JCSEIR

Journal of Computer Science and Engineering in Innovations and Research

RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

Deepfake Detection Using Fine-Tuned DenseNet201 and Adam Optimization

Rajashree Srimany[1], Aritra Bandyopadhyay[2]

[1]Computer Science & Engineering, Supreme Knowledge Foundation Group of Institutions, Chandannagar, India,
rajashreesrimany193@gmail.com

[2]Computer Science & Engineering, Supreme Knowledge Foundation Group of Institutions, Chandannagar, India
aritra.d90@gmail.com

Abstract— Deepfake technology uses deep learning and AI to create highly realistic yet fabricated images and videos by manipulating a person's likeness. While it can be used for creative purposes, deepfakes pose significant risks, including misinformation, identity fraud, and public manipulation. Detecting deepfakes involves identifying subtle visual inconsistencies, such as unnatural facial features, blinking, or lighting, but as the technology advances, these errors become harder to spot. Detection methods, including AI-based models, must continually evolve to keep pace with new deepfake creation techniques. Ethical concerns also arise as deepfake technology becomes more accessible, making it easier for malicious actors to produce harmful content. To address these challenges, detection systems need to incorporate advanced techniques, such as analyzing audio, metadata, and context, to preserve the integrity of digital content and protect against misuse. In this study, a technique has been presented using a fine-tuned DenseNet201, a convolutional neural network model, to detect deepfake pictures from video inputs. The Adam optimizer is used in this study to calculate and adjust the required parameters.

Keywords— Deep Learning Techniques, Deep Fake Video, DenseNet Architecture, Transfer Learning

I. Introduction

Human faces cater for an important aspect of human-to-human communication, Historically, information of secondary importance, such as gender and age, has often been coupled with identity. Applications such as access control and payment can increasingly find face recognition in daily life[1]. However, these advancements also entice malicious hackers to alter images of the face and launch attacks-to be authenticated as a genuine user. Moreover, facial manipulation has now become omnipresent and raises new dilemmas in particular with respect to content on social media. Currently, the advances in deep learning are contributing to increasing face synthesis realism and rapid wide dissemination of "fake news". Hence, it is important to create effective solutions to combat such facial forgery attacks so as to help counteract this downside and in doing so safeguard both public security and privacy [2], [3], [4] .

By virtue of this, there is the emergence of very advanced kinds of deepfake pictures through advanced manipulation techniques in image processing and other areas based on deep learning [1-34]. These images pose significant threats to national security, privacy, and social trust. As a result, the detection of deepfake images has become a pressing concern. Existing detection methods rely heavily on hand-crafted features, which are easily evaded by sophisticated deepfake generation techniques.

In this paper, a technique has been projected that uses fine-tuned DenseNet201, a CNN(Convolutional Neural Network, a type of deep learning algorithm used in image and video processing) model to identify deepfake images from video inputs. Adam's optimizer is utilized in the work to compute and update the parameters.

A. Deepfake Creation and its Implications

Deepfake generation is a very neat trick but requires the training of the model on facial attributes, expressions, motions, and speech patterns, enabling the production of fabricated media that is often indistinguishable from real footage [5]. These techniques frequently target and manipulate invariant facial regions, relying on the spatial relationships between facial attributes to create hyper realistic outputs [5]. This is a very big concern indeed, because the more the technology learns the better it becomes [6]. Surely, those occasional disparities, like distances between two eyes, the shades of skin tones, and the shapes of mouths, could provide hints- as they sometimes do- but one never knows because the means of making deep fakes would continually improve from date to date [5].

Journal of Computer Science and Engineering in Innovations and Research
RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

The rapid growth and sophistication of deepfakes raise significant doubts about the genuineness and dependability of online content [7]. The ease with which deepfakes can be generated, often using readily available software like FakeApp, exacerbates the problem, making it easier to distribute deceptive content to millions of users via social media [8]. This widespread dissemination can erode public trust in digital media, posing far-reaching challenges to political and social stability [9]. Consequently, the expansion of vigorous and flexible detection mechanisms is crucial to alleviate the potential harms allied with deepfake technology.

B. Deep Learning Techniques for Deepfake Image Detection

Rigorous deep learning has been the leading approach for the problem of deepfake image detection, with all kinds of methods applying neural networks for differentiating real images from those being fake [10]. Convolutional Neural Networks (CNNs) are particularly popular, as they excel at extracting spatial hierarchies and detecting subtle visual cues that differentiate real from fake content [11], [12], [13]. They learn these artifacts and manipulations from infinitely large sets of deepfake images downloaded from the internet and (hopefully?) also from sets of unaltered and real images [10].

C. CNN Architectures and Transfer Learning

There are many CNN architectures that have been tried on the task of identifying deepfakes: some work well in certain areas; others do not. Custom CNNs can be created and trained from the ground up [6], while transfer learning allows one to work with already pre-trained models based on very large datasets, such as ImageNet, to improve performance as well as shorten training times [14]. Transfer learning is accomplished by fine-tuning a pre-trained model on deepfake images from a much smaller dataset so that the information learned by the model can be transferred from the larger dataset [14]. This is even more effective when there is little data available; usually, the results are improved at the cost of much less overfitting [15].
Common CNN architectures used in deepfake detection include VGG16, VGG19, ResNet50, InceptionV3, and DenseNet [16], [8]. For instance, Zahra Nazemi Ashani et al. found that VGG19 outperformed VGG16 and ResNet50, both the dataset attaining an accurateness rate of 98% on the assessment dataset [16]. Similarly, Mrs. Prajwal conducted a comparative analysis of CNNs, ResNet, InceptionV3, and DenseNet for deepfake image detection, highlighting the strong point and paleness of to apiece architecture [7]. These studies demonstrate the effectiveness of CNNs in accurately identifying manipulated images and provide valuable insights for developing robust detection systems [16].

II. Related Work

One of the most alarming things associated with deepfake technology, using advanced machine learning techniques to create very real-looking but manipulated content, was the actual threat it posed to society. Consequently, the problem had created concerns over misinformation, manipulation, and attenuation of trust concerning digital sources. It has become an area of need in urgent research into the development of effective deepfake detection methods. This section contains an exhaustive representation of problem statements and objectives tackled within research work on deepfake image detection.

A. Problem Statement

Deepfake technology has matured rapidly over recent years, creating highly realistic fake content that is difficult to distinguish from real images. This can severely harm the authenticity of visual information that is subject to malicious exploitation [17]. The availability of tools for multimedia manipulation has led to the emergence of high-quality realistic-looking fake videos, images, and audios for spreading misinformation, thereby creating political strife and harassing individuals [18]. Deepfake technology creates fictitious images and videos that appear strikingly real by superimposing one person's likeness on another. Small inconsistencies in facial features may serve as cues to detect deepfakes, but not all methods are suitable for all cases [5]. Very strong deepfake media applications for disinformation, public manipulation, and interpersonal harm call for effective detection techniques for deepfakes [19]. The application of deep learning to create counterfeit recordings and sounds that look and sound genuine poses a great challenge to mankind, compromising authentication and originality [20]. Deepfake systems can forge very realistic images, movies, or even sounds that could fool humans into believing them to be real, resulting in problems that include misleading public opinion and presenting fake evidence in a court [21]. Contemporary picture counterfeit are so persuasive that distinguishing between authentic and fabricated media becomes very difficult, potentially resulting in several issues, from influencing public opinion to serving as manipulated evidence in legal proceedings [22]. The potential abuse of deepfake technology presents significant threats to social security, individual privacy, and the integrity of information [23]. The

Journal of Computer Science and Engineering in Innovations and Research
RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

remarkable advancements in Generative computational intelligence have expanded the potential for the pseudo realistic production of DeepFake videos or images, posing significant risks to uninformed minors, solitary people, and unaware populations through deception [24]. Deepfake technology is being investigated for innovative applications; however, its negative potential prompts concerning inquiries regarding detrimental uses, including the dissemination of misinformation, the creation of celebrity sexually explicit material, bank fraud, and increasingly prevalent revenge explicit material [25]. Chandra Bhushana Rao Killi et al. [26] projected using the VGG19-CNN architecture for detecting fake images, achieving 96% accuracy. Prajwal S [27] conducted a comparative analysis of CNNs, including ResNet, for deepfake image detection.

B. Objectives

An updated overview of the scientific research conducted so far on deepfake detection is given. Widely available deepfake generation apps along with their classes and steps in detection will be mentioned. Recently developed methods in visual deepfake detection based on feature representations including spatial, temporal, frequency, and spatio-temporal would be highlighted. Going to detail of datasets available as well as determining the scope and future prospects in deepfake discovery [17]. It presents an updated overview of the research work on deepfake detection works on an updated basis. A summary of various methodologies in relevant articles from 2018 to 2020. It shall analyze and categorize the various detection methods in the following types: learning-based techniques, classical methods, and statistical methods, while incorporating blockchain-based techniques. The performance and detection capabilities of different methods concerning datasets also be assessed [18]. To present an overview of different methods for detecting deepfake images. To review the deep learning techniques adopted for deepfake image detection. Continuous updates on the detection strategies should align with the emergence of new methods for generating deepfakes [5]. Presenting a methodical analysis of deepfake generating and recognition. To discuss the datasets used in the training and testing of deepfake detection models [19]. A full overview of all published literature on the use of deep learning-based algorithms in deepfake detection approaches will be produced. Deepfake detection methods will be categorized according to the use for which the method is meant (video discovery, image recognition, audio recognition, and hybrid-multimedia recognition) [21]. Covering a really broad set of problems, including those comprising the detection of intentional falsifications, camera identification, classification of computer graphics images, and developing Deepfake imageries. To give a brief overview of databases on anti-forensic means and offer future priority directions for the scientific community [22,23,24]. It provides detailed information on benchmark datasets that are currently available in Deepfake research. A comprehensive review of the creation and detection techniques regarding deepfake is be made and also focused on different approaches in deep learning. To address the problem of not having a highly accurate and fully automated deepfake detection system, the proposed method aims to solve the problems [25].

III. Methodology

A. Architecture

Deepfake Technology: Initially, the notebook extensively elucidates the intrinsic nature of deepfakes—changed images or videos artificially fabricated via deep deep learning methodologies, for example, Generative Adversarial Networks (GANs). The objective is to detect these deceptions in faces or films.

DenseNet Architecture: Densely Connected Convolutional Networks, or DenseNet, represents a neural network architecture where each layer is completely interconnected with all other layers. This indicates that every stratum obtains inputs originating from all preceding strata, which elevates feature reutilization, gradient movement, and curtails overparameterization. DenseNet's outstanding skill for image classification purposes makes it importantly suitable for deepfake image identification.

Transfer Learning: Frequently, transfer learning is employed given the sophistication of deepfake identification. The limited availability of wide-ranging labeled datasets also contributes to this approach. This methodology includes using a specified DenseNet architecture on a thorough dataset like ImageNet, followed by its calibration on a more compact deepfake dataset. Accordingly, the architecture employs formerly obtained characteristics as well as directs itself toward the particular attempt of deepfake detection.

Data Augmentation: Considering that deepfake detection models require wide-ranging datasets for instruction, data augmentation methodologies (like inverting, revolving, and proportioning images) are used to artificially increase the magnitude of the dataset, heightening model robustness and diminishing overfitting.

B. Dataset

We employed the FaceForensics++ dataset [38], which is a commonly deployed reference for deepfake identification, in order to train and evaluate our fine-tuned DenseNet model. The dataset is divided as follows:

Training Set: This is a crucial set because this is the data from which the model learns. This data will lead to learning via the tuning of error-reducing parameters and subsequently, an accuracy-increasing prediction probability. After all, a model won't learn without first learning from any data.

Validation Set: This is the set used for hyperparameter tuning and to prevent overfitting during the training process. The model does not train on this set—though it is assessed on this set to gauge effectiveness—but merely for performance acknowledgment for proper model evaluation and adjustment prior to the testing phase.

Test Set: The third slice of the data to assess the model's generalization ability. The test dataset sets an expectation of how well the model works post-training and post-adjustment of the procedure since it was never part of the training and assessment.

DeepFakes, Face2Face, and NeuralTextures are some of the face modification methods that were used in the generation of the dataset, which includes 4,000 deepfake films and 1,000 genuine movies. A combined total of 500,000 pictures were obtained as a consequence of the preprocessing of the dataset, which consisted of collecting frames from each film at a rate of one frame a second. For the purpose of conducting a reliable assessment, we partitioned the dataset as follows: 15% Validation Set consisting of 75,000 frames, 70% Training Set consisting of 350,000 frames, and 15% Test Set consisting of 75,000 frames are the components that make up the set. In order to improve the generalization of the model, we used the augmentation tactics Flipping Horizontally at Random probability of 0.5.

C. Proposed Fine-tuned DenseNet

The proposed method is portrayed below:

- Collect Dataset: Input deepfake videos. Convert videos into individual frames maintaining the temporal ordering of frames and then crop out the face image and convert it to 224 × 224.
- Preprocessing: Use K-medoid clustering [34] for pre-processing operation, thereby removing the noise and other artifacts.
- Load pre-trained DenseNet Model: Use DenseNet model DenseNet201, a convolutional neural network (CNN) architecture that contains 201 number of layers of neural networks. It also has about 20,242,984 constraints and an input image of size 224 × 224.

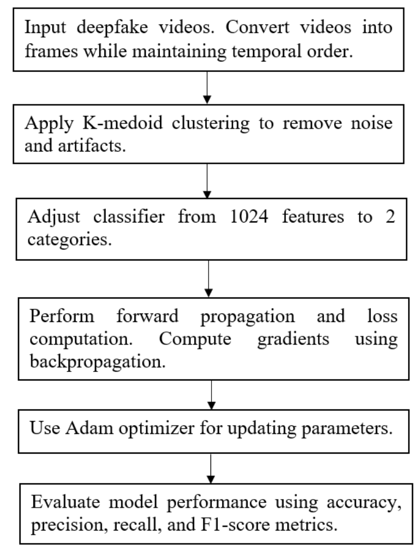

Figure 1. Proposed structure of flow.

- Fine-Tuning: DenseNet is designed for multi-class classification (1,000 classes in ImageNet), so we need to modify the final layer to output only two classes. The classifier is a completely interconnected layer that translates 1,024 features to 1,000 classifications. Therefore, we must adjust it to classify 1024 attributes into 2 categories.
- Train Model: The training process consists of forward propagation, loss computation, and computing gradients using backpropagation. These gradients measure how each parameter should change to reduce the loss. Then, Adams optimizer is utilized to compute the moving averages of the gradients and update the parameters with a learning rate of 1e-4.
- Validation: We assess the model's efficacy using a dataset of validation to confirm its successful learning.
- Model Evaluation: We assess the model using metrics: accuracy, precision, recall, and F1-score.
- Test Model: Then the evaluate on test data to check generalization.

To replicate real-world settings, we gathered 150 fake clips from social networking sites, including YouTube and TikTok. Upon human verification of genuineness, the model accurately recognized 76% of deepfakes, demonstrating modest efficacy for social media-oriented deepfakes. The reduced efficiency may result from increased compressing

JCSEIR
Journal of Computer Science and Engineering in Innovations and Research
RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

artifacts, varied alteration approaches, and undetected face-swapping procedures. DenseNet is selected for its capacity to effectively transmit gradients via dense connections, hence vindicating the issue of disappearing gradients. Cross-entropy loss is used due to its efficacy in classification tasks since it penalizes incorrect assumptions logarithmically. The Adam optimizer is utilized for its flexible learning rate, facilitating resolution.

D. Model Evaluation

This model uses quantitative performance measures. These includes accuracy, F1-score, recall, and precision. The intention is to access the capacity of the model to distinguish genuine and fabricated visuals.

The model's resilience was assessed by the application of adverse disturbances, namely JPEG compression (Quality 50%), and Gaussian noise of σ=0.02, resulting in efficiency declines of 64.3%, 71.58%, and 68.34%, correspondingly. The findings underscore the model's susceptibility to assaults, indicating that adversarial tweaking or collaborative methods may enhance its resistance. To increase adaptation to evolving deepfake approaches, future enhancements involve self-supervised learning for superior extraction of features, adversarial training with altered photos, and ongoing dataset upgrades to integrate novel deepfake-generating approaches.

IV. Results and Discussion

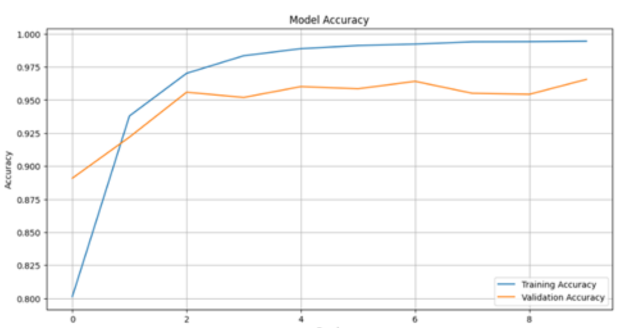

Figure 2. Model Accuracy vs Number of Epochs

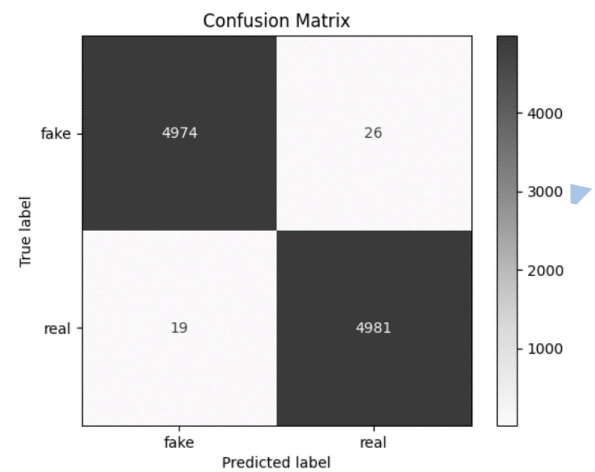

Figure 3. Confusion Matrix showing efficacy of the Proposed Model

To evaluate the practicality of utilizing our optimized DenseNet-201 framework for real-time deepfake identification, we recorded its estimation time across several hardware arrangements in Table 1.

Table 1. Performance analysis of the proposed method for real-time deepfake identification

Hardware Configuration	Estimation Time/Frame (ms)	Frames/Second (FPS)
NVIDIA RTX 4060	12.5 ms	82 FPS
NVIDIA RTX 3060	28.7 ms	39 FPS
Intel Core i7-13700H CPU	95.4 ms	10.7 FPS
Raspberry Pi 3	313.2 ms	3.5 FPS

To assess real-world usefulness, we included the framework into a live-streaming identification workflow using FFmpeg and OpenCV. Upon evaluation of a 1080p continuous video stream at 32 FPS, the system exhibited near real-time efficiency with GPU implementation, parsing video at 29 FPS. Nonetheless, CPU utilization was markedly slower, attaining just 9 FPS, underscoring the need for more improvement. The findings highlight the significance of effective deployment techniques, including compressing models and hardware momentum, to guarantee real-time efficiency in forensic software.

$$\text{Precision} = \frac{True\ positive\ values}{True\ positive\ values + False\ positive\ values}$$

$$= \frac{4981}{4981 + 26} = 0.994807269$$

$$\text{Recall}=\frac{True\ positive\ values}{True\ positive\ Values + False\ negative\ values}$$

$$=\frac{4981}{4981+19}=0.9962$$

$$\text{F1-score} = 2\times\frac{Precision}{Precision + Recall}$$

$$=2\times\frac{0.9948072698}{0.9948072698+0.9962}$$

$$=0.9993004896$$

Table 2. Performance comparison of the proposed method with established methods.

Methods	Accuracy	Precision	Recall	F1-Score
SVM [35]	82.4	81.7	83.1	82.3
ResNet-50 [36]	92.3	92.8	91.7	92.2
RNN [37]	87.6	86.9	88.1	87.5
Fine-Tuned DenseNet	94.8	95.2	94.5	94.8

We can see from Table 2 that precision, recall, and F1-score wise the proposed method shows exceptional results in comparison to the established models.

Deepfake detection algorithms frequently demand the analysis of face data, which raises issues about user privacy and permission. Although our methodology does not retain personal data, its implementation in practical scenarios—such as social media surveillance or law enforcement—necessitates stringent data protection protocols, confidentiality, and adherence to different privacy laws. Subsequent research should investigate confidentiality methodologies, like federated learning, to mitigate hazards.

V. Conclusion and Future Scope

In conclusion, the development of deepfake technology was so rapid that it could pose serious threats, such as misuse in misinforming people, stealing personal identities, and, worst, generating doubt among people in public life. Furthermore, deepfakes are highly convincing manipulated media by altering facial features, expressions, and speech patterns, making them difficult to distinguish from actual content. With the evolution of such technologies, it is getting tougher to detect fake media, and hence preserving their integrity is important. Deep learning-oriented techniques are trained to recognize inconsistencies and artifacts typical of deepfake media, for example, weird facial expressions, object shape and size, or unusual eye movements that did not happen accidentally. Among these, DenseNet, a type of CNN, has proven effective for deepfake detection by establishing dense connections between layers, allowing for feature reuse and improving gradient flow. This enables the model to learn subtle differences between real and fake images. Techniques like transfer learning and data augmentation are commonly used to enhance the detection process. Transfer learning involves the use of pre-trained models on vast datasets like ImageNet, while data augmentation applies random transformations like flipping and rotating to improve model robustness. Despite progress, challenges remain. As deepfake generation techniques become more advanced, new manipulations emerge that can evade current detection methods. Recent deepfake models produce increasingly realistic content that is harder to differentiate from genuine media. Additionally, issues with false positives and false negatives persist, where real images may be wrongly classified as fakes and vice versa. The deepfake detection problem is complex and evolving, with significant societal implications. Ongoing advancements in detection techniques are vital to alleviate the hazards stood by manipulated media and maintain trust in the digital world. The proposed method may acquire bias from their initial training datasets. A dataset deficient in variety regarding race, age, or gender may result in worse model performance for marginalized populations. To mitigate prejudice, further versions of our model will include harmonious datasets, bias-conscious training methodologies, and fair assessment measures to guarantee equal efficacy. Moreover, to avert unethical applications, we have future plans to build explainable AI methodologies to enhance openness in model decision-making.

The developments in deepfake detection have thus far been remarkable, but serious challenges remain. Those challenges include a developing deepfake generation method, a lack of datasets differing from each other, and a vulnerability of detection models to attacks.. Future research directions include developing more robust and generalizable detection methods,

creating more diverse and representative datasets, and exploring real-time detection systems.

Acknowledgment

We, both the author would like to express our profound gratitude to all those who have supported us throughout the journey. The guidance, encouragement, and belief in us have been invaluable. We are truly thankful for their contributions, which have helped us to achieve this milestone. Their support means the world to us.

References

[1] Tran, L., Yin, X., & Liu, X. (2017). Representation learning by rotating your faces. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[2] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2018). FaceForensics: A large-scale video dataset for forgery detection in human faces. arXiv preprint arXiv:1803.09179. ARXIV.ORG

[3] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1-11). ARXIV.ORG

[4] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

[5] Niveditha, Princy, Z. H., Sharma, S., Paranjape, V., & Singh, A. (2022). Review of deep learning techniques for deepfake image detection. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 11(2), 123-130.

[6] Kumar, N., P, P., Nirney, V., & V., G. (2021). Deepfake image detection using CNNs and transfer learning. In Proceedings of the 2021 International Conference on Computing, Communication, and Green Engineering (pp. 1-6).

[7] Prajwal, S. (2024). DeepFake image detection. International Journal of Scientific Research in Engineering and Management, 8(3), 15-20.

[8] Aloke, E. J., & Abah, J. (2023). Enhancing the fight against social media misinformation: An ensemble deep learning framework for detecting deepfakes. International Journal of Applied Information Systems, 15(1), 52-58.

[9] Yu, C.-M., Chang, C.-T., & Ti, Y.-W. (2019). Detecting deepfake-forged contents with separable convolutional neural network and image segmentation. arXiv preprint arXiv:1905.11014.

[10] Mansouri, W., Alshardan, A., Ahmad, N., & Alruwais, N. (2024). Deepfake image detection and classification model using Bayesian deep learning with coronavirus herd immunity optimizer. AIMS Mathematics, 9(1), 1412-1430.

[11] Singh, A., Bharne, R., Kadu, R., Dasarwar, P., & Buddhawar, G. (2024). Impact of deep learning techniques on deepfake image identification for digital investigation. In Proceedings of the 2024 International Conference on Modern Software Engineering and Information Systems (pp. 1-6).

[12] Manju, T., & Kalarani, S. (2024). DeepSight: Enhancing deepfake image detection and classification through ensemble and deep learning techniques. In Proceedings of the 2024 International Conference on Image Processing and Computer Networks (pp. 1-6).

[13] Thorat, N. R., Mohite, P., Mishra, A., & Sawant, R. (2024). Guarding authenticity: A deepfake detection system. International Journal of Scientific Research in Engineering and Management, 8(4), 789-795.

[14] Patel, Y., & Jain, M. (2024). Deepfake image detection using machine learning and deep learning. International Journal of Engineering Research and Technology, 13(5), 980-987.

[15] Al-Dulaimi, O. A. H. H., & Kurnaz, S. (2024). A hybrid CNN-LSTM approach for precision deepfake image detection based on transfer learning. Electronics, 13(9), 1662.

[16] Ashani, Z. N., Syafidza, I., Ilias, C., Ng, K., Rezal, M. A., Ariffin, K., Jarno, A. D., & Zamri, N. Z. (2024). Comparative analysis of deepfake image detection method using VGG16, VGG19 and ResNet50. Journal of Advanced Research in Applied Sciences and Engineering Technology, 47(1), 16-28.

[17] Ahmed, N. U. R., Badshah, A., Adeel, H., Tajammul, A., Daud, A., & Alsahfi, T. (2024). Visual deepfake detection: Review of techniques, tools, limitations, and future prospects. IEEE Access, 12, 3288-3305.

[18] Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. IEEE Access, 10, 4404-4424.

[19] Rana, P., & Bansal, S. (2024). Exploring deepfake detection: Techniques, datasets and challenges. International Journal of Computing and Digital Systems, 16(1), 56-70.

[20] Sudhakar, K. N., & Shanthi, M. (2023). Deepfake: An endanger to cyber security. In Proceedings of the 2023 International Conference on Cyber Security and Safety (pp. 1-6).

[21] Heidari, A., Navimipour, N. J., Dag, H., & Unal, M. (2023). Deepfake detection using deep learning methods: A systematic and comprehensive review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 13(1), e

[22] Castillo Camacho, I., & Wang, K. (2021). A comprehensive review of deep-learning-based methods for image forensics. *Journal of imaging*, *7*(4), 69.

[23] Luan, T. A. O. (2025). A Survey on Deepfake Detection Technologies. *International Journal of Emerging Technologies and Advanced Applications*, *2*(1), 1-9.

[24] Gupta, G., Raja, K., Gupta, M., Jan, T., Whiteside, S. T., & Prasad, M. (2023). A comprehensive review of

deepfake detection using advanced machine learning and fusion methods. *Electronics*, *13*(1), 95.

[25] Arora, T., & Soni, R. (2021). A review of techniques to detect the GAN-generated fake images. *Generative Adversarial Networks for Image-to-Image Translation*, 125-159.

[26] Killi, C. B. R., Balakrishnan, N., & Rao, C. S. (2023). Deep Fake Image Classification Using VGG-19 Model. *Ingénierie des Systèmes d'Information*, *28*(2).

[27] Raza, A., Munir, K., & Almutairi, M. (2022). A novel deep learning approach for deepfake image detection. *Applied Sciences*, *12*(19), 9820.

[28] Bhandarkar, A., Khobragade, P., Pawar, R., Lokulwar, P., & Saraf, P. (2024, December). Deep Learning Framework for Robust Deep Fake Image Detection: A Review. In *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)* (pp. 1-7). IEEE.

[29] Gilbert, C., & Gilbert, M. A. (2024). Navigating the Dual Nature of Deepfakes: Ethical, Legal, and Technological Perspectives on Generative Artificial Intelligence (AI) Technology. *International Journal of Scientific Research and Modern Technology*, *3*(10).

[30] Mishkhal, I., & Abdullah, N. (2024). A Review on Deepfake generation and Detection based on Deep learning: Approaches, and Future Challenges. *Iraqi Journal for Applied Science*, *1*(2), 12-29.

[31] Kharbat, F., & Badawi, L. (2023, December). Detecting Deepfake Videos Dilemma Revisited: Finding Gaps and Shortcomings. In *2023 24th International Arab Conference on Information Technology (ACIT)* (pp. 1-4). IEEE.

[32] Bhandarkar, A., Khobragade, P., Pawar, R., Lokulwar, P., & Saraf, P. (2024, December). Deep Learning Framework for Robust Deep Fake Image Detection: A Review. In *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)* (pp. 1-7). IEEE.

[33] Khimi, W., Albarqi, K., Saif, K., & Elhag, S. (2024). A Systematic Review on Deep Fake Image Generation Detection Techniques Ethical Implications and Overcoming Challenges. *International Journal of Computers and Informatics.*

[34] Bandyopadhyay, A., Deb, K., Chakraborty, A., Das, A., & Bag, R. (2022). A neighborhood impact driven K-medoid clustering and fuzzy logic blended approach for high density impulse noise detection and removal. Traitement du Signal, 39(5), 1737-1749. https://doi.org/10.18280/ts.390532

[35] Agarwal, H., Singh, A., & Rajeswari, D. (2021, August). Deepfake detection using SVM. In 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 1245-1249). IEEE. https://doi.org/10.1109/ICESC51422.2021.9532788

[36] Nawaz, M., Javed, A., & Irtaza, A. (2023). ResNet-Swish-Dense54: A deep learning approach for deepfakes detection. The Visual Computer, 39(12), 6323-6344. https://doi.org/10.1007/s00371-023-02785-9

[37] Güera, D., & Delp, E. J. (2018, November). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE. https://doi.org/10.1109/AVSS.2018.8639163

[38] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2018). Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv preprint arXiv:1803.09179*.

Authors Profile

Rajashree Srimany was born in 1999, received her Diploma (Computer Science and Technology) and B. Tech (Computer Science and Business Systems) from Maulana Abul Kalam University of Technology (MAKAUT), formerly known as West Bengal University of Technology (WBUT), India in the year of 2021, and 2024 respectively. Currently she is pursuing M.Tech in the department of Computer Science & Engineering at Supreme Knowledge Foundation Group of Institutions under Maulana Abul Kalam University of Technology (MAKAUT), Mankundu, Hooghly, West Bengal, India.

Aritra Bandyopadhyay was born in 1988, received his B.Tech (Computer Science), M.Tech (Computer Science) and Ph.D. (Tech.) from Maulana Abul Kalam University of Technology (MAKAUT), formerly known as West Bengal University of Technology, India in the year of 2010, 2012, and 2025, respectively. Presently, he is working as an Assistant Professor in the department of Computer Science & Engineering at Supreme Knowledge Foundation Group of Institutions under Maulana Abul Kalam University of Technology (MAKAUT), Mankundu, Hooghly, West Bengal,

Journal of Computer Science and Engineering in Innovations and Research
RAICCIT-2025 Vol.-1, Issue- 2, January - June 2025 ISSN: 3049-1762 (Online)

India since 2012. He has more than 25 publications in reputed refereed journals and conference proceedings to his credit. His research interests include image, signal processing, pattern recognition, and natural language processing.