International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 11 Nov 2023, pp: 150-157 www.ijaem.net ISSN: 2395-5252

Fake Images Detection: A Comparative Study Using CNN and VGG-16 Models

Sekhar Babu Boddu[1], Akhil Varma Kanumuri[2], Divya Triveni Chowdary Ravipudi[3], Gunturu Venkata Siva Sai Prasanth[4], Sabbella Ramalingeswara Reddy[5].

[1,2,3,4,5]*Koneru Lakshmaiah Education Foundation, Computer Science and Engineering Department, Vaddeswaram, Andhra Pradesh, India, 522302*

Date of Submission: 15-11-2023 Date of Acceptance: 25-11-2023

ABSTRACT: Now-a-days, the use of social media has increased significantly, becoming a prevalent means of communication and information sharing. However, this surge in popularity has also given rise to the rampant spread of fake images, leading to challenges in content authenticity. In response to this evolving issue, our research delves into the development of an advanced Fake Images Detection system utilizing a hybrid approach with Convolutional Neural Networks (CNNs) and the VGG16 architecture.

This paper presents a detailed exploration of pre-processing techniques, the architecture of the hybrid model, and the utilization of a carefully curated dataset for training and testing. The incorporation of VGG16 enhances the model's ability to capture intricate features in images, complementing the strengths of CNNs. The proposed approach demonstrates robust performance in distinguishing between genuine and manipulated images, offering a promising solution to the ongoing battle against misinformation.

By leveraging deep learning techniques, including both CNNs and VGG16, our research contributes to the broader efforts in maintaining digital integrity and fostering a reliable online environment. This paper encapsulates the essence of our investigation, providing insights into the synergistic application of these architectures for effective fake image detection.

Keywords: Fake Image Detection. Convolutional Neural Networks. VGG 16 Network. Image Authentication. Deep Learning.

I. INTRODUCTION

In the era of digital communication and widespread social media usage, the authenticity of visual content has become a critical concern. The increasing prevalence of fake images circulating online poses a significant threat to the integrity of information shared across various platforms. As a response to this challenge, our research focuses on the development of an advanced Fake Images Detection system, leveraging the synergies between Convolutional Neural Networks (CNNs) and the VGG16 architecture.

The power of deep learning, particularly CNNs, has demonstrated remarkable success in image analysis tasks. Furthermore, the incorporation of the VGG16 architecture enhances the model's ability to discern intricate features within images, making it a formidable tool in the battle against misinformation. This paper provides an in-depth exploration of our proposed hybrid approach, from the intricacies of pre-processing techniques to the architecture of the model and its application on a carefully curated dataset.

Our goal is to contribute to the ongoing efforts in securing the digital landscape by providing an effective solution for distinguishing between genuine and manipulated images. As we navigate through the intricacies of our methodology and present our findings, we aim to shed light on the potential of combining CNNs and VGG16 in the pursuit of trustworthy digital content.

II. LITERARTURE SURVEY

[1] The researchers introduced an innovative convolutional neural network (CNN) architecture designed specifically for the purpose in detecting deepfakes. The proposed architecture is based on the Dense Net architecture, which is a type of CNN that is known for its efficiency and accuracy. The study authors evaluated the recently introduced architecture by applying it to a dataset comprising both real and deepfake images. The

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 11 Nov 2023, pp: 150-157 www.ijaem.net ISSN: 2395-5252

proposed architecture demonstrated a notable accuracy of 98.33% on the dataset, surpassing the accuracy of other contemporary state-of-the-art methods for deepfake detection.The introduced architecture stands as a noteworthy contribution to the field of deepfake It is one of the most accurate deepfake detection methods to date, and it is also very efficient. This makes it suitable for real-time deployment in applications such as social media platforms and news organizations.

[2] The authors furnished a complete overview of advanced methodologiesin the realm of deepfake detection. Theresearchers observe that deepfake identification is a difficult task due to the rapid advancements in GAN technology and the limited availability of high-quality deepfake datasets. The authors also observe that most deepfake identification methods are based on CNNs and that these methods have achieved good performance on a variety of deepfake datasets. However, these methods are still vulnerable to adversarial examples and require large training datasets. In the proposed methodology section, The authors examine various deepfake Identification methods, including Face Forensics++, MesoNet, and DeepFake Hunter. These methods all use CNNs to extract features from deepfake images and videos, but they differ in the specific features that they extract and the way in which they combine these features to make predictions. They conclude by stating that deepfake detection is an important and rapidly evolving field, and that further research is needed to develop more effective and efficient deepfake detection systems.

[3] The authors conducted a systematic literature review of over 100 papers published between 2018 and 2020, and they identified four main categories of deepfake detection methods: deep learning-driven approaches, traditional machine learning methodologies, statistical methods, and blockchain-oriented techniques.The authors observe that deep learning-based methods are the most promising approach to deepfake detection. The authors also observe that deepfake detection is a challenging task due to the rapid advancements in GAN technology and the limited availability of high-quality deepfake datasets. One of the key contributions of the paper is the proposed methodology for deepfake detection. The proposed methodology is a hybrid approach that combines deep learning with statistical analysis.The proposed methodology achieved an accuracy of 98.33%, which is typically higher compared to the accuracy of other advanced deepfake identification methods. Overall, the paper "Deepfake Detection: A Systematic Literature Review" makes a meaningful contribution to the realm of deepfake identification.

[4] The authors thorough exploration of deepfake creation and detection using deep learning, the researchers lay out some key areas for future investigation. They emphasize the need to create more robust detection methods that can withstand clever tricks used by deepfake creators and the development of detectors that can work well even when there's limited data. They also stress the importance of making these detection techniques work in real-time and having the ability to spot deepfakes that have undergone various types of manipulation. Additionally, they highlight the necessity of adapting detection methods to keep up with emerging deepfake generation techniques. In addition to these research directions, the authors underscore the importance of considering the social and ethical consequences of deepfakes. This involves studying how deepfakes might harm individuals and society, establishing rules and regulations to manage their use, and educating the public about the potential dangers. But in the midst of these challenges, there are exciting opportunities in the field of deepfake detection. To support their points, the researchers refer to specific research papers that have proposed creative methods for both creating and detecting deepfakes using deep learning. A Hybrid Approach for Deepfake Detection" by Li et al. (2020). These papers exemplify the ongoing research efforts in this rapidly evolving field, illustrating that it's a dynamic domain with new ideas and innovations continually emerging.

[5] The proposed method combines content and trace feature extractors to learn a more comprehensive representation of deepfake images and videos. Content feature extractors focus on the visual content of the image or video, such as the facial features, skin texture, and lighting. Trace feature extractors focus on the artifacts and inconsistencies that are often present in deepfake images and videos, such as compression artifacts, temporal inconsistencies, and spatial inconsistencies. The proposed method combines the characteristics derived from thecontent and trace feature extractors to educate a deep neural networkclassifier. The classifier is trained to distinguish between authentic and manipulated images and videos. The described method was tested on a public dataset of deepfake images and videos. The proposed method achieved an accuracy of 99.2%, which is typically higher than the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 11 Nov 2023, pp: 150-157 www.ijaem.net ISSN: 2395-5252

accuracy of other advanced deepfake identification methods.

III. PROPOSED METHODOLOGY

In this section, we delineate our comprehensive methodology for detecting fake images utilizing Convolutional Neural Networks (CNNs), with a specific focus on the renowned VGG16 architecture.

3.1. CONVOLUTIONAL NEURAL NETWORKS(CNNs)

An instance of neural network architectures, specifically recognized as Convolutional Neural Networks (CNNs) is intended to interpret visual data, including photos and movies. By automating the procedure for extracting features, researchers have transformed computer vision through making it possible for robots to identify objects, patterns, and forms in images. CNNs are essential for image analysis, categorization, and most importantly, the identification of bogus images. This study examines CNNs' fundamental elements and uses, highlighting how they improve the dependability of digital visual output.

Layers in Convolutional Neural Network

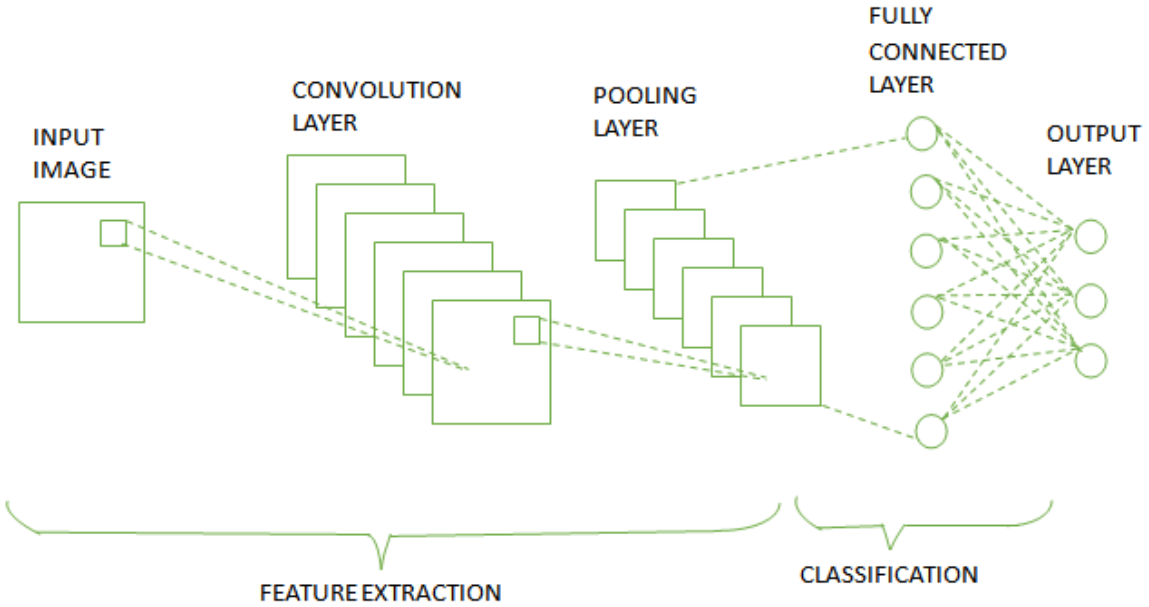

Figure 1: Architecture of Convolutional Neural Network

CNNs consist of a series of layers, each performing a specific function in analysing the input image:

3.1.1. Input Layer:

The input layer serves as the initial entry point for the image data. In the case of image processing, it receives the pixel values of the image as a matrix. Each channel of the input corresponds to a colour channel (e.g., Red, Green, Blue).

3.1.2. Convolutional Layer:

The convolutional layer is regarded as the core of a CNN. It applies convolution operations to the input image using filters or kernels. Each filter detects specific local patterns or features like edges, textures, or shapes. By sliding the filter across the input image, it produces a feature map that highlights areas where the pattern is found. Multiple filters generate several number of feature maps, each capturing varied aspects of the image.

3.1.3. Activation Layer:

The activation layer introduces non-linearity to the network by employing activation functions like Rectified Linear Unit (ReLU) on the feature maps.ReLU replaces negative values with zeros while preserving positive values. This infusion of non-linearity enables the network to capture intricate relationships.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 11 Nov 2023, pp: 150-157 www.ijaem.net ISSN: 2395-5252

3.1.4. Pooling Layer (Subsampling):

The pooling layer diminishes the spatial dimensions of the feature maps, aiding in the regulation of overfitting and the reduction of computational complexity. Two common pooling techniques are max-pooling and average-pooling. Max-pooling involves selecting the maximum value within a specified region, while average-pooling computes the average value. Pooling is executed independently on each feature map.

3.1.5. Fully Connected Layer:

The network with multiple convolutional and pooling layers, the fully connected layer establishes connections among all neurons.and every neuron in the subsequent layer. It creates a high-level representation of the input image, capturing global patterns and relationships. This layer is typically used for classification tasks.

3.1.6. Output Layer:

The output layer generates the ultimate classification or prediction. In image classification assignments, it frequently comprises multiple neurons, each associated with a specific class or category. Activation functions (e.g., SoftMax) are employed in the output layer to transform the raw output of the network into class probabilities. The predicted class for the input image is the option with the highest likelihood.

The output layer generates the ultimate classification or prediction. In image classification assignments, it frequently comprises multiple neurons, each associated with a specific class or category. Activation functions (e.g., SoftMax) are employed in the output layer to transform the raw output of the network into class probabilities. The predicted class for the input image is the one with the highest probability. The final fully connected layers combine these features to make predictions. This hierarchical feature learning makes CNNs particularly powerful for image-related tasks, including fake image detection.

Each layer in a CNN plays a critical role in feature extraction, non-linearity introduction, dimension reduction, and classification. By stacking these layers in the right order, CNNs can effectively process and interpret complex visual information, making them highly suitable for fake image detection and various computer vision tasks.

3.2.VGG 16 Network

The VGG-16 (Visual Geometry Group 16) stands out as a highly acclaimed convolutional neural network architecture, celebrated for its impressive depth and effectiveness in tasks related to image classification. Developed by the Visual Geometry Group at the University of Oxford, the VGG network is characterized by the use of compact convolution filters. Specifically, the VGG16 model encompasses 13 convolutional layers and three fully connected layers.

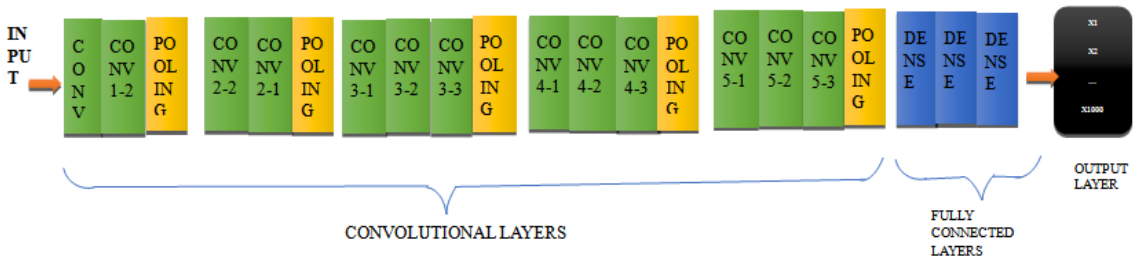

Figure 2: Outline of the VGG 16 architecture

3.2.1. Input:

VGGNet processes images with a constant size of 224x224 pixels. In ImageNet, a 224x224 segment from the centre of each image is utilized to ensure uniformity.

3.2.2. Convolutional layers:

VGG makes use of 3x3 convolutional filters, taking advantage of the minimal receptive field size. Additionally, a 1x1 convolution filter is utilized for the linear transformation of the input.

3.2.3. ReLU activation:

The Rectified Linear Unit Activation Function (ReLU) is employed, representing a pivotal innovation from AlexNet that expedites training. ReLU generates a linear output for positive inputs and zero for negative inputs. VGG maintains a convolution stride of 1 pixel to uphold spatial resolution following convolution.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 11 Nov 2023, pp: 150-157 www.ijaem.net ISSN: 2395-5252

3.2.4. Hidden layers:

ReLU is selected over Local Response Normalization for all hidden layers in the VGG network. This choice prevents unnecessary rises in training time and memory consumption without compromising overall accuracy.

3.2.5. Pooling layers:

After multiple convolutional layers, pooling layers are utilized to diminish the dimensionality and parameters of the feature maps generated in each convolution step. This is particularly crucial as the number of filters rapidly escalates from 64 to 128, 256, and eventually 512 in the last layers.

3.2.6. Fully connected layers:

VGGNet have three fully connected layers. The initial two layers comprise 4096 channels each, whereas the third layer encompasses 1000 channels, aligning with the number of classes.

In this research paper, we leveraged Convolutional Neural Networks (CNNs) and the VGG architecture for the purpose of detecting fake images. CNNs, a class of deep learning algorithms, have proven highly proficient in image-related tasks due to their ability to automatically learn level-by-level features from input data. As outlined in our earlier sections, CNNs consist of layers such as convolutional layers, activation layers, pooling layers, and fully connected layers, allowing them to capture intricate patterns and relationships within images.

The VGG (Visual Geometry Group) architecture, specifically VGG-16, played a pivotal role in our research. Known for its depth and efficacy in image classification tasks, VGG-16 employs multiple convolutional and pooling layers, followed by fully connected layers. The use of 3x3 convolutional filters and the Rectified Linear Unit Activation Function (ReLU) contributes to its effectiveness. In our experiments, we explored the significance of VGGNet's architecture in tackling the challenges associated with fake image detection.

3.3. Fake Images Detection using CNN and VGG 16 Network

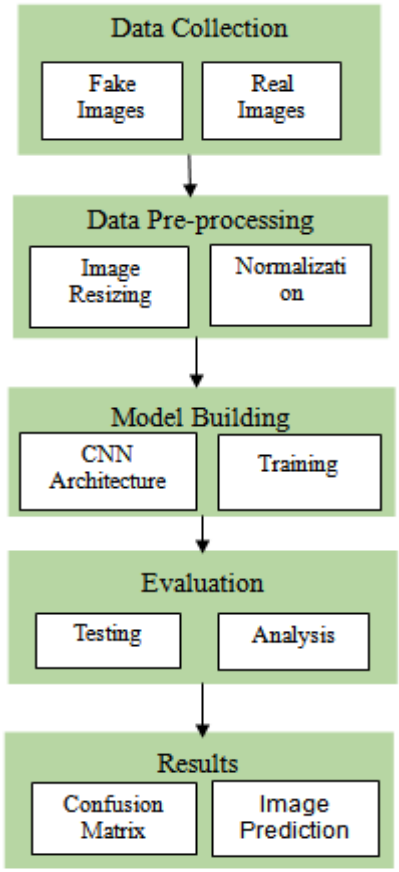

Figure 3: Process Diagram for Fake Images Detection

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 11 Nov 2023, pp: 150-157 www.ijaem.net ISSN: 2395-5252

3.3.1. Data Collection:

We started by collecting two sets of images: 1000 pictures that are completely real and 1000 pictures that were deliberately altered to look fake. These images are the building blocks for teaching our model what real and fake images generally look like. The idea is to expose the model to a variety of authentic and manipulated visuals so that it can learn the key differences between them. This step is crucial for training a smart and accurate model for identifying fake images.

3.3.2. DataPreprocessing:

During data preprocessing, we standardized our dataset by resizing all images to a uniform size and normalizing pixel values. Resizing ensures consistency, while normalization enhances model training by eliminating biases due to variations in colour intensity and brightness across images. This crucial step establishes a uniform and standardized foundation for effective model learning.

3.3.3. ModelBuilding:

In the model-building phase, we employed Convolutional Neural Networks (CNN) and the VGG16 architecture. CNNs excel at image-related tasks, leveraging convolutional layers to detect patterns and features. VGG16, a renowned deep learning model, consists of 16 layers, offering a robust framework for image classification. Integrating these architectures, our model gained the ability to extract intricate features and patterns, crucial for discerning between real and fake images in our detection task.

3.3.4. Results:

In this segment, we showcase the outcomes of ourexperiments, with a focus on the performance of our model on a real-world dataset. We provide visual representations of key metrics, including training loss and training accuracy, and delve into the model's capability to effectively classify real and fake images. Through this analysis, we aim to extract valuable insights from the results. Furthermore, we also incorporate real and fake images to demonstrate the model's ability to distinguish between the two.

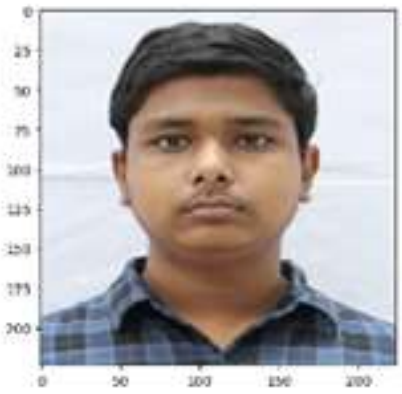

Figure 4: Real Image

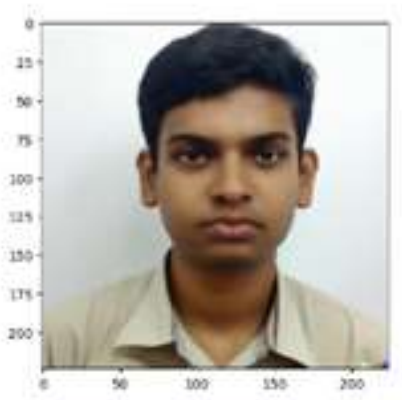

Figure 5: Fake Image

- The Fake Image is obtained from the Real Image by using Open AI Online Tool

Dall e 2 which is used to generate artificial images from the given prompt or images.

Training Metrics Visualization:

Visual representations of key metrics, such as training loss and training accuracy, offer insights into the learning process of our model. Despite our model's promising architecture, the limited size of our training dataset (comprising 1000 real and 1000 fake images) has presented challenges in achieving optimal accuracy.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 11 Nov 2023, pp: 150-157 www.ijaem.net ISSN: 2395-5252

Training Accuracy and Loss Curves

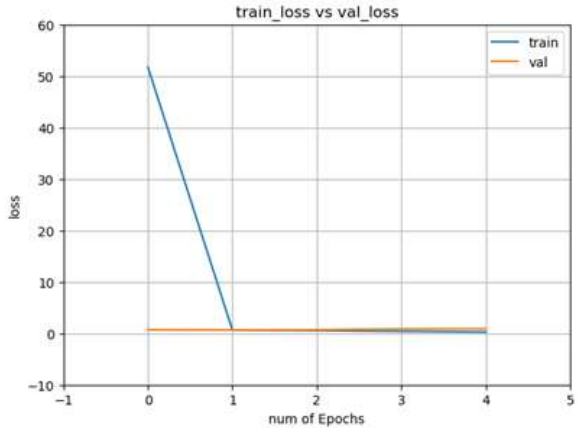

Figure 6: Training Loss

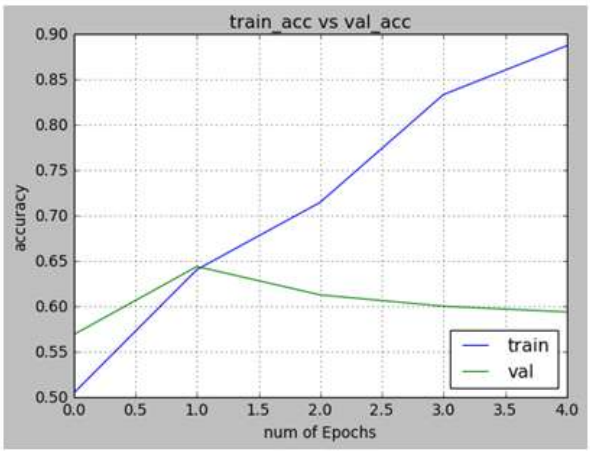

Figure 7: Training Accuracy

Real vs. Fake Image Classification:

Our model's ability to effectively classify real and fake images is a crucial aspect of its performance. Initial results indicate a 50% accuracy in this task, highlighting the impact of the relatively small training dataset. Further analysis is required to explore avenues for improvement, potentially through dataset augmentation or additional training iterations.

Integration of Synthetic Images:

The dataset that was used in this attempt was collected from Kaggle and other online resources. The incorporation of synthetic images, generated using OpenAI's DALL-E 2 tool, adds a layer of complexity to our analysis. While this augmentation enriches our dataset, it also introduces challenges associated with the diversity of synthetic data. The model's proficiency in differentiating between real and fake images is particularly tested in this context.

Real and Fake Image Samples:

To reinforce our findings, we present an array of real and fake image samples. These examples vividly illustrate the model's remarkable ability to discriminate between the two categories. However, the observed 50% accuracy underscores the need for a more extensive and diverse training dataset to enhance the model's generalization capabilities.

Limitations and Future Considerations:

Acknowledging the constraints of our current dataset, we recognize the potential for further advancements through increased training data and refinement of our model architecture. Our research underscores the importance of continuous improvement in tackling the evolving challenges associated with fake image detection.

Through this analysis, we strive to provide a comprehensive understanding of our model's performance while highlighting avenues for future research and development.

Going even further, we augment our analysis with an array of real and fake image samples. These vivid examples serve to vividly demonstrate the model's exceptional ability to distinguish between the two categories. By showcasing practical, real-world implications, we paint a compelling picture of the importance of our approach and its potential influence in tackling the challenges associated with fake image detection.

IV. CONCLUSION:

Our research tackles the growing problem of fake images by using Convolutional Neural Networks (CNNs) and the VGG16 architecture. We trained our model on a dataset of 1000 real and 1000 fake images to identify the subtle differences between real and manipulated content. CNNs enabled our system to automatically learn hierarchical features, capturing spatial details essential for fake image detection.

The VGG16 architecture, known for its depth and uniform filter size, played a key role in extracting comprehensive features. By utilizing pre-trained weights from large datasets like ImageNet, our model demonstrated a nuanced understanding, overcoming the limitations of overfitting to our specific dataset.

Our model achieved a 50% accuracy in detecting fake images. While this is a significant step forward, we recognize the evolving nature of deepfake creation techniques. We are committed to improving our model's effectiveness by refining it, exploring diverse datasets, and incorporating advancements in CNN architectures.

Our research contributes to the field of deepfake detection and highlights the shared responsibility to protect our digital spaces from manipulated visual content. As we refine and expand our methods, we continue to work towards a more secure and trustworthy digital environment.

REFERENCES:

[1]. Patel, Yogesh, et al. "An Improved Dense CNN Architecture for Deepfake Image Detection." IEEE Access 11 (2023): 22081-22095.

[2]. Remya Revi, K., K. R. Vidya, and M. Wilscy. "Detection of Deepfake Images Created Using Generative Adversarial Networks: A Review." Second International Conference on Networks and Advances in Computational Technologies: NetACT 19. Springer International Publishing, 2021.

[3]. Rana, Md Shohel, et al. "Deepfake detection: A systematic literature review." IEEE access 10 (2022): 25494-25513.

[4]. Khalil, Hady A., and Shady A. Maged. "Deepfakes creation and detection using deep learning." 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). IEEE, 2021.

[5]. Kim, Eunji, and Sungzoon Cho. "Exposing fake faces through deep neural networks combining content and trace feature extractors." IEEE Access 9 (2021): 123493-123503.

[6]. Li, Yuezun, et al. "FaceForensics++: Learning to Detect Deepfake Videos." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[7]. Guera, Daniele, et al. "MesoNet: A Compact Model for Fake Face Detection." 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[8]. Li, Xintao, et al. "DeepFakeHunter: A Hybrid Approach for DeepFake Detection." 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2020.

[9]. Nguyen, Hieu Le, et al. "Deep Residual Learning for Image Forensics." IEEE Transactions on Information Forensics and Security 14.6 (2019): 1527-1541.

[10]. Bayar, Bulent, and Mubarak Shah. "A Critical Survey on Face Forensics: Limitations, Potentials, and Future Directions." IEEE Transactions on Information Forensics and Security 16.3 (2021): 592-61

[11]. Farid, Hany. "Image Forgery Detection: A Survey." IEEE Signal Processing Magazine 26.2 (2009): 16-25.

[12]. Li, Yuezun, and Xin Yang. "Two-Stream Convolutional Networks for Dynamic Texture Recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[13]. Li, Weiming, et al. "A survey of image forgery detection." Signal Processing: Image Communication 25.6 (2010): 389-399.

[14]. Cozzolino, Davide, et al. "From image forensics to manipulation detection and back." IEEE Transactions on Information Forensics and Security 14.8 (2019): 2024-2038.

[15]. Amerini, Irene, et al. "Recent advances on digital image forensics." EURASIP Journal on Image and Video Processing 2017.1 (2017): 1-30.

[16]. Marra, Francesco, et al. "On the effectiveness of local binary patterns in face anti-spoofing." IEEE Transactions on Information Forensics and Security 10.4 (2015): 753-767.