International Journal of Emerging Trends in Science and Technology
11 (2) December 2025, 10-16
www.publishingindia.com/ijetst

Deepfake Image Detection Using GWO-Based Hypertuning and ResNet-RS Architecture-Based Classification

S. Sujanthi
Department of Computer Science and Engineering, M.Kumarasamy College of Engineering,
Karur, Tamil Nadu, India. Email: sujanthi.s@gmail.com

Abstract: Online platforms are greatly endangered by deepfakes, which are synthetic pictures created using deep generative models. This has prompted a number of research initiatives aimed at improving deepfake picture detection, with promising results on publically accessible datasets. Our analysis of eight cutting-edge detectors leads us to believe that, as a result of two new advancements, they are still not prepared for deployment. First, the possibility of an attacker creating several customized generators (to make deepfakes) has the potential to significantly increase the danger surface, especially with the advent of lightweight tools to alter big generative models. We prove that current safeguards aren't up to the task of protecting publicly accessible user-customized generative models. We go over some novel ML techniques that leverage ensemble modeling and content-agnostic features to outperform user-customized models in generalization. Furthermore, hostile deepfakes that circumvent current defenses may be created by exploiting the proliferation of vision foundation models, which are models based on machine learning trained on generalized data that can be readily adjusted to do several downstream tasks. By meticulously manipulating the semantic content of the images, we provide a straightforward adversarial strategy that uses pre-existing foundation models to create adversarial samples free of adversarial noise. We point out where various defenses are weak in the face of our assault and discuss potential ways forward that make use of adversarial training and sophisticated foundation models.

Keywords: Deepfake images, GAN model, Generative models, GWO tuning.

I. Introduction

There are currently issues with the deep forge picture recognition system, including excessive computation for identifying samples of multi-scene and multi-forgery approaches, limited generalization ability, and unknown content [1]. Built on top of advanced machine learning algorithms, deepfake makes use of deep neural networks to generate fake media that looks, sounds, and reads just like the real thing. Improvements in computing power, data availability, and deep learning models have caused this technology to progress at a fast pace. Much research has focused on the ethical considerations, societal effects, potential abuse, and legal frameworks related to Deepfake technology. In response to the difficulties caused by Deepfake material, detection methods based on deep learning methodologies have been created [2]. Making a photorealistic picture out of low-dimensional random noise is possible with the help of generative adversarial networks (GANs). The usage of such a synthetic (false) picture on social media networks, replete with objectionable information, has the potential to generate significant issues. An effective and efficient picture forgery detector is essential for the successful detection of fraudulent images. Nevertheless, traditional image forgery detectors are unable to identify GAN-based generator-generated false pictures due to the fact that these images are created and edited from the original image [3]. A lot of people are worried that digital stuff isn't real anymore because of all the deepfake photos and other modified media that have been floating around. New difficulties in picture forensics and security have arisen with the introduction of Generative Adversarial Networks (GANs) that can produce very lifelike false pictures [4]. A major threat to the authenticity of digital material in the modern information era is the ever-evolving deepfake technology [5]. These days, it takes a trained eye to tell the difference between a genuine picture and a deepfake; made photos seem so lifelike. Thanks to significant advancements in this area of technology, tools for simply creating very realistic deepfakes are now readily available to the public. Purposes such as identity theft, fraud, spreading false information, etc., make use of these altered photos [6]. Many fields have found great use for deep learning, including computer vision, machine learning, and natural language processing. It's most recent use case is the creation of deepfakes, which are very popular manipulated

movies or photos that are realistic in quality and have gained a lot of attention. Deepfake technology is being investigated for novel applications, however there are major worries about its abuse potential [7]. Since anybody may upload, delete, or change the photos on online social networks, the spread of false or colorized photographs is a major cause for worry. There has been no success in creating a model or method that can reliably identify or classify photos as real or phony as of yet [8]. Identifying deepfake media is a challenging but crucial area of study in the field of social communication. Concerns concerning its inappropriate use have arisen in light of the widespread development of deepfake methods, which has made this technology readily accessible and very effective [9]. Worries over deepfake technology's possible abuse for cybercrimes, misinformation, and propaganda have grown in tandem with its popularity. Therefore, protecting digital authenticity and preventing the spread of bogus material has made deepfake detection an important study subject [10]. The misuse of digital picture modifications to disseminate false information has become rampant. The research community has put a lot of effort on this, but much of the previous work has been performance-driven, meaning it has optimized performance using standard/heavy networks that are meant for semantic classification. Models for detecting false pictures still lack a comprehensive understanding [11]. One method of creating moving pictures is by utilizing a video editor. Here, the picture shows a person's face. This face can be animated, meaning it may be made to convey emotions in response to changes in the video frame, which might include a person speaking [12]. The pace of dissemination of deep false information, particularly in visual formats like photographs and movies, is soaring. Sophisticated DL algorithms like GANs, autoencoders, and variation autoencoders are used to produce fake material. By lowering the credibility of social media content, this false material propagates disinformation and has far-reaching societal consequences. By harnessing the capabilities of one of the DL models—CNN—we can reduce the risk associated with these approaches [13].

II. Related Work

In this research, we provide a method for identifying false photos using contrastive loss that is based on deep learning. The first step in creating the fake-real picture pairings is to use many state-of-the-art GANs. The next step is to convert the downsized DenseNet into a two-streamed network so that it may accept paired data as input. The next step is to train the common fake feature network to differentiate between actual and fake pictures using paired learning. Lastly, in order to determine whether the input picture is genuine or false, a classification layer is attached to the suggested common fake feature network. The suggested technique blew away competing state-of-the-art false picture detectors in the experiments [16]. This study presents a new method for identifying deepfake movies and pictures by employing the Xception model, a deep convolutional neural network (CNN) with outstanding picture classification skills. Gathering a varied dataset spanning several categories and include both actual and deepfake photos and videos is the first step of the project. Deepfake detection is achieved by fine-tuning the Xception model using transfer learning. Incorporating various data augmentation tactics and regularization approaches during training enhances model resilience. The suggested model achieves a test-dataset accuracy of 93.11%. Combating the proliferation of misleading and harmful deepfake information throughout the internet and social media platforms is one area where the suggested technique shows considerable possibilities. Using the features of the Xception model, this study concludes with a dependable and effective approach to deep false picture and video detection [17]. Recent advances in this field have made it possible for anybody to get tools that may be used to generate deepfakes that seem quite genuine. These altered pictures are then used for things like fraud, identity theft, spreading false information, etc. Many people use this technology to good use, therefore not everyone is bad with it. Film, animation, and video game industries are some ideal places to find some of the good applications. The necessity for efficient methods to identify altered material has arisen as a result of its harmful use. To begin, the images are pre-processed using the Error Level Analysis Algorithm. Then, they are sent to CNN for model training and fake/real picture classification [18]. This paper details the development of an AI system that can identify deepfakes in many datasets. For the purpose of deepfake detection, this research suggests deep learning neural network models that use random forests (RF) and support vector machines (SVMs) as classifiers. In this groundbreaking research, three open-source datasets—FaceForensics++, FaceAntiSpoofing, and iFakeFaceDB—were used to identify deepfakes in photos. The classifiers used were RF and SVM, which were combined with a convolutional neural network. In iFakeFaceDB, CelebA-Spoof, and FaceForensics++, the suggested technique demonstrates an accuracy of 96%, 87%, and 52%, respectively [19]. Image forgery detection is the subject of this research because of the issues it generates. In what ways some societal issues that may arise from the use of fake images include skewed public perceptions of political or religious figures, the defamation of famous persons and their work, and the potential for legal proceedings to be influenced by false images. This paper presents a method for detecting false photos using deep learning, specifically Convolutional Neural Network (CNN) architecture. The network's foundation is a CNN with Xception net-style modifications that use depthwise separable convolution layers. We employ pooling layers with dense connections to Xception output after feature map extraction to increase feature maps. Motivated by the concept of a densenet environment. However, as opposed to RGB, HSV, Lab, or any other color system, the YCbCr color system yielded a superior accuracy of 99.93% in this experiment [20].

III. Proposed Work

A major danger to the veracity of internet material has emerged with the advent of deepfakes in the current day. All businesses, especially those with contact centers, must now have deepfake detection systems to protect themselves against these frauds. However, it is still in its early stages and may be used to head off an impending issue. Therefore, deepfake detection research should be given top priority.

Within the forensics dataset known as FaceForensics++, four automated face alteration methods—Deepfakes, Face2Face, FaceSwap, and NeuralTextures—have been applied to one thousand original video sequences. All 977 movies included in the dataset include a trackable, primarily frontal face and no occlusions, which allows automated tampering techniques to create convincing fakes. The architectural analysis of the research approach is shown in Fig. 1. In order to construct the neural network methods that are used, deepfake pictures that are based on both real and fake human faces are used. Using the goal label, a dataset is constructed containing both actual and synthetic faces. Train, validation, and test sets make up the structured deepfake dataset. When training neural network algorithms, 90% of the dataset is used for training. To get the highest possible accuracy score in deepfake detection, the newly developed DFP method that was outperformed is completely hyper-parametrized. 10% of the test data is used to evaluate neural network approaches' performance on unseen test data. This innovative method successfully predicts outcomes based on unseen data. To distinguish between genuine and phony faces during deployment, an advanced method based on deep learning has been suggested; it is now in a generalized version. The architectural analysis of the research approach is shown in Fig. 1. In order to construct the neural network methods that are used, deepfake pictures that are based on both real and fake human faces are used. Using the goal label, a dataset is constructed containing both actual and synthetic faces. Train, validation, and test sets make up the structured deepfake dataset. When training neural network algorithms, 90% of the dataset is used for training. To get the highest possible accuracy score in deepfake detection, the newly developed DFP method that was outperformed is completely hyper-parametrized. Ten percent of the test data is used to evaluate neural network approaches' performance on unseen test data. This innovative method successfully predicts outcomes based on unseen data. To distinguish between genuine and phony faces during deployment, an advanced method based on deep learning has been suggested; it is now in a generalized version. Here we have a look at the neural network approaches that were used, which were based on transfer learning. Using pre-trained models for prediction is the goal of transfer learning. The capacity to apply the acquired characteristics for effective prediction is a key component of transfer learning. One way to fine-tune a network using transfer learning is to take an existing network and retrain it using the new dataset. In this part, we will examine the inner workings of the transfer learning algorithms that are used for deepfake detection. Neural network methods' architectural analysis and setup settings are run.

Utilizing a combination of RESNET-RS and convolutional neural network architecture, an innovative DFP method is suggested. Our study is the first of its kind to use a neural network architecture for deepfake detection that combines transfer learning with deep learning. The suggested model architecture is a hybrid of ResNet-RS and convolutional neural network layers. One class of artificial neural networks that has found widespread usage in image identification is the convolutional neural network family. It is the purpose of the convolutional neural networks to handle pixel data. Convolutional neural networks use multidimensional array structures to represent the picture data. Artificial neural networks are primarily designed to learn patterns from past data and provide predictions for new data that has not yet been seen. The study examines the architecture and configuration parameters of the suggested model layer. Using the suggested method, the architecture analysis shows how data flows from input to prediction layers in deepfake detection images. This design incorporates ResNet-RS and convolutional neural network hybrid layers. The innovative model architecture that was suggested was built by combining the pooling, dropout, flatten, and fully-connected layers.

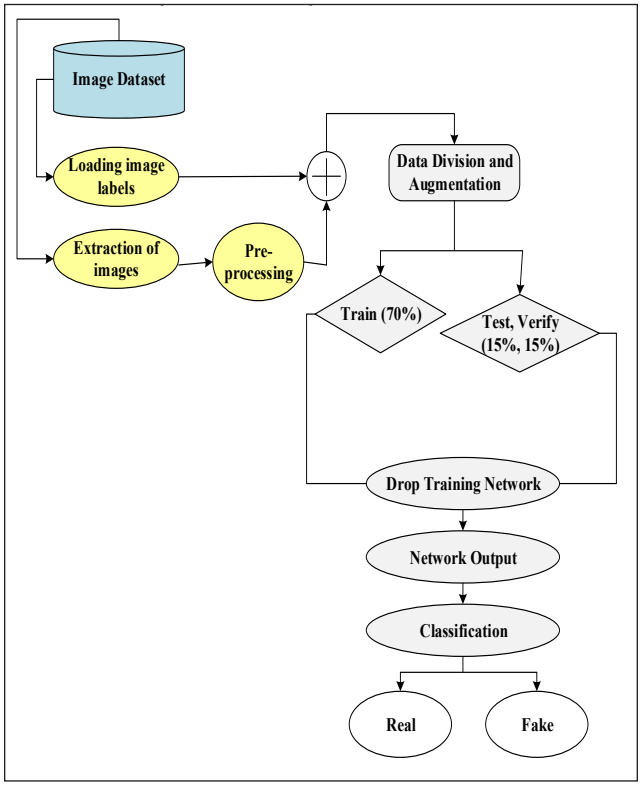

Fig. 1: Deepfake Images Detection Model

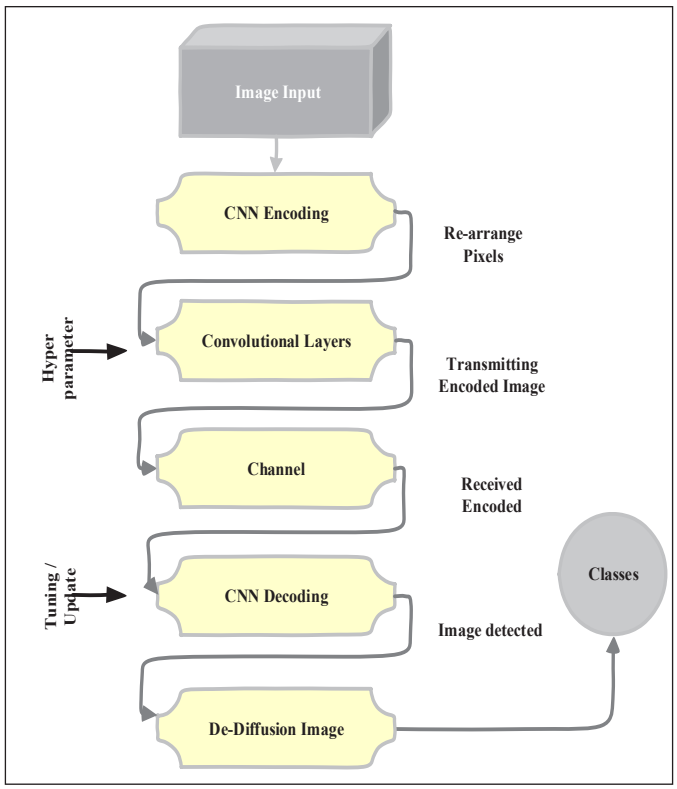

Fig. 2: GWO with ResNet-RS Model

When training and evaluating a neural network approach, hyperparameter tweaking is a cyclical process. For deepfake detection, the optimal parameters were found by applying hyperparameter tuning to the neural network algorithms that were used. Improved performance accuracy was the end outcome of using the best-fit parameters. We used the optimizer to change the neural network's weights. While training, the loss function dictated how much of a loss the neural network would experience. The metric used was the accuracy measure. There was sigmoid activation in the output layer. The neural network method goes through iterations of data learning, and each iteration is represented by an epoch.

IV. Results and Discussion

This section delves into the study findings and provides thorough thoughts about them. We take a look at the scientific assessment measures and programming tools used for neural network approaches. The neural network approaches were built using the research experimental setup. Every single one of the study's experiments was evaluated using the Python programming language. To construct the neural network methods used, the following modules were utilized: TensorFlow with version 2.8.2, and Keras with version 2.8.0. We looked at the performance-based scientific reviews of the study. Performance assessments made use of the following metrics: computation time, loss sore, accuracy score, precision score, f1 score, specificity score, and geometric mean score. At first, load the dataset by browsing 'dataset' folder. If we select the wrong folder, then it shows a prompt message to select the correct 'dataset' folder.

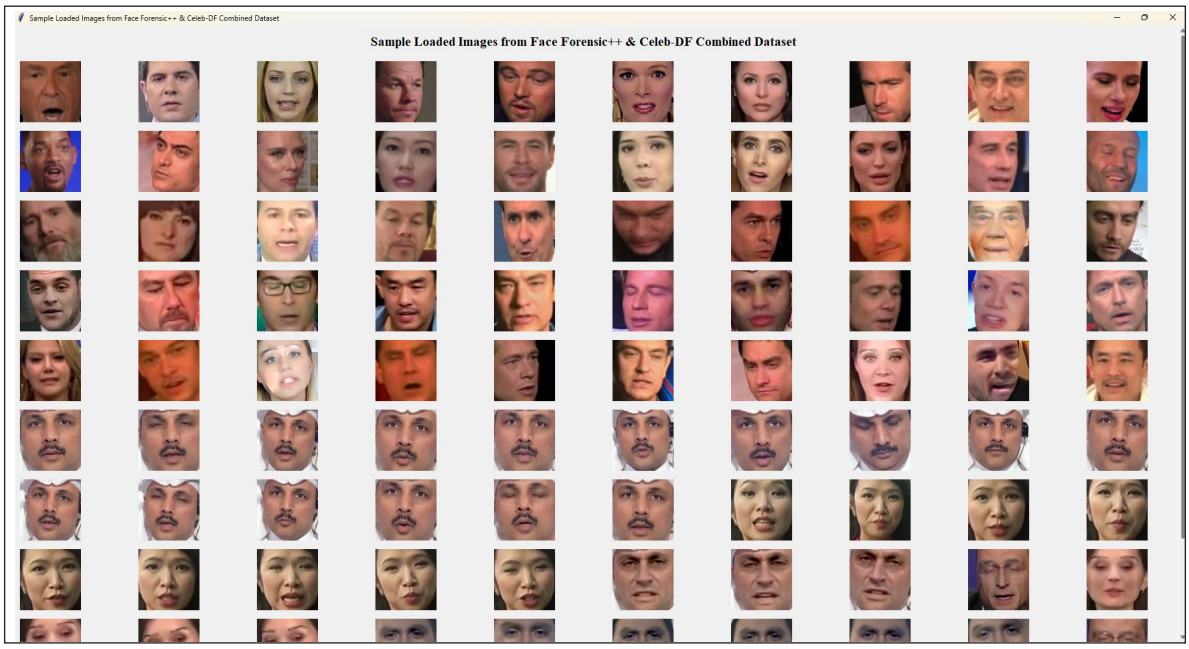

Fig. 3: Sample Images

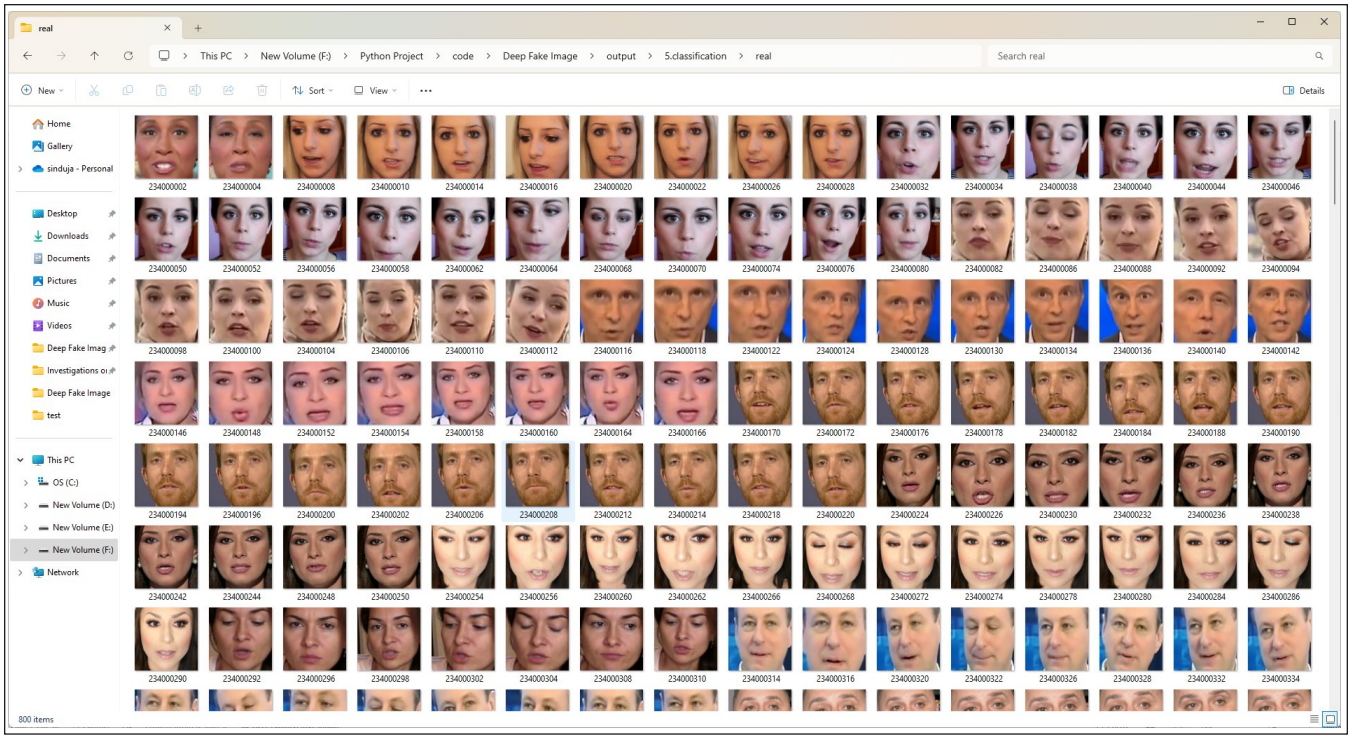

Fig. 4: Classification of Real Images

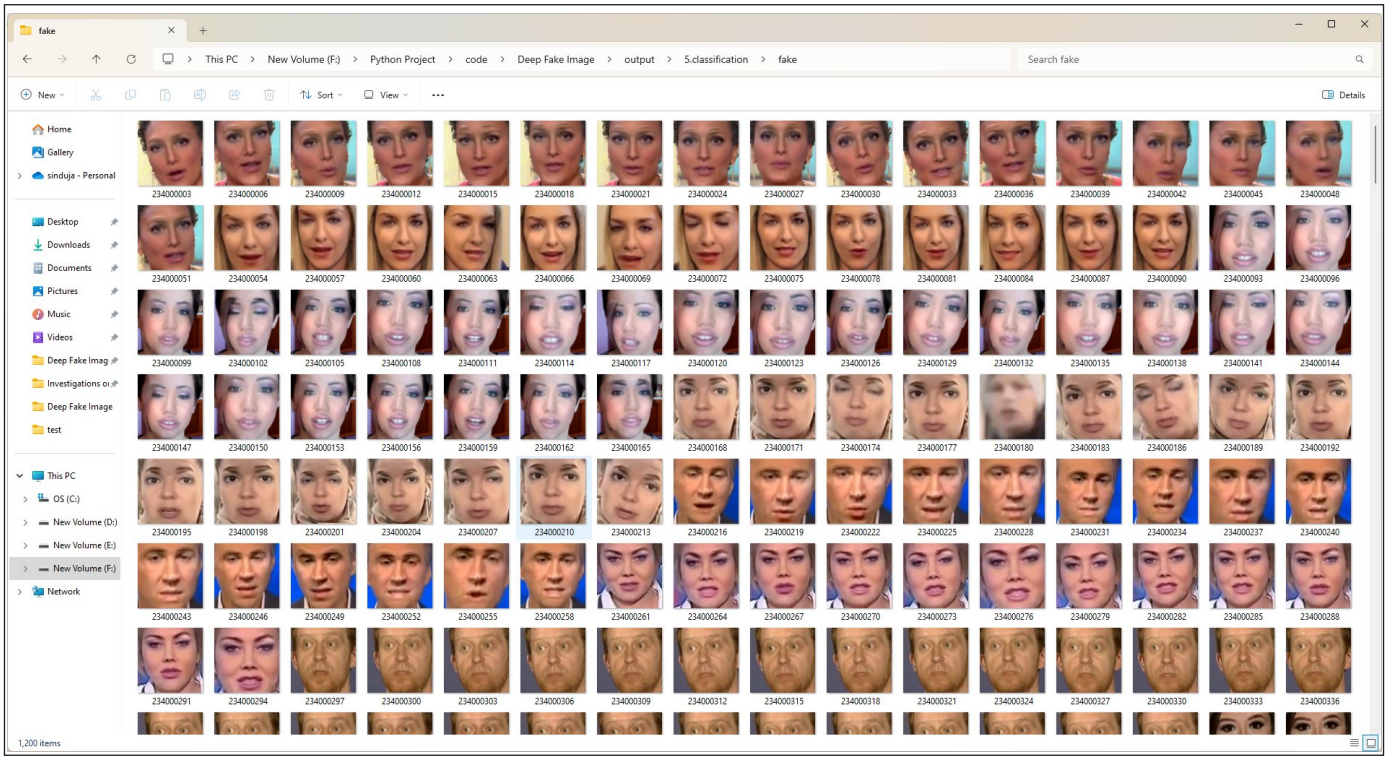

Fig. 5: Classification of Fake Images

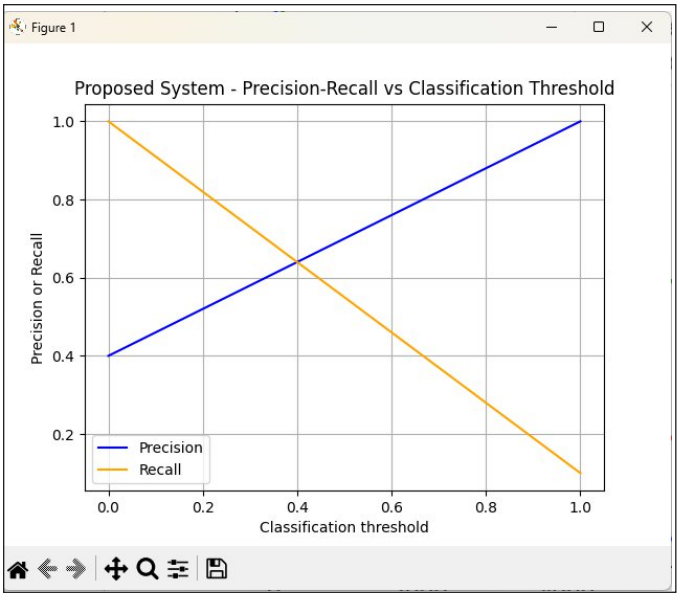

Fig. 6: Precision-Recall vs Classification Threshold

Finally, we performed the comparative performance evaluation in terms of different performance metrics like accuracy, precision, recall, f1-score, MSE, etc. Here, we compared the proposed system with exiting systems like AlexNet and CNN.

V. Conclusion

Cybersecurity experts can battle deepfake-related cybercrimes with the aid of a new deep learning-based technique to deepfake detection. This approach correctly detects deepfake material. In this comparison, the neural network approaches used are ResNet-RS and GWO. For the purpose of creating the study models, the benchmark deepfake dataset was employed, which contains both false and genuine faces. When tested using popular learning methods and cutting-edge studios, the suggested ResNet-RS and GWO method performed better. When it came to deepfake detection, the suggested model was 99% accurate. Time series analysis and confusion matrix validation both provided insight into the performance of the neural network methods used. Our suggested approach for identifying deepfakes via a secure online system will be built to use blockchain-based technologies and a cloud web system in the future.

References

[1] J. Li, and B. Li, "Deep fake image detection based on decoupled dynamic convolution," in *2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, 2022, pp. 27-31.

[2] W. Khimi, K. Albarqi, K. Saif, and S. Elhag, "A systematic review on deep fake image generation, detection techniques, ethical implications, and overcoming challenges," *International Journal of Computers and Informatics*, 2024.

[3] A. Gururaj, N. M. Ajaai, J. M. Eswaran, and B. C. Swetlin, "Deep fake image and video detection using machine learning," *International Journal of Advanced Research in Science, Communication and Technology*, 2024.

[4] S. Tiwari, A. K. Dixit, and A. K. Pandey, "Fake image detection using generative adversarial networks (GANS) and deep learning models," *Journal of Dynamics and Control*, 2024.

[5] D. Kothandaraman, S. S. Narayanan, Mohd. Iqbal M., A. Yekopalli, and S. Krishnadevarayalu S., "Deep fake image classification engine using inception-ResNet-V1 network," in *2024 International Conference on Computing and Data Science (ICCDS),* 2024, pp. 1-5.

[6] M. Usha, and B. Pradeep, "Deep fake video/image detection using deep learning," *Global Journal of Engineering and Technology Advances*, 2024.

[7] A. Jagdale, V. Kubde, R. V. Kortikar, P. A. Mote, and P. N. Rajgure, "Deepfake image detection: Fake image detection using CNNs and GANs algorithm," *International Journal of Scientific Research in Engineering and Management*, 2024.

[8] S. Selva Birunda, P. Nagaraj, S. Krishna Narayanan, K. Muthamil Sudar, V. Muneeswaran, and R. Ramana, "Fake image detection in twitter using flood fill algorithm and deep neural networks" in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence),* 2022, pp. 285-290.

[9] S. Osowski, and M. Golgowski, "Deep classifiers and wavelet transformation for fake image detection," *Journal of Telecommunications and Information Technology*, 2023.

[10] M. Berrahal, M. Boukabous, and I. Idrissi, "A comparative analysis of fake image detection in generative adversarial networks and variational autoencoders," in *2023 International Conference on Decision Aid Sciences and Applications (DASA),* 2023, pp. 223-230.

[11] J. Xu, W. Zhang, Y. Bai, Q. Sun, and T. Mei, "Flat and shallow: Understanding fake image detection models by architecture profiling," in *Proceedings of the 3rd ACM International Conference on Multimedia in Asia*, 2021.

[12] Bhanu Priya M., and J. F. Daniel, "First order motion model for image animation and deep fake detection: Using deep learning," in *2022 International Conference on Computer Communication and Informatics (ICCCI),* 2022, pp. 1-7.

[13] T. Osakabe, M. Tanaka, Y. Kinoshita, and H. Kiya, "CycleGAN without checkerboard artifacts for counter-forensics of fake-image detection," 2020.

[14] D. M. Umadevi, M. S. Barghav, and M. N. Kumar, "Deep fake face detection using efficient convolutional neural networks," in *2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN)*, 2024, pp. 344-352.

[15] C. Hsu, Y. Zhuang, and C. Lee, "Deep fake image detection based on pairwise learning," *Applied Sciences*, 2020.

[16] P. Joshi, and Nivethitha V., "Deep fake image detection using Xception architecture," in *2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST),* 2024, pp. 533-537.

[17] A. Akram, and V. Chaitanya, "A study on deep fake image detection using ELA and machine learning," *International Journal of Innovative Research in Information Security*, 2024.

[18] O. A. Al-Dulaimi, and S. Kurnaz, "Deep fake image detection based on deep learning using a hybrid CNN-LSTM with machine learning architectures as classifier," in *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2024, pp. 1-7.

[19] I. Sahib, and T. A. AlAsady, "Deep fake Image detection based on modified minimized Xception Net and DenseNet," in *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, 2022, pp. 355-360.