DeepFake Detection Using Error Level Analysis and Deep Learning

Rimsha Rafique[1], Mariam Nawaz[2], Hareem Kibriya, Momina Masood
Department of Computer Sciences
University of Engineering and Technology, Taxila
rimsha.rafique@students.uettaxila.edu.pk[1], marriam.nawaz@uettaxila.edu.pk[2]

***Abstract*—The image recognition software is used in numerous distinctive industries that include entertainment and media. The deep learning (DL) algorithms have been of great help in the development of several techniques used for creating, altering, and locating any data. The deepfake method is a photo-faking technique that includes replacing two people's faces to an extent that it becomes very difficult to identify it with a naked eye. The convolution neural network (CNN) models including Alex Net and Shuffle Net are used to recognize genuine and counterfeit face images in this article. The technique analyzes the performance and working of all distinctive algorithms using the real/fake face recognition collection from Yonsei University's Computational Intelligence Photography Lab. The first step in the process starts by the normalizing of pictures then the Error Level Analysis is carried out before it is put into several difference CNN models. Then the in-depth features are extracted from the CNN models utilizing the Support Vector Machine and the K-nearest neighbor methods. The most perfect accuracy of 88.2% of Shuffle Net via KNN was analyzed while Alex Net's vector had the accuracy of 86.8%.**

Keywords—Deep Learning, Machine Learning, CNN, Deepfake Detection, SVM, KNN

I. INTRODUCTION

Media forensics has attracted researchers much interest in recent years because of the growing issues around ***DeepFakes*** [1]. The term "Deepfake" is based on a DL technique in which fake videos and images are created by changing the face of one person with another person. Deepfake technology finds many applications that are beneficial in the media industry like lip sync, face swapping, and de-aging people. Even though the advancement in DL and deepfake technology has several helpful applications in business, entertainment, and the film industry, but on the other hand they also conjointly serve malicious purposes and build individuals' struggle to believe what's real [2].

Several different forms of applications including computer vision, data mining, natural language processing is now using DL. The picture forensic technique is the one used to know if an image is genuine or not, however, it tends to get tough when the number of images is more than 100. [3] Different forms of approaches are opted for in computer vision and image processing to differentiate between real and fake images.

Deep learning is a subset of Machine Learning (ML) and Artificial Intelligence (AI) that employs previously acquired knowledge to solve problems. The data used by DL models are both structured and unstructured. Transfer Learning works on pre-trained models. These models are trained on large data sets and improve the model's efficiency by making them more intelligent. It is actually enhancement of learning novel information from the previous model or task that has been learned [3].

We present two strategies for detecting deep fakes utilizing ELA and DL techniques in this paper. In ELA digital data, i.e., images are compressed at a positive degree of image quality, and then we take the difference of this compressed data with original data. The pictures from the dataset are first normalized by scaling them to 255 x 255 pixels, and then ELA is applied to the scaled images. After that, we send these pictures to CNN models like Alex Net and Shuffle Net, which have been fine-tuned. These deep feature vectors are also provided to SVM and KNN. The following sections go through the specifics of each of these approaches.

Section II highlights previous research on the topic, whereas Section III illustrates the proposed method for deep-fake detection. Section IV details the dataset and experimental results, and finally, section V concludes the whole study.

II. RELATED WORK

Recently various researchers proposed DL based deepfake detection techniques. For example, In [4] the authors proposed a new technique to detect real and manipulated videos. The proposed technique used a YOLO face detector to extract face area from video frames and then pass these faces are passed to the InceptionResNetV2 CNN model to extract visual features. These graphical features are passed to the XGBoost classifier to identify original and manipulated videos. The proposed model achieved 90.73% accuracy on the merged dataset CelebDF-FaceForencics++ (c23).

Khalil et al. [5] proposed an iCaps-Dfake method for deepfake image and video detection. This method compromises two techniques Local Binary Patterns (LBP) and CNN-based modified High-Resolution Network (HRNet), along with an application of capsule neural networks (CapsNets). The proposed system uses DeepFake Detection Challenge-Preview (DFDC-P) dataset to train the model whereas testing is done on DFDC-Preview and Celeb-DF datasets.

Wang and Dantcheva [6] studied deep learning approaches based on 3DCNN (3D ResNet, 3D ResNeXt, and I3D) in detecting manipulated videos. Faceforensics++ dataset was used to perform experiments on three different manipulation techniques and report the true classification rate (TCR) in all experiments. The proposed work put our attention to work on other deepfake manipulation techniques that can overcome its limitations. More, in the future, we plan

2021 4th International Conference on Computing & Information Sciences (ICCIS) | 978-1-6654-9441-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCIS54243.2021.9676375

to develop new deepfake-detection approaches, which emphasize appearance, motion, and pixel-level based generated noise, to get the better of manipulation algorithms.

In addition to traditional deepfake detection techniques, a Hybrid approach was announced to detect deepfake images effectively. In [7] authors proposed a hybrid approach called pairwise learning to detect deepfake images. In the first step, the proposed approach used generative adversarial networks (GANs) for fake image creation and generation. Then pairwise learning is used to take the discriminant information among the fake and real images by using the most popular common fake feature network (CFFN).

III. PROPOSED METHOD

The proposed method introduces a novel technique to detect original and fake images. Initially we preprocessed the images by resizing them to 225 x 225 and then apply ELA to find the compression ratio between real and fake images. These images are supplied to the CNN models to detect real and manipulated samples. The flow of the proposed technique is shown in Fig.1

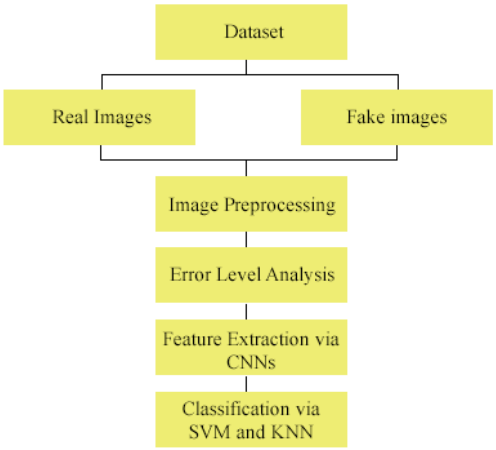

Fig. 1. Proposed Method

A. Preprocessing

The original dataset used to conduct this study consists of RGB images of different sizes. Before forwarding these images to CNN models the real and fake images are resized to 255 x 255 to decrease the computational load.

B. Error Level Analysis

After preprocessing ELA was used to check the compression ratio of images because compression level of both real and fake images are always different. In ELA images are compressed at a certain quality level and resaved. Then difference between the original image and resaved image is computed. Assume that the original image is denoted as (O_I) and resaved image (R_I) then ELA of the image is measured according to Eq.1

$$ELA = O_I - R_I \qquad (1)$$

The ELA techniques work well with lossy images such as jpg. Our dataset contains lossy images of the JPG extension. To convert images into ELA we preprocessed them and compressed both original and manipulated sample images at 85% compression level. In the end, we make the difference between preprocessed and resaved images and save them with .png extension. Some sample images from the dataset are shown in Fig. *2*

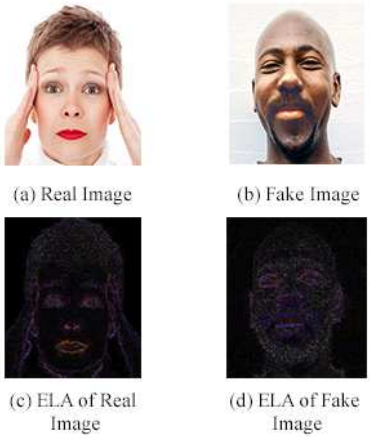

Fig. 2. Sample of some images from the dataset

C. Convolutional Neural Network

CNN is a deep learning-based approach that brought significant developments in the field of computer vision. In our proposed method we used two DL models such as Alex Net and Shuffle Net for the deepfake image detection problem. The frequently used technique in DL is transfer learning, which targets to improve learning and transfer knowledge from one model to another [8].

To train DL models from scratch a lot of resources are required, so to solve resource utilization and computational complexity problem the transfer learning frameworks are used [3]. The journal transfer learning architecture of our model is shown in Fig. *3*

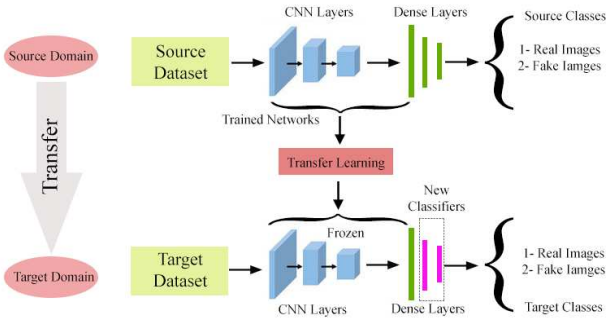

Fig. 3 Transfer Learning architecture of our model

In our proposed model we pass the ELA image to two pre-trained deep learning models which are Alex Net and Shuffle Net. These layers of these two models are adjusted according to our requirements. The input layers are changed according to the size of pre-trained images (255 x 255) and the last layers are adjusted according to the classes of the dataset. These CNN architectures are trained by using different parameters settings as shown in TABLE *1*. The most suitable parameters are presented in bold.

TABLE 1. DIFFERENT PARAMETERS USED

Parameters	**Settings**
Mini Batch	**4**,6
Epochs	**10**,20
Optimizer	**Rmsprop**, Adam
Learning Rate	**0.0001**, 0.0003

In the end, deep CNN features were classified using SVM and KNN classifiers. Both classifiers belong to the supervised learning technique. SVM and KNN are popular algorithms and used by many researchers in machine learning. SVM is used in high dimensional spaces and is memory efficient [9]. Whereas the calculation time of KNN is very fast and gives high accuracy.

IV. EXPERIMENTAL SETUP AND RESULTS

A. Dataset

In our research, we use a publically available deepfake dataset "Real and Fake Face detection" by Computational Intelligence Photography Lab, Yonsei University [10]. The dataset consists of total 2041 images that are divided into two classes real and fake images. The distribution of images is 960 fake images and 1081 real images. Some images of both categories from the dataset are shown in Fig. *4*

(a) Sample of Real Images

(b) Sample of Fake Images

Fig. 4. Real and Fake images

B. Performance Metrics

To evaluate the efficiency and performance of the proposed system we use different evaluation metrics that are Accuracy (Acc), Precision (Pre), Recall (Rec), and F1 Score. The equations used to calculate these performance metrics are mentioned below

$$Acc = {Tp + Tn}/{Tp + Tn + Fp + Fn} \quad (2)$$

$$Pre = {Tp}/{Tp + Fp} \quad (3)$$

$$Rec = {Tp}/{Tp + Fn} \quad (4)$$

$$F1\ Score = 2 * \frac{Pre*Rec}{Pre+Rec} \quad (5)$$

In the above equations Tp stands for True Positives, Tn represents True Negatives, Fp denotes False Positives, and Fn False Negatives.

C. Results

In our proposed method we use two DL models i.e Alex Net and Shuffle Net to detect deepfake images. The accuracy obtained from end-to-end classification models Alex Net and Shuffle Net is 60.5% and 60.9% respectively.

The deep feature vector we obtained is then supplied to two classifiers SVM and KNN. Alex Net obtained 86.8% and 86.1% accuracy via KNN and SVM respectively, whereas the accuracy obtained by Shuffle Net via KNN and SVM is 88.2% and 87.9%. The detailed results in terms of precision. Recall, F1 Score and accuracy for both models are shown in Fig. *5* and Fig. *6*

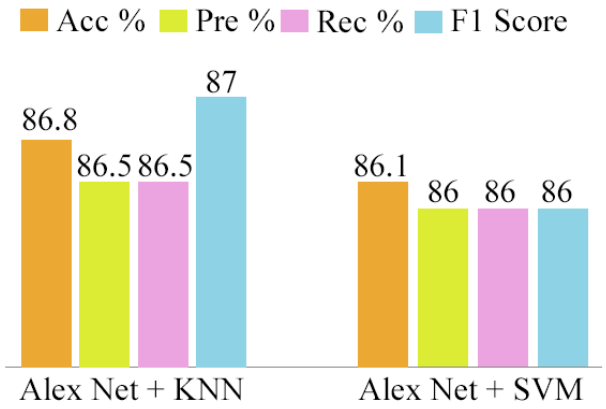

Fig. 5 Results of Alex Net via KNN and SVM

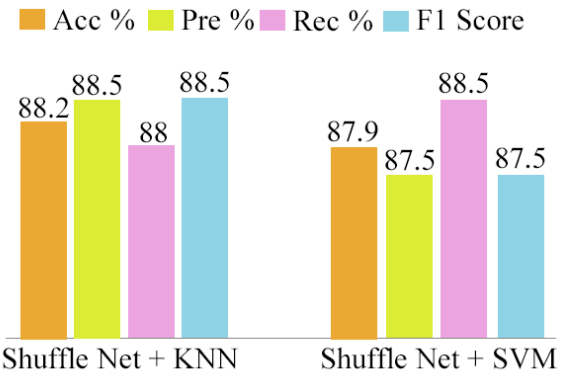

Fig. 6 Results of Shuffle Net via KNN and SVM

D. Result Comparison

We compared the results of our proposed method with other state of art methods that are mentioned in TABLE *2*. From the comparison, we see that our proposed technique gives the best result on the "Real and Fake Detection" dataset and outperforms in terms of validation accuracy. The proposed system not only achieved better performance than the existing

systems but also utilizes lightweight and efficient models to detect deep fakes from images.

TABLE 2. COMPARISON OF PROPOSED METHOD WITH EXISTING TECHNIQUES

Ref.	Year	Technique	Dataset	Acc %
[11]	2019	Logistic Regression model & Multilayer Perceptron	Private Database	85.1
[12]	2019	VGG 16	CASIA Version 2.0	88.46
[13]	2020	Alex Net KNN	Real & Fake Face Detection	55.8
		QIEA-FS		57.9
		IQIEA-FS		58.3
[14]	2021	ResNet-152	Real & Fake Face Detection	76.79
Proposed Method	**2021**	**Alex Net + KNN**	**Kaggle "Real & Fake Face Detection"**	**86.8**
		Alex Net + SVM		**86.1**
		Shuffle Net + KNN		**88.2**
		Shuffle Net + SVM		**87.9**

V. CONCLUSION

Due to robust performance, the DL techniques are widely used in deepfake detection both in videos and images. In this paper, we introduced ELA-based DL techniques to detect real and fake images. The images from the dataset are preprocessed then apply ELA to check the image compression ratio after that pass image to two CNN models i.e. Alex Net and Shuffle Net for image classification. In the end, the deep feature vector is forwarded to KNN and SVM classifiers for final output. The proposed system achieved the highest accuracy using the comparison of Shuffle Net and KNN i.e. 88.2%. Our proposed system is lightweight, robust, and efficient.

REFERENCES

[1] R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, "DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance," *arXiv preprint arXiv:2004.07532,* 2020.

[2] A. Mehra, "Deepfake detection using capsule networks with long short-term memory networks," University of Twente, 2020.

[3] H. Kibriya, R. Rafique, W. Ahmad, and S. Adnan, "Tomato Leaf Disease Detection Using Convolution Neural Network," in *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*, 2021, pp. 346-351.

[4] A. Ismail, M. Elpeltagy, M. S Zaki, and K. Eldahshan, "A New Deep Learning-Based Methodology for Video Deepfake Detection Using XGBoost," *Sensors,* vol. 21, p. 5413, 2021.

[5] S. S. Khalil, S. M. Youssef, and S. N. Saleh, "iCaps-Dfake: An Integrated Capsule-Based Model for Deepfake Image and Video Detection," *Future Internet,* vol. 13, p. 93, 2021.

[6] Y. Wang and A. Dantcheva, "A video is worth more than 1000 lies. Comparing 3DCNN approaches for detecting deepfakes," in *FG'20, 15th IEEE International Conference on Automatic Face and Gesture Recognition, May 18-22, 2020, Buenos Aires, Argentina.*, 2020.

[7] C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, "Deep fake image detection based on pairwise learning," *Applied Sciences,* vol. 10, p. 370, 2020.

[8] H. Kibriya, M. Masood, M. Nawaz, R. Rafique, and S. Rehman, "Multiclass Brain Tumor Classification Using Convolutional Neural Network and Support Vector Machine," in *2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*, 2021, pp. 1-4.

[9] V. Jakkula, "Tutorial on support vector machine (svm)," *School of EECS, Washington State University,* vol. 37, 2006.

[10] C. I. a. P. Lab. (2019,). *Real and Fake Face Detection (ed.)*. Available: https://www.kaggle.com/ciplab/real-and-fake-face-detection

[11] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, 2019, pp. 83-92.

[12] I. B. K. Sudiatmika and F. Rahman, "Image forgery detection using error level analysis and deep learning," *Telkomnika,* vol. 17, pp. 653-659, 2019.

[13] H. Mittal, M. Saraswat, J. C. Bansal, and A. Nagar, "Fake-Face Image Classification using Improved Quantum-Inspired Evolutionary-based Feature Selection Method," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 989-995.

[14] K. Chandani and M. Arora, "Automatic Facial Forgery Detection Using Deep Neural Networks," in *Advances in Interdisciplinary Engineering*, ed: Springer, 2021, pp. 205-214.