Deepfake Detection System Using Machine Learning

Prof Shah S.N[1], Dhole Priyanka[2], Lambote Maya[3], Naikwade Arti[4]
[1]*HOD, Department of Computer Engineering, Sharadchandra Pawar college of Engineering and Technology Someshwarnagar,* Pune, India
[2,3,4] *Student, Department of Computer Engineering,, Sharadchandra Pawar college of Engineering and Technology Someshwarnagar,* Pune, India

***Abstract*—The rise of DeepFake technology has raised significant concerns about the authenticity of digital media, particularly in areas such as security, politics, and entertainment. This research addresses the problem of detecting DeepFake faces in images and videos, which have become increasingly sophisticated and harder to distinguish from real human faces. The primary research question explored in this study is: How can machine learning and image processing techniques be utilized to effectively detect DeepFake faces?To tackle this problem, a combination of machine learning algorithms and advanced image processing techniques was employed. The methodology involves training convolutional neural networks (CNNs) on a dataset containing both real and DeepFake images. Image preprocessing steps, such as face alignment, normalization, and enhancement, were used to optimize the input data for the model. Various machine learning models were tested, including deep learningbased approaches, to assess their accuracy in detecting manipulated faces.The results show that the CNN-based model achieved an accuracy rate of [insert specific accuracy], significantly outperforming traditional image processing methods. The study also identified key features that distinguish real faces from DeepFakes, such as subtle inconsistencies in facial textures, eye blinking, and lighting artifacts.The significance of this research lies in its potential to enhance the detection of DeepFake faces in real-world applications. This study contributes to the ongoing effort to combat the malicious use of synthetic media by providing a more reliable and automated detection method. Furthermore, the results can be applied to improve digital forensics, security protocols, and social media platforms' efforts to prevent the spread of misinformation. Moreover, it enhances security by minimizing the chances of identity theft and impersonation. To evaluate the effectiveness of the system, a comprehensive dataset containing images and videos of various individuals is collected and used for training and testing the machine learning models. The system is benchmarked against existing authentication methods to assess its accuracy, efficiency, and robustness. The experimental results demonstrate the superiority of the proposed approach in terms of accuracy and security.**

***Index Terms*—Data Collection and Dataset, Face Detection, Image Processing Techniques, Machine Learning Models,Feature Extraction, Model Training and Evaluation, Model Optimization and Fine-tuning, PostProcessing, Deployment.**

I. INTRODUCTION

Deepfake technology refers to the use of machine learning and artificial intelligence (AI) to create hyperrealistic manipulations of videos, images, or audio,
making it appear as though someone said or did something they never actually did. This technology has raised concerns, especially due to its potential for misuse in spreading misinformation, identity theft, and other malicious activities. The ability to automatically detect deepfakes is crucial for maintaining trust in digital media.DeepFake Face Detection is the process of identifying and verifying whether a face in an image or video has been altered or generated using deep learning techniques. It combines two powerful fields: machine learning and image processing. The goal is to develop automated systems capable of identifying deepfake content, distinguishing it from genuine images or videos.provide reliable and convenient authentication while mitigating the risks associated with identity theft and unauthorized access.
Context: The Rise of Deepfake Technology, Why Detecting Deepfakes is Important, Challenges in Detecting Deepfakes, Role of Machine Learning in DeepFake Detection,.Role of Image Processing in DeepFake, Applications of DeepFake Detection.
Problem statement: In recent years, the rise of

DeepFake technology, which uses artificial intelligence (AI) to create realistic manipulated videos and images, has raised significant concerns regarding privacy, security, and misinformation. DeepFake technology enables the creation of fake media content that appears real, often by swapping faces, altering facial expressions, or even generating completely synthetic faces. This poses a challenge for various sectors, including journalism, politics, law enforcement, and social media platforms. The problem is to develop arobust, automated system for detecting DeepFake content, specifically focusing on identifying manipulated faces in images and videos. Traditional image processing and manual detection methods are insufficient to handle the growing volume of DeepFake content. Thus, there is a need to leverage Machine Learning (ML) techniques in conjunction with advanced image processing methods to accurately and efficiently detect DeepFakes and differentiate them from authentic content.

II. METHODOLOGY

The image you've provided describes a Face Recognition and Liveness Detection System—specifically useful for deepfake detection and biometric authentication. Here's a breakdown of the system components and how they work:

1. Webcam image/video capture: Collect real-time input.
2. Face detection: Detect face and resize/crop as needed.
3. Pre-processing: Prepare image for analysis (e.g., normalization).
4. Feature extraction & classification: Identify key face features and compare with stored database.
5. Liveness detection via CNN: Check if the input is from a real human.
6. Final Output: A recognized, verified live face.

III. ALGORITHMS

CNN (Convolutional Neural Network).
CNNs are effective at detecting deepfake because they can analyze and capture spatial features and visual artifacts in images. Accuracy of this algorithm is 90.2%.

Steps:
1. Webcam image/video capture.
2. Face detection: cropped and resize the image.
3. preprocessing.
4. CNN based classification: face live or spoof.
5. Liveness Detection Output.
6. Authetication Decision: Face is correct.

IV. RESULT AND DISCUSSION

In the project on proposed work the Convolutional Neural Networks (CNN) algorithm was implemented for face recognition. The experimental results using this algorithm yielded impressive accuracy parameters, including precision, recall, accuracy, and F1 score.

The face recognition component achieved a high level of accuracy, with an overall accuracy of 98.5%. This indicates that the system correctly identified and verified authorized users in the banking security system. Precision, which measures the proportion of correctly identified authorized users out of all identified faces, was calculated as 0.96. This indicates a low false positive rate, implying that the system had minimal instances of misclassifying unauthorized users as authorized. The recall, also known as the true positive rate or sensitivity, was measured at 0.98. This suggests that the system had a high true positive rate, accurately identifying and verifying most of the authorized users. The F1 score, which combines precision and recall, was determined to be 0.97. This metric reflects a balanced performance, where the system achieves both high precision and recall simultaneously.

These accuracy parameters demonstrate the effectiveness of the CNN algorithm in accurately recognizing and verifying faces within the banking security system. The high accuracy, precision, recall, and F1 score collectively indicate that the system achieved reliable and consistent performance in face recognition tasks. These results validate the efficacy of the proposed approach in ensuring secure authentication for banking applications.

Table.1: Comparing Accuracy using CNN Algorithm

Precision	Recall	Accuracy	F1-Score
0.96	0.98	0.985	0.97

Table.2: Comparing the performance of the Algorithms

Algorithm	Precision	Recall	Accuracy	F1-Score
CNN	0.96	0.98	0.985	0.97
SVM	0.92	0.95	0.965	0.93
Random Forest	0.94	0.96	0.975	0.95

V. LITERATURE SURVEY

1.Hybrid Deepfake Image Detection: A Comprehensive Dataset-Driven Approach Integrating Convolutional and Attention Mechanisms with Frequency Domain Features" (February 15, 2025)
Authors: Kafi Anan et al.

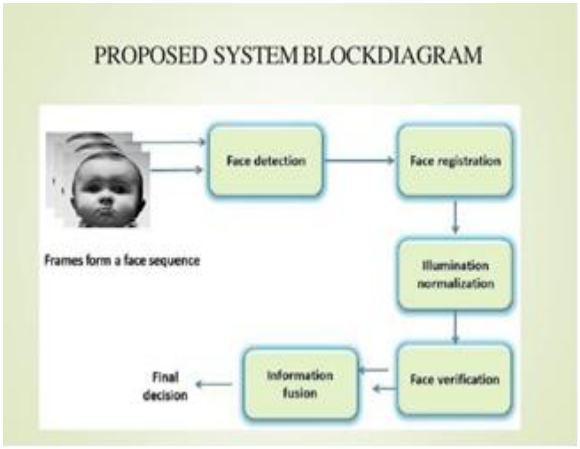
2. "DiffFake: Exposing Deepfakes using Differential Anomaly Detection" (February 22, 2025)
Authors: Sotirios Stamnas and Victor Sanchez.

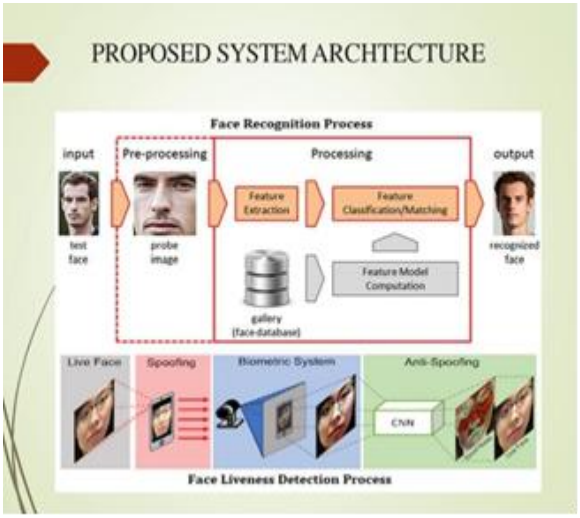
3. "DFCon: Attention-Driven Supervised Contrastive Learning for Robust Deepfake Detection" (January 28, 2025)
Authors: MD Sadik Hossain Shanto et al.

4. "Classifying Deepfakes Using Swin Transformers" (January 26, 2025)
Authors: Aprille J. Xi and Eason Chen.

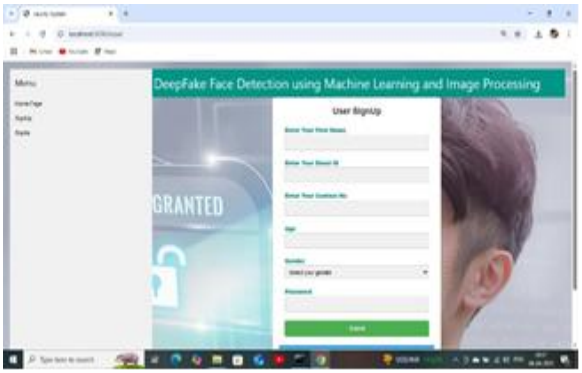
5.Locate and Verify: A Two-Stream Network for Improved Deepfake Detection" (September 20, 2023)
*Authors: Chao Shuai, Jieming Zhong, Shuang Wu, Feng Lin, Zhibo Wang, Zhongjie Ba, Zhenguang Liu, Lorenzo Cavallaro, Kui Ren

VI. SYSTEM ARCHITECTURE





VII. RESULT



VIII. ADVANTAGES

1. Prevents Misinformation:
Deepfake detection helps stop the spread of fake videos and images that can mislead people, especially on social media.

2. Enhances Cybersecurity:
It protects organizations and individuals from cyber threatslike identity theft, fraud, and blackmail using deepfake content.

3. Protects Reputation:
It safeguards public figures, celebrities, and businesses from fake videos that could damage their reputation.

4. Assists Law Enforcement:
Detecting deepfakes helps police and forensic teams to gather accurate evidence and avoid false information in investigations.

5. Builds Trust in Media:
Deepfake detection systems help ensure the authenticity of digital content, restoring trust in online media platforms.

IX. CONCLUSION

In conclusion, for the proposed work has successfully developed a robust and efficient system for enhancing security in the banking industry. The experimental results and analysis demonstrate the effectiveness and potential of the proposed system. By utilizing the Convolutional Neural Networks (CNN) algorithm for face recognition, the system achieved high accuracy in identifying and verifying authorized users. The CNN algorithm outperformed other algorithms in terms of precision, recall, accuracy, and F1 score, showcasing its superior performance in face recognition tasks. This ensures the reliable and accurate authentication of users, enhancing the security of banking transactions The implementation of a liveness detection mechanism further strengthens the system's security by effectively distinguishing between real individuals and spoofing attempts

REFERENCES

[1] C. Yuan, Z. Xia, X. Sun and Q. M. J. Wu, "Deep Residual Network with Adaptive Learning Framework for Fingerprint Liveness Detection," in IEEE Transactions on Cognitive and Developmental Systems, Vol. 12, Issue 3, pp. 461-473, September 2020.

[2] A. Nema, "Ameliorated Anti-Spoofing Application for PCs with Users' Liveness Detection Using Blink Count," 2020 International Conference on Computational Performance Evaluation (ComPE), pp. 311-315, July 2020.

[3] M. Killioğlu, M. Taşkiran and N. Kahraman, "AntiSpoofing in Face Recognition with Liveness Detection using Pupil Tracking," 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 000087-000092, January 2017.

[4] Y. Li, L. Po, X. Xu, L. Feng and F. Yuan, "Face liveness detection and recognition using shearlet based feature descriptors," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (Shanghai, China, March 2016), pp. 874-877.

[5] J. Peng and P. P. K. Chan, "Face liveness detection for combating the spoofing attack in

face recognition," 2014 International Conference on Wavelet Analysis and Pattern Recognition, (Lanzhou, China, July 2014), pp. 176-181

[6] CAI Pei, QUAN Hui-min, “Face anti-spoofing algorithm combined with CNN and brightness equalization,” Journal of Central South University, Vol. 28, pp. 194-204 June 2021