iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78691

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

VeriSight Sentinel: Deepfake Detection and Reporting System

Sakshi Yadav[1], Vaishnavi Misal[2], Princy Singh[3], Jitesh Bhoir[4], Prof. Vijayalaxmi Tadkal[5]
Department of Computer Science Engineering (AIML), Bharat College of Engineering, Mumbai University, Badlapur, India

Abstract: ***This project focuses on Deepfake Detection and Reporting, addressing the growing concern of synthetic media manipulation using artificial intelligence. Deepfakes, which use deep learning techniques to alter or generate realistic audio, video, or images, pose serious threats to privacy, misinformation, and digital trust. The proposed system employs Convolutional Neural Networks (CNN) and Machine Learning algorithms to detect inconsistencies in facial expressions, blinking patterns, and pixel-level artifacts that are often present in deepfake content. A trained model analyzes input media to identify forged elements with high accuracy. Once detected, the system can automatically generate a detailed report highlighting the probability of manipulation, affected regions, and potential sources. This tool can be integrated into social media platforms, news verification systems, and digital forensics. By combining detection and real-time reporting, the project aims to strengthen digital media integrity, promote responsible AI usage, and empower users to identify and report deepfakes effectively and efficiently.***

Keywords: Deepfake Detection, Synthetic Media Manipulation, Artificial Intelligence (AI), Deep Learning, Convolutional Neural Networks (CNN), Machine Learning Algorithms, Facial Expression Analysis, Digital Forensics, Misinformation Prevention, Privacy Protection, Digital Media Integrity, Real-Time Reporting System, Social Media Verification, AI-based Media Analysis

I. INTRODUCTION

In the digital era, the emergence of Artificial Intelligence (AI) has revolutionized how humans create, share, and consume content. Among its many applications, one of the most controversial and rapidly evolving developments is Deepfake technology. A deepfake refers to synthetic media — whether video, audio, image, or text — that has been digitally altered or generated using deep learning techniques to make it appear authentic. These manipulations are powered primarily by Generative Adversarial Networks (GANs) and autoencoders, which learn to mimic human-like features, voices, and behaviours with exceptional accuracy.

While the use of deepfake technology has promising applications in entertainment, education, and accessibility, it has also introduced severe ethical, legal, and social challenges. Malicious use of deepfakes has become increasingly prevalent, from spreading misinformation and defaming individuals to committing financial fraud and manipulating public opinion. As deepfakes become more sophisticated, distinguishing real from fake media is becoming extremely difficult for humans and even for traditional verification systems. This has led to an urgent demand for automated, AI-powered detection systems capable of identifying and flagging manipulated content. The project "VeriSight Sentinel" addresses this critical challenge by developing a Deepfake Detection and Reporting System that utilizes Machine Learning (ML), Deep Learning (DL), and Generative AI (GenAI) to identify forgeries across different media types. The proposed system not only detects deepfakes but also provides detailed analytical reports, making it a powerful tool for maintaining digital media integrity and public trust.

II. PROBLEM DEFINITION

The rapid advancement of deepfake technology has made it increasingly difficult to distinguish between authentic and manipulated digital media. Deepfakes can be used to spread misinformation, commit fraud and damage reputations. Existing detection systems are often limited to specific media types and lack secure reporting mechanisms, highlighting the need for an integrated AI-based solution capable of accurately detecting and reporting deepfake content.

III.OBJECTIVE

The main objective of the VeriSight Sentinel – Deepfake Detection and Reporting System is to develop an AI-based platform capable of detecting manipulated digital media such as images, videos, audios and text. The system aims to identify deepfake content using advanced machine learning and deep learning algorithms while generating detailed reports that highlight manipulation probability, suspicious regions, and authenticity indicators.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

IV.SCOPE OF PROJECT

The scope of the project includes designing and implementing a secure deepfake detection system that can analyze multiple forms of media including images, videos, audio and text. The system integrates deep learning models with a web-based interface to provide real-time analysis, report generation and secure data storage, helping organizations, researchers and users verify the authenticity of the digital content.

V. LITERATURE SURVEY OF THE PROJECT

Sr. No.	Title of the Paper	Author	Methodology	Disadvantages / Future Scopes
1	FaceForensics++: Learning to Detect Manipulated Facial Images and Videos	Niessner et al.	Comprehensive dataset of manipulated facial videos; utilized CNN for detection benchmarks	Primarily focused on facial video manipulation.
2	A Survey on Face Manipulation and Deepfake Detection Techniques	Tolosana et al.	Categorized diverse facial manipulation techniques; provided a broad review of detection methodologies	Mainly emphasized face swap detection challenges.
3	DeepFake Detection Challenge	Dolhansky et al.	Large-scale video dataset; established a crucial benchmark for algorithmic deepfake detection.	Inherent dataset bias; challenges with noisy face extraction.
4	Deepfakes: Trick or Treat?	Kietzmann et al.	Examined the broader societal impact of deepfakes and the evolving detection challenges.	Lacked in-depth technical analysis of detection algorithms.
5	Practical Guide to Deepfake Detection	Paravision AI	Focused on real-world applications of AI-based approaches for facial deepfake detection.	Did not cover audio or text-based deepfakes

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

VI. PROPOSED SYSTEM

The methodology of VeriSight Sentinel involves developing a secure, authenticated and AI-driven system capable of detecting deepfakes across images, videos, audio and text. For Image detection, CNN models like ResNet and EfficientNet identify pixel-level inconsistencies. Video detection uses a CNN -LSTM hybrid model to analyze both spatial and temporal patterns. The Audio module employs spectrogram-based CNNs to detect cloned or AI-written content. The system backend is built with Flask, and the frontend with React.js, ensuring smooth interaction. All detection reports and user data are securely stored in SQLite database, accessible only to authenticated users. Additionally, Generative AI is integrated to provide explainable results and enhance user understanding.

VII. METHODOLOGY

A. Algorithms used for Image Module

Convolutional Neural Network (CNN) : Convolutional Neural Network are widely used for analyzing visual data such as images and videos frames. CNN automatically extracts important features like edges, textures and patterns using convolutional and pooling layers,. In this system, CNN identifies pixel-level inconsistencies, lighting mismatches and texture artifacts commonly found in manipulated or deepfake images.

B. Algorithms used for Video Module

Long Short-Term Memory (LSTM) : Long Short-Term Memory (LSTM) networks analyze temporal dependencies — the motion and sequence of frames over time. LSTM is a type of Recurrent Neural Network (RNN) that can remember long-term patterns, making it ideal for analyzing video data where consecutive frames are correlated. An LSTM consists of memory cells and gating mechanisms (input, forget, and output gates) that control how information flows through the network. This allows it to retain essential motion cues while ignoring irrelevant noise. In the context of deepfake detection, the LSTM observes facial expressions, head movements, and blinking sequences across multiple frames. It learns to identify unnatural temporal transitions, such as delayed lip movements or frame jittering caused by manipulation.

CNN – LSTM Hybrid Architecture : The combined CNN–LSTM model leverages the strengths of both spatial and temporal analysis. The CNN acts as the feature extractor for visual data, while the LSTM acts as the sequence analyzer that models motion patterns. Together, they detect not only per-frame artifacts but also inconsistencies across multiple frames. The final classification layer applies a Softmax function to output probabilities indicating whether the video is “real” or “fake.” This hybrid structure enables VeriSight Sentinel to operate efficiently on compressed or low-quality videos, such as those shared on social media, where manipulation artifacts are more subtle. The system is trained using benchmark datasets like FaceForensics++ and DeepFake Detection Challenge (DFDC) to ensure robustness against diverse manipulation techniques.

C. Algorithms used for Audio Module

Spectrogram-Based Convolutional Neural Network (CNN) : The Spectrogram-CNN model serves as the core algorithm for detecting deepfake audio. Since a spectrogram can be treated like an image, CNNs are highly effective at identifying unique patterns within it. The convolutional layers extract local frequency-based features such as pitch contour, timbre, and energy distribution, while pooling layers downsample the data to retain only the most relevant details. Activation functions like ReLU (Rectified Linear Unit) introduce non-linearity, enabling the model to learn complex relationships between sound frequencies.

D. Algorithms used for Text Module

Transformer – Based NLP Model (BERT) : The primary algorithm used in the text detection module is BERT (Bidirectional Encoder Representations from Transformers), developed by Google. BERT is a Transformer-based model that understands the context of words by processing text bidirectionally — meaning it reads the sentence from both left to right and right to left simultaneously. This bidirectional understanding allows BERT to capture semantic meaning and contextual relationships with high precision. - The model is fine-tuned on labeled datasets containing both human-written and AI-generated text samples. During training, BERT learns the statistical and contextual signatures of fake content. At the classification stage, it outputs a probability score indicating the likelihood that the input text is machine-generated. This probability score is then displayed to the user in the results section, along with an explainable summary provided by Generative AI.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VIII. SYSTEM ARCHITECTURE

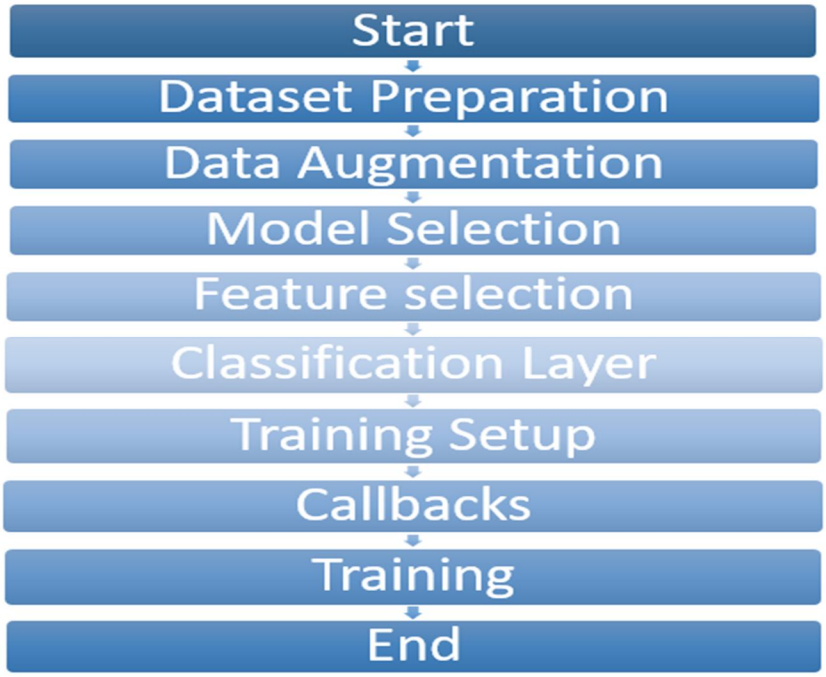


FIG.1.

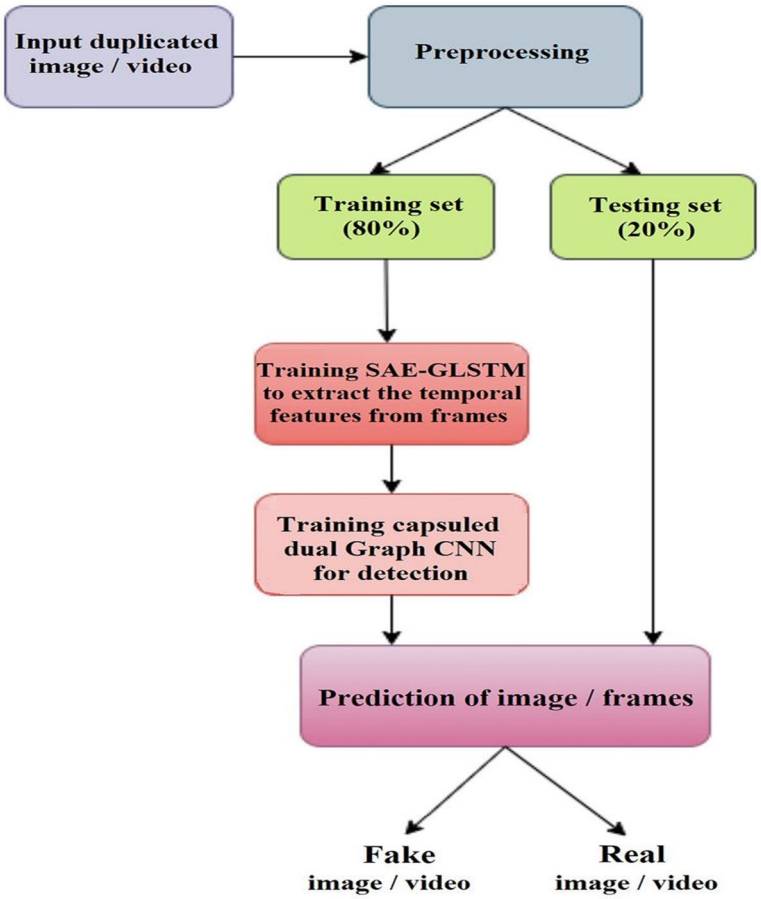


FIG.2.

IX. RESULT ANALYSIS

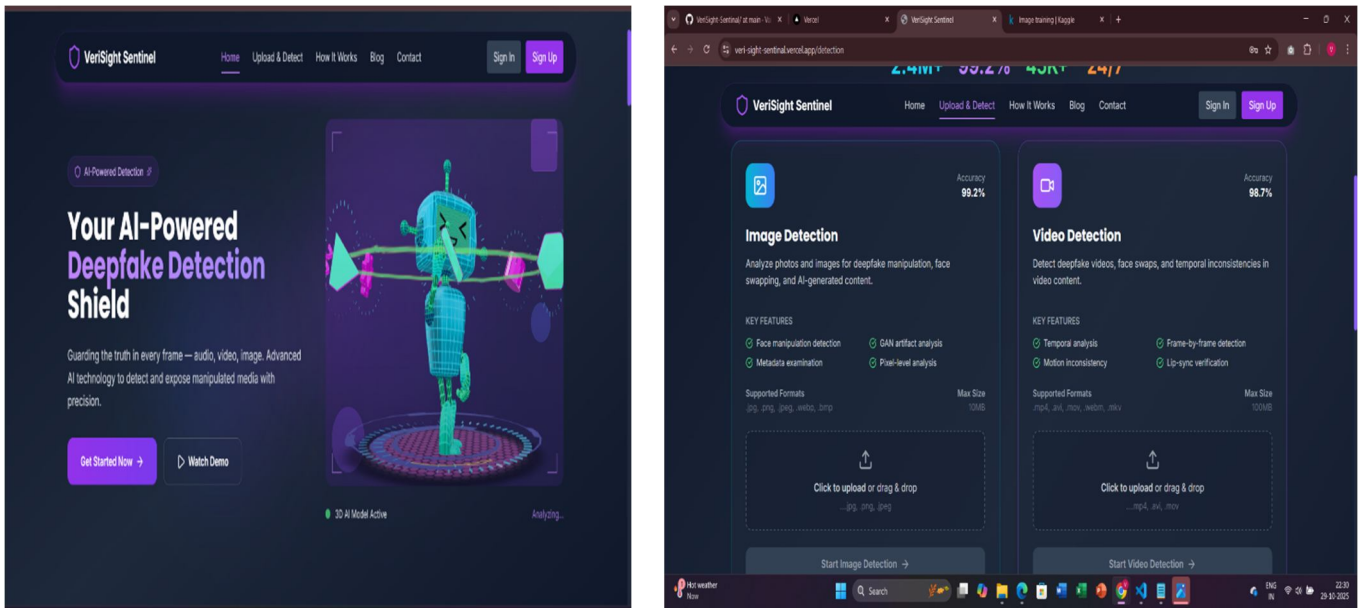


FIG.3. HOME PAGE

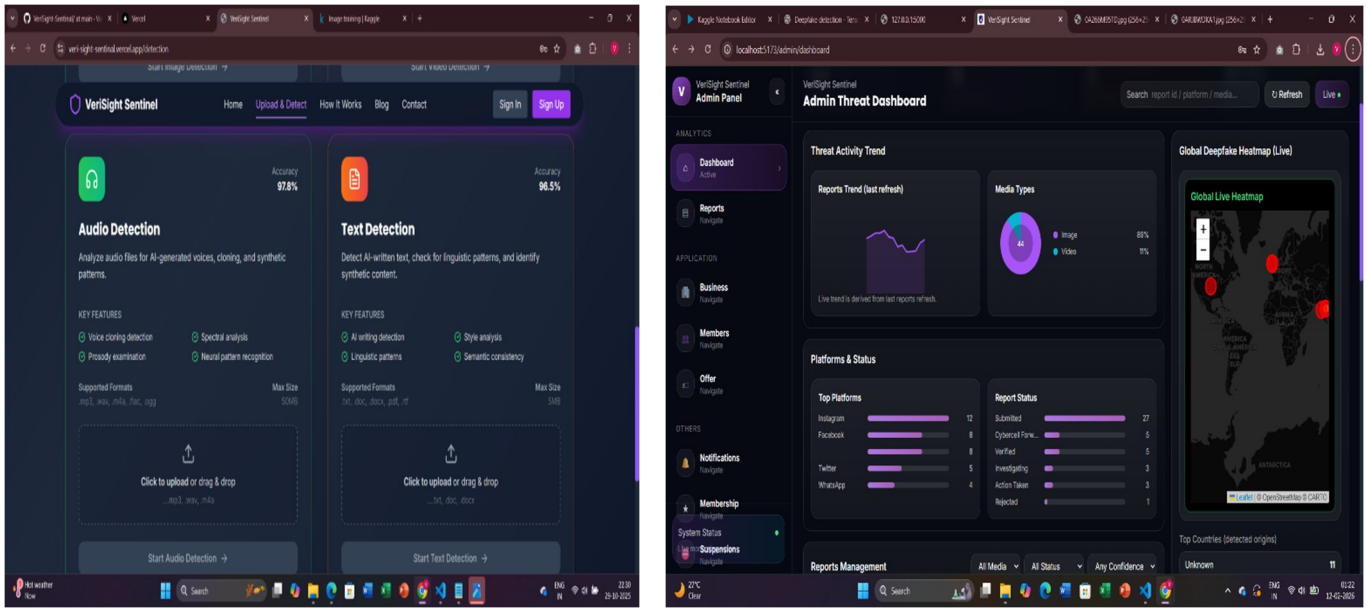


FIG.4. DASHBOARD ANALYSIS

X. APPLICATIONS

1) Media Verification – Can be used by journalists and media houses to verify the authenticity of images before publication.
2) Cybersecurity and Digital Forensics – Helps law enforcement and forensic teams detect image-based digital fraud or identity manipulation.
3) Social Media Monitoring – Can assist social media platforms in identifying and flagging deepfake or fake image content.
4) Corporate Security – Useful for companies to detect forged visual content that could harm brand reputation or spread misinformation.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

5) Educational Institutions – Can be used in academic research and awareness programs to demonstrate how deepfakes are detected.
6) Legal Evidence Verification – Supports courts and investigators in validating the authenticity of digital image evidence.
7) Government and Defense – Assists agencies in identifying fake propaganda or manipulated visual data that threaten public security.

XI. CONCLUSION

The project VeriSight Sentinel – Deepfake Detection and Reporting System aims to develop a secure, authenticated, and AI-driven platform capable of detecting deepfakes across multiple media types. Currently, the Image Deepfake Detection Module has been successfully implemented using CNN, ResNet, and EfficientNet models, achieving accurate results in identifying manipulated images. The system also ensures secure storage of results using SQLite and allows authenticated access for users. The project work is still ongoing, and further development will focus on integrating video, audio, and text detection modules to make VeriSight Sentinel a complete, multimodal deepfake detection and reporting system.

XII. ACKNOWLEDGMENT

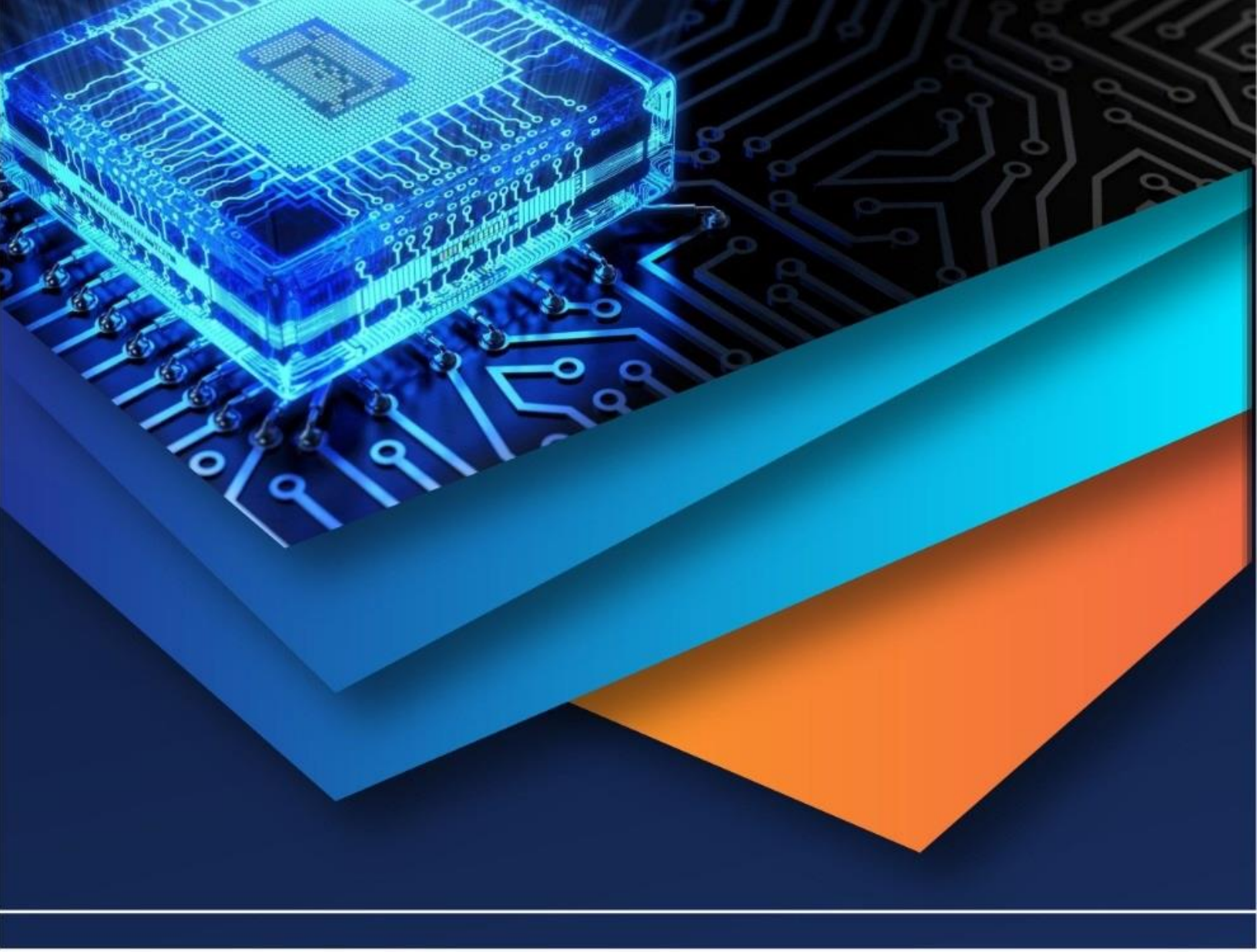
It is an immense pleasure for us to present the project report on “Verisight Sentinel: Deepfake Detection And Reporting System” expressing my heartfelt gratitude to all those who have generously offered their valuable suggestions towards the completion of this report.

It's rightly said that we are built on the shoulders of others for all our achievements. The credit goes to my guide Prof. Vijayalaxmi Tadkal Department of Computer Science & Engineering (AIML), Bharat College of Engineering, Badlapur whose positive attitude, moral support, and encouragement led to the success of the report. Her generous help, excellent guidance, lucid suggestions, and encouragement throughout the course of this work have greatly helped me in the successful completion of this work.

REFERENCES

[1] Deressa Wodajo and Solomon Atnafu. Deepfake video detection using convolutional vision transformer. arXiv preprint arXiv:2102.11126, 2021.
[2] Irene Amerini, Leonardo Galteri, Roberto Caldelli, and Alberto Del Bimbo. Deepfake video detection through optical flow based cnn. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pages 0–0, 2019.
[3] David Guera and Edward J Delp. Deepfake video detection using recurrent ¨ neural networks. In 2018 15th IEEE international conference on advanced video and signal-based surveillance (AVSS), pages 1–6. IEEE, 2018.
[4] Nicolo Bonettini, Edoardo Daniele Cannas, Sara Mandelli, Luca Bondi, Paolo ` Bestagini, and Stefano Tubaro. Video face manipulation detection through ensemble of cnns. In 2020 25th International Conference on Pattern Recognition (ICPR), pages 5012–5019. IEEE, 2021.
[5] Hany Farid. Photo forensics. MIT press, 2016.
[6] Lorant. Lincoln-picture-story-his-life. www.amazon.com, 1969.
[7] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1125–1134, 2017.
[8] Mart´ın Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), pages 265–283, 2016.
[9] F. Chollet et al. Keras. https://keras.io, 2025.
[10] Ayush Tewari, Michael Zollhofer, Hyeongwoo Kim, Pablo Garrido, Florian Bernard, Patrick Perez, and Christian Theobalt. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pages 1274–1283, 2017.
[11] Grigory Antipov, Moez Baccouche, and Jean-Luc Dugelay. Face aging with conditional generative adversarial networks. In 2017 IEEE international conference on image processing (ICIP), pages 2089–2093. IEEE, 2017.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)