WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Neural Shield: GAN Augmented ResNet for Deepfake Defense

K. Aarthi, K. Deepa

Information Technology, Kingston Engineering College Vellore, India
aarthikarthikeyan1@gmail.com, deepakalingarayan@gmail.com

ABSTRACT –

Human brains can be able to, P. 1 differentiate the features of faces, utilisation of sophisticated technology and artificial intelligence is confusing the disparity between real and edited pictures. The Digital editing has evolved to the extent that there are applications, now available, that can be used by the consumer.

They made up very real looking false faces, and it is more difficult so that human beings can distinguish between the real and the made ones. This is why it is possible to use such techniques as deep learning.

And being put more and more in use as a means of separating the real and artificial faces, more durable and accurate results. To identify the face of fraud, This paper presents a groundbreaking hybrid deep learning relationship, which combines the potentials of Generative Adversarial .

Having combined the powerful generative properties of GANs and the discriminative properties of RESNET, the model suggested represents a new way of identifying real or artificial faces. The performance of the hybrid model is obtained through a comparative analysis with given pretrained models like VGG16 and RESNET 50. The findings reveal that the hybrid model is significantly more effective in false face detection and one must have made a significant progress in the area of facial image recognition and authentication. The results of using a benchmark dataset demonstrate that the given model achieves excellent performance indicators, such as precision 0.79, recall 0.88, F1-score 0.83, accuracy 0.83, and ROC AUC Score 0.825. The inferences of the study show that the hybrid model yields impressive results of detecting the fake face especially in terms of accuracy, precision and economy of the memory. This resolves the issue of more complicated fake faces generation methods by marrying the generative power of GANs and the discriminative properties of RESNET, this holds a lot of potentials in identity verification, social media content moderation, cybersecurity and many others, thus the study aims to contribute to the field of false face identification. Under such conditions, it is important to understand how to correctly distinguish between genuine and manipulated faces. It is important to note that the suggested model uses Channel-Wise Attention Mechanisms and adds it to the feature extraction phase of RESNET50, making it even more effective and enhancing its performance.

Index Terms - RESNET, generative adversarial networks, deep learning, real and fake faces, face detection, channel-wise attention.

I. INTRODUCTION

Digital methods of adjustment concerning images and movies presenting false facial expressions have aroused high levels of public criticism in the current times [1]. The name Deepfake stands out to define the visuals, audio, and the videos that are realistic in sound and highly realistic, yet produced by the artificial intelligence machines [2]. Recent advancements in deepfake creation currently make deepfake more realistic and easier to produce. Deepfake has been a major menace to society and we have a right to privacy so we have to come up with deepfake detection methods to counter it. the following issues [3], [4]. In December 2017, a user under the alias Deepfakes [5] developed pornography videos with photos and videos where real faces were manipulated with artificial ones through the help of publicly available artificial intelligence app. Deepfakes is a consumer of the social media Instagram site [6]. Deepfaking is the replacement of the images (namely faces) of a person through artificial intelligence algorithms. One form of synthetic media is called the deepfake; the sort of synthetic media that uses deep learning software to create fake movies, recordings, and/or photographs. It involves replacing faces of individuals in a photograph or video with the images of another individual so as to create a believable copy in the view of deceiving their audience or distorting the original meaning of content [7]. Most of the deepfake detection methods are based on features and machine learning. Some of the ongoing challenges in detecting deepfakes are deepfake generated images, insufficient high-quality datasets, and insufficient benchmarks. Future deepfake detection trends can consist of effective, efficient and systematic deepfake detection methods and efficient datasets [8]. The technology of GANs has allowed creating almost indistinguishable realistic images of faces that appeared to be close to the vision of real faces [9]. The two components of a Generative Adversarial Network consist of the generation process and an object called discriminator that work together to produce fake photos that may be difficult to distinguish between real photos. The fake pictures are generated by the generator as the discriminator is trained such that it differentiates fake pictures and real pictures [10]. The generator attempts to generate more believable photos so as to fool the discriminator during training and the discriminator tries to identify false images more effectively. GANs are applied into making pictures of human beings, animals, and

products, and they can also be employed in making fake pictures used in bad intentions [11]. What is more devastating, human beings have a hard time detecting such persuasive deep fake videos, audios, and pictures. Thus, it is essential, urgent, and obligatory to distinguish between a real media and deepfakes. Thus, the necessity to develop a really efficient model that would help to provide the certain identification of the real and altered photos should emerge. Owing to the recent upsurge in the possibility of fraudulent activities, various ways on how to detect phony face photos have been created in a bid to resolve this situation [12]. Such methods can be grouped as follows: some methods are based on characteristics that have been created manually and depend on the statistical features of the photos. The other segment involves the application of deep learning architecture which makes use of state of the art neural networks to determine patterns and features of the photos [13]. This paper will have six major sections. In Section I, the research challenge is mentioned and deep significance of the topic and the objectives of the study are highlighted. Sections II show an overview of the relevant background information and relevant studies. Section III presents materials and methods to be used. Section IV offers the proposed model along with details of the architecture, design, and implementation. Section V shows the results of the implementation and discuss them. Section VI draws the conclusions, contributions to the area and the future work.

II. LITERATURE SURVEY

A 1. Face Warping Artifacts

This paper aimed at identifying artifacts based on the comparison of generated face areas with the adjacent ones with the help of the specific Convolutional Neural Network model. The technique exploits the fact that existing deepfake methods produce low-resolution images, which should be transformed to the resolution of a source video, which introduces artifacts. This approach, however, does not take into account the study of the temporal frames.

B 2. Eye Blinking Detection

The method via this paper proposed eye blinking to be the major indicator to detect deepfakes. The video frames of people with eye blinking were analyzed using Long-term Recurrent Convolution Network (LRCN). Nevertheless, the authors have considered the possibility to synthesize blinking by modern Deepfake generators and it is not enough to use this parameter when using them. Other symptoms such as wrinkles on the face as well as positioning of the eyebrows should be added.

C 3. Deepfake Detection using Capsule Networks

It used capsule networks to detect altered images and videos to be applicable in situations such as replay attack detection and synthetic video analysis. Training of the model involved the addition of some random noise that enhanced performance on their dataset although it might result in a lower accuracy in real-time. This is because our model does not involve such noise to get a better generalization.

D 4. ImageNet Pre Trained and RNN

The study combined the temporal recognition of the video frame due to the use of Recurrent Neural Networks (RNN) with the spatial insight based on ImageNet pre-trained model. They showed great potential, but their data (HOHO) did not cover as many types of videos (its base included only 600 such videos). As opposed to this, we base our model on a diverse, large-scale training set.

E 5. Synthetic Proportion Videos with Biological Signals The approach presented in this paper has used the alternative to get the biological signal (e.g, heartbeat) in real and counterfeit videos by utilizing facial areas. Photoplethysmography (PPG) maps were used in the analysis of space and time coherence. The model was a hybrid of the Support Vector Machines (SVM) and CNNs. This was very much effective but application of a differentiable loss that relied on signal processing was a challenge noted.

III. Proposed Process of the System

In this section, the author proposes an elaborate description of the intended model and its procedure of how the two types of faces can be identified as real and fake. The model proposed correctly addresses a primary problem of face recognition that is recognised in this paper. The model will be useful in other areas and aspects such as security and criminal investigation because it would be capable of distinguishing between real and fake photos. The main attributes of the suggested model, which are the application of machine learning, the use of deep learning, and features of the spatial domain will be outlined shortly. The research techniques employed in the research will also be discussed and any weaknesses will be discussed as well.

A. DATASET

This analysis was performed on the dataset presented in [1], The name of the dataset is The Real and Fake Face Detection dataset is a popular benchmark dataset includes 2, 041 face images where 1, 081 images are referred to as a real image e.g. Figure 3(a) and 960 images are referred to as a fake image e.g. Figure 3(b). The evaluation of the effectiveness of the different face detection models is carried out through the benchmark. making the distinction between true and false images. The fake images in this dataset are also generated using different methodologies of the digital image manipulation such as face swap, face2face, and Deepfakes.

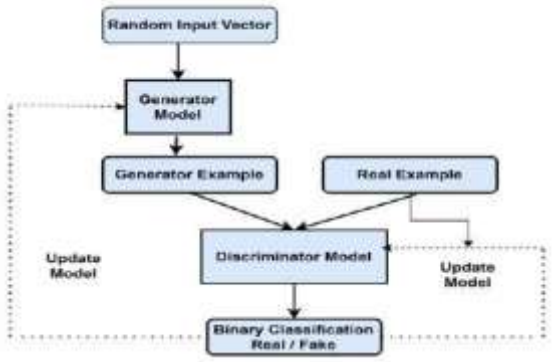

FIGURE 1. Architecture of GAN

B. SUGGESTED MODEL STAGES AND ARCHITECTURE

The six phases which are proposed in the present architecture are as follows. Figure 3 reveals the overall architecture of the proposed model. The initial stage is preprocessing of data. Initially, data cleaning is implemented as to verify the dataset by observing the corrupted data and mislabeled data as images. Any distorting or erroneous pictures are discarded so that data integrity must be maintained, and in no case must the model be trained on the wrong or noisy examples. This is followed by data augmentation to expand the dataset in order to make it larger and more varied. Normal augmentations are flipping, scaling and random crops. This is one of the ways that makes the model become and more able to generalize better on unseen data.

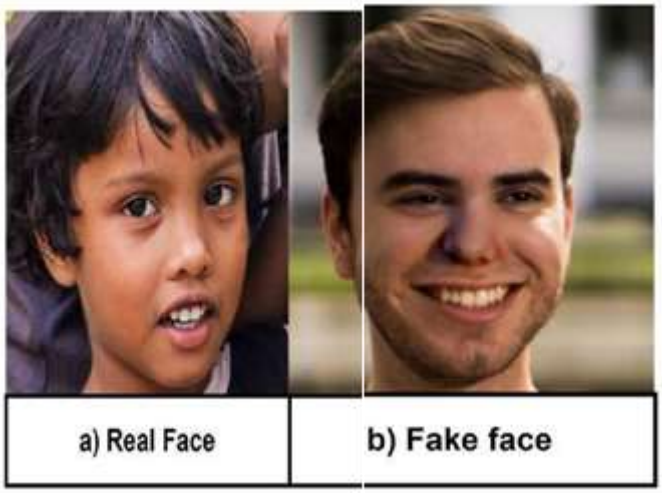

FIGURE 2. Examples of real and fake faces used in the training phase.

This is among the factors that render the model to become and are better able to generalize on unseen data. Then, all the pictures are downsized to a consistent size which can be phrased to the deep learning model. Models of deep learning, like ResNet, usually need picture images of a set dimension. Twenty four hundred by 224 is standard. Finally, normalization of the pixel values of the images is done to scale them to a desired level. Scaling the pixel values over the range [0, 1] is the most typical way of doing it. This process assists the acceleration of convergence of the model in the training process and eliminates the problem associated with various scales of the pixel values. Within the scope of the investigated research, the off-the-shelf ResNet model architecture was changed in an attempt to make it best suited to a particular analytical task. First, a basic framework was chosen, which is the pre-trained Convolutional Neural Network (CNN), i.e., the ResNet50 model. The deviation from the common approach to delete only the very last layers connected with the aim of classification was carried out in order to customize the model to feature extraction goals. The starting point of the feature extraction process is a ResNet50 model as a basic representation of the Convolutional Neural Network (CNN). Conversely, the ResNet backbone includes attention mechanisms, as opposed to the traditional means of discarding the last classification layers. Specifically, attention modules are inserted following specific convolutional layers in order to draw attention-weighted feature representations. This can be characterized as the model is able to focus on the relevant characteristics on the face and artifacts of the manipulation, and these attention modules dynamically adapt the relevance of different spatial areas in the feature maps. Instead, substantial improvement was done to increase training efficiency and accuracy in predicting. This has included the introduction of more convolution layers in the ResNet structure with the fine-tuning of the structures to pick the tricky features in the image dataset. Also kernel sizes were optimised to better capture nuances subtleties patterns on the data. Use of innovative regularization methods such as dropout and batch normalization was utilized to reduce the risk of overfitting so as to increase the generalization of the model. The effectiveness of such alterations was strictly tested by trial-and-error experiments, and the evidence about their significant effects on training performance and prediction was empirical. Then the altered ResNet architecture with customized changes was used in feature extraction hence strong feature representations. Such extracted features with custom-incorporated modifications were used as an input in further analytical processes, such as classification, clustering, and feature similarity analysis.

By means of designed improvements, this work stands out among the standard ResNet framework due to its extended effectiveness and flexibility in the context of the desired analytical sphere. The third step implies creating synthetic faces by GAN. This is achieved by training a GAN to produce life-like fake face pictures. The input of generator network is random noise that is converted into fake face images. The discriminator network attempts at distinguishing between actual and bogus faces. GAN is then trained on a mix of adversarial and reconstruction loss to make sure the fake faces generated are realistic.

The fourth stage entails the offered hybrid model.

To begin with, it inputs the feature representations got after the CNN. Some subsequent layers (e.g. fully connected layers) are then added onto the feature representation of CNN. The result of the extra layers will then be fed to the discriminator network of GAN. Training The combined model is then trained with the frozen CNN and the updated GAN discriminator along with other layers. In case of necessity, the whole hybrid model can be optimised. The fifth step is Training where a labeled set of images of genuine and counterfeit faces is utilized

for training. Apt loss function (e.g. binary cross-entropy) is used to train a hybrid model to distinguish between real and fake faces. In line with this the weights of the hybrid model are upgraded based on backpropagation and gradient descent. The sixth step is Evaluation where the working of the hybrid model is measured on a distinct validation or test set. Accuracy, precision, recall and F1score Metrics are the measures that are used to assess the effectiveness of the model in the detection of fake faces.

IV. DISCUSSIONS, RESULTS

The results of numerous experiments obtained and the proposed model are presented in this section, however, first, this section shortly describes various measures that are applied to estimate the performance of these models. Sensitivity (Recall): sensitivity is the ability of the test to identify the true positives. To express it differently, the sensitivity is the likelihood with the help of which a test will identify a positive case. T = TP/ (TP + FN). (1) TP: True Positive FN: False Negative. Precision: In precisions, the ratio of the favorable predictions which are favourable is used as a value Precision = TP/TP + FP. Where F P is False Positive. Accuracy: accuracy is an estimation of the percentage of predictions that are right, without concern about whether these are negative or negative.

Accuracy = (TP+TN)/(TP+FP+ FN + TN). F1 Measure: is an average of precision and recall with the precision and recall weighted. It is computed as an harmonic mean of precision and recall.

F 1Measure = 2 (precision Recall)/(precision + Recall). (4)

To ensure the stability and feasibility of the proposed hybrid deep learning model, the suitable data partitioning plan is used, where 70 percent of the data are allocated to training, and 30 percent to validation. This kind of partitioning made it possible to have a good assessment of the performance of our model, where it gets the benefit of training on the majority of the data, but also being subjected to a critical test on unseen parts of a different data. We also applied a k-fold crossvalidation (CV) method during training, where it split the training data into multiple folds and trained the model and checked it in a loop.

The capability of the model

This approach further ensured that generalization across different chunks of training data was done. The test-set amounting to 30 percent was used as a separate dataset that was not manipulated in any way during the modelling process in order to approximate real-life conditions and enhance the practicability of the model. The goal of data splitting and k-fold CV integration is to strengthen the reliability of the conclusions as well as emphasize the effectiveness of the model in a diverse set of scenarios. Several experiments were used on the data in this research to compare the obtained outcomes with the suggested model. To begin with we used the Deep convolutional VGG16

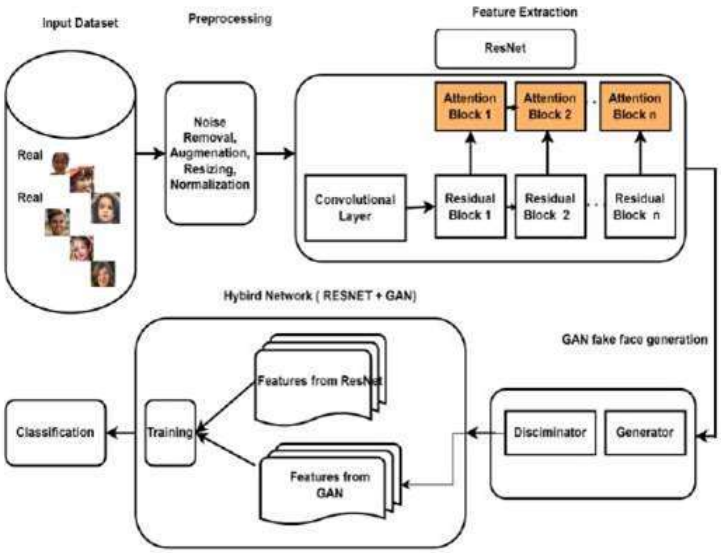

FIGURE 3. The overall architecture of the proposed model.

TABLE 1. VGG16 results.

Precision	Recall	F1-score	Accuracy	ROC AUC score
0.6232	0.6224	0.6226	0.6260	0.6421

TABLE 2. ResNet-50 results.

Precision	Recall	F1-score	Accuracy	ROC AUC score
0.7264	0.7265	0.7260	0.7263	0.6991

Its simple yet effective architecture known as a neural network, is put into use. There are 16 weight layers with convolutional as well as fully connected layers. It is similar to a repetitive structure of small 3 x 3 convolutional filters after max-pooling layer. However, the usage of VGG16 constitutes its primary contribution since the model establishes how deep networks should be utilized in classifying images. The findings of the VGG16 network are presented in Table 1 and in Figure 4. The results generated by the second experiment are presented in Table 2 and Figure 5, in this experiment ResNet-50 is applied. categorize the real faces and imposter faces The outcomes of the third study, which produced the results based on the proposed model, which is the hybrid of the ResNET-50 and Table 3 and Figure 6 state the GAN algorithm. The hybrid model tries to establish the ideal when using 400 images, the best value was obtained as the image from GAN was taken over 400 images as the starting point.generated

TABLE 3. Results of the proposed model.

Precision	Recall	F1-score	Accuracy	ROC AUC score
0.7916	0.8824	0.8345	0.8298	0.825

A ratio of 70 % and 30 % training and testing was applied. %. The model has been fine-tuned on 100 epochs, which each lasted about 10 hours, on a workstation with a Spectrum Blue, single-GPU NVidia, a 16-GB RAM, and a 6-core i7 Intel processor. The nevis of the parameters, namely, an input size, (224, 224) and a batch were selected. of 64, was guided by rigorous experimentation. Custom layers seamlessly integrated into the model augmented its discriminative capabilities, leveraging the robustness of the ResNet-50 architecture pretrained on ImageNet. Additionally, data augmentation techniques, such as rotation, width and height changes, and horizontal flips, were employed to enhance the model's ability to identify complex elements in facial photographs. The utilization of the Adam optimizer with binary crossentropy loss contributed to improved accuracy. Despite the ResNet model's known demand for a substantial number of parameters, resulting in a bulky size, the proposed

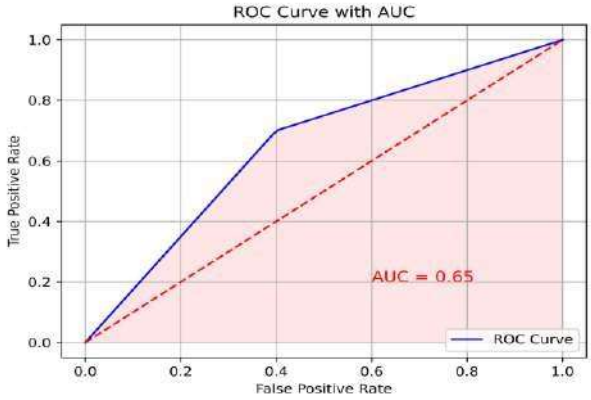

FIGURE 4. False positive rate vs true positive rate for VGG 16 network.

TABLE 4. Architecture of VGG16 and ResNET-50.

Architecture	VGG16	ResNET-50
Batch size	64	64
Number of Epochs	100	100
Learning rule	1e-4	1e-3
Optimizer	SGD	Adam
Total parameters	134 M	25.6 M

In addition to introducing a generative component, the addition of a GAN allows the model to identify fake faces that already exist as well as potential variants or new instances of synthetic faces that might appear in the future. Second, the GAN component improves the model's generalization over a wide variety of fictitious face variants by introducing a novel type of data augmentation during training. The hybrid model gains exposure to a wider dataset by producing realistic synthetic faces. This can potentially mitigate the risk of overfitting and enhance its resilience in real-world situations

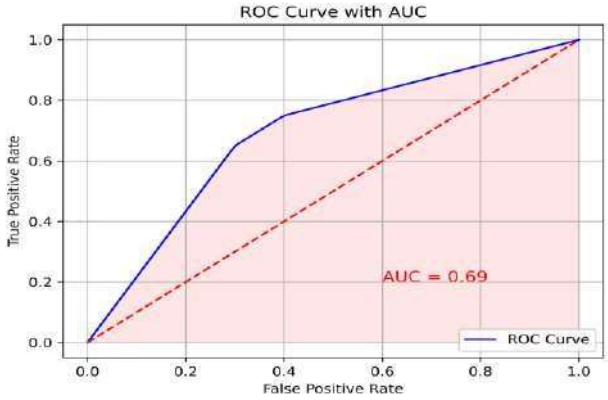

FIGURE 5. False positive rate vs true positive rate for ResNet-50 network.

model structure and parameters are the result of repeated refinement and experimentation and thus, are the optimal that could be achieved by extremely fine tuning. Recognizing that there can even be more improvements with respect to its overall accuracy and research results, the future research will focus on employing optimization strategies or heuristic algorithms to seek an optimal set of parameters in a systematic fashion by applying them to the ResNet model. Based on the above outcomes, it can be observed that the outcomes that have been provided by ResNET- 50 are superior to VGG16 because of residual connection. ResNET 50 connection mixed with GAN algorithms was proposed as a means to improve the performance of the ResNET-50 with the application of the hybrid proposal. The accuracy of the proposed model in table 3 was more than 83 percent, and this is superior to the scores recorded by the ResNET- 50 network by almost 10 percent. Figure 7 indicates a comparative analysis between the proposed hybrid and VGG16 and ResNET-50 in general. Table 5 is a comparison between the proposed model in this research and an implementation of a model with ResNET 18 proposed in [1]. Compared to historical models, the third one, a hybrid one including both the specifics of both RESNET50 and GAN algorithm, has also shown improved

results when contrasted to more mainstream models and independent deep models such as VGG16, or RESNET50. It is possible to attribute the noteworthy success of the hybrid model to several significant features, which, in turn, enhance its detectability of fake faces. On the one hand, the first one is that the hybrid model inherits the benefits of both generative and discriminative components. The discriminative core is RESNET50 that has gained a reputation of deep and efficient feature extracts. Consequently, the model is able to differentiate between tiny details and patterns associated with both genuine and artificial facial features. Besides inserting a generative element, the incorporation of GAN will enable the model to recognize fake faces that currently exist. In that the occurrences of the false faces can be quite random or rather dynamic. Besides, the success of the hybrid model allows highlighting the importance of considering the full context of false face detection. The GAN images used in the hybrid model allow it to more clearly comprehend the nuances of facial formation, facial expression, and realistic appearance features, which are important when distinguishing between advanced fake faces that might not be recognized under the simple machine learning models, although deep learning systems such as VGG16 and RESNET50 are adept at capturing details. Also, the success of the hybrid model means that limitations observed in the traditional models can be resolved through a judicious combination of the discriminative approach and the generative approach. The current finding points to the importance of hybrid frameworks to the boundaries of being precise and reliable in detecting fake faces and establishes new research directions. In short, the third model is superior to the others since it has a synergetic use of generative ability of a GAN and discriminative capability of RESNET50. Such unique combination not only enhers feature discrimination, but also is a new means of addressing how to handle the issues brought up by continuously evolving techniques of creating fake faces. The performance of the hybrid model can be valuable when it comes to future work and study of the next image for identifying fake faces and the research and application of artificial intelligence. Generally, poor results are also attained in most of the instances in using a single architecture such as the RESNET50 or VGG16 as compared to a hybrid. comprising GAN and RESNET50 (Residual network) model

TABLE 5. Comparison between the proposed model and model that uses ResNET 18.

Evaluations	Proposed hybrid model	Model in (1) ResNet 18
Precision	0.7916	0.79
Sensitivity	0.8824	0.73
Accuracy	0.8298	0.77

This benefit is presented by the properties of the two components in combination. Since it is a deep network, RESNET50 finds a lot of success in deriving features of pictures in detail, GANs can however generate synthetic data samples that approximate the training sample. The hybridization integrates the feature extracting capacity of RESNET50 in combination with the data generating capacity of GANs in order to distinguish between counterfeit and authentic. extra synthetic data, resolving problems with insufficient training data and improving the model's capacity to generalize to new data. The retrieved characteristics are also refined by the GAN's ability to differentiate between real and produced data, potentially improving their suitability for classification tasks. The hybrid

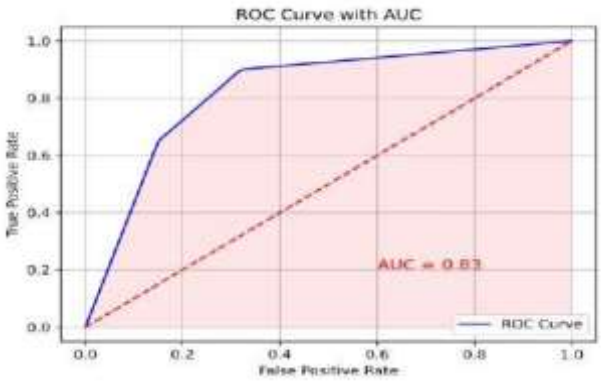

FIGURE 6. False positive rate vs true positive rate for the proposed model.

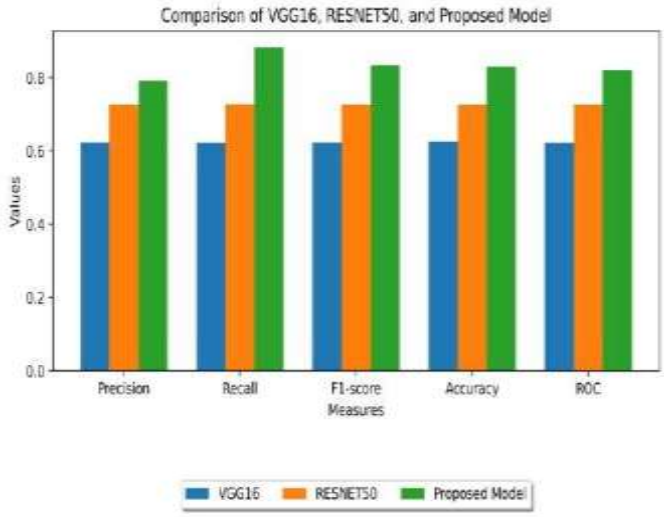

FIGURE 7. Comparison between proposed model vs VGG16 and ResNET-50.

The model captures underlying data distributions by utilizing GANs for unsupervised pretraining, leading to more efficient feature extraction during the next fine-tuning stage. This method is very useful in situations where there are noisy or unbalanced datasets. To improve class separation during classification, the GAN can produce synthetic samples for minority classes or clean noisy training data. The hybrid technique is also effective for domain adaptation tasks where there are distribution mismatches between the source and target domains. The model can improve its performance in the target domain by adapting features from the source domain to it. A powerful ensemble effect results from the interaction

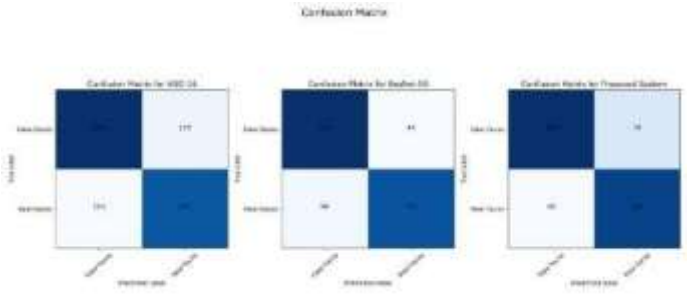

FIGURE 8. Confusion matrix.

of GANs including RESNET50. Once the learnt discriminative features of the RESNET50 model are coupled with the diversity brought in by GAN generated data, the model performance is often enhanced. It should be emphasized that the efficiency of this hybrid method is subject to a series of factors, among which one can distinguish qualities of datasets, task complexity, the effectiveness of the process of training GAN, and others. In the case of a hybrid GAN-RESNET50 model, to identify whether it works actually performs

When tested and analyzed, they have shown to be better than standalone models such as RESNET50 or VGG16 when used on a specific task. The reasons why the performance of the suggested hybrid deep learning model is the way it is are numerous, and methods of continually improving its performance are continually being sought after. The first success of the model could be explained by the smart integration of GANs and the RESNET architecture used in it: it gathers the benefits of the mentioned technologies. The synergistic effect of generative ability of GAN and discriminating capacity of RESNET enhances discriminating capabilities between fake and real faces of the model. precise method for distinguishing real from fake facial photos, the study made use of the features of the RESNET architecture after applying Channel-Wise Attention Mechanisms and GANs. On a benchmark dataset, the suggested model performed superbly, obtaining high precision, recall, F1-score, accuracy, and ROCAUCscore. These findings highlight the model's efficiency and dependability in the critical task of detecting fake faces. The contribution is significant because it has the potential to be used in many other fields, such as cybersecurity, identity verification, and social media content control. In these domains, the ability to discriminate between real and altered faces is crucial, and our hybrid model provides a potent tool for tackling this problem. Future research in the field of fake face detection should focus on a few crucial areas to further improve the capabilities of hybrid deep learning models. For example, new deep learning architectures should be investigated, and optimization techniques should be investigated to increase the model's precision, recall, and overall accuracy. Increased detection performance can be facilitated by stateof-the-art structures and well calibrated parameters. Finally; future work could certainly explore cross database evaluations to further validate the generalizability of the proposed model across different datasets and scenarios.

V. CONCLUSION

In the article, the authors suggest a new deep learning hybrid architecture to address the emerging issue of fake face detection in the age of deepfake technology and more advanced technologies in manipulating photographs. To create a proper and accurate model of telling the difference between the true and fake photos of the face, the study utilized the features of the architecture of the RESNET following the implementation of Channel-Wise Attention Mechanisms and GANs. The proposed model was outstanding in a benchmark dataset with high precision, recall, F1-score, accuracy, and ROCAUC score. These results evidence the model efficiency and reliability

in the tricky mission to identify artificial faces. The contribution is major since it maybe applied in numerous other areas including cybersecurity, identity authentication, control of social media contents. The discrimination of real and manipulated face in such domains is very essential and our hybrid model can give a powerful tool against this issue. The next steps in research in the sphere of fake face detection must be devoted to several important aspects in order to enhance the level of the results. potentials of combination deep learning models. To give an example, new deep learning models ought to be explored, and optimization strategies have to be examined to enhance precision, recall, and overall performance of the model. Stateof the art structures and optimal parameters calibration can help make the detection perform better. Lastly; perhaps, the future work can definitely test cross database assessments in order to prove whether there is greater generalizability of the proposed model on other datasets and situations.

VI. REFERENCES

1. N. A. S. Eldien, R. E. Ali, and F. A. Moussa, ‘‘Real and fake face detection: A comprehensive evaluation of machine learning and deep learning techniques for improved performance,’’ in IEEE MTT-S Int. Microw. Symp. Dig., Jul. 2023, pp. 315–320.
2. Y. Zhu, C. Zhang, J. Gao, X. Sun, Z. Rui, and X. Zhou, ‘‘High-compressed deepfake video detection with contrastive spatiotemporal distillation,’’ Neurocomputing, vol. 565, Jan. 2024, Art. no. 126872.
3. A. Gandhi and S. Jain, ‘‘Adversarial perturbations fool deepfake detectors,’’in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2020, pp. 1–8.
4. M. M. El-Gayar, M. Abouhawwash, S. S. Askar, and S. Sweidan, ‘‘A novel approach for detecting deep fake videos using graph neural networks,’’ J. Big Data, vol. 11, no. 1, p. 22, Feb. 2024.
5. O. B. Newton and M. Stanfill, ‘‘My NSFW video has partial occlusion: Deepfakes and the technological production of non-consensual pornography,’’ Porn Stud., vol. 7, no. 4, pp. 398–414, Oct. 2020.
6. W.-D. Zhou, L. Dong, K. Zhang, Q. Wang, L. Shao, Q. Yang, Y.-M. Liu, L.-J. Fang, X.-H. Shi, C. Zhang, R.-H. Zhang, H.-Y. Li, H.-T. Wu, and W.-B. Wei, ‘‘Deep learning for automatic detection of recurrent retinal detachment after surgery using ultra-widefield fundus images: A singlecenter study,’’ Adv. Intell. Syst., vol. 4, no. 9, Sep. 2022, Art. no. 2200067.
7. A. M. Almars, ‘‘Deepfakes detection techniques using deep learning: A survey,’’ J. Comput. Commun., vol. 9, no. 5, pp. 20–35, 2021.
8. X. Chang, J. Wu, T. Yang, and G. Feng, ‘‘DeepFake face image detection based on improved VGG convolutional neural network,’’ in Proc. 39th Chin. Control Conf. (CCC), Jul. 2020, pp. 7252–7256.
9. Y. Fu, T. Sun, X. Jiang, K. Xu, and P. He, ‘‘Robust GAN-face detection based on dual-channel CNN network,’’ in Proc. 12th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI), Oct. 2019, pp. 1–5.
10. N.-T. Do, I.-S. Na, and S.-H. Kim, ‘‘Forensics face detection from GANs using convolutional neural network,’’ in Proc. ISITC, 2018, pp. 376–379.
11. F. F. Kharbat, T. Elamsy, A. Mahmoud, and R. Abdullah, ‘‘Image feature detectors for deepfake video detection,’’ in Proc. IEEE/ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA), Nov. 2019, pp. 1–4.
12. J. Parmar, S. Chouhan, V. Raychoudhury, and S. Rathore, ‘‘Open-world machine learning: Applications, challenges, and opportunities,’’ ACM Comput. Surv., vol. 55, no. 10, pp. 1–37, Oct. 2023.
13. B. Bayar and M. C. Stamm, ‘‘A deep learning approach to universal image manipulation detection using a new convolutional layer,’’ in Proc. 4th ACM Workshop Inf. Hiding Multimedia Secur., Jun. 2016, pp. 5– 10.
14. B. Chesney and D. Citron, ‘‘Deep fakes: A looming challenge for privacy, democracy, and national security,’’ Calif. L. Rev., vol. 107, p. 1753, Jan. 2019.
15. K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.
16. O. A. Montesinos López, A. Montesinos López, and J. Crossa, ‘‘Fundamentals of artificial neural networks and deep learning,’’ in Multivariate Statistical Machine Learning Methods for Genomic Prediction. Cham, Switzerland: Springer, 2022, pp. 379– 425.
17. X. Wu, D. Hong, J. Chanussot, Y. Xu, R. Tao, and Y. Wang, ‘‘Fourierbased rotation-invariant feature boosting: An efficient framework for geospatial object detection,’’ IEEE Geosci. Remote Sens. Lett., vol. 17, no. 2, pp. 302–306, Feb. 2020.

18. C. Clarke, J. Xu, Y. Zhu, K. Dharamshi, H. McGill, S. Black, and C. Lutteroth, ‘‘FakeForward: Using deepfake technology for feedforward learning,’’ in Proc. CHI Conf. Hum. Factors Comput. Syst., Apr. 2023, pp. 1–17.

19. S. Solaiyappan and Y. Wen, ‘‘Machine learning based medical image deepfake detection: A comparative study,’’ Mach. Learn. Appl., vol. 8, Jun. 2022, Art. no. 100298.

20. S. Tufail, H. Riggs, M. Tariq, and A. I. Sarwat, ‘‘Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms,’’ Electronics, vol. 12, no. 8, p. 1789, Apr. 2023.

21. P. Theerthagiri and G. B. Nagaladinne, ‘‘Deepfake face detection using deep InceptionNet learning algorithm,’’ in Proc. IEEE Int. Students’ Conf. Electr., Electron. Comput. Sci. (SCEECS), Feb. 2023, pp. 1–6.

22. R. Chauhan, ‘‘Deep learning-based methods for detecting generated fake faces,’’ Authorea Preprints, 2023.

23. D. AbdElminaam, N. Sherif, Z. Ayman, M. Mohamed, and M. Hazem, ‘‘DeepFakeDG: A deep learning approach for deep fake detection and generation,’’ J. Comput. Commun., vol. 2, no. 2, pp. 31–37, Jul. 2023.

24. [24] M. A. Arshed, S. Mumtaz, M. Ibrahim, C. Dewi, M. Tanveer, and S. Ahmed, ‘‘Multiclass AI-generated deepfake face detection using patchwise deep learning model,’’ Computers, vol. 13, no. 1, p. 31, Jan. 2024.

25. [25] F. M. Salman and S. S. Abu-Naser, ‘‘Classification of real and fake human faces using deep learning,’’ Tech. Rep., 2022.

26. J. C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proença, and J. Fierrez, ‘‘GANprintR: Improved fakes and evaluation of the state of the art in face manipulation detection,’’ 2019, arXiv:1911.05351.

27. Z. Zhang, Z. Lei, M. Omura, H. Hasegawa, and S. Gao, ‘‘Dendritic learning-incorporated vision transformer for image recognition,’’ IEEE/CAA J. Autom. Sinica, vol. 11, no. 2, pp. 539–541, Feb. 2024.

28. H. M. T. Khushi, T. Masood, A. Jaffar, S. Akram, and S. M. Bhatti, ‘‘Performance analysis of state-of-the-art CNN architectures for brain tumour detection,’’ Int. J. Imag. Syst. Technol., vol. 34, no. 1, Jan. 2024, Art. no. e22949.

29. E. Hassan, M. S. Hossain, A. Saber, S. Elmougy, A. Ghoneim, and G. Muhammad, ‘‘A quantum convolutional network and ResNet (50)- based classification architecture for the MNIST medical dataset,’’ Biomed. Signal Process. Control, vol. 87, Jan. 2024, Art. no. 105560.

30. H. Wang and L. Ma, ‘‘Image generation and recognition technology based on attention residual GAN,’’ IEEE Access, vol. 11, pp. 61855–61865, 2023.