DeepScan

A Deepfake Video Detection System

Nighila Ashok
Asst. Professor of Computer Science Department
Universal Engineering College
Thrissur, India
nighila@uec.ac.in

Adithya Ajith
Student, Computer Science Department
Universal Engineering College
Thrissur, India
adithyanonu@gmail.com

Aparna Shaju
Student, Computer Science Department
Universal Engineering College
Thrissur, India
aparnakanatil2002@gmail.com

Arjuna Chandran V V
Student, Computer Science Department
Universal Engineering College
Thrissur, India
arjunachandranvv8546@gmail.com

Fahmi Fathima T S
Student, Computer Science Department
Universal Engineering College
Thrissur, India
fahmifathimats@gmail.com

***Abstract*—Deepfake is defined as a multimedia content synthetically modified or created through automatic (or barely controlled) machine learning models. The rise of deepfake technology points out the importance of accurate detection methods. In this article, we propose a deepfake detection system based on Long Short-Term Memory (LSTM) networks and the ResNext architecture. Users can upload videos for examination, which determines if they are legitimate or fake. LSTM evaluate face motions, gestures, and expressions, whereas ResNext identifies and extracts facial features and landmarks. Additionally, we provide users with an option to report suspected deepfake videos via email, facilitating community involvement in identifying fraudulent content. Moreover, our platform includes a directory of legal advocates, enabling users to seek legal support tailored to their location and needs. In conclusion, our deep learning-based deepfake video detection project represents a vital step in addressing the growing threat of digital manipulation.**

Keywords*— *Deepfake, Artificial Intelligence (AI), Deep Learning, Video Analysis, Facial Recognition, Video Authentication, Detection.

I. INTRODUCTION

Deepfake technology in digital communication has posed a significant challenge to the validity of visual content. Deepfakes, powered by AI algorithms, may modify videos to create convincingly changed settings that mix reality and fabrication. Misinformation, identity theft, and loss of trust in visual media grow more likely as these manipulations gain popularity and accessibility. Our deepfake video detection project aims to combat deceptive information and promote truth and integrity in the digital age. Deepfake, a term coined to represent multimedia information that has been synthetically manipulated or manufactured using advanced

AI algorithms, makes it difficult to distinguish between the reality and fabrication. As deepfake technology advances, the requirement for effective detection methods becomes more critical to combating the spread of manipulated information and protecting against its potentially damaging results.

The paper describes an innovative approach to deepfake detection that employs modern deep learning algorithms. Specifically, we present a deepfake detection system that takes advantage of Long Short-Term Memory (LSTM) networks and the ResNext 50 architecture. LSTM networks excel at capturing temporal dependencies within sequential data, making them ideal to analyze the dynamic nature of video footage. Meanwhile, ResNext 50 serves as a powerful feature extractor, capable of discovering and extracting subtle patterns suggestive of deepfake manipulation from visual data.

The fundamental goal of our proposed approach is to offer consumers with a dependable method of distinguishing between authentic and fake videos. Users can upload films to our platform for analysis, where the LSTM network analyzes facial motions, gestures, and expressions, and the ResNext 50 architecture identifies and extracts facial characteristics and landmarks that are critical for identifying manipulation. This combination approach improves the accuracy and effectiveness of our deepfake detection technology, allowing users to make more educated decisions about the credibility of multimedia material they encounter.

Users are encouraged to submit suspected deepfake films by email, which promotes community involvement in discovering and flagging fraudulent content. Furthermore, to assist persons harmed by deepfake instances, our platform includes a complete directory of legal advocates. This website allows users to get legal assistance based on their geographic area and unique needs, empowering them to handle the legal difficulties associated with deepfake-related issues.

II. Related Works

D. Pan, L. Sun, R. Wang, X. Zhang, and R. O. Sinnott [1] investigate deepfake technology, which has brought about societal concerns, including election biasing. They use advanced neural network architectures, such as Xception and MobileNet, to identify deepfake movies using automated classification tasks. Our evaluation uses datasets from FaceForensics++, including a variety of deepfake scenarios provided by four major technologies. The results show that Xception and MobileNet are effective at detecting altered information, with accuracy rates ranging from 91% to 98%.

Buslaev, Iglovikov, and colleagues [2] explore Albumentations, an open-source image augmentation toolkit that addresses key concerns in existing frameworks. Albumentations prioritizes speed, flexibility, and inclusivity to address overfitting concerns in deep learning models. It supports both basic and advanced transformations. Notable features include seamless connection with other libraries, adaptability for practitioners, and a user-friendly design for customizing augmentation pipelines. The research provides real examples from several computer vision tasks to demonstrate Albumentations' efficiency in improving model robustness and generalization.

Luca Guarnera, Oliver Giudice, and Battiato [3] investigate the revolutionary influence of the convergence of AI and Image Processing, citing examples such as FACEAPP, which uses sophisticated Generative Adversarial Networks (GANs) to make realistic Deepfakes. These synthetic media provide a significant difficulty in multimedia forensics due of their authentic appearance. The research describes a novel Deepfake detection method based on the Expectation-Maximization (EM) algorithm, with an emphasis on the extraction of Convolutional Traces (CT) left by GANs during image production. The suggested method yields an outstanding overall classification accuracy of over 98%, demonstrating robustness to multiple attacks and excelling in real-world circumstances.

According to Luisa Verdoliva [4], significant improvements in multimedia content generation have made it difficult to distinguish between real and synthetic media. The internet has made it easier to create convincing phony images and movies, posing a big issue. The democratization of these instruments raises worries about everything from manipulating public opinion to engaging in fraud or extortion. To combat this rising threat, the study provides a thorough analysis of approaches for visual media integrity verification, with a particular emphasis on detecting manipulated images and videos. The paper explores data-driven forensic methods aimed at addressing the challenges posed by this form of fabricated media content.

Tolosana, Rodriguez, Fierrez, Morales, and Garcia [5] discuss how the convergence of unrestricted access to extensive public databases and the rapid evolution of deep learning techniques, particularly Generative Adversarial Networks (GANs), has resulted in the creation of remarkably realistic fake content, which has significant societal implications, especially in an era dominated by the spread of fake news. The survey goes further into four different types of facial manipulation: entire face synthesis, identity swap (DeepFakes), attribute manipulation, and expression swapping. For each manipulation category, the report describes the methodologies utilized, the availability of public databases, and critical criteria for evaluating false detection approaches. This article focuses on the most recent DeepFakes, highlighting technological advancements and obstacles in establishing efficient detection algorithms.

Guarnera, Giudice, Nastasi, and Battiato [6] discuss the widespread issue of Deepfakes, which use deep learning algorithms to replace faces in photos and videos. This study presents a concise review of Deepfake picture generating technologies and undertakes a forensic investigation to assess

their authenticity. Despite advanced approaches, the article emphasizes the ongoing issue of recognizing bogus Deepfake images. The paper proposes a proactive approach using frequency domain analysis to address the growing threat of Deepfake technology, recognizing the limitations of traditional forensic procedures. This unique technology identifies frequency irregularities, providing a more reliable way to determine face authenticity compared to existing methodologies.

Le, Nguyen, Yamagishi, and Echizen [7] discuss the growing worry of deepfake media in today's quickly evolving technology ecosystem. The work addresses the issues of altered faces in social media by introducing two new countermeasures: multi-face forgery detection and segmentation in real-world scenarios. Multi-face forgery detection is more sophisticated than classic deepfake recognition since it detects falsified faces in the context of multiple human faces in natural scenarios. The study introduces the OpenForensics dataset, a large-scale dataset developed for multi-face forgery detection and segmentation. It includes face-wise annotations to aid research in this field. The study evaluates innovative methods for detecting and segmenting instances on a new dataset.

Chintha et al. [8] propose employing recurrent convolutional structures to detect deepfakes in audio and video recordings. The authors emphasize the necessity of identifying deepfakes, which can propagate false information and confuse people. The discovery of generative adversarial networks has made it possible for anyone to make deepfakes. Detecting deepfakes is becoming crucial for journalists, social media platforms, and the general public. This research presents a method for detecting deepfakes in audio and video recordings using recurrent convolutional architectures. Deepfakes pose a growing threat of deception and misinformation. The method uses convolutional latent representations, bidirectional recurrent structures, and entropy-based cost functions to extract relevant information and detect spatial and temporal signatures of deepfake content.

According to Juan Hu and Xin Liao [9], the use of technologies to create deep fake films is quickly expanding. These films can be easily manipulated without leaving noticeable evidence. While forensic detection in high-definition video datasets has shown promising results, the forensics of compressed films warrants additional investigation. In reality, compressed videos are prevalent on social media platforms such as Instagram, Wechat, and Tiktok. As a result, determining ways to identify compressed Deepfake movies becomes an important topic. In this research, we present a two-stream technique for assessing compressed Deepfake films on both the frame and temporal levels. Because video compression introduces a lot of redundant information into frames, the suggested frame-level stream gradually prunes the network to keep the model from fitting the compression noise.

Yuval Nirkin and Lior Wolf [10] created a unique method for detecting face swapping and identity modifications in photos. DeepFake is a technique that modifies face features in photos to match the surrounding environment while preserving the original image. Manipulation can cause variations between the altered face and its immediate surroundings. To do this, we used two specialized networks. One network recognizes faces within specific semantic boundaries, while the other identifies contextual elements such as hair or ears. By evaluating recognition signals from these networks, we may identify disparities and discover falsified images.

Gen Li and Xianfeng Zhao [11] introduce a novel DeepFakes forensics approach called forensic symmetry. This approach identifies whether two symmetrical face patches have similar or different natural features. We propose a multi-stream learning structure consisting of two feature extractors. The first feature extractor extracts symmetry features from front-facing photos. The second feature extractor extracts similarity features from side face photos. Natural features refer to both symmetry and similarity features. The forensic symmetry system puts symmetrical face patches into angular hyperspace to measure their inherent differences. Face photos are more likely to be tampered with when their natural features differ significantly.

Yukai Wang and Chunlei Peng [12] emphasize the growing necessity to detect false faces, particularly with advanced technology such as GANs. Detecting false faces requires a large number of samples and databases for computers to learn from. There are limited databases available for recognizing false faces in near-infrared light, which is extensively employed in security systems. The publication introduces "ForgeryNIR," an extensive collection of genuine and modified faces. This collection of over 50,000 genuine and modified faces aims to train computers to recognize phony faces in near-infrared light, simulating real-world conditions. Faces are altered to be more realistic.

Maryam Taeb, Hongmei [13] Chi covers deepfakes, which are very realistic yet fake media generated by strong computer algorithms. Algorithms analyze large volumes of data to modify films or photos, including switching faces or objects. Fake content can disrupt public discussions and jeopardize human rights. Deepfakes are utilized in legal situations to deceive and alter evidence and court decisions. Detecting altered media is vital in digital forensics as it plays a significant part in judicial proceedings. There is a great need to build technology that accurately distinguishes between real and falsified material.

Fatima Maher and Samy S. Abu-Naser [14] use artificial intelligence (AI) and deep learning approaches to recognize authentic and phony faces.
This introduction to AI covers deep learning, machine learning, and neural networks, highlighting their potential to solve real-world problems by mimicking human cognitive functions. The project uses deep learning, a subset of machine learning, to generate models for identifying real and modified facial photos. The goal is to develop an effective solution for this critical classification.

Chih-Chung Hsu [15] raises concerns about the exploitation of synthesized images created by generative adversarial networks (GANs), which can produce realistic but unsuitable content. Fake photos on social media constitute a big hazard and require sophisticated detection tools. Traditional picture forgery detectors struggle to recognize GAN-generated images due to their intricate manipulation. The suggested technique uses deep learning and contrastive loss to detect fraudulent images accurately. The methodology uses advanced GANs to produce pairings of fake and real photos.

III. Proposed System

Our model combines CNN and RNN. We used the Pre-trained ResNext CNN model to extract frame-level information, then trained an LSTM network to categorize the video as deepfake or immaculate. The Data Loader is used to load video labels into the training model.

Deepfake technology has generated worries regarding manipulation of visual media, highlighting the necessity for solid detection tools. Deepfakes use advanced machine learning techniques to create convincing but faked images or videos, sometimes superimposing one person's face on another or modifying information to trick viewers. To address this issue, researchers and practitioners are experimenting with advanced neural network architectures such as ResNext and LSTM. These strategies seek to develop reliable algorithms capable of distinguishing genuine information from altered or fabricated media.

ResNext, a convolutional neural network, succeeds in detecting subtle patterns in images, making it effective for feature extraction tasks. LSTM, a form of recurrent neural network, can handle sequential input by keeping long-term dependencies, making it ideal for processing video frames or extracted features. A deepfake detection method relies on the combination of ResNext for feature extraction and LSTM for temporal modeling. Using pre-trained ResNext models to extract high-level features from images or frames and input them into LSTM networks might help identify modified content by learning patterns and temporal connections within sequences. This hybrid approach provides a comprehensive analysis of visual media, detecting small alterations or discrepancies that may suggest deepfake manipulations. These models aim to detect deepfakes with reliability and accuracy through rigorous training, evaluation, and testing on various datasets.

The application's User Interface is built using the Django framework. The initial page of the user interface features a tab for browsing and uploading videos. The model processes the uploaded video and makes predictions. The model gives an output indicating whether the video is real or false, as well as its confidence. If the recognized video is fraudulent, the user can file a complaint via email and contact the advocates using the information provided based on their location.

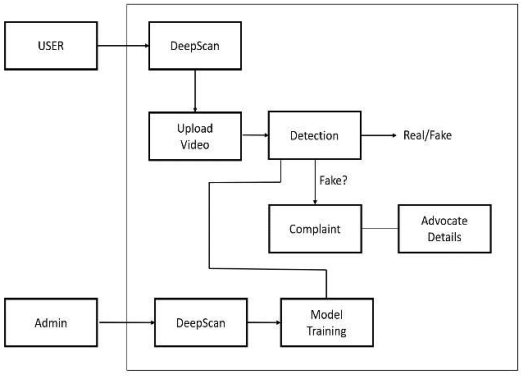

Fig. 1. System Architecture

A. Data Collection

The FaceForensic++ dataset is a well-known and extensively utilized resource in the field of deepfake detection. It is designed primarily to facilitate research and development in the field of identifying altered facial photos and videos. This dataset contains a wide range of films featuring both real and synthetic faces, with the synthetic faces created using a variety of deep learning approaches, including generative adversarial networks. The FaceForensic++ dataset contains a wide range of modified movies, including deepfake videos created via face swapping, face reenactment, and facial expression manipulation.

The DFDC collection contains videos from a variety of sources, including movies, television series, and social media platforms, assuring diversity in both content and quality. The modified movies in the collection are created using powerful AI algorithms, including deep learning techniques like generative adversarial networks (GANs), resulting in extremely realistic deepfake content. Each video in the DFDC dataset is painstakingly tagged to reflect its authenticity state, allowing researchers to train and evaluate deepfake detection models efficiently.

Celeb-DF offers both legitimate and deepfake films created with advanced AI techniques like generative adversarial networks (GANs). Deepfake videos use face swapping and other forms of facial modification to generate realistic but fictional footage. Each video in the Celeb-DF dataset is precisely annotated to indicate its authenticity state, allowing researchers to design and test deepfake detection models more efficiently.

B. Model Training

In training a deepfake detection model using datasets such as FaceForensic++, DFDC, and Celeb-DF, we use a combination of LSTM and ResNext architectures to successfully capture temporal correlations and subtle signals indicating manipulation. Initially, the datasets are

preprocessed to guarantee uniform resolution and format, and relevant characteristics are extracted. The first steps in the preprocessing of the video are to split the video into frames. After splitting the video into frames the face is detected in each of the frame and the frame is cropped along the face. Later the cropped frame is again converted to a new video by combining each frame of the video. The process is followed for each video which leads to creation of processed dataset containing face only videos. The frame that does not contain the face is ignored while preprocessing. The datasets are then separated into training, validation, and testing sets to ensure an even distribution of legitimate and modified samples.

C. *Detection*

In deepfake detection, the LSTM and ResNeXt models play separate roles in detecting alterations within videos. The LSTM, which is well-known for its ability to analyze sequential data, focuses on temporal dependencies. It examines the sequence of frames, identifying abnormal motion patterns and anomalies that indicate deepfake modifications. By recording long-term context, LSTM may detect small changes across numerous frames. ResNeXt, on the other hand, specializes in spatial feature extraction, which involves analyzing individual frames for visual discrepancies.

D. *Prediction*

Our deepfake detection model uses a combination of LSTM and ResNeXt architectures to estimate the authenticity of uploaded movies. Our strategy is based on extracting significant aspects from video data, such as frames, optical flow, and facial landmarks, in order to capture both temporal and spatial qualities. Our model design uses an LSTM network to assess temporal correlations and trends in the video sequence, while a ResNeXt network extracts spatial features from individual frames.

E. *Details of Advocates*

Incorporating information about legal advocates into the deepfake detection platform offers users with critical help and direction in dealing with any legal consequences of deepfake situations. The database of legal advocates includes professionals who specialize in fields such as media law, intellectual property rights, and digital privacy. Users can search this database using the platform, filtering advocates based on their geographic area.

IV. Result and Discussion

Following the introduction of our deepfake detection system based on LSTM networks and ResNext architecture, we saw promising results in properly distinguishing between real and altered multimedia content. After extensive testing, our system displayed excellent precision in evaluating face motions, gestures, and expressions, as well as competent extraction of facial characteristics and landmarks. The combination of LSTM and ResNext proved beneficial in delivering a thorough analysis of video content, allowing for robust detection of deepfake abnormalities.

In addition, the inclusion of a reporting tool encouraged community participation, boosting the overall effort to identify and resolve instances of fraudulent information. Furthermore, the presence of a database of legal advocates was extremely helpful in guiding consumers through the legal intricacies involved with deepfake occurrences. Overall, our system's performance constitutes a significant step forward in minimizing the growing threat of digital manipulation, emphasizing the need of precise detection methods and joint efforts in ensuring online integrity. Continued development and optimization of our system has the potential to increase its effectiveness and contribute to a safer digital environment for all users.

Our suggested model, which combines CNN and RNN using LSTM and ResNext, can detect whether an uploaded video is genuine or not. The performance of our model is illustrated in Figure 2. The combined performance of LSTM and ResNext models in deepfake detection demonstrates a comprehensive approach that combines the benefits of both architectures to improve detection accuracy and robustness. The combined statistics graph depicts the performance of multiple computer vision models across epochs, including tasks like object detection, image segmentation, and image translation. Each model's performance metric is displayed against the number of training epochs, allowing for a detailed comparison of their learning progress over time.

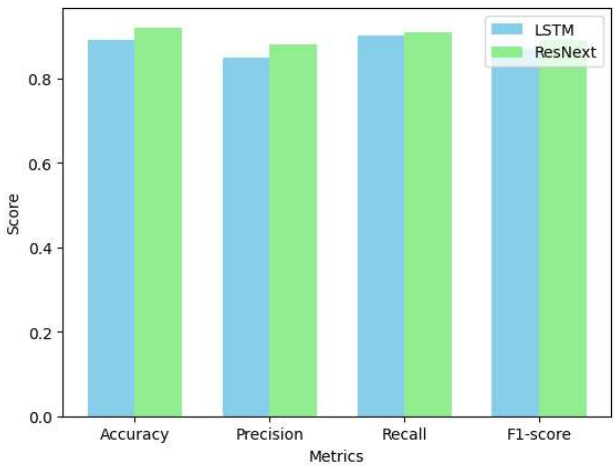

Fig. 2. Performance graph of LSTM and ResNext

V. Conclusion

Advancements in AI and computer vision are transforming how humans interact with images. Deepfakes are becoming

increasingly common these days. Deepfake is a type of synthetic media that uses AI and machine learning to change existing photos, videos, or audio content. Long Short-Term Memory (LSTM) networks and ResNext-50 operate together to detect deepfake videos, providing a strong framework for countering altered information.

Our proposed deepfake detection solution, which is powered by LSTM networks and the ResNext architecture, represents a significant step forward in ensuring digital integrity. By providing consumers with a simple way to submit videos for review, we enable them to make informed decisions about the legitimacy of multimedia content. The combination of LSTM's ability to analyze facial movements and ResNext's ability to extract features results in a complete and effective detection method. Furthermore, the inclusion of a reporting feature on our platform promotes community interaction, hence increasing joint efforts to discover and combat fraudulent information.

VI. REFERNCES

[1] D. Pan, L. Sun, R. Wang, X. Zhang and R. O. Sinnott, "Deepfake Detection through Deep Learning," 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK, 2020, pp. 134-143, doi: 10.1109/BDCAT50828.2020.00001.

[2] Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. "Albumentations: Fast and Flexible Image Augmentations". Information 2020, 11, 125. [CrossRef].

[3] Guarnera, L.; Giudice, O.; Battiato, S. "Fighting Deepfake by Exposing the Convolutional Traces on Images". IEEE Access 2020,8, 165085–165098. [CrossRef].

[4] Verdoliva, L. "Media Forensics and Deepfakes: An Overview". IEEE J. Sel. Top. Signal Process. 2020, 14, 910–932. [CrossRef]

[5] Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J. "Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detec tion". Inf. Fusion 2020, 64, 131–148. [CrossRef].

[6] Guarnera, L.; Giudice, O.; Nastasi, C.; Battiato, S. "Preliminary forensics analysis of deepfake images". In Proceedings of the 2020 AEIT International Annual Conference (AEIT), Catania, Italy, 23–25 September 2020; pp. 1–6. [CrossRef].

[7] Le, T.; Nguyen, H.H.; Yamagishi, J.; Echizen, I. "OpenForensics: Large Scale Challenging Dataset For Multi-Face Forgery Detection And Segmen tation In-The-Wild". In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV),Online, 11–17 October 2021; IEEE Computer Society: Los Alamitos, CA, USA, 2021; pp. 10097–10107. [Cross Ref].

[8] Chintha, A., Thai, B., Sohrawardi, S. J., Bhatt, K., Hickerson, A., Wright, M., Ptucha, R. (2020). "Recurrent convolutional structures for audio spoof and video deepfake detection". IEEE Journal of Selected Topics in Signal Processing, 14(5), 1024–1037. https://doi.org/10.1109/jstsp.2020.2999185.

[9] Hu, J., Liao, X., Wang, W., Qin, Z. (2022b). "Detecting Compressed Deep fake Videos in Social Networks Using Frame-Temporality Two-Stream Con volutional Network". IEEE Transactions on Circuits and Systems for Video Technology, 32(3), 1089–1102. https://doi.org/10.1109/tcsvt.2021.3074259.

[10] Nirkin, Y., Wolf, L., Keller, Y., Hassner, T. (2022). "DeepFake detection based on discrepancies between faces and their context". IEEE Transac tions on Pattern Analysis and Machine Intelligence, 44(10), 6111–6121. https://doi.org/10.1109/tpami.2021.3093446.

[11] Li, G., Zhao, X., Cao, Y. (2023). "Forensic symmetry for DeepFakes. IEEE Transactions on Information Forensics and Security", 18, 1095–1110. https://doi.org/10.1109/tifs.2023.3235579.

[12] Wang, Y., Peng, C., Liu, D., Wang, N., Gao, X. (2022). "ForgeryNIR: Deep Face Forgery and Detection in Near-Infrared Scenario". IEEE Transactions on Information Forensics and Security, 17, 500–515. https://doi.org/10.1109/tifs.2022.3146766.

[13] Taeb, M.; Chi, H. "Comparison of Deepfake Detection Techniques through Deep Learning". J. Cybersecur. Priv. 2022, 2, 89-106. https://doi.org/10.3390/jcp2010007.

[14] Fatima Maher Salman, Samy S. Abu-Naser "Classification of Real and Fake Human Faces Using Deep Learning".

[15] Chih-Chung Hsu, "Deep Fake Image Detection Based on Pairwise Learning".