IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

BEYOND THE ILLUSION: A MACHINE LEARNING APPROACH TO DEEPFAKE DETECTION

AARTHINIVASINI.R.B(M.E)[1],KEERTHIGA.D[2],KULASI.K[3],SANGAMITHRA.M.K[4], NISHIHA.T[5]

ASSISTANT PROFESSOR[1],STUDENT[2,3,4]

DEPARTMENT OF INFORMATION TECHNOLOGY[1]

MEENAKSHI COLLEGE OF ENGINEERING,CHENNAI,TAMIL NADU,INDIA[1]

Abstract : As the prevalence of deepfake videos continues to escalate, there is an urgent need for robust and efficient detection methods to mitigate the potential consequences of misinformation and manipulation. This abstract explores the application of Long Short-Term Memory (LSTM) networks in the realm of deepfake video detection. LSTM, a type of recurrent neural network (RNN), has proven to be adept at capturing temporal dependencies in sequential data, making it a promising candidate for analyzing the dynamic nature of videos. The research delves into the intricacies of utilizing LSTM architectures for the detection of deepfake videos, emphasizing the significance of understanding temporal patterns inherent in manipulated content. The proposed methodology involves preprocessing of video data, including the creation of high-quality training datasets and the application of data augmentation techniques to enhance model generalization. The training process and optimization strategies specific to LSTM networks are explored to achieve optimal performance in deepfake detection. Evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the model's effectiveness in distinguishing between genuine and manipulated content. The abstract also addresses challenges and limitations inherent in deepfake detection, including mitigating false positives and negatives, and discusses potential avenues for future research to enhance the robustness of LSTM-based detection systems.
The findings of this research have implications for real-world applications, particularly in the context of social media platforms and video hosting services, where the integration of LSTM-based deepfake detection can contribute to a safer and more secure online environment.

I. INTRODUCTION

1.1 INTRODUCTION

Deep-fakes, generated by sophisticated artificial intelligence algorithms, have the ability to seamlessly replace faces and manipulate audio, blurring the lines between truth and fiction. As the technology behind deep-fakes advances at an unprecedented pace, the implications for misinformation, identity theft, and malicious content creation become increasingly profound. LSTM, a type of recurrent neural network (RNN), has proven to be adept at capturing temporal dependencies in sequential data, making it a promising candidate for analyzing the dynamic nature of videos. The research delves into the intricacies of utilizing LSTM architectures for the detection of deepfake videos, emphasizing the significance of understanding temporal patterns inherent in manipulated content. Machine Learning is a system of computer algorithms that can learn from example through self-improvement without being explicitly coded by a programmer. Machine learning is a part of artificial Intelligence which combines data with statistical tools to predict an output which can be used to make actionable insights.The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results. Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input and uses an algorithm to formulate answers.

A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendation.Machine learning is also used for a variety of tasks like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on.

1.2 NEED OF THE STUDY.

The suggested approach trains a neural network to discriminate between real and fake faces using deep learning techniques and an extensive dataset of actual and fake face images. The system was tested using a number of metrics, and the results demonstrated that it is more effective at identifying deep fakes. The evaluation outcomes implied that, even when endured with fraudulent faces, the suggested system can recognize faces with high accuracy of 91.82% and a F1 score of 91% which has been developed using an easily understandable LSTM network rather than complex techniques like SVM, Boosting, Transfer Learning and other advanced architectures

II. RESEARCH METHODOLOGY

Evaluation metrics such as accuracy, precision, recall, and F1 score are employed to assess the model's effectiveness in distinguishing between genuine and manipulated content. The abstract also addresses challenges and limitations inherent in deepfake detection, including mitigating false positives and negatives, and discusses potential avenues for future research to enhance the robustness of LSTM-based detection systems. The findings of this research have implications for real-world applications, particularly in the context of social media platforms and video hosting services, where the integration of LSTM based deepfake detection can contribute to a safer and more secure online environment.

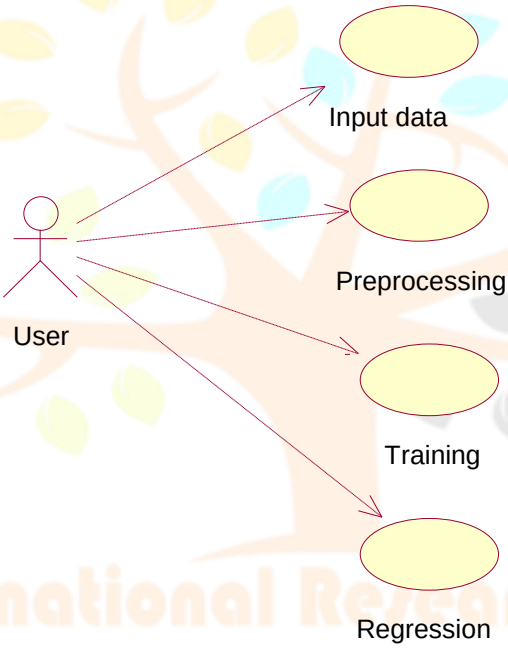


FIG 1 - USE CASE DIAGRAM

III. SYSTEM IMPLEMENTATION

3.1 DATA MINING

Collect deepfake videos created using different voice manipulation.This ensures the model is exposed to a variety of manipulation methods.Data Pre-processing Emphasis on the significance of high- quality training data. The role of data pre-processing techniques to ensure the LSTM model is trained on a diverse and representative dataset.

3.2 PRE PROCESSING

Pre-processing can be used to boost system performance before the feature extraction process. A variety of processes are involved in pre-processing a picture, such as face alignment and identification, lighting, posture, occlusion, and data augmentation correction. Images that are realistic include different sizes, stances, zooms, lighting, noise, etc. The Data Augmentation approach is employed to make the network resilient to these often-occurring effects. The network will experience these effects during training by rotating input pictures at various angles, flipping images along various axes, translating/cropping or padding the images and to simplify the complicated pixel values.

3.3 FEATURE FUSION

After the image data has been pre-processed ,it Combine the spatial features extracted by the CNN from individual frames with the temporal features learned by the RNN. This fusion of spatial and temporal information enhances the model's ability to detect subtle patterns indicative of deepfake manipulation.

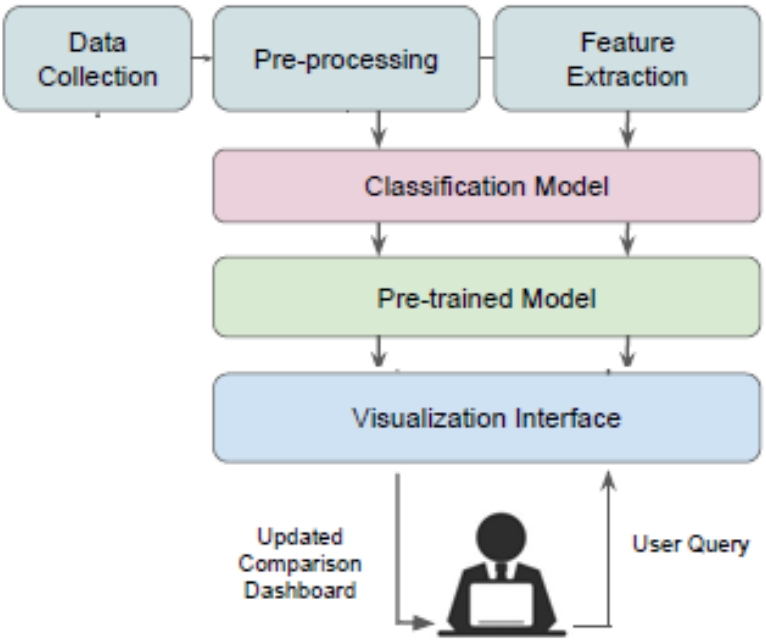
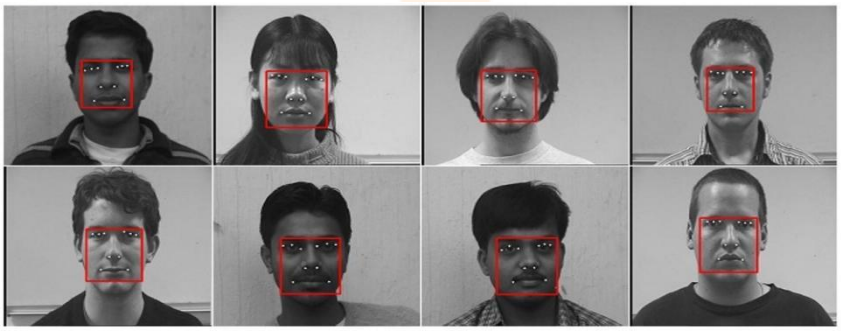


FIG 2 - SYSTEM ARCHITECTURE

IV. RESULTS AND DISCUSSION

4.1 EVALUATION METRICES

Key evaluation metrics such as accuracy, precision, recall and F1 score. It is tailored for assessing the efficacy of LSTM-based deepfake prediction. Evaluation metrics are quantitative measures used to assess the performance and effectiveness of a statistical or machine learning model. This evaluation provide insights into how well the model is performing and help in comparing different models or algorithms.



D

FIG 3 - FEATURE DETECTION

FIG 4 - FEATURE EXTRACTION

CONCLUSIONS

There are numerous applications in various fields such as security, entertainment, forensics, journalism, healthcare, sports, marketing and interactive platforms. This project broadens the forensics department's perspectives and ideas. Facial Manipulations such as Identity Swap, Expression Swap, Attribute Manipulation and Entire Face Synthesis in security systems can be recognized. The future

scope for deep fake face detection involves a multidisciplinary approach, incorporating advancements in computer vision, machine learning, and psychology to develop more robust, accurate, and reliable detection techniques. Deep Fakes present a number of obstacles, and developing effective detection and mitigation techniques will need continued research collaboration and cooperation between academia, business, and policymakers.

REFERENCE

1) DeepFake Detection for Human Face Images and Videos –Asad Malik, Minoru Kuribayashi, Sani M. Abdullahi, Ahmad Neyaz Khan IEEE February 22, 2023.

2) Zhiqing Guo, Gaobo Yang, Jiyou Chen, Xingming Sun (2021) "Fake face detection via adaptive manipulation traces extraction network" in Computer Vision and Image Understanding-Volume 204.

3) Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A Efros. CNN generated images are surprisingly easy to spot...for now. In IEEE Conference on Computer Vision and Pattern Recognition, 2020.

4) Carlini, N. and Farid, H. (2020) "Evading deep fake-image detectors with white-and black box attacks" in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.

5) Richard Zhang et al., (2018), "Making Convolutional Neural Networks Shift Invariant Again" in ICML 2019.

6) Zhou, P., Han, X., Morariu, V. I., and Davis, L. S. Learning rich features for image manipulation detection Conference on Computer Vision (ICCV), 2018.