# **IMPROVING DEEPFAKE DETECTION IN VIDEOS USING DEEP NEURAL NETWORKS**

**Abstract**

Deepfakes—synthetic media created using sophisticated machine learning techniques—present serious issues across a range of domains. To lessen the detrimental effects of deepfakes, it is essential to comprehend them and develop detection techniques. Through an examination of several models, observations of existing gaps in the literature, and an examination of current methodologies, this research seeks to determine the most effective methods for detecting deepfakes. Deepfakes, a chilling fusion of "deep learning" and "fake," represent a paradigm shift in the realm of synthetic media. These AI-generated creations seamlessly manipulate video and audio, enabling the uncanny superimposition of one person's face or voice onto another. Fueled by the relentless march of deep learning techniques, particularly Generative Adversarial Networks (GANs), deepfakes are becoming increasingly sophisticated, blurring the lines between reality and fabrication with terrifying ease.

**Introduction**

Deepfakes pose a significant threat by enabling the creation of highly realistic manipulated videos that can be used to spread misinformation, damage reputations, and erode trust in media. Current deepfake detection methods, while effective to some degree, still face challenges in accurately identifying deepfakes, particularly as deepfake creation techniques continue to evolve.

This research aims to address this critical gap by exploring and evaluating the potential of deep neural networks for improved deepfake detection in videos. The objective is to develop a robust and generalizable deep learning model that can effectively distinguish between real and deepfake videos, even in the face of advancements in deepfake creation methods.

* 1. ***What are Deepfakes?***

Deepfakes are synthetic media made with deep learning algorithms; they are mainly films or images. These algorithms alter or substitute the original information with fake film that is frequently identical to real footage. Because deepfake technology is so sophisticated, there is a chance that it will propagate false information, erode public confidence in the media, and jeopardize security and privacy.

***1.2 Why Is It necessary to detect deepfakes?***

The spread of deepfakes intensifies worries about digital manipulation, distortion, and disinformation already present. Deepfake detection is crucial for maintaining information integrity, shielding people from harm, and maintaining the authority of visual content in a variety of fields, such as politics, journalism, entertainment, and security.

***1.3. Research in Deepfakes***

One of the most important areas of research regarding deepfakes is the creation and detection of deepfakes. Several studies have shown various techniques for deepfake creation as well as detection. The authors of the study categorized deepfake methodologies into four groups, including deep-learning-based, classical machine-learning-based, statistical, and blockchain-based techniques. And according to their findings deep-learning-based methods perform better than all.

***1.4 Impact of Deepfakes on society***

Deepfake has an ethical concern with both positive and negative effects on society.

***Positive impact*** *–* Deepfakes are used in creating the best version of videogames, electronic communications, AR/VR oriented clothing and e-commerce are just few applications the holds benefit from deepfake invention. Now businesses can hire phony models and actresses to display fashionable attire on a wide range of models with various height, weight, skin color. Realistic-sounding and smart-looking assistants and enhanced telepresence in online games and virtual chat environments. This promotes improved online communication and interpersonal relationships.

***Negative Impact*** ***-*** Deepfake creates numerous amounts of security issues Usually, the targeted audience for deepfakes are famous celebrities and politicians.

* “deepfakes” derives its name from being used to create pornographic material on purpose. To put this into context, VOX research [81–84] indicates that 96% of the deepfake movies in online pornography in 2020 were produced to harm the reputations of their victims.
* Deepfakes can be used by political enemies to sway the people and to foster mistrust. Barack Obama was caught on camera in 2008, 2012, and 2016 stating that individuals in hard-hit areas frequently turned to religion and guns.

***1.5 An Approach for deepfake detection***

Recent approaches have suggested convolutional neural networks (CNN) as an effective fit for DF detection models. Usually, pre-trained CNN models are applied on single frames, while other approaches have considered recurrent convolutional networks where frames can be grouped to form the decision. In addition, some approaches consider facial expression patterns to capture fake content.

Most CNN-based approaches are black boxes, where the models are overfitting. In other cases, the validation, testing, and training split are not uniformly distributed, which leads to different interpretations of the same datasets under different operating conditions. For example, a DF detection model on the Facebook DF detection challenge dataset is proposed. The model scored an average precision of 82.56% on these datasets, but the performance drastically drops to 65.18% on the validation dataset, as it is collected from various sources. Thus, a generalization through CNN on one dataset does not hold a cross-performance on another dataset.

The inconsistencies can be mitigated through an effective deep CNN (D-CNN) model that can address the cross-domain interpretability while maintaining the robustness and generalizability of the DF detection scheme, which would yield a high accuracy through an effective ensemble to the proposed CNN approaches.

**METHODS AND RESULTS**

We developed a website to compare the performance of various models, including MesoNet, ResNeXt, LSTM, and EfficientNet. These0 models were tested on the DFDC dataset, and their results were analyzed.

**MesoNet** - a CNN-based model, was implemented due to its simplicity and effectiveness in detecting manipulated media. This model achieved an accuracy of 57% on the DFDC dataset. While the performance is moderate, it highlights the need for more sophisticated models to improve detection accuracy.

**ResNeXt with LSTM -** ResNeXt, an advanced CNN architecture known for its scalable design, demonstrated improved performance with an accuracy of 84.45%. The model's enhanced feature extraction capabilities significantly contributed to its higher accuracy compared to MesoNet.

**EfficientNet -** a transformer-based model, exhibited the highest accuracy at 90.37%. This model combines efficiency and accuracy, making it particularly suitable for deepfake detection. Its superior performance underscores the potential of transformer-based architectures in addressing the challenges posed by evolving deepfake techniques.

**EVALUTION OF MODEL**

The comparison of these models revealed several key insights:

**Complexity vs. Performance:** There is a clear correlation between model complexity and detection performance. More advanced architectures like ResNeXt and EfficientNet outperformed simpler models like MesoNet.

**Temporal Features:** Models that can capture temporal features, such as LSTM, are crucial for video-based deepfake detection, highlighting the importance of sequential data analysis in improving detection accuracy.

**Scalability:** EfficientNet’s architecture demonstrated that scalability and efficiency are critical for achieving high accuracy while maintaining computational feasibility.

**Implementation on Our Website-** Our website provides a platform for users to upload videos and receive deepfake detection results based on the aforementioned models. The user-friendly interface allows for easy comparison of model performance and visualization of detection outcomes. Additionally, the website features educational resources on deepfake technology and detection techniques, fostering awareness and understanding among users.

**CONCLUSION**

Deepfake detection is a critical area of research due to the increasing sophistication of synthetic media. Our study underscores the need for centralized datasets, human involvement, real-time detection capabilities, and generalizable models. By leveraging deep neural networks, particularly transformer-based models like EfficientNet, we can enhance the robustness and accuracy of deepfake detection methods. The development and deployment of our website further demonstrate the practical application of these models, providing a valuable tool for combating the threats posed by deepfakes.

**FUTURE WORK**

* Developing Centralized Datasets: Establishing standardized benchmarking datasets to facilitate fair comparisons and improve model evaluations.
* Human-in-the-Loop Systems: Incorporating human expertise to enhance automated detection systems through contextual analysis and fact-checking.
* Real-Time Detection: Addressing the trade-off between accuracy and computational efficiency to develop real-time detection systems.
* Model Generalization: Enhancing the generalizability of detection models to adapt to new and evolving deepfake techniques.

**AUTHOR –**

**SARGAM TYAGI (E22CSEU1583)**

**SATVIK KARAN (E22CSEU1332)**