iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Hybrid ResNet-CNN Framework for Real-Time Deepfake Image and Video Detection

Amuthavalli G, Dr. I. Shahanaz Begum, Poonggazhal TTK
Dept of Computer Science and Engineering MIET Engineering college Tiruchirappalli , India

Abstract - Due to the rise of online digital communication channels, the exchange of image and video information on the web is at an all-time high, leading to issues related to authentication. With the recent improvements in AI technologies, the concept of deepfakes has emerged, allowing to produce very realistic altered images and videos that can mimic actual human beings and events. Such forged data can be utilized in a number of ways, including but not limited to misinformation, identity theft, cyber crimes, political manipulation, and other malicious activities against individual people and society. As such, there has been a need to develop solutions that would allow detecting such alterations and identifying deepfakes. This study aims to offer an advanced deep-learning based solution for detecting deepfake images and videos. The CNN layer performs the task of extracting both low-level and high-level spatial features, whereas the ResNet structure solves the issue of vanishing gradients and enables deep learning by the model. Compared to the standard way of learning, the proposed method of developing a model gives rise to fewer complications and resolves itself much faster than traditional means. Results from numerous studies conducted on many popular datasets support conclusions that the proposed system outperforms conventional methods in both accuracy and false positive rate for identifying deepfake videos as well as requiring very little time to provide an output once a video has been provided as input. The proposed model also performs well across the range of scenarios based on the variation of how inputs are altered (e.g., the level of resolution, the amount of compression used, the type of alterations made to the input). The proposed model is perfectly suited for integration into many types of applications such as social media monitoring, law enforcement, digital verification and cyber security, since the proposed model can detect deepfake videos in real-time.
KEYWORDS: Authenticity, Cybersecurity, Deepfake, Detection, Forensics, ResNet, Videos

I. INTRODUCTION

The advancement of digital communication technology has provided means to create and share large volumes of visual content across a number of different online platforms. While different advances in digital communication have improved ease-of-use/connection, the availability of digitally manipulated (forged) images and videos created from advanced artificial intelligence algorithm (known as deepfakes) raises concerns over the authenticity of digital content. Such manipulation can be used to deceive individuals, as well as for criminal activities like identity fraud and reputation damage, among other illicit uses, necessitating an automated system for detecting and monitoring for fraudulent media to enhance the safety of digital content. However, current forgery detection methods use either manual analysis or low-level feature extraction techniques to detect forgery and therefore do not provide any protection from sophisticated manipulation methods, such as those employed through new technologies like Generative Adversarial Networks (GANs), which represent a significant advancement in the creation of forged media and have therefore far surpassed the ability of existing methods to address.. In addition, there are many issues related to scalability, performance, and flexibility within current forged detection models. These models have difficulty scaling with respect to both size of the database they are being run on (i.e., increasing size from 1,000 entries in a SQL database to potentially millions) and adapting for the type of manipulation performed.Within this structure, new intelligent frameworks for providing offender detection will be created. Therefore, the study developed plans for a new innovative technology (like how google maps illustrates how bad or good traffic is) for providing an intelligent open-ended flexible scalable reliable forgery detection method by incorporating dug-up data from all available research papers and from social media digital records using the combination of ResNet architecture and CNN architecture for deep learning by providing functional and high-level visuals that represent the changes made to original images along with manipulations made to create original images as well as being able to be used for real time detection purposes by adding the option for supporting detection of multiple formats.The use of these advanced technical applications will allow for the establishment of a calibrated and credible/accurate method for deepfake detection.

A. Problem statement

Due to rapid advancements in both digital media technologies and AI, it has become increasingly easy to create and spread manipulated images or videos, commonly referred to as deepfakes. Because the resulting counterfeit material can be exceptionally real looking to human viewers, many potential problems arise, including misinformation, identity theft, damage to one's reputation, cybercrime, etc. Historically speaking, the ways to detect such counterfeit photographs were primarily through manual inspections and basic image processing methods, but these methods do not function well against false images created using modern AI technologies. Automated methods for detecting deepfakes suffer from similar issues as well: extremely high false positive rates; inefficient operation; high costs to use; inability to simultaneously analyze multiple types of deepfake media; etc. Furthermore, existing systems often exhibit low levels of performance in real-time settings thus rendering them ineffective for use cases such as social media surveillance and cyber security. Consequently, it is essential that an intelligent solution which automatically detects counterfeit images and videos is developed; this solution must be efficient, secure, and highly reliable.

B. Dataset details

This dataset used for deepfake detection using images and video is comprised of both real photos and videos (from reputable networks) as well as photos and videos that have been manipulated by deepfake technology using a variety of different methods. The images are comprised of many different people and scenes, filmed with different lighting conditions, facial expressions, camera angles, backdrops, and image resolutions to create diversity while training the model. On the manipulation side of the data, all images and videos were made using advanced technology including Face-Swap, Face-Reenactment, Lipsyncing/Face Synthesis (GANs). While videos have been cut into single frames to permit efficient extraction of spatial features from each frame of video, all data was pre-processed using methods such as image resizing, pixel value normalization, and rejection of samples that were poorly constructed. Other techniques to augment the data included rotating, flipping, brightening, cropping, etc., will improve the generalizability of the model and reduce the risk of overfitting. The complete dataset is then split into three different datasets: training dataset, validation dataset, and test dataset. All three datasets will measure the accuracy of the system when detecting deepfake videos.

C. Objectives

The primary aim of the research is to create a deep learning-based system to detect manipulated images (deepfakes) which is reliable and intelligent in operation. The system will detect forged digital media by analysing visual inconsistencies as well as the hidden manipulation techniques that humans cannot detect with the naked eye. Other aims of the research include The use of a hybrid CNN/ResNet model. The project will compare the differences between authentic and forged media based on facial expression, light conditions, texture distortion, and unnatural integration. The focus of this process will be to ensure accurate detection with a very low computational cost and fast processing speed. The ability of the system to support various media formats will provide a comprehensive and extended application of the project for both still images and video media. The additional objective of the project is to provide the model with greater generalisation capabilities to adapt to differing data sets, resolution levels and different types of manipulation. Additionally, the project aims to reduce to a minimum the total number of false positive/negative detection results so that accurate results may be provided. One of the goals is to advance cybersecurity and digital forensics by developing automated media authentication solutions. Additionally, it is important to assist social media companies in detecting and removing malicious fake information on their platforms. Moreover, it is essential to develop a scalable system that can be adapted to future techniques used to generate deepfakes. It is also necessary to minimize reliance on human interventions in verification processes using intelligent automation systems. Furthermore, it is crucial to gain users' trust in online communications by ensuring the authenticity of the information.

II. RELATED WORK

Rafique, Rimsha, and colleagues [1] have performed an analysis of how digital image manipulation, particularly deepfake video technology, can be detected and classified using both Error-Level Analysis (ELA) and machine learning. Deepfake video manipulation has recently become common on social media and other online platforms. By using ELA, a digital forensics expert can identify areas of an image where the quality has been altered by comparing them with areas that have not been altered. Machine learning models can then be trained to identify whether the content of an image is "real" or "fake." The goal behind using these two techniques together is to increase efficiency through the combination of image forensics (via ELA) and automation via machine learning. ELA can identify areas of digital images that have been manipulated but that are impossible for the human eye to see.

The use of machine learning combined with ELA will increase the accuracy of classifying manipulated images as either genuine or fake. According to the authors, combining conventional forensics with artificial intelligence would be very effective at identifying deepfake videos. Therefore, the research presented in this study will add considerable utility to modern digital forensics systems for identifying forged digital images. The authors believe that the method outlined in this study will also be very useful for identifying forged digital images in future research.

Heidari, Arash, et al. [2] composed a thorough and organized review of several different types of approaches for deepfake detection with an emphasis on using deep learning based algorithms. The authors describe technological advancements that have occurred rapidly over the past few years and some of the negative implications that could potentially be a result. There are numerous approaches for detecting deepfakes, some of which utilize different types of machine learning and deep learning techniques. Many model architectures also exist including CNNs, RNNs, LSTMs, Vision Transformers, etc. The authors compare comparative benchmark datasets that are typically used for both training and evaluating models to determine which are generally better suited for performing deepfake detection (e.g., FFD, Celeb-DF, and many others). Additionally, the authors review commonly used performance indicators including accuracy, precision, recall, and various others that may help achieve more robust results when performing deepfake detections. The authors discuss several challenges associated with detection, including the problems associated with adversarial attacks, the inherent nature of models not being able to generalize well, and the inability of many models to be implemented in real-time. Lastly, the authors mention some of the most pressing concerns associated with multi-modal deepfake detection techniques, which not only include video/image deepfake detection systems, but also audio deepfake detection systems as well as the potential usage of explainable AI within deepfake detection technologies. Finally, the authors conclude the paper with suggestions for areas needing further investigation. Overall, this review serves as an excellent resource for individuals just beginning to research deepfake detection technologies, providing them with a brief overview of the state-of-the-art techniques available today.

Wang, Tianyi, et al. [3] This study provides an overview of the state-of-the-art deepfake detection with particular attention to their reliability. The authors of this paper argue that high precision of a system is not sufficient in its reliable operation in practice without reliability, trustworthiness and robustness, which are other variables that must be presented to accomplish the reliability and trustworthiness of the system in practice. The authors discuss several different types of attacks against deepfake videos, including spatial, temporal, physiological, and frequency domain attacks. The authors also examine the robustness and reliability of different deepfake models in the face of previously unseen attacks and low-quality videos. Additionally, the authors discuss various aspects of deepfake detection such as confidence estimation and calibration, and the ability to provide a justification for deepfake detection using model interpretability. There is a classification of existing deepfake detection techniques using various dimensions of reliability. The authors analyzed the datasets and benchmarking methods for deepfake detection. The authors also discussed some ethical considerations related to deepfake detection, most notably fairness issues, and stated that there is a lack of standardization in the benchmarking processes. The authors indicated that the overall reliability of deepfakes can also be improved by the availability of a variety of features. Finally, they asserted that deepfakes should be used in a trusted manner in a judicial and security environment.

Suratkar, Shraddha, and Faruk Kazi [4] In this paper, a novel approach is proposed for detecting deepfake videos by implementing a transfer learning technique. Importance of creating a simple learning process is emphasized by the researchers without compromising accuracy. The use of pre-trained models for deep learning can be to extract features of deepfakes in videos. In addition, using transfer learning can provide a means for the machine to learn from large amounts of previously trained data sets. Therefore, large volumes of deepfake data is not required to train the model; rather, each frame will be studied separately for signs of manipulation. There are many characteristics of deepfake videos that will be examined for this purpose, including the presence of mismatched facial features and/or unnatural facial expressions. One particular benefit of this approach is its speed of convergence. The experimental results demonstrate that the approach generates satisfactory results for different benchmark data sets. The authors of this paper also operate with an understanding of the efficiency of transfer learning for real-time deepfake detection, as well as with an understanding of how well this process will work in resource-constrained environments. Compressed videos and lower resolutions are some examples of the issues cited by the authors.

Patel, Yogesh., et al. [5] This paper presents a detailed discussion on the creation and detection of deepfakes using a case study and various challengesThe authors provide insight into how generative models (e.g., GANs and Autoencoders) create deepfakes by demonstrating multiple examples of actual abuse (e.g., identity fraud, political propaganda). In addition, the authors discuss different methods of detecting deepfakes, both by using machine learning and by utilizing forensic evidence. They compare detection methods based on images, videos, or audio for advantages and disadvantages.

A major finding of the paper is that there are many challenges to detecting deepfakes due to advances in generating them. The paper identifies additional challenges related to ethical/legal/social issues, such as data sets, adversarial attacks, and privacy. The authors encourage collaborative efforts among researchers, governmental agencies, and industries to generate increasingly sophisticated deepfake technology. The authors place emphasis on raising public consciousness about the possible risks of misleading digital media. Finally, the authors also provide recommendations for developing an ethical framework for using artificial intelligence technologies. Readers of this article will gain an appreciation for the complexities associated with generating deepfakes.

Ustubioglu, Beste., et al [6] The authors of this paper present an independent attack audio deepfake detection system based on cochleagram images. The research emphasizes the detection of audio forgeries instead of visual deepfakes. The biological architecture of cochleagrams, which are images representing sound frequencies, saw an increase in their utilization for audio segment analysis and being classified as such. The development of dynamic thresholding provided an avenue for the detection of the suspicious changes to speech patterning through their use within the classification process. Because of the 'attack-independence' characteristic associated with this approach, it is therefore not reliant upon any specific attack algorithm.This makes the model more adaptable and of greater utility concerning the various methods of manipulating audio. The capacity of machine learning models to accurately classify both legitimate and fraudulent audio segments has been established through experimentation. The demonstrated efficacy of the proposed method has been reported.The relevance of the topic along with its threats to society in terms of audio deepfakes, particularly with voice cloning and synthetic-speech platforms, is expounded upon within the context of research being presented. Furthermore, the research has implications for multimodal forgery detection and opens future possibilities in the areas of image and video forensics.

Wang, Zhi, Yiwen Guo, and WangmengZuo et al [7] In the paper, deepfake detection techniques using an adversarial game have been presented. The detection method is built on an adversarial learning process that operates in opposition to the forger's method. This model contains a framework from which the detector learns from the forger's most difficult and challenging images. Through learning from a forger's most difficult image, the detector will become more robust against a wider range of advanced deepfakes. Additionally, the framework emphasizes learning of discriminative forensic features from forged faces. This will increase the detector's ability to generalize when applied to new forgery algorithms. Furthermore, this will also help resolve the overfitting issues that have been associated with model-based forgery detection. Results from the experimental work demonstrated that the proposed method is superior to traditional methods.

Rana, Md Shohel, et al., [8] In this paper, we have performed a literature review on deepfake detection. We examine a wide range of articles related to the topic under consideration. The deepfake detection approaches are divided into image-based detection, video-based detection, audio-based detection, and multimodal detection. Several popular deep learning methods, including CNN, RNN, and GAN based detectors are discussed in the article. Furthermore, several important datasets that have been used in this area are mentioned. Various challenges associated with deepfake detection, such as data imbalance, lack of generalization, and adversarial attacks, are highlighted. A comparison is made between metrics used in earlier literature reviews. The rapid evolution of techniques for generating deepfakes is also emphasized. Authors stress the importance of building efficient and reliable detectors. Additionally, ethical implications of deepfakes are explored. Overall, the paper can be considered a useful guide for future researchers.

Huang, Yihao, et al., [9], The present paper describes the development of Fakelocator, which is a robust method for detecting localization in GAN-generated face manipulations. While existing solutions perform fake or non-fake media classification, the suggested approach focuses on identifying manipulated regions specifically. The neural network used by the framework is able to highlight suspicious faces' parts like eyes, mouths, and skin contours. Such ability allows for a better understanding of the process of the media manipulation for forensic purposes. Furthermore, Fakelocator has been designed with consideration of multiple GAN-generated face manipulation types. The proposed technique is based on deep neural networks and allows for pixel-level predictions. Experimental results have shown an accurate detection of manipulation and proper localization of the region within the image. Moreover, the proposed method works perfectly with compressed and resized images. In addition to accurate detection, the paper stresses importance of the explainability factor in detecting deepfakes. The localization of manipulated regions makes users believe in the system's performance and findings.

Nirkin, Yuval, Yosi Keller, and Tal Hassner et al., [10] The current paper introduces a new version of the FSGAN, which is a subject-agnostic face swap and reenactment framework that can produce very realistic face swapping and reenactment results without relying on a particular subject's identity. The framework incorporates some advanced GAN architectures for generating outputs and supports face reenactments, where the expressions of one person can be transferred to another face. The framework generates very realistic outputs that are hard for detectors to detect. This paper is important because it demonstrates advanced forgery generation capabilities.

Studying such forgery generation frameworks can help researchers develop more accurate detection systems. The authors assess the effectiveness of the framework based on realism and identity metrics. The framework achieves impressive results under various datasets and subjects. The paper reveals how rapidly the synthetic media generation field evolves. In addition, the paper brings to light ethical issues associated with the misuse of such frameworks. The output faces produced by the FSGANv2 framework can be used to train detectors.

III. PROBLEM STATEMENT

Due to the rapid advancement of AI, as well as new and advanced technologies in the digital media industry today, producing realistic media content has never been easier than it is now with deepfake technology. As a result of the ease with which realistic media content is created, the ability to disseminate it across multiple platforms has significantly increased within our social networks or news websites, or through any other type of digital communication. It is becoming increasingly difficult for individuals to determine what is real and what is fake when searching for information on the Internet. The content can be used for misleading individuals to spread false information; mislead individuals to commit identity theft; mislead individuals for political purposes; commit financial crimes; and commit cybercrimes. The traditional methods for verifying the authenticity of media content relied on either a manual review/a manual analysis of each item or through a basic, cursory forensic analysis, which are generally slow, unreliable, and inefficient when compared to AI-based manipulative techniques. Automated tools that are available for detecting deepfake technology experience certain issues when it comes to their performance, scalability, high computing power utilization, and their ability to generalize across different datasets. Most, if not all, of the current technologies available have demonstrated a lack of real-time detection technology associated with practical applications such as monitoring social media channels, performing digital forensic investigations, or management of cybersecurity systems to enhance security in these systems. Furthermore, advancements in image and video production through technologies like GANs and face swapping have made deepfakes hard to detect. For this reason, there is a significant demand for a reliable framework that utilizes deep learning to identify manipulated content and maintain authenticity.

IV. PROPOSED METHODOLOGIES

The proposed system provides a state-of-the-art robust deep learning framework that can efficiently detect deepfake images and videos. The architecture of the system combines CNN (a hybrid model of ResNet) to extract features and classify inputs; hence, it can effectively identify inputs based on classification by using features extracted from the input data. Depending on whether an image or video was input, existing methods for detecting deepfake content typically process images and videos as separate entities. The proposed method, however, allows processing both types of input (images and videos) as one combined system. For example, the proposed model can recognize manipulation signs, such as facial distortion, unnatural facial expressions, poor lighting, unnatural eye movement, blurred edges or blending artifacts. To ensure optimal system performance (speed) characteristics, the proposed model has an inverted residual block and uses linear bottlenecks. Additionally, this lightweight design will allow the model to be scalable throughout the deployment process. The proposed model will be trained with an extensive data set consisting of both genuine and modified (deepfake) inputs using a variety of manipulation techniques, resolutions, and compression formats. During the training process, the CNN layers identify both low and high level image pattern information, whereas the ResNet layers use residual connections for performing deep learning on data received from its CNN layers. The system detects minute clues of tampering that humans cannot identify manually. The generalization capability of the model further enhances to ensure accurate detection of novel deepfake attacks. In addition, the shortened inference process time of the proposed model is ideal for applications in real-time environments, such as social media sites, forensics investigation, and cybersecurity systems. The system reduces false predictions using optimized classification and confidence scoring approaches. It can accommodate high volumes of digital data uploaded in real-time scenarios because of its scalability. The suggested framework compensates for the limitations of the existing methodologies in terms of multi-formatted support, accuracy, and speed.

V. METHODOLOGY

1) IMAGE AND VIDEO UPLOAD AND TRAINING

This module acts as the interface through which images and videos can be uploaded for verification purposes. After the data is uploaded, it undergoes a preliminary preprocessing stage to make sure it meets all the criteria that the neural network needs. In this process, videos need to undergo processes of frame rate standardization, resolution adjustments, and frame extraction, whereas image data will need only resizing and normalization.

This module handles the training data set as well, organizing both the real and forged data for use as training data for the deep learning model. This requires uploading various labeled instances of both types of media to help the system learn the difference between them.

2) INPUT PROCESSING

Once uploaded, the input processing stage is responsible for making both video and image inputs ready for feature extraction. In the event of videos, the system divides the video into frames or clips, based on the level of granularity needed for analysis. Images are directly fed in after being processed. Normalization is carried out by balancing the pixel intensities and scaling the images to the size demanded by the CNN architecture. In cases where there are multiple scenes within the video, scene detection can be carried out to extract relevant segments.

3) FEATURE EXTRACTION

The feature extraction module is the backbone of the system that uses a Convolutional Neural Network based on ResNet architecture to extract deep features from images as well as frames of videos. The hierarchical architecture of ResNet helps to identify complex spatial patterns in images as well as temporal inconsistencies in frames of videos. These features include artifacts like lighting mismatch, unusual face texture, image blending problems, and inconsistencies in the frames indicating forgery. Other modules used between architectures such as inverted residual blocks and linear bottlenecks aid in retaining spatial features while minimizing dimensions and complexity.

4) FORGERY DETECTION

In the forgery classification process, the features obtained from the above extraction process are analyzed for determining whether the image or video is real or fake using the forgery detection module. A classifier that is based on a neural network is then applied to classify the data by detecting irregularities that are characteristic of manipulations of images and Deepfake videos. The classifier performs its operations by checking for spatial inconsistencies for images and both spatial and temporal inconsistencies for videos. Techniques for ensuring generalization and preventing overfitting have been incorporated in the classifier.

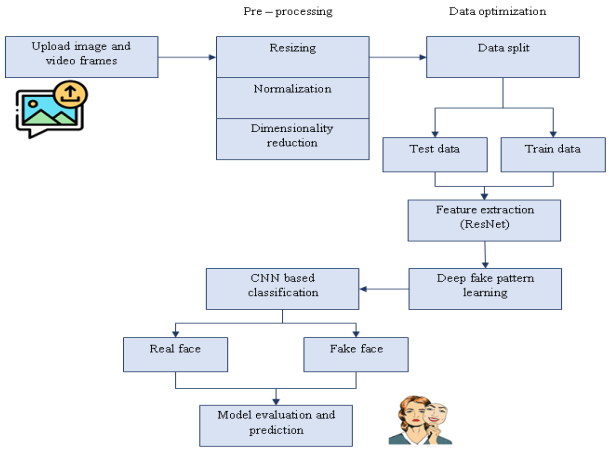


Figure 1: Proposed methodology

VI. IMPLEMENTATION DETAILS

The process of implementing the proposed system for detecting deepfake images can be done through employing the strategy of deploying the deep learning approach which includes the software packages, pre-processing, training model, and deploying model phases. In the process of deployment of the deep learning model, Python programming language would be used along with software libraries including TensorFlow, Keras, OpenCV, NumPy, Pandas, and Scikit-learn, for manipulation, pre-processing, and modeling tasks. The process begins with the collection of datasets consisting of the original images and manipulated ones from credible sources, and then arranging them in respect to their use for either training, validation, or testing methodológía. The videos will be preprocessed using OpenCV software wherein frames will be extracted from the sequence at equal interval; each frame is treated as an image instance that needs to be classified.

The preprocessing process involves resizing the images and frames, normalization, and cleansing of the corrupt instances to values of zero to one.The hybrid architecture is implemented using CNN layers to extract spatial feature information and ResNet network architecture as the framework for gradient transfer. Activation function like Rectified Linear Unit (ReLU) is used in hidden layers whereas the output layer makes use of Sigmoid/Softmax function for predicting whether the image/video is real or fake. Hybrid architecture implementation is done using the Adam optimizer where binary cross-entropy function is used to evaluate cost functions using epoch and mini-batch. Performance of the model can be assessed using accuracy and loss graph for validation as well as adopting early stopping algorithm to prevent overfitting. The evaluation metric for model performance includes Accuracy, Precision, Recall, F1-Score and Confusion Matrix after training the model. In terms of implementation, the trained model is implemented in the user interface to allow users to upload their images and videos for prediction of whether it is real or fake. When videos are uploaded, they are automatically transformed into frames and analyzed independently to determine their authenticity.

VII. EXPERIMENTAL RESULTS

The experimental assessment conducted for the proposed methodology for detecting deepfakes shows very efficient results in recognizing manipulated images and videos. When assessing the hybrid model built based on CNN with ResNet architecture, a set of training, validating, and testing datasets made up of actual and manipulated images were applied. It was found out that during the training phase, there was a consistent convergence of the model with decreased loss values and increased validation accuracy at different iterations. On the other hand, in terms of efficiency, the model demonstrated high accuracy coupled with excellent precision, recall, and F1-score. These results show that the model can significantly minimize the cases of both false positives and false negatives.This model has achieved success in recognizing inconsistencies, distortions in textures, improper blending of edges, and inappropriate lighting while performing tests using images. For video-based test samples, this model can identify whether the video is deepfake or not by analyzing different frames from the video and studying the temporal inconsistency in the videos. The use of ResNet architecture has facilitated faster convergence. Furthermore, the proposed model has also been able to perform exceptionally well in generalization for different test samples. The inference time has also been less for the proposed model, thus making it capable of performing real-time predictions for manipulated images and videos uploaded via uploading facility. The comparison between baseline models has proven the superiority of the proposed model over other machine learning models.

Performance Metric	Existing System (%)	Proposed System (%)
Accuracy	88.40	96.85
Precision	86.90	95.72
Recall	85.75	96.10
F1-Score	86.30	95.91
Specificity	87.20	96.45
Detection Rate	84.95	97.08
AUC Score	89.10	97.32
Inference Speed	81.50	93.65

Table 1: Performance Comparison Table

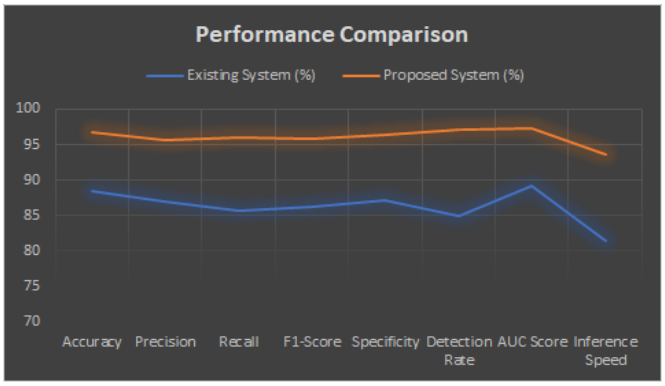


Figure 2: Performance metric chart representation

The comparative analysis of the system performance shows that the proposed model prevails among criteria used for detecting deepfakes. Even though the existing system provided decent results based on poor feature extraction ability, relatively low generalization ability, and significant vulnerability to current manipulation techniques, the hybrid CNN model equipped with ResNet proved superior in accuracy compared to its competitor. In particular, the model had excellent performance related to both the classification of original and manipulated images with improved accuracy. The improvement in precision means that the suggested system generated fewer false positives and provided accurate results. Increased recall also means that the model effectively classified original deepfakes and did not ignore any cases of manipulation. Thus, an increase in the F1 score can also be observed, indicating improved precision and recall rates simultaneously. Finally, higher specificity scores mean that the new model was more successful in identifying actual examples of media. In fact, the detection rate was greatly enhanced since the new model could distinguish small elements like blending errors, facial distortion, and lighting problems.The enhanced value of the area under curve indicates that the classifier is far more competent in distinguishing between authentic and generated faces. Apart from increasing accuracy, the speed was significantly higher because of the optimized architecture and residual learning, which make the system perfect for real-time applications. The use of preprocessing techniques, data augmentation, and residual connections has made the model more robust for future use on different image datasets. Overall, this comparison demonstrates that the proposed method outperforms the conventional technique.

VIII. OUTPUT RESULT

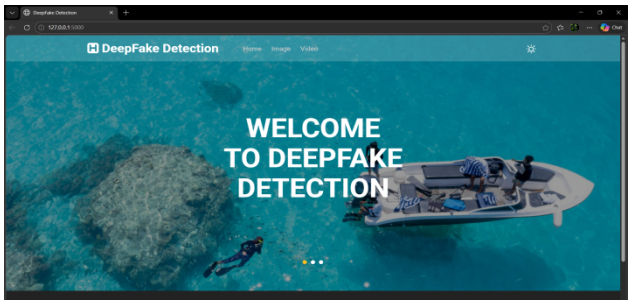


Figure 3: Home page

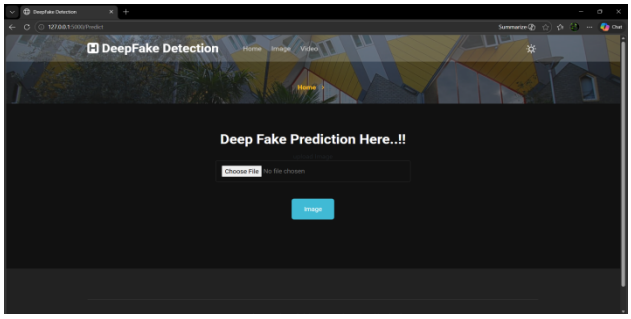


Figure 4: File upload & prediction Module

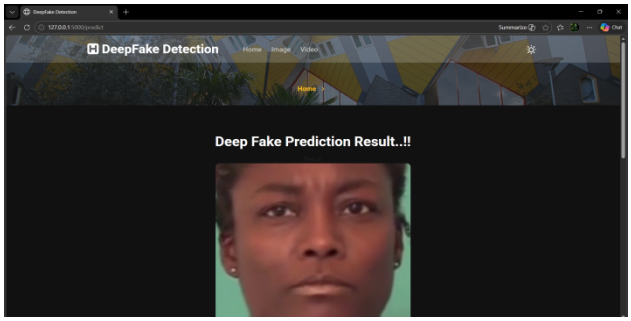


Figure 5:DeepFake Prediction Result for Single Image Input

Figure 6:Prediction Confidence Score Visualization - 90.15% Fake

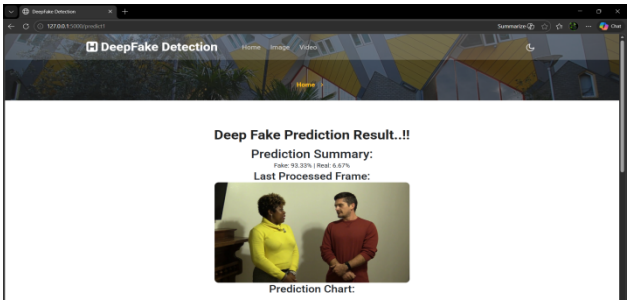


Figure 7:DeepFake Detection Result for Video Input with Frame Analysis

IX. CONCLUSION

From the results(fig 3 to 7) obtained through the research, it is clear that the proposed framework of deep learning is a plausible approach in detecting deepfake images and videos in the present-day digital world. Through the integration of Convolutional Neural Networks and ResNet architectures, the system has enabled the detection and learning of various attributes related to images and videos by identifying any alterations made to their characteristics like inconsistencies in facial expressions, unnatural fusions, and distorted lighting and textures. This has led to the creation of an accurate, efficient, and fast classifier of genuine and fake media content. Furthermore, the application of residual connections has significantly helped improve the training speed and reduce computational costs compared to other traditional techniques. In contrast to the existing systems for authentication, the proposed system will resolve all of the limitations posed by those methods in terms of scalability, robustness, and efficiency. Moreover, the proposed system will help combat various problems associated with synthetic media due to its capability of automating the authentication process. The findings from this study help establish trustworthiness within digital communication systems. In general, this project demonstrates the significance of using deep learning models to counter forgery attacks on images.

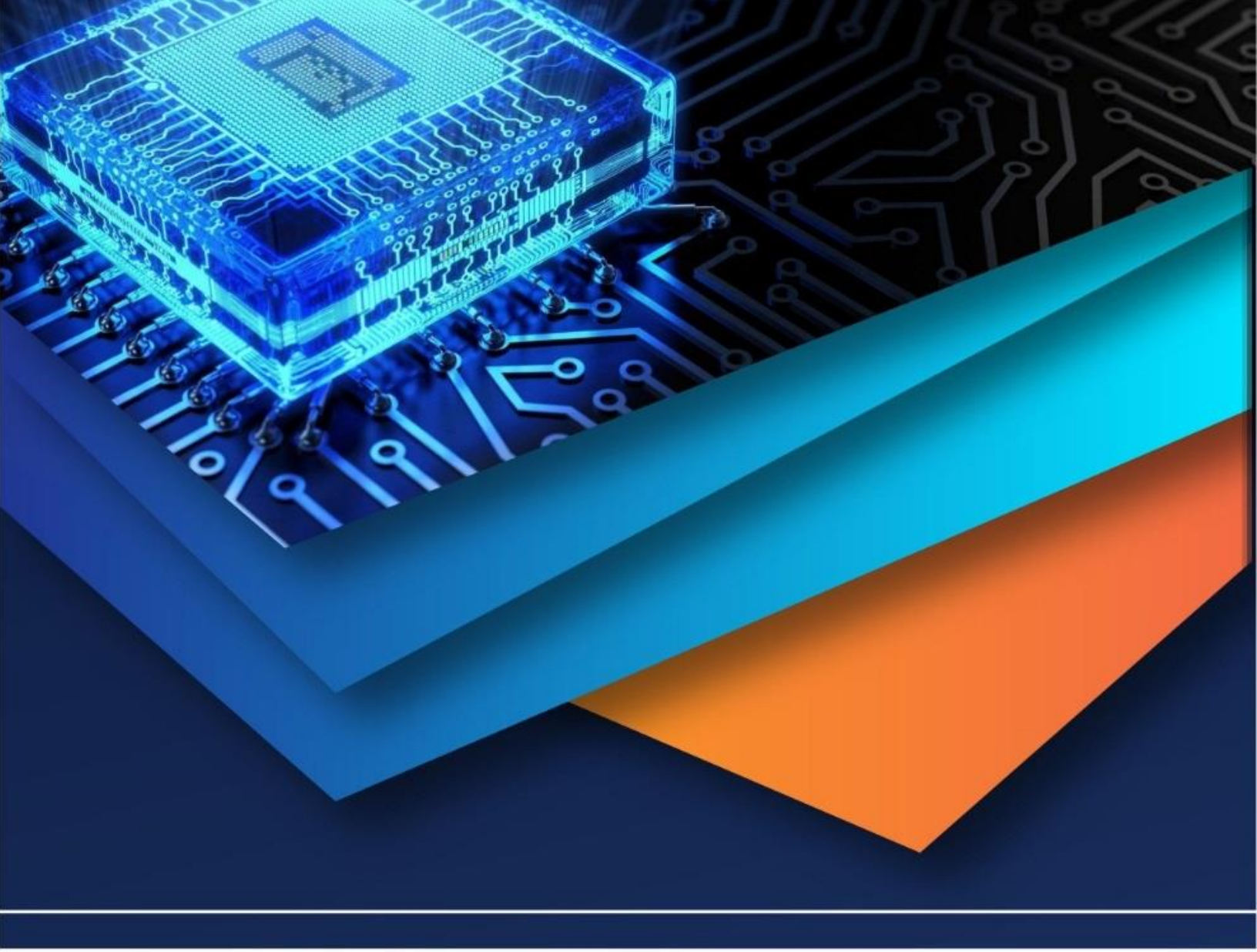
X. FUTURE WORK

Research that can be carried out on this subject in future could include methods to improve the performance and capabilities of deepfake detector technology for keeping up with new technologies in the forgery process. The latest technologies in deep learning models, such as using transformer architecture and attention mechanisms, can be utilized in addition to the existing convolutional neural network (CNN) and ResNet models in order to detect forgery. Another topic that can be investigated in future in relation to forgery is multimodal forgery. It involves the detection of forgery by analyzing different attributes such as vision, language, and audio data. Training of the algorithm with more data may increase its performance and address any biases in the results of forgery detection. The detection of advanced forgery techniques such as diffusion forgery and forged web content can also be included in the future development of this project. In addition, methods that explain how algorithms make decisions and indicate suspicious areas may be used. Finally, in upcoming versions of the project, methods that facilitate continuous learning of the model can be applied.The software can also be embedded in the websites for social networking, monitoring, and cybersecurity applications. All these aspects will ensure that the software becomes more intelligent, scalable, and faster to adapt to future threats.

REFERENCES

[1] Rafique, Rimsha, et al. "Deep fake detection and classification using error-level analysis and deep learning." Scientific reports 13.1 (2023): 7422.
[2] Heidari, Arash, et al. "Deepfake detection using deep learning methods: A systematic and comprehensive review." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 14.2 (2024): e1520.

[3] Wang, Tianyi, et al. "Deepfake detection: A comprehensive survey from the reliability perspective." ACM Computing Surveys 57.3 (2024): 1-35.
[4] Suratkar, Shraddha, and Faruk Kazi. "Deep fake video detection using transfer learning approach." Arabian Journal for Science and Engineering 48.8 (2023): 9727-9737.
[5] Patel, Yogesh, et al. "Deepfake generation and detection: Case study and challenges." IEEE Access 11 (2023): 143296-143323.
[6] Ustubioglu, Beste. "An attack-independent audio forgery detection technique based on cochleagram images of segments with dynamic threshold." IEEE Access 12 (2024): 82660-82675.
[7] Wang, Zhi, Yiwen Guo, and WangmengZuo. "Deepfake forensics via an adversarial game." IEEE Transactions on Image Processing 31 (2022): 3541-3552.
[8] Rana, Md Shohel, Mohammad Nur Nobi, Beddhu Murali, and Andrew H. Sung. "Deepfake detection: A systematic literature review." IEEE Access (2022).
[9] Huang, Yihao, Felix Juefei-Xu, Qing Guo, Yang Liu, and Geguang Pu. "Fakelocator: Robust localization of GAN-based face manipulations." IEEE Transactions on Information Forensics and Security 17 (2022): 2657-2672.
[10] Nirkin, Yuval, Yosi Keller, and Tal Hassner. "FSGANv2: Improved subject agnostic face swapping and reenactment." IEEE Transactions on Pattern Analysis and Machine Intelligence 45, no. 1 (2022): 560-575.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY