INTERNATIONAL JOURNAL OF Telecommunications
AIR DEFENSE COLLEGE

Secure Facial Verification: A hybrid model for detecting Spoof Attacks with ResNet50-DenseNet121

Citation: ElSayed , A.; Hikal ,N.; Sakr , N. ; Takieldeen, A.
Title. Inter. Jour. of Telecommunications, ***IJT'2024***, *Vol. 04, Issue 02, pp. 01-11, 2024.*
Editor-in-Chief: Youssef Fayed.
Received: 30/05/2024.
Accepted: 18/07/2024.
Published: 18/07/2024.
Publisher's Note: The International Journal of Telecommunications, IJT, stays neutral regard-ing jurisdictional claims in published maps and institutional affiliations.

Aya ElSayed*[1], Noha A. Hikal[2], Nehal A Sakr [3] and Ali E. Takieldeen [4].
Faculity of Artificial Intelligence , Delta University for Science and Technology , Gamsa, Egypt [1]; aya.sayed5777@gmail.com.
nformation Technology Department, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt [2]; dr_nahikal@mans.edu.eg
Information Technology Department, Faculty of Computers and Information, Mansoura University, Mansoura, Egypt [3]; nehal_sakr@mans.edu.eg.
IEEE Senior Member, Faculty of Artificial Intelligence, Delta University for Science and Technology [3]; a_takieldeen@yahoo.com.

Abstract : The present study introduces a novel hybrid deep learning model, leveraging the synergies inherent in the amalgamation of ResNet50 and DenseNet121 architectures. This fusion aims to effectively tackle the formidable task of detecting spoof attacks. Spoof attacks pose a significant threat to digital systems and networks, where adversaries attempt to deceive systems by impersonating legitimate users or sources. The proposed hybrid model aims to enhance detection accuracy and robustness against various spoof attacks by leveraging the complementary features of ResNet50 and DenseNet121. Integrating these architectures creates a unified framework that effectively captures local and global input data features, enabling more comprehensive detection capabilities. The problem of detecting spoofing attacks is stated as a classification task, and we train the hybrid model using large-scale datasets comprising fake and real data samples. The experimental results illustrate the superior performance of the proposed hybrid model in comparison to individual SVM, KNN, CNN, and RNN models, highlighting its efficacy in mitigating the risks associated with spoof attacks in digital systems and networks.

Keywords: Hybrid Model, Deep learning techniques, Spoof attacks, ResNet50, Densenet121**.**

1. Introduction

A hybrid deep learning model is a powerful approach combining different architectures' strengths to enhance performance and effectiveness. This research paper presents a pioneering hybrid methodology for detecting spoofing attacks using a combination of ResNet50 and DenseNet121 architectures. Detecting fake attacks is a significant challenge in cyber security, as attackers constantly develop their techniques to fool systems and networks .The hybrid model aims to improve the efficiency and robustness of spoof attack detection by leveraging the unique features of both ResNet50 and DenseNet121.

Integrating ResNet50 and DenseNet121 architectures allows a hybrid model with superior data assimilation capabilities to capture local and global features in the input data [1,2]. We formulate parody attack detection as a classification task and train the hybrid model on comprehensive datasets containing various real and fake data samples.

The experimental results provide evidence for the enhanced performance of the proposed hybrid model. over individual ResNet50 and DenseNet121 models in terms of accuracy and performance. This highlights the effectiveness of hybrid approaches in cybersecurity and their potential to strengthen defenses against sophisticated spoof attacks. Furthermore, this study addresses future challenges and trends in spoof attack

detection, providing insights for further research. Overall, The results represent a significant advance in cyber security technologies, contributing to ongoing efforts to enhance the detection and mitigation of spoofing attacks in digital systems and networks.

Both ResNet50 and DenseNet121 have demonstrated exceptional performance on benchmark datasets. such as ImageNet [3]. Their success can be attributed to their innovative architectural designs, which address critical challenges in deep learning, including training convergence, feature representation, and parameter efficiency. ResNet50 and DenseNet121 are two influential deep-learning architectures that have significantly advanced the field of computer vision. Their effectiveness in capturing complex image features and achieving high classification accuracy has made them indispensable tools in various applications, ranging from medical imaging to autonomous driving [4,5].

In this implementation, the chart below transforms the hidden layer's output, which is essentially a more chaotic and intricate representation of the input data plus the linearly compounded bias component for each covert node. The Ikeda map's borders are as follows:

$$u = 1 + c(x\cos y - y\sin x). \quad (1)$$

$$v = c(x\sin y + y\cos x). \quad (2)$$

When the scaling parameter is c = 0.92, the data sources are x and y, and the outcomes are u and v.Using the ens uing linear transformation of the hidden layer(h),the network's output (y) is obtained.

$$y = W2h + b2 \quad (3)$$

The matrix W2 has dimensions M x N, while the vector b2 possesses M elements. The weights and biases undergo update during training using backpropagation and an α learning rate. For the result layer, the discrepancy (e) is defined as follows:

$$e = t - y \quad (4)$$

where t is the desired outcome.The error e for the hidden layer is defined as follows:

$$e(i) = e\partial x/\partial u. \quad (5)$$

where the fractional subsidiary of u with respect to x is represented by ∂x/∂u, which is given by:

$$\partial u / \partial x = c(cosy - xsiny) \quad (6)$$

The weights and biases are modified in the following manner:

$$\Delta 2 = \alpha eh \quad (7)$$

$$W = W\,2 + \Delta 2 \quad (8)$$

$$b\,2 = b\,2 + \alpha e \quad (9)$$

$$\Delta 1(i) = \alpha e(i)x. \quad (10)$$

$$W1(i) = W1(i) + \Delta 1(i). \quad (11)$$

$$b1(i) = b1(i) + \alpha e(i) \quad (12)$$

The key components are the following: input data (x), hidden layer output (h), W1(i) weight matrix for hidden n ode (i), b1(i) bias term for hidden node (i), Δ2 matrix, and Δ1(i) matrix (for each hidden node (i)) Weights and biases in the chaotic neural network design are initialized at random By guaranteeing each input image's uniqueness and unpredictability during training, this improves the network's capacity for learning.

The subsequent sections of this paper are organized as follows: First, formulate a comprehensive problem. The solution using the hybrid model is presented in Section II. Related studies pertaining to the hybrid model are explored in Section III: simulation outcomes and methods and material, which are in Section IV. Lastly, Section V encompasses the concluding remarks of this study, and the analysis is detailed.

2. Problem Formulation

This section introduces a hybrid model to address the weaknesses of face recognition systems against spoof attacks. Initially, there were several problems with previous models in recognizing fake and real faces.

2.1. Convolutional Neural Networks (CNN)

CNNs may struggle to deal with changes in lighting, angles, and noise, reducing efficiency in recognizing real and fake faces in different environments. [6]

$$y = f(Wx + b). \quad (13)$$

2.2. Recurrent Neural Networks (RNN)

RNNs may be ineffective in handling long data sequences such as videos, which can result in loss of important information or distinctive movements indicating the presence of spoofing attacks. [7,8]

2.3. k-Nearest Neighbors (KNN)

KNN may face challenges in dealing with large amounts of data or high dimensions, as it may require a long time to compute distances between samples, thus becoming inefficient in terms of time and memory. [7,8,9]

2.4. Support Vector Machine (SVM)

SVM may struggle with non-linear cases or unbalanced data, affecting its capacity to accurately discern genuine and fabricated facial characteristics. [8,9]

$$Y = sign\ (Wtx + b) \quad (14)$$

These are the problems that previous models CNN, RNN, KNN, and SVM may encounter in recognizing fake and real faces, indicating the need to develop hybrid models or use advanced techniques for more effective and accurate face recognition.

3. Literature review

In the literature, several approaches have been proposed to identify spoofing assaults with photos, repeated movies,and synthetically created masks.the two primary types of current Face anti-spoofing techniques are static approaches and dynamic approaches, which vary according on the cues used.

Static approaches mostly rely on examining textural variations in surface reflection and material contrasts between real and fake facial pictures. Maatta et al. [8] created concatenated histograms by using multi-scale local binary patterns (LBP) to extract texture from 2D photos. To distinguish between real and artificial faces, these histograms were loaded into an SVM classifier. In order to identify print attack spoofing, their study demonstrated the effectiveness of concatenating three LBP descriptors with various configurations. They achieved a 0% Half Complete Blunder Rate (HTER) on the Print Assault dataset in a later study [9] by introducing a score-level combination strategy that makes use of LBP, a histogram of located slopes, and Gabor wavelets derived from surrounding face picture blocks.

Pereira et al. [10] created LOP-
TOP, which considers three symmetrical planes converging the focus point of a pixel in the XY, XT, and YT dir ections, where T addresses the time hub, in order to integrate temporal data defining dynamic facial structures for face anti-spoofing. According to their testing, on the Replay Assault dataset, multi-goal LBP-TOP with an SVM classifier achieved the greatest decreased HTER of 7.6%.

Dynamic approaches have an impact on liveliness cues throughout the film, such as head, lip, and eye movem ents. Among the novel approaches are "nosy communications," in which clients are expected to follow clear instructions. Using a non-nosy approach, Container et al. [11] identified eye squinting as unfriendly to mocking by creating a constrictive irregular field to illustrate different stages of eye flickering. By combining the 3D characteristics of face images with the eye flickering and mouth development, Kollreider et al. presented a method in [12] for handling identification of mocking attacks. In the IJCB facial mocking rivalry [12].

certain members utilized facial expressions such as head motions and eye flashing to determine liveliness.

Shreve et al. introduced a worldwide stain metric in [13] to detect inconspicuous face parts in video successions for identifying caricaturing attacks. The metric was derived using optical stream designs on facial localities. Bharadwaj suggested enhancing small- and large-scale highlights for faces that are resistant to ridicule in a follow-up study [14].

4. Method And Material

The hybrid model consists of ResNet50 and DenseNet121 architectures as the main model for training. The operational process of our suggested approach for identifying face spoofing is outlined in Fig.1. The input data encompass three sets of image information acquired through an RGB camera. Initially, the RGB camera identifies and captures the facial area [15].

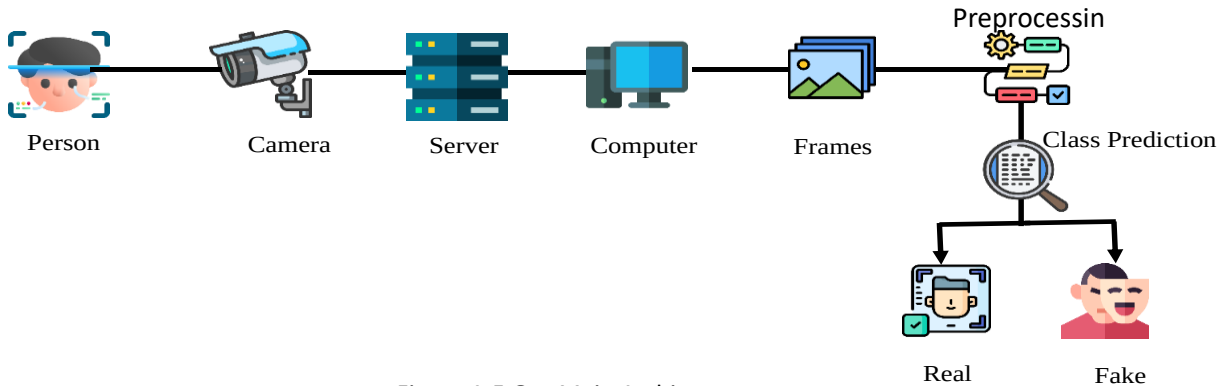

Figure 1. Hybrid Model Architecture

The designated facial area encompasses three specific regions, as shown in Fig 2:

1. The entirety of the facial area.
2. The region surrounding the nose.
3. The region surrounding the eyes.

Figures 1 and 3 were focused on taking the features into account, as shown in Figure 3 and Figure 4 [16].

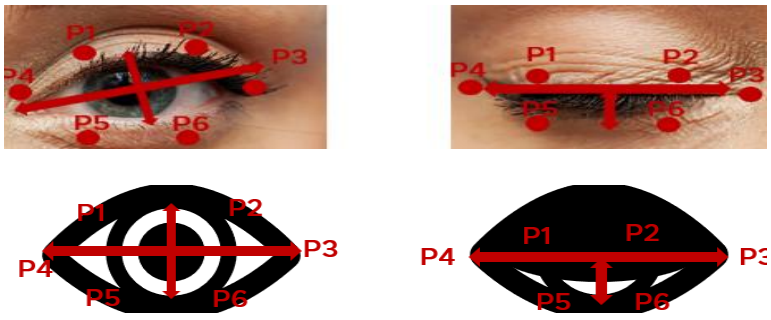

Figure 2. Example of Blink Dimension.

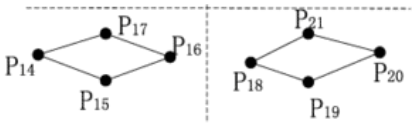

Figure 3. Eye Data Points

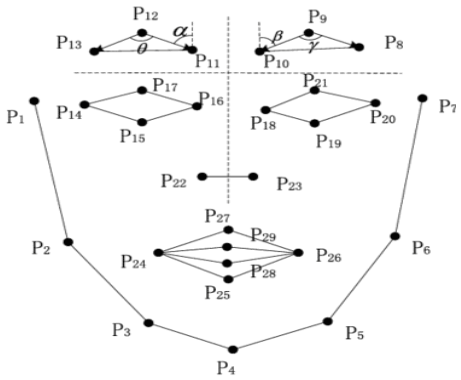

Figure 4. Facial Data Points.

A fusion of two Neural Network Architectures, Resnet 50 and Densenet121, is employed in this model for recognizing spoofing or deceptive activities, including deepfake videos, voice impersonation, and digital impersonation. Let's explore the functions of each network within this model. Resnet 50 is a Convolutional Neural Network (CNN) architecture, as shown in Fig.5, mainly focusing on image tasks. It excels in image classification, object -detection, and analyzing information derived from images or video frames [6]. Specifically, it concentrates on features, gestures, and other visual indicators [2].

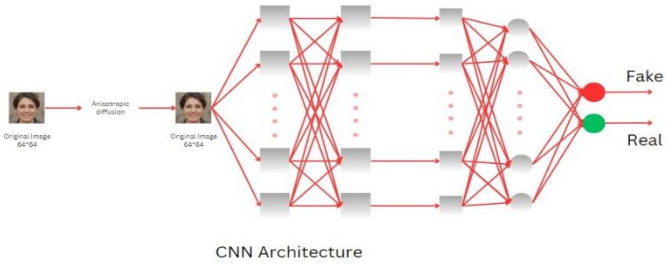

Figure 5. CNN Architecture

Resent 50 can extract high-level features from input images. It generates valuable representations in distinguishing between content and manipulated images or videos. It can capture patterns and variations present in data, as shown in Fig. 6. It is An essential capability for detecting deepfake videos or other forms of image-based spoofing [1]. ResNet50 and DenseNet121 are deep learning models trained on large datasets for image classification tasks. ResNet50 is a 50-layer Residual network trained on the ImageNet dataset, exhibiting top-tier performance. State-of-the-art results on many picture classification benchmarks have been attained by it [5].

ResNet50 helps alleviate the vanishing gradient issue and is renowned for its capacity to handle extremely deep networks. It is excellent at capturing aspects that are hierarchical in pictures [2]

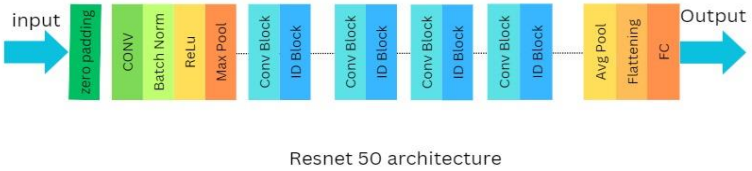

Figure 6. ResNet50 Architecture

The ImageNet dataset was also used to train the 121-layer convolutional neural network known as DenseNet121. [1]. It has shown to be effective in image classification tasks and has achieved high accuracy on various benchmarks. Regarding Spoofing Detection, Resnet50 and densenet121 can be applied to ResNet50 (Residual Network). It introduces residual learning, the concept of utilizing shortcut connections, enabling the bypassing of one or more layers within the network [5]. DenseNet121 (Densely Connected Convolutional Network) employs densely connected blocks, ensuring that each layer receives input from all preceding layers [3]. DenseNet50 fosters the reuse of features and optimizes feature propagation through the network [16]. It often requires fewer parameters than traditional architectures, potentially leading to better parameter .

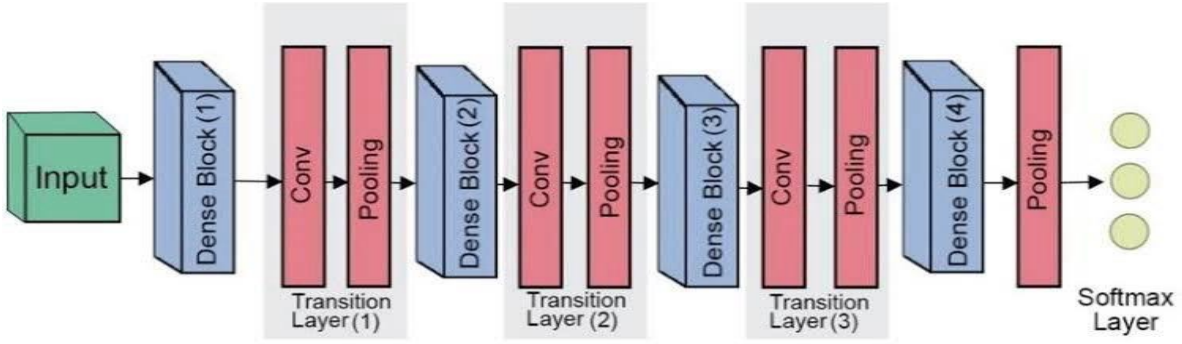

Figure 7. DenseNet121 Architecture

5. Result And Analysis

A. Amplifying Facial Expressions

To make small changes in movies that could be difficult to notice with the unaided eye, a technique called Eulerian motion magnification [15] has been developed. In [16] With this method, little movements in actual video scenes are successfully emphasized.

Motion magnification approaches use a regular video sequence as input, analyze it spatially, and then filter the frames temporally. The objective of motion magnification is stated as follows, assuming that the observed changes in intensity over time are represented by a displacement function δ(t), where I(x, y, t) is the video frame at position (x, y) and time t and I(x,y, t) = f(x + δx(t),y + δy(t)), with δx (t) and δy (t) as displacement functions in the x and y directions. improve quiet movements in films to increase their visibility

$$I(x, y, t)=f(x+(1+\alpha)\delta x(t) ,y +(1+\alpha)\delta y(t)) \quad (15)$$

A first-order Taylor series expansion around the x and y axes may be used to describe the video I given a magnification factor α.

$$I(x, y, t)\approx f(x, y)+\delta x(t)\, \partial f(x ,y) / \partial x +\delta y(t)\, \partial f(x ,y) / \partial y \quad (16)$$

After applying a comprehensive temporal bandpass filter to the input video I at each (x, y) point, define B(x,y,t). By doing this, every element that isn't f(x, y) is guaranteed to be removed. The mathematical model of filter B(x, y, t) is represented by the following:

$$B(x, y, t) = \delta x(t)\partial f(x, y)/\partial x + \delta y(t)\partial f(x, y)/\partial y \quad (17)$$

As a result, the updated video I(x, y, t) has the following characteristics:

$$I(x, y ,t)=I(x, y ,t)+\alpha B(x ,y, t) \quad (18)$$

By integrating equations 15, 16, 17, and 18, we obtain the ultimate depiction of the motion-expanded video, designated as I.

$$I(x, y, t) = f(x, y) + (1 + \alpha)$$
$$[\delta x\ (t)\ \partial f(x ,y)/ \partial x +\delta y(t)\ \partial f(x, y)/ \partial y\ ] \quad (19)$$

A comparison of Equations 16 and 19 suggests that the spatial displacement δ(t) of the local image f(x,t) at time t has been amplified to a magnitude of (1+α). The magnification level is significantly influenced by the chosen filter as well as the magnification factor α. To find the optimal value for α, we visually evaluate the films that were generated using the training dataset in this study.

We used Accuracy in the main performance metric: compare the proposed model with different models employed to detect facial spoofing.

Formula: $$Accuracy = \frac{TP+TN}{T+TN+FP+FN} \quad (20)$$

Accuracy (97.8%):

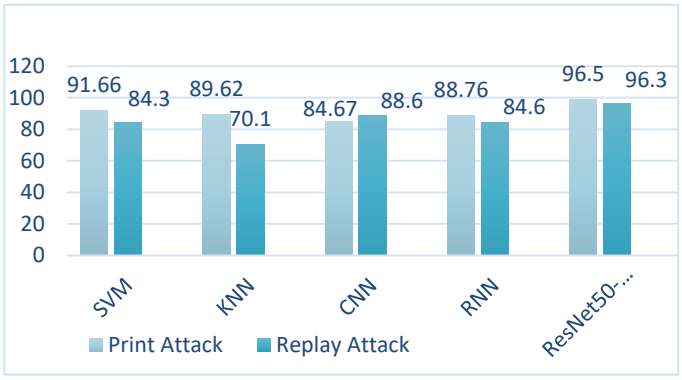

Figure 8. Accuracy against Various Spoof Attacks

Accuracy=TP + TN Total Predictions Accuracy=Total Predictions TP +TN

Explanation:

Accuracy is a critical metric for evaluating the performance of machine learning models, as it measures the overall correctness of predictions. In the given context, an accuracy rate of 98.7% indicates that the model correctly predicted 98.7% of the instances, as depicted in Figure 9. Various models demonstrate different accuracy rates: Support Vector Machine (SVM) achieves 87.98%, k-Nearest Neighbors (KNN) reaches 79.86%, Convolutional Neural Networks (CNN) attain 86.63%, and Recurrent Neural Networks (RNN) achieve 86.68%. Notably, advanced architectures such as ResNet and DenseNet exhibit superior performance with an accuracy of 97.8%. These results underscore the effectiveness of these models in accurately predicting outcomes, with ResNet -50 showing particularly high accuracy, making them suitable for applications requiring high precision, shown in Figure.9

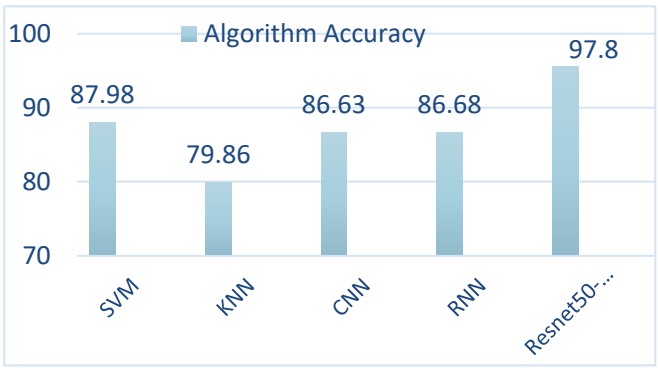

Figure 9. Efficiency of Hybrid Algorithm Compared to Others

B. The F1 scores for training, validation, and test data

The process of discriminant channel selection involves analyzing a Pixel Hop with dimensions 3 × 3 × 3. In Fig.10, the x-axis shows the channel index, and the y-axis displays energy percentage and F1 scores for train, validation, and test data. Blue lines represent energy percentages, while red, magenta, and green lines depict F1 scores. Four distinct settings are considered in this analysis.

F1 =2·TP/2·TP+FP+FN

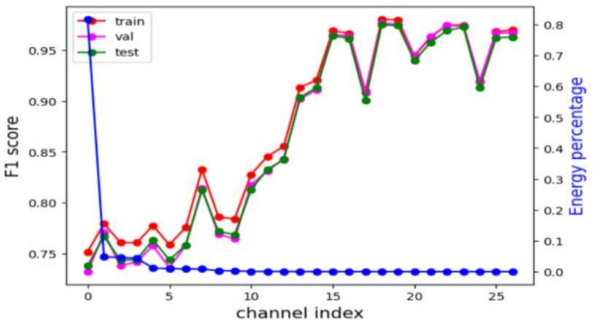

Figure .10 F1 Score without perturbation

C. The Epoch vs Accuracy

In Figure.11, as the number of epochs increases, you observe whether the model's accuracy improves, plateaus, or potentially degrades. A rising accuracy curve indicates that the model is learning well.

x-axis shows a number of epochs in model .

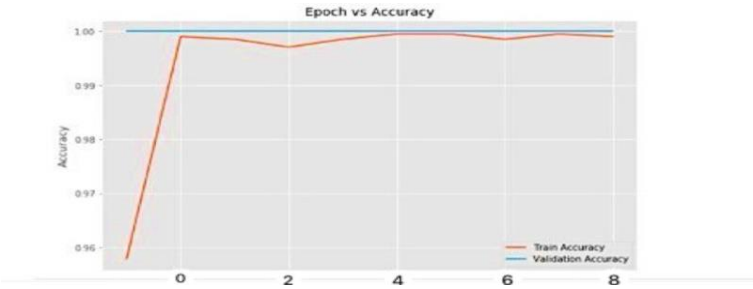

Figure.11 Epoch vs Accuracy

D. Epoch vs. Training Loss

In Figure.12, the training loss should decrease over epochs, indicating that the model is improving its prediction ability. A decreasing loss suggests that the model is converging and learning the patterns in the training data, number od epochs in traning loss .

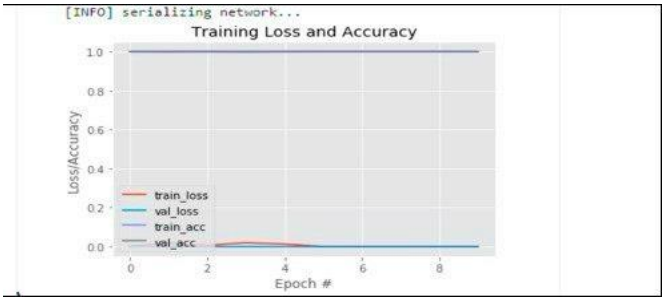

Figure.12 Epoch vs. Training Loss

E. Confusion Matrix

A confusion matrix is a table of data that provides the performance characteristics of a classification system. It is particularly useful when the procedure splits instances into two or more classes. The four entries in the matrix are False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN).

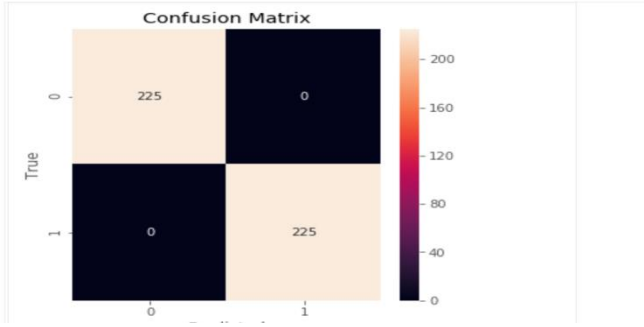

Figure.13 Confusion Matrix

F. Precision(0.995):

Formula: Precision=TPTP + FP Precision=TP + FPTP [5,6,7]

Explanation:

Precision focuses on the accuracy of positive predictions. A precision of 0.995 suggests that 99.5% of the positive predictions were correct.

G. Recall (0.982):

Formula: Recall=TPTP + FN Recall=TP + FNTP [5,6,7]

Explanation:

Recall is a measure of how well the model can find all relevant events. The model identified 98.2% of actual positive cases correctly, with a recall of 0.982.

H. Matthews Correlation Coefficient (MCC) (0.977):

Formula:

MCC=TP×TN−FP×FN(TP + FP)(TP + FN)(TN + FP)(TN + FN)MC C=(TP + FP)(TP + FN)(TN + FP)(TN + FN)TP×TN−FP×FN [5,6,7]

Explanation:

MCC provides a correlation coefficient between the expected and observed classes. A substantial correlation is shown by a 0.977 value.

I. Receiver Operating Characteristic (ROC) AUC (0.988):

Explanation: The ROC AUC provides a measure of the model's class discrimination ability. With a significant trade-off between true and false positive rates, an AUC of 0.988 indicates great discriminating power.

x-axis show performance matrics across folds.

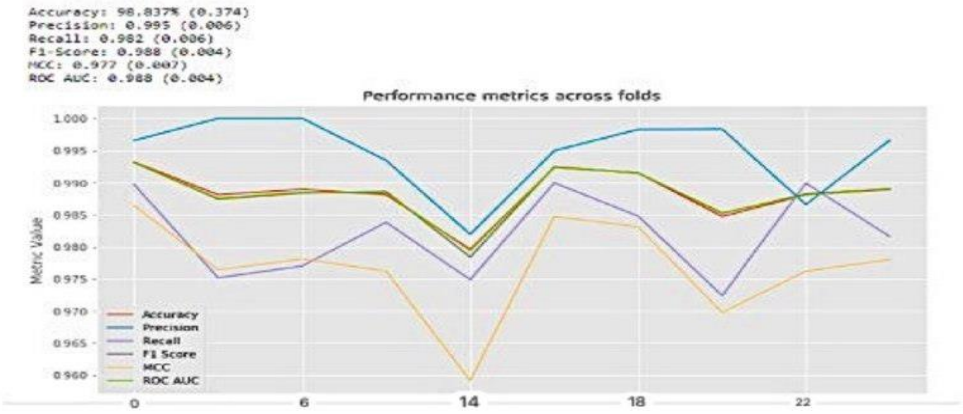

Figure.14 Performance Metrics across folds

6. CONCLUSION

The wide range of materials used in different spoofing devices means that existing face anti-spoofing techniques sometimes struggle to achieve strong generalization. It is essential to improve face anti-spoofing's generalization capabilities for real-world application systems. This work aims to strengthen the resilience of face anti-spoofing by focusing on a more general feature: fine-grained movements across video frames. To extract dense and spatial characteristics from input photos, a hybrid model of ResNet50 and DenseNet121 is utilized. The hybrid model incorporates Eulerian motion magnification and attention techniques to ensure a full capture of dynamic changes. A number of commonly used techniques are applied for comparison analysis, and intra- and inter-test evaluations are carried out on two difficult datasets to show the efficacy of the suggested approach.

7. References

[1] Li, Bin. "Facial expression recognition by DenseNet-121." Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems. Academic Press, (2022). 263-276.

[2] Zhu, Yanjia, et al. "Tinaface: Strong but simple baseline for face detection." arXiv preprint arXiv:2011.13183 (2020).

[3] Nandy, Abhilash. "A densenet based robust face detection framework." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. (2019).

[4] Yu, Su-Gyeong, et al. "Face Spoofing Detection Using DenseNet." International Conference on Intelligent Human Computer Interaction. Cham: Springer International Publishing, 2020.

[5] Lee, Jia-Rou, Kok-Why Ng, and Yih-Jian Yoong. "Face and facial expressions recognition system for blind people using ResNet50 architecture and CNN." Journal of Informatics and Web Engineering 2.2 (2023): 284-298.

[6] Yu, Su-Gyeong, et al. "Effect of Facial Shape Information Reflected on Learned Features in Face Spoofing Detection." Journal of Internet Technology 23.3 (2022): 517-525.

[7] Zhou, C. Gao, F. Chen, C. Li, X. Li, F. Yang, Y. Zhao, Face anti-spoofing based on multi-layer domain adaptation, in 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE, (2019), pp. 192–197

[8] J. Hernandez-Ortega, J. Fierrez, A. Morales, J. Galbally, Introduction to face presentation attack detection, in: Handbook of Biometric AntiSpoofing, Springer, 2019, pp. 187–206.

[9] S. Jia, G. Guo, Z. Xu, Q. Wang, Face presentation attack detection in mobile scenarios: A comprehensive evaluation, Image and Vision Computing 93 (2020) 103826.

[10] Moon, Youngjun, Intae Ryoo and Seokhoon Kim. "Face anti-spoofing method using color texture segmentation on fpga." Security and Communication Networks 20211-11, (2021)

[11] Pazo, D. Jim'enez-Cabello, E. V'azquez-Fern'andez, J. L. AlbaCastro, R. J. L'opez-Sastre, Generalized presentation attack detection: a face anti-spoofing evaluation proposal, in: 2019 International Conference on Biometrics (ICB), IEEE, 2019, pp. 1–8.

[12] S. Saha, W. Xu, M. Kanakis, S. Georgoulis, Y. Chen, D. P. Paudel, L. Van Gool, Domain agnostic feature learning for image and video-based face anti-spoofing, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2020, pp. 3490–3499.

[13] Y. Jia, J. Zhang, S. Shan, X. Chen, Single-side domain generalization for face anti-spoofing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8484–8493.

[14] S. Fatemifar, M. Awais, S. R. Arashloo, J. Kittler, Combining multiple one-class classifiers for anomaly-based face spoofing attack detection, in: 2019 International Conference on Biometrics (ICB), IEEE, 2019, pp. 1–7.

[15] Haroun, R. F., A. E. Takieldeen, E. H. Abdelhay, and M. A. Mohamed. "Performance Evaluation of QoS for VoIP and Video Streaming over LTE Networks." communications 1 (2018): 3.

[16] Z. Yu, C. Zhao, Z. Wang, Y. Qin, Z. Su, X. Li, F. Zhou, G. Zhao, Searching central difference convolutional networks for face anti-spoofing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5295–5305.