Deep fake face detection using K-Nearest Neighbor and deep learning

*Dr. Joshua C Nwokeji, **Mir Yousuf Sultan, ***Rahul Singh Brijesh, ****Sai Pramod Reddy. Gunda
Dept of Computers & Information Sciences
Gannon University, Erie, PA

Abstract—**In the age of the modern world, machine learning and computer vision is playing hard, it is going to be the future of technological advances. One such advancement is going to be facial recognition. Facial recognition is crucial since it will replace biometric authentication in the future for identity cross-verification. But there is a problem with facial recognition. The world is making its move towards advancement, it is also getting on the worst side to alter the advancement. Facial recognition can be altered using open-source applications. Altering the face is nothing but changing the facial features at pixelated level. A lot of work is done in this area but most of them worked with SVM classifier with models like VGG19, RestNet50, Xception, and LBP (Linear Binary Pattern). In this paper we propose to conduct a comparison study of several CNN model like RESNET50, LBP(local binary patter),VGG-19 and Xception with KNN classifier to find out the better-performing model and their accuracy. The result will be the detection and classification of fake and real images from the dataset. The used dataset is from the "Real and Fake Face identification" deepfake dataset from Yonsei University's Computational Intelligence Photography Lab and the size is 2100 images which are around 418MB.**

Keywords—SVM ,KNN ,CNN ,Fake face Detection, Deep Learning.

1. INTRODUCTION

Fake face detection is something that is used to detect faces using an algorithm after the changes made to the face. It is a vast topic that needs attention in the modern world because it can lead to a major security crisis. There are a lot of people in the world, and each one has a unique identity.. Previously, we used to confirm the identity by checking the name manually then later comes the biometric authentication system, and now the latest identity authentication system is face detection. It is a very advanced system that can authenticate a person's identity just by checking the image or video. The images can be used to fake the person's identity and can be used anywhere like banks, public places, and many more. The most common applications are image analysis, pattern recognition, etc. There are some methods like the Background of Deepfake, which has three phases Face-Synthesis, Face-swap, Facial- attributes and expression [1]. They are methods to manipulate the face using the GANs (Generative Adversarial Networks). Face Synthesis in this stage a phony image is used to replace the original. and generates changes in hair, freckles, etc. Face Swap, the fake face is detected using the Face Swap computer graphics concept. The skin color, hair color, age, gender, and expressions like joyful, sad, furious, and so on are among the facial characteristics and emotions. Face detection can be used in various departments like security and law enforcement, schools, and crime departments for crime detection, forensics, etc. [2]

Detecting the fake face is a complex task when we have less amount of data. To increase the efficiency of detection we need some robust computer vision models like ResNet50, LBP, VGG16, VGG19 to extract the features from the image . We passed the extracted data through KNN to find out the maximum efficiency through metrices F-1 Score, precision and Mean Square error. Through this, we can achieve two goals. One, we can successfully detect the fake and real faces and two, we can find out the most efficient model that can be used with KNN. A subset of machine learning known as supervised learning uses labeled datasets to accurately train and categorize the anticipated data. [3].

Face recognition using PCA and K-NN was presented by Z. Haiyang et al. in their paper published in Nature Communications [4]. Because lighting and positioning can readily impair facial recognition, the outcome is not particularly satisfying [4]. With the use of technologies like morphing, Snapchat, Computer Generated Face Image (CGFI), Generative Adversarial Networks, and Face2Face, it is now possible to effortlessly generate modified photos and videos in real-time [5]. The personation attack can be created using the enactment technique, which involves transferring the source actor's facial emotions to a target actor, creating the modified photographs and videos. This facial sample that was created using such techniques is known as a fake face [6]. The choice of the CNN architectures AlexNet, VGG19, and ResNet50 is based on recent research revealing extremely high performance for various tasks [7].

Our research is to compare between K-NN algorithm and some CNN predefined models like VGG16, VGG19, and ResNet50. The main aim is to detect the fake face and parallelly know the model that gives the best efficiency with KNN. Another way of increasing efficiency is to give the maximum amount of input data for training and testing the data. Training and testing the data plays a major role in detecting the image more accurately. Once we extract the features, we store them securely and use them for training the model. The efficiency of the model depends on the type of dataset that we give. Here we took the data from an open source for training the simple model, if we are getting good results, we can implement large chunks of data.

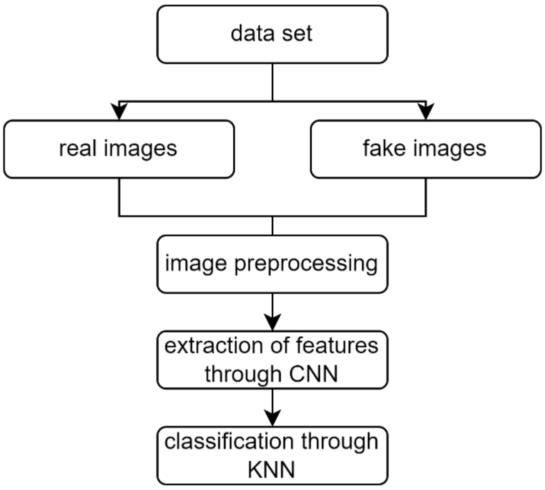
2. PROBLEM STATEMENT

AI innovation and advancement have proven to be both beneficial and detrimental. Face recognition techniques have evolved promptly with the help of Artificial Intelligence. Detecting fake faces has been a major challenge (fake faces vs. real faces). As the quality of fake faces improves, trained models become increasingly ineffective at detecting novel fake faces because the corresponding training data is deemed out of date and continues to be upgraded using a variety of techniques such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), CGI, Face2face, could generate photorealistic face images that have been entirely or partially modified [8, 9]. This work addresses only digital manipulation attacks, to automatically detect manipulated faces and localize modified facial regions. In Digital Face Manipulation there are four types of methods: expression-swap, identity-swap, attribute-manipulation, and entire face-synthesis. This method can extract the expression and transfer it to another person's images using RGB cameras. Identity replacement technology refers to any technology that is used to conceal all or a portion of a person's identity, whether in person or virtually [12]. For Example, Techniques like Face2face, which uses a renowned person's website photo to swap it out with another person's photo where they never appeared, and deepfake, which uses deep learning to conduct face swapping, are examples. Techniques [12]. The deep Fake Technique used an Autoencoder that fed images which creates images with the same resolution as the input by encoding the images using 1024 variables and 64 * 64 * 3 = 12, 288 characteristics. [11].

From our knowledge, Different algorithms were used by different researchers for feature extraction and classification. Multiple Feature extraction and classification techniques were used such as Local Directional Position Pattern (LDPP), and Principal Component Analysis (PCA). Furthermore, the research work proposed techniques such as CNN (Xception, Resenet50, VGG16, VGG19, and so on) Sajjad et al. [4] used HOG and the Uniform Local Ternary Operator to extract features and combine them into a single feature vector. A feature vector is classified with a multiclass SVM using a one-to-one approach and a rest-to-one approach.

We were going to propose a classification model i.e., a KNN classifier which is trained on a set of labeled (known) faces, finds the k most similar faces in its training set (images with the closest facial features under Euclidean distance), then conducts a majority vote (perhaps weighted) on their label to determine who is depicted in an ambiguous image.The Eigen feature extracted algorithm must be used to extract the features KNN classifier generates a response that indicates whether or not the test image is authentic. Due to similarities in the traits of the faked photos, the accuracy of the KNN classifier is decreased throughout the detection phase. The metrics that we are going to use i.e. F1 score, Mean-Squared Error.

3. THEORITICAL FRAMEWORK


data set
real images
fake images
image preprocessing
extraction of features through CNN
classification through KNN

*RQ1.*which CNN model performs better with KNN algorithm?
RQ2. As coming up with multiple CNN models Is the KNN algorithm reliable?

4. RESEARCH METHODOLOGY

A. Data Collection:

We use the "Real and Fake Face identification" deepfake dataset from Yonsei University's Computational Intelligence Photography Lab for our research [10]. The images in the dataset are in RGB format which includes 2041 total photos that are split into two categories: real photographs and fraudulent images. 960 fake photos and 1081 genuine photographs make up the distribution of the images. Fig. 1 displays a few photos from the collection representing both categories.

B. Data Analysis:

In this method, we preprocessed the images by resizing them into 225x225 and passed it through different CNN models to extract the features through the different layers of the model and then pass the extracted data through the classification algorithm KNN.

Image preprocessing:

The RGB photos in the original dataset, which was used to carry out this study, come in various sizes. To reduce the computational load, the real and fake images are reduced to 255 x 255 before being sent to CNN models. Fig. 1 displays a few photos from the collection representing both categories.

(a) Sample of Real Images

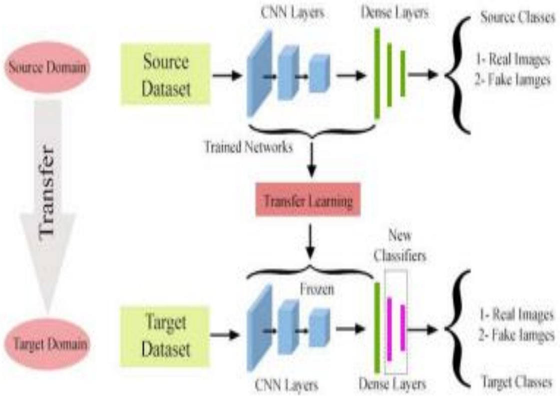
(b) Sample of Fake Images

Fig. 2. Real and Fake images

Convolutional Neural Networks:

CNN is a deep learning method with numerous layers, and over time, DL has made tremendous progress that has aided researchers in understanding deep neural networks. In our proposed methods we are taking four DL models like RESNET50, LBP(local binary patter),VGG-19 and Xception for the deep fake detection problem Transfer learning, which aims to enhance learning and transfer knowledge from one model to another, is a frequently employed DL technique.[13].

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 model of our architecture is shown in the figure below.



In our suggested model, four deep learning models RESNET50, LBP (local binary patter),VGG-19, and Xception are fed with the preprocessed images(225x225). The size of our preprocessed images determines how the input layers are modified . KNN classifiers were used to categorize the retrieved features. The KNN algorithm operates quickly and provides highly accurate results.

We evaluate and compare the performance of KNN with all four CNN models like RESNET50,LBP(local binary patter),VGG-19 and Xception on the basses of F1 score and mean square error.

References:

[1] c. U. A. A. K. V. N. B. K. H. Š. T. P. Suganthi ST, "Deep learning model for deep fake face recognition and detection. 2022.

[2] H. S. S. R. V. M. Bhatt, *Emerging Covariates of Face Recognition*, Delhi: Indraprastha Institute of Information Technology, 2014.

[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] H. Zhang, *The Research of Face Recognition Based on PCA and KNN*, China: Suzhou University, 2012.

[5] M. M. S. C. T. J.Thies, "Face2face: Realtime face capture and reenactment of rgb videos," in *IEEE*, 2016.

[6] R. R. C. B. A. Khodabaksh, "A taxonomy of audiovisual fake multimedia content creation technology," 2018.

[7] R. R. K. R. Ali Kudabaksh, *Fake Face Detection Methods: Can they be Generalized?.*

[8] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In ICLR, 2014

[9] Pavel Korshunov and Sebastien Marcel. DeepFakes: a new ´ threat to face recognition? assessment and detection. arXiv preprint arXiv:1812.08685, 2018.

[10] Luan Tran, Xi Yin, and Xiaoming Liu. Representation learning by rotating your faces. TPAMI, 2018.

[11] Darius Afchar, Vincent Nozick, Junichi Yamagishi, Isao Echizen. MesoNet: a Compact Facial Video Forgery Detection Network. arXiv:1809.00888v1 [cs.CV] 4 Sep 2018.

[12] Hao Dang, Feng Liu, Joel Stehouwer, Xiaoming Liu, Anil Jain Department of Computer Science and Engineering Michigan State University, East Lansing MI 48824. On the Detection of Digital Face Manipulation. In CVPR, 2020.

[13] 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.