



Optimization Accuracy of Deepfake Face Detection using DenseNet121 Technique

Shelke Sanket Babasaheb¹, Prof. Suresh S. Gawande²

M. Tech. Scholar, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal¹

Guide, Department of Electronics and Communication, Bhabha Engineering Research Institute, Bhopal²

Abstract: The rapid advancement of generative artificial intelligence has enabled the creation of highly realistic deepfake facial images and videos, creating significant challenges for digital media authentication, privacy protection, and cybersecurity. Accurate and robust deepfake detection is therefore essential for identifying manipulated facial content. This study proposes an optimization-based deepfake face detection framework using the DenseNet121 deep learning architecture. The proposed approach utilizes DenseNet121 to automatically extract discriminative hierarchical features from facial images and classify them into authentic and deepfake categories. The methodology includes dataset collection, image preprocessing, face-region extraction, data augmentation, feature learning, model training, and performance evaluation. DenseNet121 is selected because its dense connectivity facilitates effective feature reuse and enables the propagation of information and gradients across network layers, which can improve feature representation while reducing redundant parameters. To optimize detection performance, appropriate training parameters and classification settings are systematically adjusted, and the resulting model is evaluated using accuracy, precision, recall, F1-score, and confusion matrix analysis. The proposed framework is intended to improve the reliable identification of subtle facial manipulation artifacts while maintaining efficient deep feature extraction. The study also considers practical challenges such as image quality variation, compression, illumination changes, and differences between manipulation techniques. The expected outcome is a robust DenseNet121-based detection model capable of achieving improved classification accuracy and better generalization for deepfake face detection. The proposed approach can support applications in digital forensics, social-media security, identity protection, and trustworthy multimedia authentication.

Keywords: Deepfake Detection, DenseNet121, Deep Learning, Face Manipulation, Transfer Learning, Image Classification, Computer Vision, Digital Forensics.

I. INTRODUCTION

The rapid development of Artificial Intelligence (AI), particularly Generative AI and deep learning, has significantly transformed the process of creating and manipulating digital media. One of the most notable outcomes of this development is deepfake technology, which can generate highly realistic facial images and videos by replacing, modifying, or synthesizing human faces. Although deepfake technology has several legitimate applications in entertainment, filmmaking, education, and digital communication, its misuse has created



serious concerns related to misinformation, identity theft, privacy violations, fraud, cybercrime, and damage to personal reputation. Consequently, the development of accurate and reliable deepfake face detection techniques has become an important research problem in computer vision and digital forensics [1, 2].

Deepfake face detection aims to automatically distinguish between authentic facial content and artificially manipulated content. Traditional detection techniques generally depend on manually designed features, such as texture inconsistencies, abnormal facial boundaries, color variations, compression artifacts, and irregularities around the eyes and mouth. However, modern deepfake-generation methods are capable of producing increasingly realistic images, making these manually designed features less effective. Deep learning provides a promising solution because it can automatically learn complex and hierarchical representations directly from facial images. Convolutional Neural Networks (CNNs) have therefore been widely adopted for deepfake detection because of their ability to capture spatial patterns and subtle visual artifacts [3].

Among the different CNN architectures, DenseNet121 has significant potential for deepfake face detection. DenseNet121 introduces dense connectivity between layers, where each layer receives feature information from preceding layers. This architecture promotes feature reuse, improves gradient propagation, and enables the network to learn both low-level and high-level representations efficiently. In deepfake detection, such characteristics can help identify subtle manipulation artifacts that may occur in facial texture, skin regions, edges, facial boundaries, and other local structures. DenseNet121 can also be employed through transfer learning, where weights learned from a large-scale image dataset are adapted to the deepfake classification task. This approach can reduce training requirements and provide effective feature representations even when the task-specific dataset is comparatively limited [4, 5].

The performance of deepfake detection is strongly influenced by several factors, including dataset quality, preprocessing, image resolution, class distribution, data augmentation, learning rate, batch size, optimizer, and model architecture. Therefore, simply applying a deep learning model may not provide optimal performance. Optimization of the detection framework is required to improve classification accuracy and ensure that the model can distinguish genuine and manipulated facial content under different conditions. Data augmentation techniques such as rotation, flipping, scaling, cropping, and brightness variation can increase the diversity of training samples and reduce overfitting. Similarly, suitable hyperparameter selection and transfer-learning strategies can improve the learning capability of DenseNet121 [6].

A major challenge in deepfake face detection is the generalization problem. A model may achieve high accuracy on images generated using manipulation techniques represented in its training dataset but may perform poorly when exposed to unseen manipulation methods. In addition, real-world images and videos may contain compression, noise, changes in illumination, different facial poses, occlusions, and variations in image quality. These factors can remove or distort the subtle artifacts learned by the detector. Hence, an effective deepfake detection model should not only achieve high training and testing accuracy but

should also demonstrate strong precision, recall, F1-score, and robustness under diverse conditions [7, 8].

In this study, an Optimization Accuracy of Deepfake Face Detection using DenseNet121 Technique is proposed. The framework involves facial image acquisition, preprocessing, data augmentation, DenseNet121-based feature extraction, model training, optimization of relevant parameters, and binary classification into real and deepfake categories. The effectiveness of the proposed model is evaluated using standard performance measures,. The objective is to develop an accurate and robust deep learning-based framework capable of identifying subtle facial manipulation patterns and improving the reliability of automated deepfake detection.

II. DENSENET121 TECHNIQUE

DenseNet121 is a deep convolutional neural network (CNN) architecture that belongs to the Dense Convolutional Network (DenseNet) family. It was introduced to improve information propagation, feature reuse, and gradient flow in deep neural networks. Unlike conventional CNN architectures, where each layer generally receives input only from the immediately preceding layer, DenseNet connects each layer to every subsequent layer within a dense block. In DenseNet121, these dense connections enable the network to preserve and reuse previously learned features throughout the architecture. This property makes DenseNet121 particularly suitable for image classification and computer vision applications, including deepfake face detection.

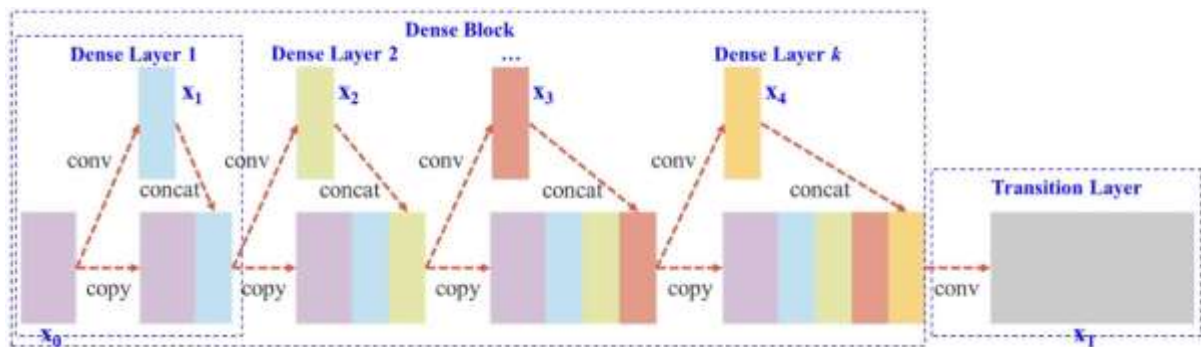


Figure 1: DenseNet technique

The fundamental principle of DenseNet121 is dense connectivity. If the output of a previous layer is represented as x_l , the input to a subsequent layer is formed by concatenating the feature maps produced by all preceding layers. Mathematically, the operation can be represented as:

$$x_l = H_l([x_0, x_1, x_2, \dots, x_{l-1}])$$

where $H_l(\cdot)$ represents the transformation performed by the l^{th} layer and $[x_0, x_1, \dots, x_{l-1}]$ denotes the concatenation of feature maps from earlier layers. This

connectivity allows the network to reuse features instead of learning similar representations repeatedly. Consequently, DenseNet121 can achieve effective feature learning with comparatively fewer parameters than many conventional very deep CNN architectures [9, 10].

DenseNet121 consists of dense blocks and transition layers. The dense blocks contain multiple convolutional layers connected through the dense connectivity mechanism. Each layer typically performs batch normalization, activation, and convolution operations. The transition layers are positioned between dense blocks and are responsible for reducing the spatial dimensions and controlling the number of feature maps. Pooling operations are also used to gradually reduce the spatial resolution of feature representations. The network terminates with a global average pooling operation followed by a classification layer. For deepfake face detection, the original classification layer can be replaced with a binary classification layer that predicts whether a facial image is real or manipulated [11].

In the proposed deepfake detection framework, DenseNet121 is used to extract hierarchical facial features from input images. Initially, facial images are resized to a suitable input dimension and normalized before being provided to the network. During the early layers, the model learns low-level features such as edges, contours, textures, and intensity variations. Deeper layers progressively learn more complex facial characteristics and manipulation-related patterns. Dense connectivity allows these low-level and high-level features to be reused across the network, which can help detect subtle artifacts introduced during face synthesis or face swapping [12, 13].

Another important advantage of DenseNet121 is its suitability for transfer learning. A DenseNet121 model pre-trained on a large image dataset can be adapted to the deepfake detection task by replacing and fine-tuning its final classification layers. Transfer learning reduces the requirement for training a deep network entirely from scratch and can improve convergence when the available deepfake dataset is limited. Fine-tuning selected deeper layers further allows the network to adapt its learned representations to specific characteristics of manipulated facial images.

For the proposed system, the final DenseNet121 classifier can be formulated as a binary classification problem:

$$P(y = 1 | x) = \sigma(Wf(x) + b)$$

where $f(x)$ represents the feature vector extracted by DenseNet121, W and b are the classification parameters, and σ is the sigmoid activation function. The output probability can be compared with a predefined threshold to classify the input image as Real or Deepfake [14].

III. PROPOSED METHODOLOGY

The proposed methodology for Optimization Accuracy of Deepfake Face Detection using DenseNet121 Technique is designed to automatically distinguish authentic facial images

from deepfake-generated facial images. The complete framework consists of dataset collection, image preprocessing, face detection, data augmentation, DenseNet121-based feature extraction, transfer learning, hyperparameter optimization, classification, and performance evaluation. The primary objective is to obtain a highly accurate and generalized deepfake detection model by utilizing the feature-reuse capability of DenseNet121.

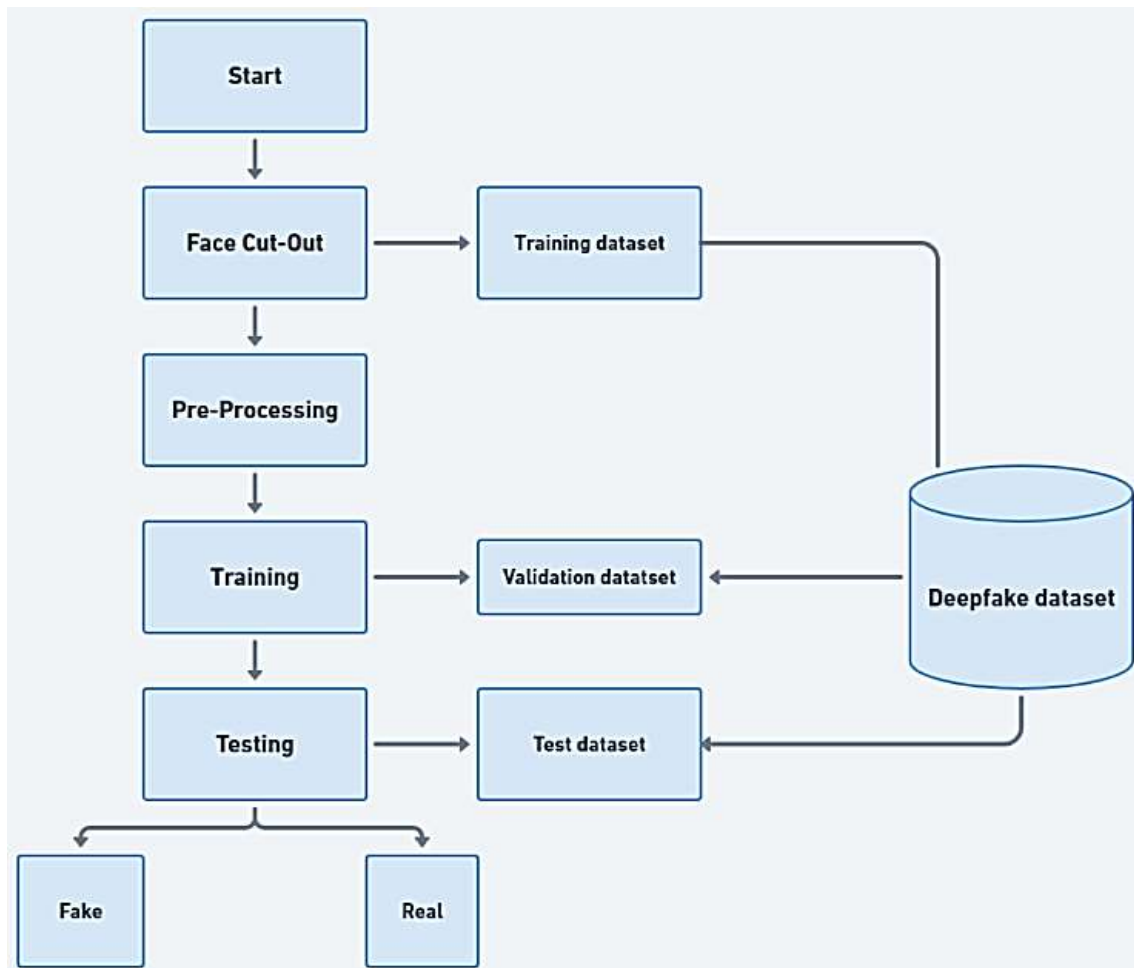


Figure 2: Flow Chart

A. Dataset Collection

The first stage involves collecting a dataset containing two major classes: Real and Deepfake facial images. The dataset should contain sufficient variations in facial identity, pose, illumination, background, image quality, and manipulation techniques. Real images represent authentic facial content, whereas deepfake images contain artificially generated or manipulated faces. The collected dataset is divided into training, validation, and testing subsets. A suitable division, such as 70% training, 15% validation, and 15% testing, can be adopted to evaluate the performance of the proposed model objectively.

B. Image Preprocessing

The collected images may differ in size, resolution, illumination, and quality. Therefore, preprocessing is performed before supplying the images to DenseNet121. First, facial regions are identified and cropped to remove unnecessary background information. The extracted faces are resized to the input dimensions required by DenseNet121, commonly 224×224 pixels. Pixel values are subsequently normalized to improve the stability and convergence of the training process. Preprocessing ensures that the model focuses primarily on facial characteristics and manipulation artifacts rather than irrelevant image information.

C. Data Augmentation

Deepfake datasets can contain limited variations or class imbalance, which may result in overfitting. To improve the diversity of training samples, data augmentation is applied to the training set. Operations such as horizontal flipping, rotation, scaling, cropping, brightness variation, and zooming can be used. These transformations generate slightly modified training samples while preserving their class labels. Data augmentation improves the model's ability to recognize manipulated facial features under different viewing and imaging conditions.

D. DenseNet121-Based Feature Extraction

After preprocessing, the facial images are supplied to the DenseNet121 network. DenseNet121 contains multiple dense blocks in which each layer receives feature maps from all preceding layers. This dense connectivity promotes feature reuse and improves gradient propagation. During training, the network learns different levels of facial representations. Initial layers capture low-level characteristics such as edges and textures, while deeper layers identify complex facial structures and manipulation-related patterns. These learned features are particularly useful for detecting subtle artifacts associated with face synthesis and face swapping.

E. Transfer Learning

To improve learning efficiency, transfer learning is employed. A DenseNet121 model pre-trained on a large-scale image dataset is used as the initial feature extractor. The original classification layer is removed and replaced with a task-specific classification layer for distinguishing Real and Deepfake classes. Initially, the earlier DenseNet121 layers can be frozen to preserve general visual features, while the newly added classification layers are trained. Subsequently, selected deeper layers can be unfrozen and fine-tuned using the deepfake dataset. This strategy enables the model to adapt general image representations to the specific characteristics of deepfake facial manipulation.

F. Classification Layer

The extracted high-level features are passed to the classification section. Global Average Pooling can be applied to reduce the feature maps into a compact feature representation. A fully connected layer followed by a dropout layer can then be used to reduce overfitting. Finally, a sigmoid activation function is employed for binary classification.

$$P(y = 1) = \frac{1}{1 + e^{-z}}$$

where z represents the output of the final classification layer. A probability close to 1 indicates a deepfake image, whereas a probability close to 0 indicates an authentic image.

G. Model Optimization

The optimization stage is introduced to improve the classification accuracy of DenseNet121. Important training parameters, including learning rate, batch size, number of epochs, dropout rate, optimizer, and fine-tuning depth, are adjusted based on validation performance. The Adam optimizer can be employed because of its adaptive learning capability. Binary cross-entropy is used as the loss function for the two-class classification problem:

$$L = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})]$$

where y represents the actual class and $\hat{y}$ represents the predicted probability.

Early stopping and learning-rate scheduling can also be incorporated to prevent overfitting and improve convergence. The optimal parameter configuration is selected based on validation accuracy and loss.

H. Training and Validation

The optimized DenseNet121 model is trained using the training dataset, while the validation dataset is used to monitor the learning process and select the best-performing model. During each training epoch, forward propagation generates predictions, the loss function calculates the classification error, and backpropagation updates the network weights. The process continues until the model reaches stable validation performance or the predefined stopping condition is satisfied.

IV. SIMULATION RESULT

The graph presents the training and validation accuracy of the proposed DenseNet121 model over six epochs (0–5) for deepfake face detection. The training accuracy starts at approximately 93.1% in Epoch 0 and increases consistently throughout the training process, reaching around 98.7% by Epoch 5. This continuous improvement indicates that the DenseNet121 model is effectively learning discriminative features from the training dataset and progressively improving its ability to distinguish between real and deepfake facial images. In contrast, the validation accuracy begins at approximately 93.2%, increases to 93.7% in Epoch 1, and then improves substantially to 96.6% in Epoch 2 and approximately 97.5% in Epoch 3. After Epoch 3, the validation accuracy decreases to approximately 95.7% in Epoch 4 and shows a slight recovery to about 95.8% in Epoch 5.

The highest validation accuracy of approximately 97.5% is achieved at Epoch 3, indicating that this epoch provides the best validation performance among the training iterations shown. After Epoch 3, the training accuracy continues to increase while validation accuracy declines, producing a growing gap between the two curves. This behavior suggests the beginning of overfitting, where the model becomes increasingly specialized to the training data and its generalization performance starts to decrease. Therefore, Epoch 3 can be considered an appropriate checkpoint for selecting the model based on validation accuracy.

Overall, the results demonstrate that DenseNet121 provides strong learning capability for deepfake face classification, achieving high training accuracy while maintaining validation accuracy above 95% during the later epochs. The observed reduction in validation performance after Epoch 3 also indicates that techniques such as early stopping, dropout, data augmentation, or learning-rate scheduling could further improve the generalization capability of the proposed model.

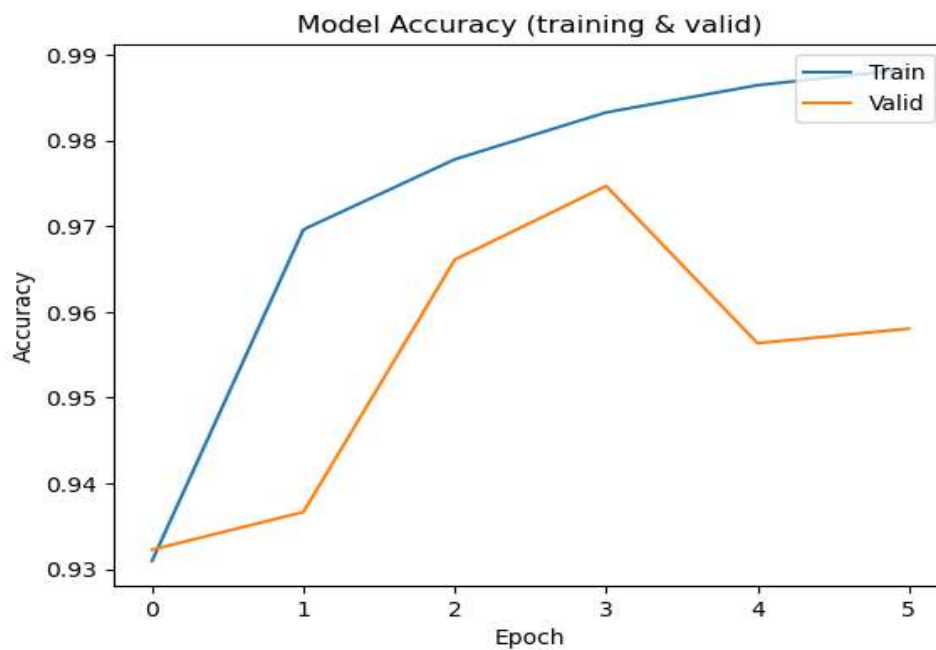


Figure 3: Training and Validation Accuracy

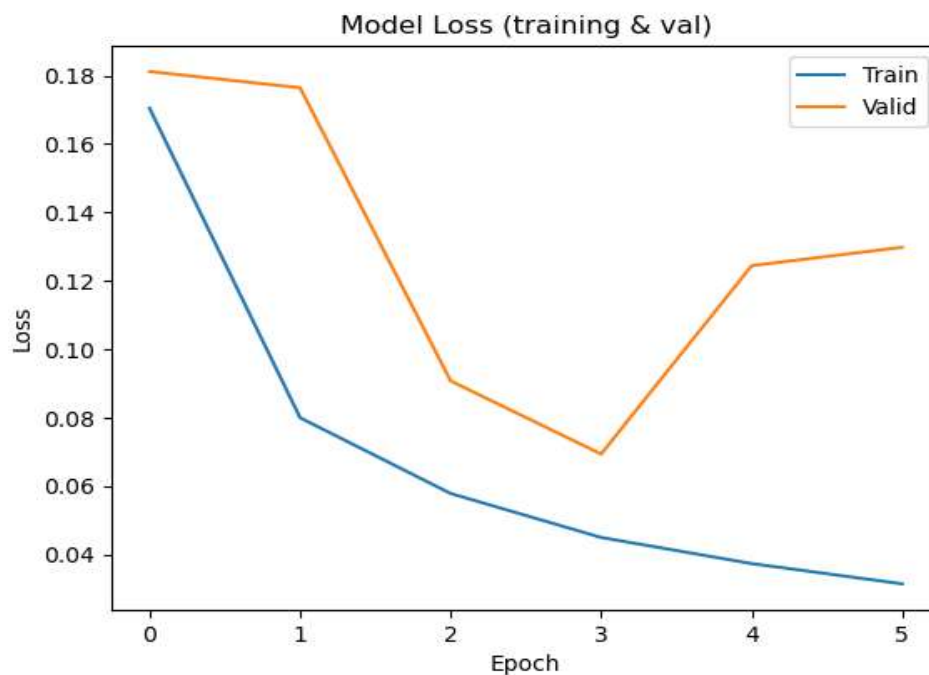
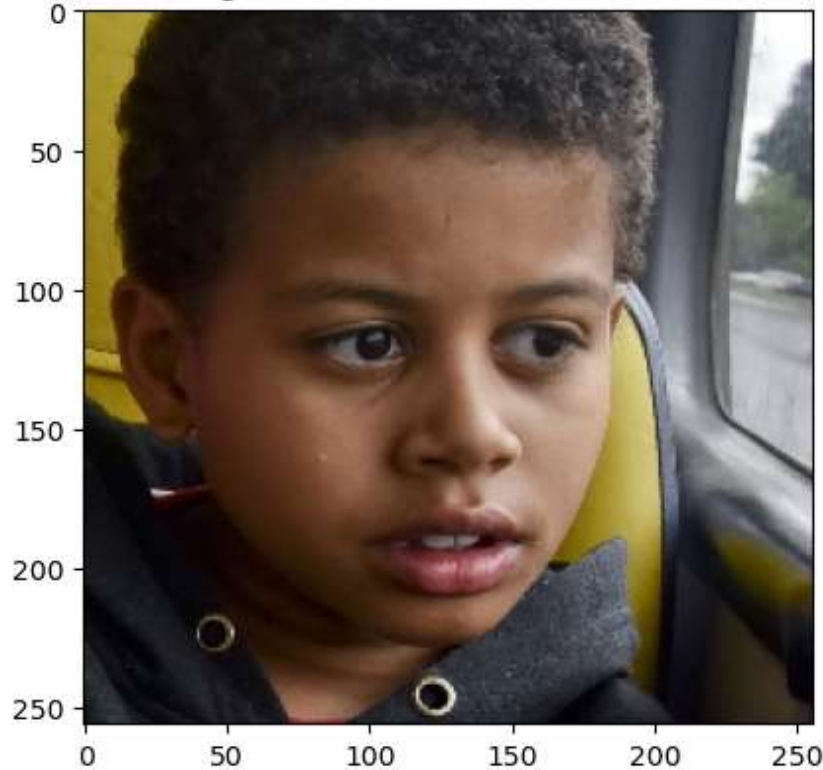


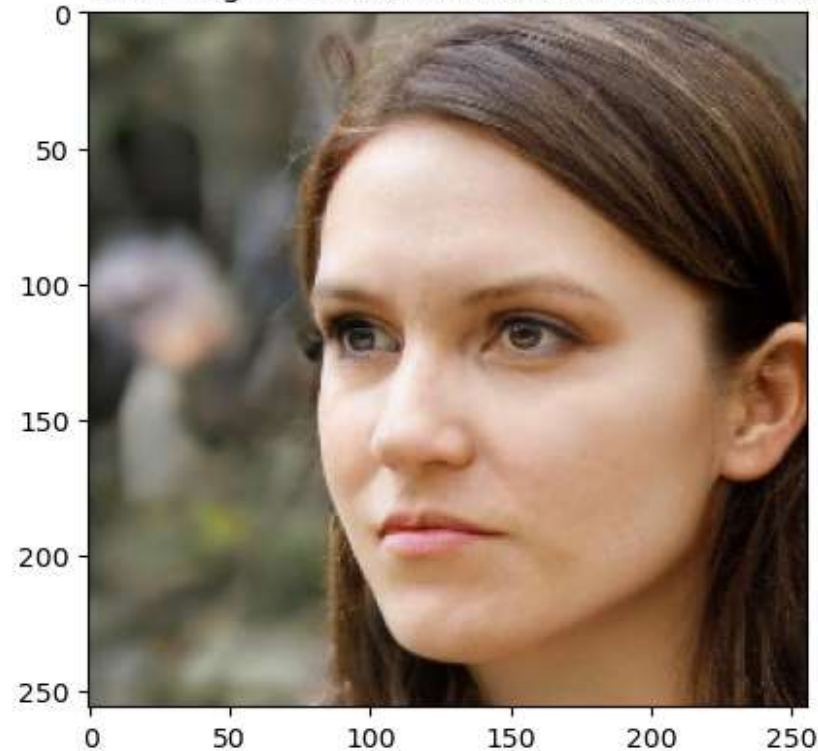
Figure 4: Training and Validation Loss

Qualitative Evaluation:

This image is 0.00% Fake and 100.00% Real.



This image is 100.00% Fake and 0.00% Real.



The confusion matrix illustrates the classification performance of the proposed DenseNet121-based deepfake face detection model for two classes, represented by 0 and 1, where the exact class labels should correspond to the dataset definition (e.g., Real and Deepfake). The model correctly classified 9,644 samples of Class 0 and 9,824 samples of Class 1, as indicated by the diagonal elements of the matrix. These results demonstrate that the model performs very well for both categories. On the other hand, 356 Class 0 samples were incorrectly classified as Class 1, while 176 Class 1 samples were incorrectly classified as Class 0. Thus, the model produced a total of 19,468 correct predictions out of 20,000 samples, corresponding to an overall accuracy of approximately 97.34%.

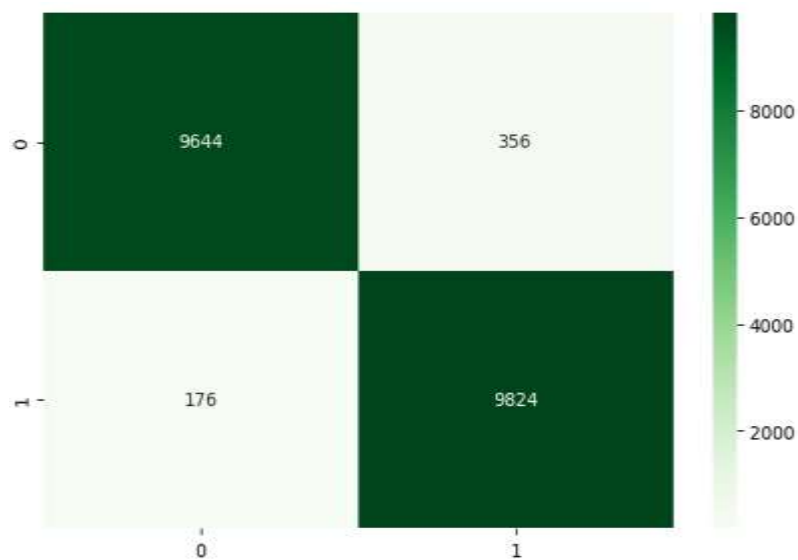


Figure 5: The test set is used to generate a confusion matrix

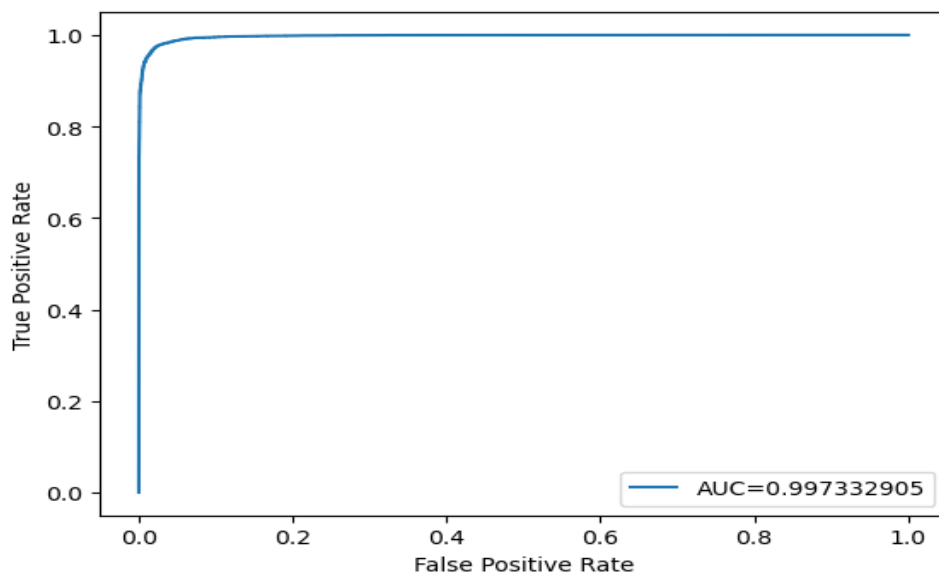


Figure 6: ROC Curve for True and False Predictions



The relatively small number of misclassified samples indicates that the proposed DenseNet121 model has a strong ability to distinguish between authentic and manipulated facial images. The lower number of errors for Class 1 compared with Class 0 also indicates slightly better recognition of Class 1. Overall, the confusion matrix confirms the high classification capability of the proposed approach, while the remaining false predictions indicate that further optimization, data augmentation, and testing on diverse and unseen deepfake samples could potentially improve the model's generalization performance.

V. CONCLUSION

The proposed study, “Optimization Accuracy of Deepfake Face Detection using DenseNet121 Technique,” demonstrates the effectiveness of deep learning for identifying manipulated facial images. The DenseNet121 architecture successfully learns discriminative facial features through its dense connectivity and feature-reuse mechanism, enabling effective classification of real and deepfake samples. The training results show that the model accuracy progressively improves from approximately 93.1% to 98.7%, indicating successful learning of relevant facial representations. The validation accuracy reaches its highest value of approximately 97.5% at Epoch 3, after which a slight decline indicates the beginning of overfitting.

The confusion matrix further demonstrates the strong classification capability of the proposed model, with 9,644 Class-0 samples and 9,824 Class-1 samples correctly classified. Out of 20,000 test samples, 19,468 were correctly classified, resulting in an overall accuracy of approximately 97.34%. Only 356 Class-0 samples and 176 Class-1 samples were misclassified. These results indicate that DenseNet121 can effectively distinguish authentic facial content from manipulated deepfake content.

Overall, the proposed approach provides a reliable foundation for deepfake face detection and digital media forensics. However, further improvements can be achieved through stronger data augmentation, early stopping, hyperparameter optimization, and evaluation using diverse and unseen deepfake datasets. Future work can also investigate DenseNet121-based hybrid CNN-Transformer models, temporal video analysis, explainable AI, and lightweight architectures for improving generalization and enabling real-time deepfake detection.

REFERENCES

- [1] Ali Raza, Kashif Munir and Mubarak Almutairi, “A Novel Deep Learning Approach for Deepfake Image Detection”, Apply Science, pp. 01-15, 2022.
- [2] Zobaed, S.; Rabby, F.; Hossain, I.; Hossain, E.; Hasan, S.; Karim, A.; Hasib, K.M. Deepfakes: Detecting forged and synthetic media content using machine learning. In Artificial Intelligence in Cyber Security: Impact and Implications; Springer: Berlin/Heidelberg, Germany, 2021; pp. 177–201.



- [3] Thambawita, V.; Isaksen, J.L.; Hicks, S.A.; Ghouse, J.; Ahlberg, G.; Linneberg, A.; Grarup, N.; Ellervik, C.; Olesen, M.S.; Hansen, T.; et al. DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. *Sci. Rep.* 2021, 11, 21869.
- [4] Ahmed, M.F.B.; Miah, M.S.U.; Bhowmik, A.; Sulaiman, J.B. Awareness to Deepfake: A resistance mechanism to Deepfake. In *Proceedings of the 2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, Taiz, Yemen, 4–5 July 2021; pp. 1–5.
- [5] Y. Nirkin, L. Wolf, Y. Keller, T. Hassner, Deepfake detection based on discrepancies between faces and their context, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44 (10) (2021) 6111–6121.
- [6] Brian Dolhansky, Russ Howes, Ben Pflaum, Nicole Baram, and Cristian Canton Ferrer. The Deepfake Detection Challenge (DFDC) Preview Dataset. *arXiv eprints*, page arXiv:1910.08854, 2019.
- [7] Brian Dolhansky, Joanna Bitton, Ben Pflaum, Jikuo Lu, Russ Howes, Menglin Wang, and Cristian Canton Ferrer. The DeepFake Detection Challenge Dataset. *arXiv e-prints*, page arXiv:2006.07397, 2020.
- [8] Md Shohel Rana, Mohammad Nur Nobi, Beddhu Murali, and Andrew H Sung. Deepfake detection: A systematic literature review. *IEEE access*, 10:25494–25513, 2022.
- [9] Jin, B.; Cruz, L.; Gonçalves, N. Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis. *IEEE Access* 2020, 8, 123649–123661.
- [10] Lewis, J.K.; Toubal, I.E.; Chen, H.; Sandesera, V.; Lomnitz, M.; Hampel-Arias, Z.; Prasad, C.; Palaniappan, K. Deepfake Video Detection Based on Spatial, Spectral, and Temporal Inconsistencies Using Multimodal Deep Learning. In *Proceedings of the 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Washington DC, DC, USA, 13–15 October 2020; pp. 1–9.
- [11] Lee, H.; Park, S.H.; Yoo, J.H.; Jung, S.H.; Huh, J.H. Face recognition at a distance for a stand-alone access control system. *Sensors* 2020, 20, 785.
- [12] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen. Attgan: Facial attribute editing by only changing what you want. *IEEE Transactions on Image Processing*, 28(11): 5464–5478, 2019.
- [13] Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., & Nahavandi, S. Deep learning for deepfakes creation and detection, 2019, arXiv preprint arXiv:1909.11573.
- [14] Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. Mocogan: Decomposing motion and content for video generation. pages 1526–1535, 2018.
- [15] Guha Balakrishnan, Amy Zhao, Adrian V. Dalca, Frédo Durand, and John Guttag. Synthesizing images of humans in unseen poses. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.