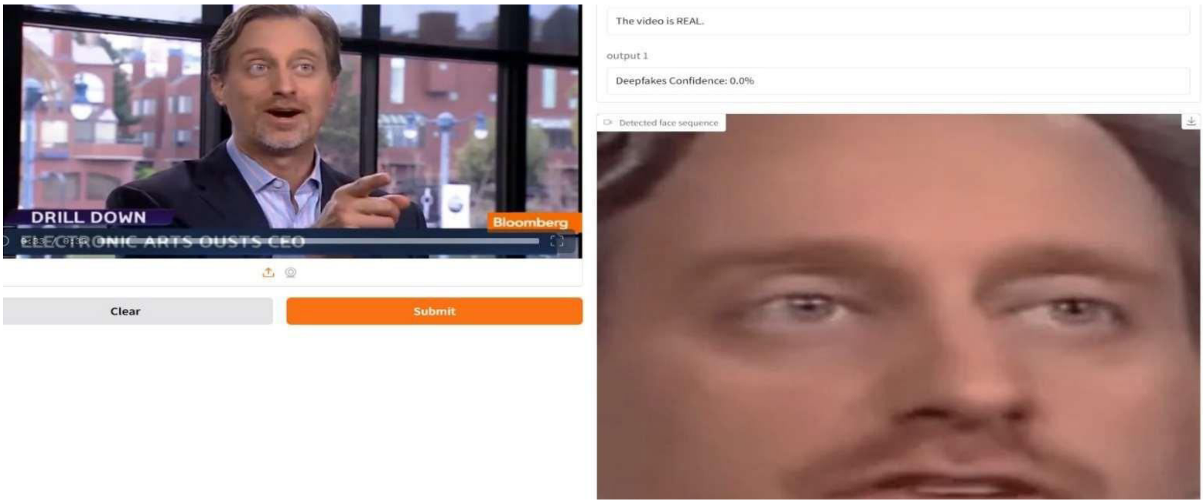


127.0.0.1:7860

The video is FAKE.

output 1

Deepfakes Confidence: 100.0%



The video is REAL.

output 1

Deepfakes Confidence: 0.0%

V. CONCLUSION AND FUTURE SCOPE

1.1 Conclusion

The results of successful deepfake detection and localization extends far beyond mere expectation. This proposed model using DenseNet121 architecture and binary masking has yielded promising results., exhibiting a remarkable accuracy of 99.43%. Through extensive experimentation and analysis, we have demonstrated the efficacy of our approach in accurately identifying deepfake content and localizing manipulated regions within multimedia files. The DenseNet121 model exhibited robust performance in classifying the images as genuine or manipulated, achieving high accuracy rates and effectively distinguishing between authentic and deepfake content. Furthermore, the incorporation of binary masking techniques enabled precise localization of manipulated regions within media files, providing valuable insights into the areas affected by synthetic alterations. By identifying and isolating these regions, our method facilitates the mitigation of potential risks associated with deepfake dissemination, including misinformation, privacy breaches, and targeted attacks. Overall, our project contributes to the ongoing efforts in combating the spread of deepfake content and safeguarding the integrity of digital media platforms. The combination of DenseNet121 classification and binary masking localization offers a promising approach to detecting and mitigating the impact of synthetic media manipulation, empowering users with the tools needed to verify the authenticity of multimedia content and preserve trust in digital communication channels. In conclusion, our project represents a significant step towards addressing the challenges posed by deepfake technology, emphasizing the importance of collaborative efforts in developing innovative solutions to protect the integrity and authenticity of multimedia content in the digital age.

1.2 Future Scope

Future research directions may focus on enhancing the scalability, efficiency, and robustness of deepfake detection and localization systems. Additionally, exploring the integration of multimodal analysis techniques and advanced adversarial detection methods could further strengthen the resilience of anti-deepfake solutions in combating evolving threats in the digital landscape. One potential direction is the integration of multi-modal analysis techniques to augment the capabilities of the detection system. By incorporating additional modalities such as audio and text, it may be possible to create more comprehensive models capable of detecting sophisticated deepfake content across various media formats. This could enhance the robustness of the system and improve its performance in identifying manipulated content that may evade traditional visual-based detection methods. Moreover, ongoing advancements in deep learning architectures and algorithms offer opportunities to improve the performance and efficiency of the detection system further. Continual refinement and optimization of the DenseNet121 model, as well as exploration of novel architectures tailored specifically for deepfake detection, could lead to significant advancements in the field. Additionally, leveraging techniques such as transfer learning and domain adaptation could enable the adaptation of pre-trained models to new domains and scenarios, enhancing the versatility and applicability of the detection system across different contexts. In conclusion, the project on deepfake detection and localization presents a rich landscape for future research and development, with numerous opportunities for innovation and advancement. By exploring these avenues, researchers can contribute to the ongoing efforts to combat the proliferation of deepfake content and safeguard the integrity of digital media platforms in an increasingly complex and dynamic landscape.

REFERENCES

[1]. LIY, C.M., InIctuOculi, L.: Exposingaicreated fakevideosbydetectingeyeblinking. In: IEEE WIFS (2018)
[2]. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation (1997)
[3]. Dang, H., Liu, F., Stehouwer, J., Liu, X., Jain, A.K. On the detection of digital face manipulation. In: IEEE CVPR (2020)
[4]. Nguyen, H.H., Yamagishi, J., Echizen, I.: Capsule-forensics: Using capsule networks to detect forged images and videos. In: IEEE ICASSP (2019)
[5]. Nguyen, H.H., Fang, F., Yamagishi, J., Echizen, I.: Multi-task learning for detecting and segmenting manipulated facial images and videos. In: IEEE BTAS (2019)
[6]. Dang, H., Liu, F., Stehouwer, J., Liu, X., Jain, A.K.: On the detection of digital face manipulation. In: IEEE CVPR (2020)
[7]. Kong, C., Chen, B., Li, H., Wang, S., Rocha, A., Kwong, S.: Detect and locate: Exposing face manipulation by semantic-and noise-level telltales. IEEE TIFS (2022)
[8]. Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F., Guo, B.: Face x-ray for more general face forgery detection. In: IEEE CVPR (2020)

[9]. arXiv:2306.17075 [cs.CV]

[10]. Budati, M., Karumuri, R. An intelligent lung nodule segmentation framework for early detection of lung cancer using an optimized deep neural system. Multimed Tools Appl (2023). https://doi.org/10.1007/s11042-023- 17791-8

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com