National Research Journal of Information Technology & Information Science
Volume No: 12, Issue No: 2, Year: 2025 (July- December)
PP: 544-550
ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

A DEEP LEARNING FRAMEWORK FOR REAL AND FAKE IMAGE DETECTION USING CNN AND LSTM NETWORKS

Neha Saini

Department of Computer Application, Shanti Devi Arya Mahila College Dinanagar

ABSTRACT

The creation of digital media has been transformed by the quick development of artificial intelligence (AI) and deep learning, which has made it possible to produce extremely lifelike synthetic images and movies known as "deepfakes." Deepfake technology poses serious problems with disinformation, identity theft, cybercrime, political manipulation, and digital fraud, despite its many positive uses in entertainment, education, healthcare, and the production of digital material. Consequently, the creation of precise and dependable deepfake detection systems has grown in importance as a field of study in multimedia forensics and computer vision.This research proposes a deep learning-based system for recognizing real and false photos by integrating effective image preprocessing approaches with convolutional neural networks (CNNs) for spatial feature extraction. The suggested method uses Long Short-Term Memory (LSTM) networks to capture temporal relationships between successive video frames for video-based deepfake identification. Data gathering from publicly accessible deepfake repositories, preprocessing with Error Level Analysis (ELA), image normalization, data augmentation, feature extraction, model training, and performance assessment are all part of the methodology. The robustness of the suggested model is assessed using a number of benchmark datasets, such as FaceForensics++, Celeb-DF, DFDC, CIFAKE, and CASIA.Standard assessment criteria like accuracy, precision, recall, F1-score, and confusion matrix analysis are used to evaluate the performance of the suggested system. The findings of the experiments show that deep learning models can detect tiny visual abnormalities and temporal inconsistencies introduced during the creation of deepfakes with excellent detection accuracy across various modification strategies. The report highlights future research paths like explainable AI, lightweight detection models, multimodal learning, and real-time deployment while also discussing present problems like adversarial attacks, limited dataset diversity, and the quick growth of generative models. Overall, by offering an efficient deepfake detection framework, this study helps to enhance the dependability, security, and authenticity of digital material.

Keywords: Deepfake, Image Forgery Detection, Artificial Intelligence, Deep Learning, CNN, GAN, Deep Learning Techniques, LSTM

1. INTRODUCTION

With the development of artificial intelligence, digital image editing has become more complex. Realistic synthetic images that are frequently indistinguishable from real photographs can be produced using contemporary AI-based image generating algorithms. Deep learning algorithms like Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models are used to create these altered images, which are referred to as "deepfakes." Deepfake technology poses serious concerns even while it has beneficial uses in gaming, education, entertainment, and healthcare. Fake photos can be used for internet fraud, identity theft, criminality, political propaganda, and disinformation. As a result, creating trustworthy techniques for identifying phony images has emerged as a key focus of computer vision research. In addition to reviewing deep learning techniques for

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

distinguishing between real and false photos, this study looks at popular datasets, talks about model architectures, and identifies areas for further research.

The goal of this research is to use deep learning techniques to create reliable deep fake face detection systems for video footage. This research aims to detect and differentiate deepfake films from real ones by taking advantage of the temporal dependence of video sequences. By providing a reliable method for identifying and stopping the propagation of deepfake content, this research seeks to improve the security and legitimacy of digital media platforms. The rapid advancement of sophisticated deep neural networks and the abundance of data have made it more challenging to discern between faked and authentic images and videos.Our work contributes to the growing body of research in deep learning and enhances the understanding of sophisticated media manipulation techniques.

The scope includes all of the necessary elements for developing, testing, and putting into practice a trustworthy deepfake face detection system that makes use of deep learning techniques. The following are the main components of the project scope:

- Compiling an extensive and varied dataset from several sources that includes both real and deepfake videos.
- . To ensure optimal performance during model training and evaluation, the video frames are preprocessed to preserve consistency in dimensions, format, and overall visual quality.
- The goal is to use deep learning to create and construct a sophisticated deepfake detection system. Particular goals consist of:
    - Building an LSTM-Based Detection Model: Constructing a deep learning architecture that can identify deepfake modifications by examining temporal patterns in video sequences.
    - Dataset Preparation: The LSTM model is trained and tested using a sizable dataset that includes both real and altered movies. This helps to guarantee the model's resilience and dependability in a variety of situations.
    - Temporal Analysis of Video Sequences: Sequential video frames are examined using LSTM networks to capture face expressions and movements that are frequently difficult for deepfakes to effectively replicate.
    - Model Design and Implementation: To improve characteristic recognition and diagnosis performance, an LSTM architecture tailored for temporal analysis is designed, and extra neural network layers are included as needed.
    - Training and Optimization: The preprocessed dataset is used to train the LSTM model, and hyperparameters are carefully adjusted to optimize detection accuracy and guarantee optimal model performance.
    - Preventing Overfitting: Throughout the training phase, methods like data augmentation and regularization are used to enhance the model's generalization and prevent overfitting.

EVALUATION AND VALIDATION

A different validation dataset is used to thoroughly assess the accuracy, robustness, and generalization performance of the trained LSTM model. The LSTM-based method is compared to other contemporary deepfake detection methods, highlighting the advantages and disadvantages of each. Key performance indicators like accuracy, precision, recall, and

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

F1-score are used to measure the model's efficacy. To make sure the model works effectively in a variety of scenarios, its efficacy is also examined across a range of deepfake generating techniques and video quality levels. Deepfake detection's ethical ramifications, such as privacy issues and possible effects on free speech, are also taken into account. The goal of this project is to create a reliable, cutting-edge deepfake detection technique based on LSTM networks by carefully considering both technological and ethical issues. These days, deep learning is the norm in machine learning, which has led to innovations like deepfake that jeopardize national security and privacy. Deepfake uses AI algorithms to replace a person's likeness in media. Even though they are frequently employed for political or comedic purposes, they create moral questions about privacy and false information, which calls for legislative action and increased public knowledge.

DETECTING REAL AND FAKE IMAGE

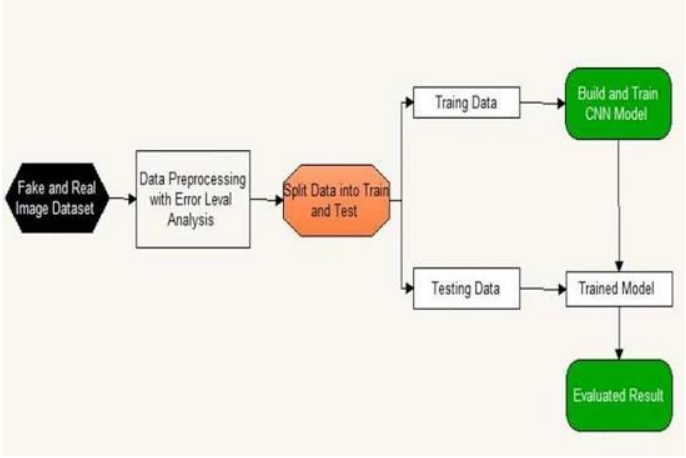


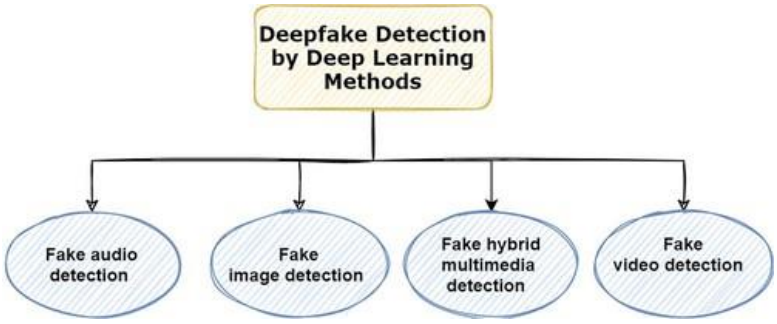
Fig.1:Block diagram for detecting real and fake image

This architecture diagram is based on the suggested system. The dataset should be gathered first, and then it should be preprocessed using Error Level Analysis (ELA). After that, the data should be divided into train and test categories. The Convolutional Neural Network (CNN) will use the training data to train the model. We utilize the trained model to assess the outcome for our test data after it has been trained.

Gathering a diversified dataset comprising samples of both real and fake photos is a critical first step in the process of developing an image classification model to differentiate between real and fake images. After that, this dataset is preprocessed using techniques including scaling, normalization, data augmentation, and label preparation to guarantee consistency and model compatibility. To speed up the learning process, the preprocessed data is then divided into training subsets for actual and fake images, which are then fed into the model. A different, unseen testing dataset is used to thoroughly assess the model's performance after the training phase. After that, the algorithm makes predictions about whether fresh, unseen photos are authentic or fraudulent, and these forecasts are classified appropriately. Lastly, metrics such as overall correctness, precision, recall, and confusion matrices are used to determine the model's accuracy. This allows for a comprehensive evaluation of the model's performance and permits modifications as necessary.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

DEEPFAKE DETECTION USING DEEP LEARNING METHODS



5.1 Fake image detection

As discussed in Section 2, identifying phony faces is the most difficult issue in the field of image forgery detection. False personas can be created on social media platforms by using fake photos, which makes it possible to steal personal information illegally. For example, it is possible to create potentially dangerous photographs of celebrities with objectionable content using the bogus picture generator. We'll look at DL methods for spotting fake images in this part. Deepfake's development significantly lowers the bar for face-manufacturing techniques. In particular, Deepfake uses GANs to substitute a different person's face for the original image's face. Because the GAN models were trained on 10 out of 100 images, they are more likely to produce realistic faces that can be accurately spliced into the original image. Appropriate post-processing can improve the realism of the image. Additionally, face-swap innovation has been applied in a number of situations, such as multimedia synthesis, privacy protection, and other cutting-edge applications, as DL has grown.

5.2 Fake video detection

One of the hardest problems is spotting deepfakes in videos. Deepfake creates incredibly complex face-swap videos with just one GPU and a lot of training data. Many computer enthusiasts have posted live-performance recordings with manufactured stars in place of regular people's faces to the short videosite, which has caused controversy. There have been issues with this strategy. Before this technique, movies were thought to be reliable and could even be used as video evidence for multimedia forensics. In the internet age, video deepfake technology has put public confidence at risk. Some even believe that this development could impede societal progress. Thanks to the rapid expansion of open-source datasets, significant advancements in the study of topics like GANs, and notable technological advancements in the field of high-speed computing, creating and modifying fake videos has become relatively easy in terms of the cost of producing a manipulated sequence. By superimposing a politician's face over that of a target actor, a deepfake movie could be used to propagate false political propaganda and defamation. Additionally, it could be used to discredit, humiliate, or divide significant political institutions, jeopardizing national peace, by disrupting the delicate balance of peace between people, governments, and nations globally.

5.3 Fake sound detection

In daily life, machines produce most of the voices. People are increasingly using voice to manage daily tasks as technology becomes more automated. Virtual assistants such as Google Assistant, Alexa, Siri, Bixby, and others are increasingly using artificial or machine-generated voices. Even well-known individuals, such as politicians and celebrities, fall victim to these technologies despite having access to them. The main cause for concern is Deepfake, a cutting-edge technique that primarily uses GAN to produce synthetic sounds to resemble

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

real individuals. Because there are so many different ways to make synthetic speech, the task of identifying it is extremely difficult.. However, synthetic speech can be produced using straightforward cut-and-paste methods that accomplish waveform concatenation, which are frequently offered as open-source toolkits. The source of the speech stream or the bandpass filter model can also be used to build it. Recent years have seen the introduction of numerous CNN-based techniques for producing fake audio. These produce extremely realistic results that are difficult to distinguish from actual speech and human listeners.

5.4 Fake hybrid multimedia detection

Deepfakes, which are algorithmically changed images, sounds, movies, and other information, have the potential to generate massive social upheaval when combined with social media virality. To verify the legitimacy of digital content, anti-disinformation tools like deepfake detection algorithms or immutable metadata are necessary. Having well-established techniques to recognize, assess, and defeat deepfake digital content is crucial when working with multimedia deepfake technology. Multimedia content includes things like movies, artwork, photos, sound recordings, and more. Even with a trustworthy, secure, and dependable system for tracing the history of digital content, achieving this goal can be difficult. This section will look into DL techniques for identifying hybrid multimedia content. Several artificial intelligence algorithms are combined in fake hybrid multimedia detection to identify modified or fake digital content. To expose hidden fakes, it examines text, image, audio, and video combinations. Compared to examining only one kind of media, this method is far more accurate.

PROPOSED METHODOLOGY

The robustness of CNNs for spatial feature retrieval and LSTM (Long-Short-Term-Memory) networks for image temporal sequence analysis are leveraged in the suggested methods for deepfake image or video detection. The goal of this combining model is to accurately depict the temporal dynamics and spatial inconsistencies that are characteristic of deepfakes.

3.1. Data Collection and Preprocessing Dataset Compilation: Gather an extensive collection of photographs and videos from various sources that are both authentic and fraudulent. To guarantee robustness, this dataset should contain a variety of people, settings, and manipulation methods. Data Augmentation: Using techniques like rotation, cropping, flipping, and scaling to increase the variety of the training data during CNN model training also improves the model's capacity for generalization. Preprocessing: Adjust the size and format of the photos and videos. Create a series of images for temporal analysis by extracting frames from films at regular intervals.

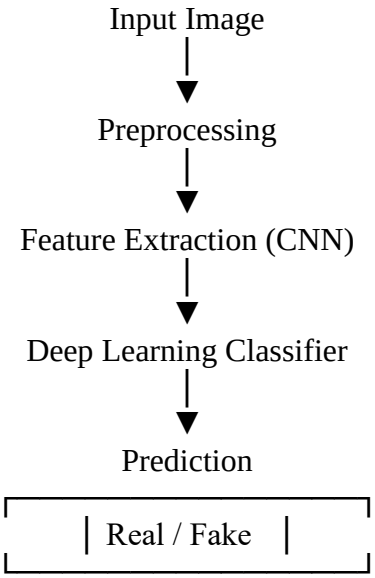
3.2. Spatial Feature Extraction Using CNNs CNN Architecture: Create a CNN architecture specifically designed for picture feature extraction. To meet the unique needs of deep fake detection, common architectures like VGG16, ResNet, or Inception can be modified. CNN training: To learn spatial characteristics that distinguish between real and altered images, train the CNN using the preprocessed dataset. Utilize a sizable labelled dataset and, if needed, implement strategies like transfer learning. High-level feature maps can be extracted from the CNN's intermediate layers. The spatial data represented by these feature maps will be utilized in the next temporal analysis.

3.3. Temporal Analysis Using LSTMs Sequence Formation: Sort the CNN's retrieved feature maps into sequences that match the video frames. The temporal evolution of a video's features is represented by each sequence. LSTM Architecture: Create an LSTM network to handle feature map sequences. The temporal patterns and dependencies that are typical of

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

deepfake films will be recognized by the LSTM. Training the LSTM: To learn temporal dynamics, train the LSTM using feature map sequences. To avoid overfitting and enhance generalization, employ suitable loss functions and regularization strategies.

SYSTEM ARCHITECTURE



DATASET

Commonly used datasets include:

Dataset	Description
FaceForensics++	Face manipulation dataset
Celeb-DF	High-quality celebrity deepfakes
DFDC	Facebook DeepFake Detection Challenge dataset
CIFAKE	AI-generated and real image dataset
CASIA	Image tampering dataset

FUTURE WORK

Future research should focus on:

- Lightweight models for mobile devices
- Explainable AI for detection transparency
- Multimodal detection using images and text
- Federated learning for privacy-preserving training
- Detection methods for diffusion-generated images
- Real-time deployment on social media platforms

CONCLUSION

Artificial intelligence's growing sophistication has made it easier than ever to create incredibly lifelike fake photos and movies. Deepfake technology raises severe issues about disinformation, criminality, identity theft, privacy infringement, and public trust, even while it has many benefits in fields including entertainment, education, healthcare, and digital media production. Thus, it is now crucial to build effective and precise deepfake detection

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

methods in order to preserve the authenticity and integrity of digital content.With a focus on convolutional neural networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal connection analysis in video sequences, this study examined deep learning-based methods for identifying authentic and fraudulent images. Comprehensive data preprocessing, such as Error Level Analysis (ELA), picture normalization, and data augmentation, is incorporated into the suggested methodology. Feature extraction, model training, and performance evaluation come next. Benchmark datasets that are accessible to the public, like FaceForensics++, Celeb-DF, DFDC, CIFAKE, and CASIA, were shown to be useful tools for creating reliable detection systems.The suggested methodology shows how combining CNN and LSTM architectures may successfully detect temporal inconsistencies and spatial manipulation artifacts found in deepfake material. Deep learning methods offer dependable and scalable solutions for differentiating real photos from modified information, according to performance evaluation using accuracy, precision, recall, F1-score, and confusion matrix analysis. However, the continued development of generative models, especially GANs and diffusion models, poses persistent problems that need for more flexible and broadly applicable detection techniques.Future studies should concentrate on creating explainable and lightweight deepfake detection models that may be used on social media sites and mobile devices. The robustness and usefulness of deepfake detection systems can further be increased by including multimodal learning, transformer-based architectures, federated learning, and real-time detection capabilities. In conclusion, deep learning continues to be one of the most promising technologies for preventing the manipulation of digital media and ensuring the security, legitimacy, and authenticity of online information in a world that is becoming more and more AI-driven.

REFERENCES

1. Goodfellow, I., et al. (2014). Generative Adversarial Networks. Advances in Neural Information Processing Systems (NeurIPS).
2. Rossler, A., et al. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. IEEE International Conference on Computer Vision (ICCV).
3. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
4. Dosovitskiy, A., et al. (2021). An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. International Conference on Learning Representations (ICLR).
5. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. International Conference on Machine Learning (ICML).
6. He, K., et al. (2016). Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
7. Wang, S., et al. (2020). CNN-generated Images Are Surprisingly Easy to Spot… for Now. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
8. Verdoliva, L. (2020). Media Forensics and DeepFakes: An Overview. Proceedings of the IEEE, 109(5), 736–759.
9. S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro
10. H. Farid. A Survey Of Image Forgery Detection. IEEE Signal Processing Magazine, 26(2):26-25, 2009.
11. Mallet, J., Dave, R., Seliya, N., and Vanamala, M., 2022, November. A deep learningbased approach for detecting deepfakes. In Proceedings of the 9th International Conference on Soft Computing & Machine Intelligence.
12. S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, H. Li, Protecting world leaders against deep fakes., in: CVPR Workshops, 2019, pp. 38–45.