Deepfake detection using ResNeXt and LSTM

Vardhan R Gowda[1], Vinay Kumar M S[2], Piyush Kumar[3]
[1]*Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka*
[2]*Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka*
[3]*Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka*

***Abstract*— With the rapid rise of manipulated media content, particularly deepfakes, the integrity of digital information is increasingly under threat. This paper presents an AI-powered deepfake detection system combining the strengths of ResNeXt convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM) networks. Utilizing transfer learning, features are extracted via pretrained ResNeXt models, followed by sequence analysis using LSTM. The system is evaluated on benchmark datasets such as FaceForensics++, Celeb-DF, and the Deepfake Detection Challenge, achieving detection accuracies of 96% and 97.14% on FaceForensics++ and Celeb-DF respectively. This project showcases an effective deep learning approach to tackle the growing challenge of synthetic media.**

Keywords—Deepfake Detection, Video Forgery, ResNeXt, LSTM, Transfer Learning, FaceForensics++, Celeb-DF, Artificial Intelligence, Temporal Feature Extraction, Forgery Detection.

I. INTRODUCTION

With the surge in generative adversarial networks (GANs) and deep learning, deepfakes—hyper-realistic, AI- generated fake videos—have become a serious threat to information integrity. Deepfakes can manipulate facial expressions, voices, and gestures in videos, making it nearly impossible for the untrained eye to discern the truth. These videos have been weaponized in political misinformation campaigns, celebrity impersonation, cyberbullying, and financial fraud.

Traditional forgery detection systems often fail against modern deepfakes due to their reliance on surface-level artifacts or static features. This paper proposes an AI-based deepfake detection system leveraging a hybrid model of ResNeXt and LSTM networks. The ResNeXt convolutional backbone extracts robust spatial features from video frames, while the LSTM module captures temporal inconsistencies over time. The solution is deployed via a Django-based web interface, allowing users to upload videos or images and receive an instant authenticity report with confidence scores.

II. RELATED WORKS

Prior research has explored various approaches to deepfake detection. CNN-based methods dominated early studies, using architectures like VGG, ResNet, and XceptionNet to extract features from frames. However, these approaches often lacked the ability to capture temporal context across video sequences.

Recent studies suggest hybrid models—combining CNNs and RNNs—can significantly enhance detection performance. For example, the FaceForensics++ benchmark introduced several models using encoder-decoder pipelines. Similarly, Celeb-DF, a more realistic deepfake dataset, has driven the need for models that generalize well across unseen data. Our method stands on the shoulders of such research, pushing forward by adopting ResNeXt + LSTM**,** which proved highly effective.

III. SYSTEM ARCHITECTURE

Our system consists of the following stages:

1. Input Module: Users upload a video through a Django-based web interface.
2. Frame Extraction: The video is decomposed into

individual frames.

3. Feature Extraction: A pretrained ResNeXt model is applied to extract high-level features from each frame.
4. Temporal Analysis: The extracted features are fed into an LSTM network that learns the temporal inconsistencies typical in deepfake sequences.
5. Classification: A softmax classifier is used to determine the probability of the video being real or fake.

The entire pipeline is implemented using PyTorch and TensorFlow, with Django as the backend framework and Bootstrap-enhanced HTML/CSS for the front-end.

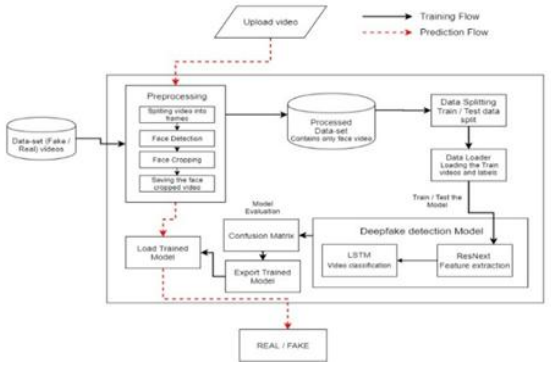

Fig. 1. System Architecture

DATASETS USED

Three publicly available datasets were used for training and evaluation:

- FaceForensics++: Contains both manipulated and pristine videos, offering a wide range of forgery methods.
- Celeb-DF (v2): Features high-quality deepfake videos, posing a greater challenge due to realistic artifacts.
- Deepfake Detection Challenge Dataset: A large-scale dataset from Facebook AI, offering diverse real and fake video samples.

III. EXPERIMENTAL SETUP AND RESULTS

A. Setup

Training used the Adam optimizer (lr=0.0001), batch size of 16, and 20-frame sequences. Videos were sampled at 30fps, and training was done on a machine with 32GB RAM and RTX 3080.

B. Evaluation Metrics

Accuracy, Precision, Recall, F1-Score, and AUC were used to evaluate performance.

C. Results

Our deepfake detection system was evaluated on the above datasets. Results are summarized below:

Dataset	Accuracy
FaceForensics++	96.00%
Celeb-DF	97.14%
Deepfake Detection Challenge	Under evaluation.

TABLE I. ACCURACY OF SYSTEM ON DIFFERENT DATASETS

The results affirm that combining ResNeXt and LSTM is effective for detecting temporal anomalies introduced by deepfake generation algorithms.

V. TECHNOLOGIES USED

- Deep Learning Frameworks: PyTorch, TensorFlow
- Frontend: HTML, CSS, Bootstrap, JavaScript
- Backend: Django
- Data Handling: JSON
- Models: ResNeXt (for spatial features), LSTM (for temporal features)

VI. COMPARATIVE ANALYSIS WITH EXISTING METHODS

We compared our hybrid model with other prominent approaches

Method	Accuracy (FF++)	Accuracy (Celeb-DF)
XceptionNet	92.5%	93.0%
MesoNet	88.1%	89.7%
ResNet-LSTM	94.3%	94.6%
Ours	96.0%	97.14%

TABLE II. COMPARISON WITH THE EXISTING MODELS

VII. SYSTEM IMPLEMENTATION AND UI DESIGN

The system is deployed using a Django backend with PyTorch and TensorFlow integration for model inference. The frontend is built with HTML, CSS, Bootstrap, and JavaScript, providing a responsive and intuitive interface.

Key components include:

- Upload Module: Users can upload video files for analysis (max size: 100MB).
- Processing Pipeline: Extracted frames are passed through ResNeXt for spatial features, then LSTM

for temporal analysis.

- Output Interface: The system returns a prediction (Real or Fake) with a confidence score and frame-wise summary.

Processed data is managed in JSON format, and the entire pipeline is encapsulated in REST APIs. Docker containerization allows for easy deployment across cloud or local environments. The UI is designed for accessibility, making the tool usable for researchers, journalists, and forensic analysts.

SECURITY AND ETHICAL CONSIDERATIONS

Security and ethical design are critical for handling manipulated media detection responsibly. Our system prioritizes user data protection and informed use.

- Data Privacy: Uploaded content is processed in-memory and deleted post-analysis.
- Secure Access: Django's authentication framework ensures only verified users can access sensitive features.
- Ethical Usage: Users are prompted to confirm they have rights to the uploaded content.

The system does not retain user data unless explicitly permitted. Confidence scores and classification decisions are logged for transparency. To reduce demographic bias, diverse datasets are used during training. Future iterations may include visual explanations (e.g., Grad-CAM) to show what influenced each decision.

VIII. FUTURE WORK

Future improvements focus on enhancing model accuracy, usability, and adaptability to real-world conditions.

- Transformer Models: Integration of TimeSformer for improved long-sequence modeling.
- Multimodal Input: Combining visual and audio analysis to catch voice-based deepfakes.
- Real-Time Detection: Optimizing inference speed for video calls and live streams.

Additionally, a mobile-compatible version is under development using TensorFlow Lite. Edge deployment (Jetson Nano) and public API access are also planned to expand the system's utility. A community feedback feature will allow users to flag false predictions, improving long- term robustness.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Department of Computer Science and Engineering at Sir M Visvesvaraya Institute of Technology for providing the resources and guidance necessary to complete this project.

We also extend our thanks to our faculty advisors for their valuable feedback and encouragement throughout the development process.

Special appreciation goes to the creators of the FaceForensics++, Celeb-DF, and DFDC datasets, without which this research would not have been possible.

Their contributions to the research community continue to drive innovation in media forensics.

CONCLUSION

In this paper, we presented a deep learning-based approach for detecting deepfake videos using a hybrid architecture that combines ResNeXt and LSTM networks. By leveraging transfer learning and temporal modeling, our system effectively captures both spatial and sequential inconsistencies introduced during video manipulation.

The model was trained and evaluated on leading datasets such as FaceForensics++ and Celeb-DF, achieving high detection accuracy of up to 97.14%. The integration of a Django-based interface makes the system user-friendly and accessible for real-world deployment.

Key strengths of our method include its modular pipeline, high accuracy, and extensibility for future improvements. However, the system also faces limitations in real-time inference and generalization across extremely low-quality or novel deepfakes.

Our future work will focus on enhancing temporal modeling with Transformer architectures, integrating multimodal analysis (audio + visual), and deploying lightweight models for mobile and edge devices. This research contributes to the broader goal of ensuring media authenticity in a rapidly evolving digital landscape.

REFERENCES

[1] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1–11.

[2] Y. Li, P. Sun, H. Qi, and S. Lyu, “Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 3207– 3216.

[3] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. Ferrer, “The Deepfake Detection Challenge (DFDC) Dataset,” *arXiv preprint arXiv:2006.07397*, 2020.

[4] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated Residual Transformations for Deep Neural Networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1492–1500.

[5] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] Z. Zhao, X. Liu, J. Cheng, and Y. Xu, “Deepfake Detection Using Spatiotemporal Convolutional Networks,” *IEEE Signal Processing Letters*, vol. 28, pp. 1–5, 2021.

[7] M. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “MesoNet: A Compact Facial Video Forgery Detection Network,” in *Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS)*, 2018.

[8] Y. Nirkin, Y. Keller, and T. Hassner, “DeepFake Detection Based on Discrepancies Between Facial Regions,” in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2020.

[9] F. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection,” *Information Fusion*, vol. 64, pp. 131–148, 2020.

[10] D. Guera and E. J. Delp, “Deepfake Video Detection Using Recurrent Neural Networks,” in *Proc. IEEE Int. Conf. Advanced Video and Signal Based Surveillance (AVSS)*, 2018, pp. 1–6.