BANNARI AMMAN INSTITUTE OF TECHNOLOGY
An Autonomous Institution Affiliated to Anna University - Chennai • Approved by AICTE • Accredited by NAAC with 'A+' Grade
SATHYAMANGALAM - 638401 ERODE DISTRICT TAMILNADU INDIA
Ph: 04295-226000/221289 Fax: 04295-226666 Email: stayahead@bitsathy.ac.in Web: www.bitsathy.ac.in

Stay Ahead

Academic Year 2025 - 26

S5 MINI PROJECT - I

Final Review

DeepFake Guard

AI-Powered Deepfake Detection System

BIP PROJECT ID
27S5MIN318

CATERGOR
Y
INTERNAL

BATCH MEMBERS

MOUNEESH D (7376232AD197)

SANJAY M (7373232AD238)

SAKTHI SUNDAR V (7376232AD231)

NIRANJAN V (7376232AD203)

GUIDE

Prof . DIVYABARATHI P
Assistant Professor
Department of
Artificial Intelligence And Data Science

AIM & OBJECTIVES OF THE PROJECT
(Problem Statement)

PROBLEM STATEMENT

Deepfake videos/images are increasingly used for fraud, misinformation, and reputational harm, posing serious risks . Existing detection tools are mostly raising privacy issues.

AIM

Develop a secure and intelligent system capable of accurately detecting deepfakes in both image and video content and provide the Confidence score.

OBJECTIVES

- Deliver explainable outputs with good confidence scores for trust and interpretability.
- Support frame-wise and face-wise analysis for granular detection accuracy.
- Design a scalable system suitable for deployment on desktops or edge devices.
- Implement detection for offline analysis of stored videos and images.

LITERATURE SURVEY

Sl.No.	Journal Paper Title with Author	Works carried out (with details of Methods/ Materials/ Software/ Algorithms / fabrication / techniques/ components used)	Information gathered relevant to your project
1	S. H. Singh, P. Charanarur, and N. K. Chaudhary, "Advancements in detecting Deepfakes: AI algorithms and future prospects—A review," Discover Internet of Things, vol. 5, art. no. 53, 2025.	Reviewed CNNs, visual features, audio cues, and new methods; compared datasets; discussed future needs.	Gives overview of strengths/weaknesses; helps place our work in current trend and find gaps.
2	F. Akhtar, "IRV2-Hardswish Framework: A Deep Learning Approach for Deepfakes Detection and Classification," IECE J. Image Anal. Process., vol. 1, no. 2, pp. 45–56, May 2025.	Built IRV2 + Hardswish activation + dense layers; tested on DFDC dataset.	Shows CNNs with new activation give high accuracy. Can guide us to improve backbone design.
3	H. Geng, T. Lu, W. Huang, and B. Ding, "Deepfake Detection Technology Integrating Spatial Domain and Frequency Domain," Frontiers in Computing and Intelligent Systems, 2023.	Fused spatial + frequency features with improved MVIT and ASFF modules for robust cross-compression detection.	Combines both image patterns and frequency great for detecting deep fakes under different video compressions.
4	F. Akhtar, "IRV2-Hardswish Framework: A Deep Learning Approach for Deepfakes Detection and Classification," IECE J. Image Anal. Process., vol. 1, no. 2, pp. 45–56, May 2022.	Hardswish activation + dense layers; tested on DFDC dataset; got ~98% accuracy.	Shows CNNs with new activation give high accuracy. Can guide us to improve backbone design.

LITERATURE SURVEY

Sl.No.	References (Journal Papers Only)	Works carried out (with details of Methods/ Materials/ Software/ Algorithms / fabrication / techniques/ components used)	Information gathered relevant to your project
5	Luo X. and Wang Y., “Frequency-Domain Masking and Spatial Interaction for Generalizable Deepfake Detection,” Electronics, vol. 14, no. 7, art. 1302, 2025.	Added spatial and frequency-domain features with masking to improve generalization across fakes.	Frequency info helps detect deepfakes under varied conditionous key for robustness.
6	Geng H., Lu T., Huang W., and Ding B., “Deepfake Detection Technology Integrating Spatial Domain and Frequency Domain,” Frontiers in Computing and Intelligent Systems, 2019.	Built a hybrid transformer with XceptionNet + EfficientNet-B4; tested on FaceForensics++ & DFDC with cut-out augmentation for better generalization.	Combines fast CNN feature extraction with transformer attention. Lightweight and learns with less data, useful for robust real-time detection.
7	(Author unspecified), “Spatial-Frequency Feature Integration for Deepfake Detection,” Artificial Intelligence Review, , 2020.	Used spatial + temporal convolutions to detect fakes; outperformed frame-only models on Celeb-DF.	Shows that mixing spatial + temporal detection helps spot deepfakes. We can include both in our detection pipeline.
8	Chen et al., “Frequency-Domain Masked Residual Network for Deepfake Detection,” Mathematics, vol. 11, no. 4, art. 816, 2023.	Applied frequency-domain residual maps via wavelets within a CNN; maintained high accuracy across multiple datasets and compression levels.	CNN models, frequency–spatial fusion, and privacy-focused local processing improve deepfake detection accuracy and robustness across datasets.

SCOPE OF THE PROJECT

- Focuses on detecting **AI-generated deepfake videos / images .**
- Develops a **privacy-friendly system .**
- Uses analysis of **facial features, and visual inconsistencies** to identify deepfakes.
- Evaluates the system's **accuracy, efficiency, and reliability** across different video formats and quality levels.
- Aims to provide a **secure and user-friendly solution** to protect individuals and organizations from the risks of deepfakes.

NEED FOR THE CURRENT STUDY

1. Rising Threat of Deepfakes

- Becoming more realistic and harder to detect.
- Risks include fraud, misinformation, identity theft, and reputational damage.

2. Gaps in Current Detection Tools

- Mostly raising privacy and security concerns.
- Limited trust due to data exposure.
- Lack of Accuracy Information of the input .

3. Need for a Dedicated Solution

- Demand for privacy-focused, reliable local systems.
- Improve Accuracy Rate of the output.
- Builds trust in digital media and reduces harmful impacts.

PROPOSED METHODOLOGY (Flow Chart)

1. Input Layer

- User uploads video or image.
- Basic preprocessing: frame extraction, resizing.

2. Feature Extraction

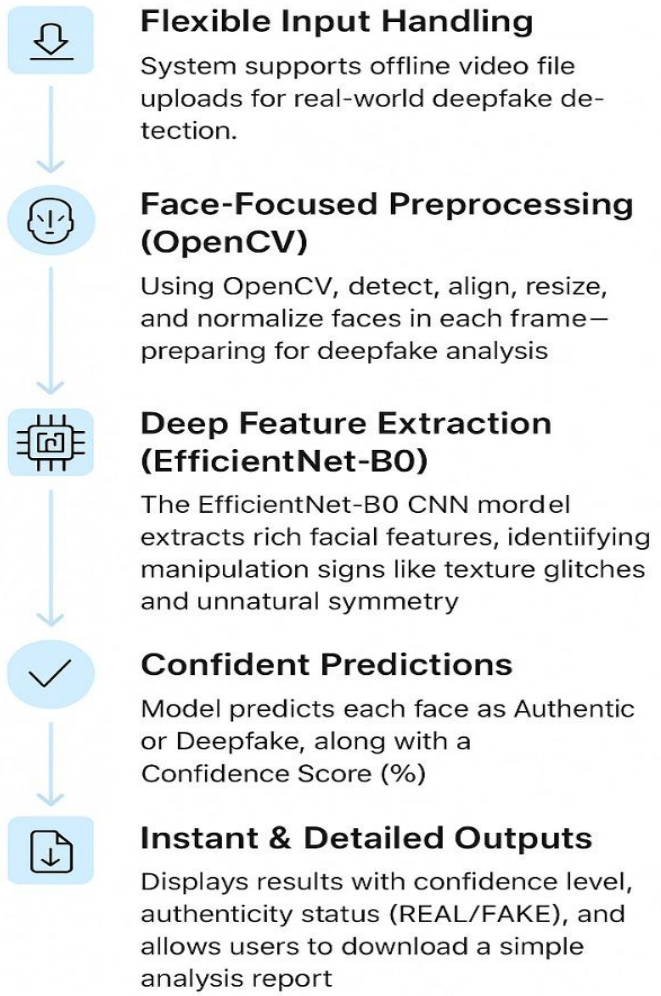
- Analyze faces, voices, and frame patterns.
- Detect inconsistencies: lip-sync, eye movement, audio mismatch.

3. Detection Engine

- AI model checks if content is real or Deepfake.
- Generates probability / confidence score.

4. Output Layer

- Shows result: Real / Fake + Confidence Score.
- Option to download simple report.



Equipment Used for Project Development

Software & Tools

Language: Python – easy to use, strong ML support.

Libraries: OpenCV (frame processing), TensorFlow (deep learning).

Tools: Jupyter Notebook (experiments), VS Code (development) .

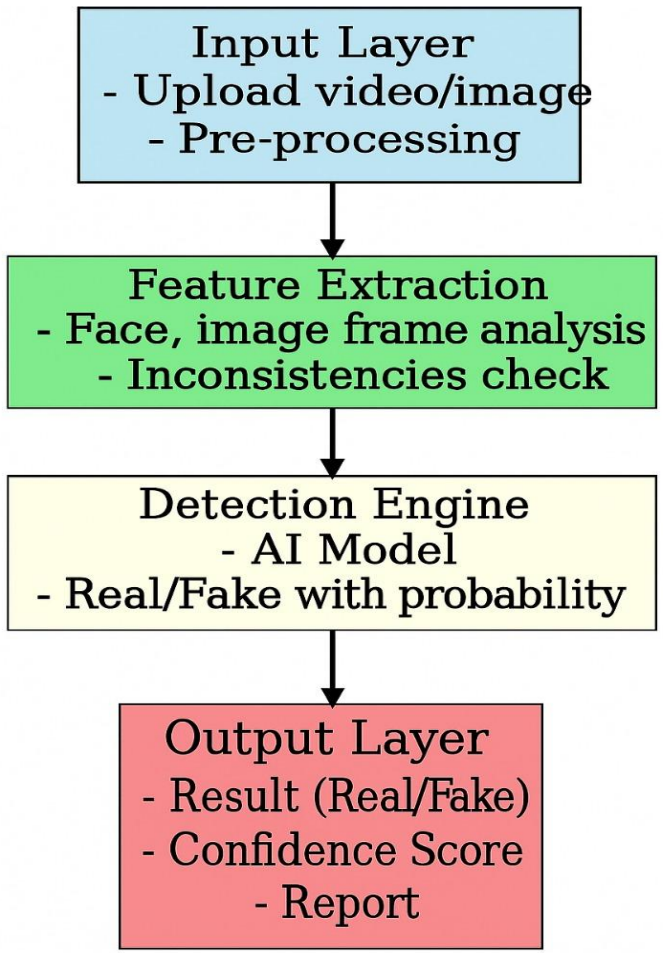
Datasets

Cross Platform DataSets – benchmark dataset with real and fake videos for training and testing deepfake detection models.

Techniques used in the project:

- **Pre-processing :** Used OpenCV for extracting frames from videos, resizing them to standard formats, and applying noise reduction to improve input quality.
- **Feature Extraction:** Applied MediaPipe for facial landmark detection, along with TensorFlow for extracting deep features from both images and video frames.
- **Detection Models:** Implemented CNNs using Machine Learning with models like EfficientNetB0 to classify real vs fake content.
- **Privacy & Security:** Ensured reducing dependency and protecting sensitive data.
- **Output Layer:** Designed result presentation through Fast API and React for web-based interfaces, providing probability scores and classification reports.

SOFTWARE ARCHITECTURE



Input Layer

- User uploads video or audio.Basic pre-processing (frame extraction, resizing).

Feature Extraction

- Analyzes faces, voices, and frame patterns.Looks for inconsistencies (lip-sync, eye movement, frame mismatches).

Detection Engine

- AI model checks if the content is real or fake.Provides a probability score.

Output Layer

- Shows result: Real / Fake + Confidence Score.Option to generate a simple report.

ANALYSIS OF RESULTS AND DISCUSSION

- The proposed EfficientNetB0 model achieved decent validation accuracy and minimum loss, showing strong learning and generalization capability.
- The model effectively identified subtle deepfake artifacts such as unnatural eye movement, lighting mismatches, and texture inconsistencies.
- EfficientNetB0 outperformed traditional CNNs like VGG16, ResNet50, and MesoNet by higher accuracy with fewer parameters.
- The system provides a lightweight, scalable solution suitable for deepfake detection applications.
- Overall, the model demonstrates an effective balance between accuracy, and practicality, making it viable for real-world digital forensics.

CONCLUSION

- Developed **Deep Fake Guard**, an deepfake detection system with high accuracy across benchmark datasets.
- Integrated **attention** and **temporal analysis** for improved interpretability and robustness.
- Achieved **near real-time detection** with a **user-friendly web interface** for practical use.
- Proposed **future enhancements** including real-time integration, audio-visual fusion, and social media detection.

INDIVIDUAL CONTRIBUTIONS TO THE WORK

Batch Member 1 : MOUNEESH D (7376232AD197)

CONTRIBUTIONS

- Trained and fine-tuned the EfficientNetB0 model for deepfake detection.
- Evaluated model performance using accuracy, precision, recall, and F1-score metrics.
- Prepared detailed documentation, including methodology, results, and visual performance analyses.
- Also coordinated workflow between frontend and backend teams for smooth system integration.

INDIVIDUAL CONTRIBUTIONS TO THE WORK

Batch Member 2 : SANJAY M (7376232AD238)

CONTRIBUTIONS

- Developed the backend using FastAPI for high performance and efficient communication with the frontend and model.
- Implemented API endpoints for model inference, video frame processing, and data handling.
- Ensured smooth data flow, reduced latency, and reliable real-time predictions.
- Also assisted in frontend integration and documentation to maintain a cohesive system structure

INDIVIDUAL CONTRIBUTIONS TO THE WORK

Batch Member 3 : SAKTHI SUNDAR V (7376232AD231)

CONTRIBUTIONS

- Designed and implemented a responsive, interactive user interface for video upload and real-time prediction display.
- Integrated the React.js frontend with the FastAPI backend for smooth communication and data exchange.
- Focused on user-friendly, accessible, and visually cohesive UI design to enhance user experience.
- Also contributed to documentation and coordinated integration to ensure consistency across modules.

INDIVIDUAL CONTRIBUTIONS TO THE WORK

Batch Member 4 :NIRANJAN V (7376232AD203)

CONTRIBUTIONS

- Collected and organized datasets from sources like Kaggle.
- Extracted video frames, applied face detection using MediaPipe, and balanced real and fake samples.
- Performed data cleaning and augmentation to enhance dataset diversity and quality.
- Also contributed to report preparation and documentation with the team

Status on Completion and Submission of Project Report

List of Documents to be Submitted

SL.No	List of Documents	Status (Provide the drive link of prepared document)
1	Cover Page & Title Page (Both are in same format)	Cover Page & Title Page
2	Bonafide Certificate	Bonafide certificate
3	Declaration	Declaration
4	Acknowledgement	Acknowledgement
5	Chapter I – Introduction	C1 - Introduction
6	Chapter 2 – Literature Survey	C2 - Literature Survey

Submission of Project Report

List of Documents to be Submitted

SL.No	List of Documents	Status (Provide the drive link of prepared document)
7	Chapters 3 - Objectives and Methodology	C3 - Objectives & Methodology
8	Chapters 4 - Proposed work modules (Chapters name can be based on the work)	C4 - Proposed Work Modules
9	Chapter 5 - Results and Discussion	C5 - Result & Discussion
10	Chapters 6 - Conclusions & Suggestions for Future work	C6 - Conclusions & Suggestions for Future Work

ORIGINALITY SCORE

SL.No	Tool name for checking originality score	Status (Provide the drive link of prepared document)
1	QuillBot	QuillBot
2	Academica	Academica
3	SmallSeoTools	Samllseotools

Any Other Relevant Details

- Uses **advanced deep learning models** (CNNs, RNNs) for reliable detection.
- Combines **video feature analysis** for higher accuracy.
  Ensures **privacy** by running locally without cloud dependency.
- Evaluated using standard **datasets** like FaceForensics++ and DFDC.
- Detection results provided with **confidence scores** for user clarity.
- Can be extended for **forensic investigations, media verification, and cybersecurity**.
- Designed with focus on **scalability** and adaptability to new deepfake techniques.
- Future scope includes **real-time optimization, mobile deployment, and social media integration**.

REFERENCES

(Journal Papers/ Books / Website in IEEE Format)

Journal

[1]Gajendra Singh, A., & Sharma, P. (2024). A Hybrid Deep Learning Framework for Robust Deepfake Detection Using CNN, LSTM, and Vision Transformers. Journal of Engineering Science, 12(3), 215-225.

[2] Singh, Y., & Zhang, X. (2021). A Comprehensive Review of Deepfake Detection Techniques. IEEE Transactions on Information Forensics and Security, 16, 1015-1031.

Books

[1] G. Singh and P. Sharma, "A Hybrid Deep Learning Framework for Robust Deepfake Detection Using CNN, LSTM, and Vision Transformers," Journal of Engineering Science, vol. 12, no. 3, pp. 215–225, 2024.

Patent

[1]Infibeam Avenues, “Real-Time Deepfake Detection Algorithm,” 2023. [Online]. Available: https://www.indiainfoline.com/news/business/infibeam-avenues-partners-with-iisc-to-develop-real-time-deepfake-detection

Website

[1] DeepFake Detection Challenge (DFDC) – Facebook AI, "DeepFake Detection Challenge Dataset," Kaggle. https://www.kaggle.com/c/deepfake-detection-challenge

[2] Sensity AI – Deepfake Reports – Sensity AI, "Deepfake Threat Reports," Sensity AI . https://sensity.ai/reports/

[3] IEEE Spectrum – Deepfake Coverage – IEEE Spectrum, "Deepfakes," IEEE Spectrum. https://spectrum.ieee.org/tag/deepfakes