ISSN 2277-2685
IJESR/Jan-Mar. 2026/ Vol-16/Issue-1/227-235
Pendota Shylika *et. al.*, /International Journal of Engineering & Science Research

Deep Fake Detection Using Deep Learning

Dr. Sakam Nagi Reddy[1], Pendota Shylika[2], Maggidi Srija[3], Mothe Pravalika[4], Pesaru Srikanth[5]
[1]Associate Professor, Department of Electronics and Communication Engineering, Teegala Krishna Reddy Engineering College, Hyderabad, India.
[2,3,4,5]B.Tech Students, Department of Electronics and Communication Engineering, Teegala Krishna Reddy Engineering College, Hyderabad, India.

nagireddysakam@tkrec.ac.in, pendotashylika05@gmail.com, srijapatel969@gmail.com, pravalikareddmothe28@gmail.com, srikanthreddypesaru@gmail.com
Article Received 22-02-2026, Accepted 14-03-2026

Abstract- *This project tackles the increasing threat posed by deepfake technology, which enables the creation of highly convincing but deceptive images and videos. Deepfakes can spread misinformation and compromise digital trust, making effective detection methods crucial. Our solution is a comprehensive deepfake detection system designed to analyze both image and video content with high accuracy. For video analysis, we utilize Gated Recurrent Units (GRUs), a type of recurrent neural network well-suited for modeling temporal sequences. GRUs capture subtle temporal patterns and inconsistencies across video frames that often indicate manipulation. For images, we employ Convolutional Neural Networks (CNNs), focusing on two well-known architectures: VGG16 and MobileNet. These CNNs excel at extracting detailed visual features, enabling reliable classification of images as either genuine or deepfakes. To make the detection accessible and user-friendly, we implement a web application where users can upload their images or videos for instant analysis and receive classification results in real time. This combination of temporal modeling with GRUs and spatial feature extraction through CNNs offers a robust approach to detecting sophisticated deepfake content. Our system is designed not only for accuracy but also for ease of use, aiming to empower individuals and organizations to verify digital media authenticity effectively. Ultimately, this project contributes toward mitigating the impact of deepfakes, helping to preserve the integrity and trustworthiness of digital media in an era where manipulation techniques are rapidly evolving.*

Keywords: GRU, CNN, VGG16, MobileNet, Fake, Real

INTRODUCTION

The rise of deepfake technology has introduced significant challenges in verifying the authenticity of digital media. Deepfakes produce highly realistic yet fabricated images and videos, which can be exploited to spread misinformation, manipulate public opinion, and undermine trust in digital content. Detecting such manipulated media accurately and efficiently is thus essential to safeguard information integrity. This project proposes a comprehensive deepfake detection system that analyzes both images and videos using advanced deep learning techniques. For videos, we utilize Gated Recurrent Units (GRUs), which are well-suited for capturing temporal dependencies and detecting inconsistencies across frames—key indicators of manipulation. For image analysis, Convolutional Neural Networks (CNNs), specifically VGG16 and MobileNet architectures, are employed to extract intricate visual features and classify content as authentic or fake. To enhance accessibility, we develop a user-friendly web application enabling real-time uploading and analysis of media, thereby providing immediate feedback on authenticity. By combining these powerful methods, the system targets multiple levels of deepfake detection, improving robustness against diverse manipulation techniques. This approach not only leverages temporal and spatial features but also incorporates modern neural network architectures that have demonstrated high effectiveness in related image and video recognition tasks.

A. Motivation

The increasing sophistication and accessibility of deepfake generation tools have created an urgent need for reliable detection mechanisms. Existing solutions often focus on either images or videos, lacking a holistic approach. By integrating GRUs and CNNs within a unified framework, this work aims to provide a robust detection system that addresses the nuances of both media types. The web application further democratizes this technology, empowering individuals and organizations to verify content easily and confidently. Moreover, the project is motivated by the growing social, political, and economic implications of misinformation fueled by deepfakes.

By providing a practical tool that can be widely adopted, we seek to curb the negative impacts of manipulated media. The adaptability of the models also allows for continuous improvement as new deepfake techniques emerge, making this solution future-proof. Our approach highlights the importance of proactive detection in maintaining public trust and supporting digital literacy in an increasingly complex media landscape.

B. Scope

This work covers data preprocessing, feature extraction, model training, and evaluation of both temporal and spatial detection components. It also involves designing a seamless user interface for practical deployment. Our system explores optimal parameter tuning for GRU and CNN models and assesses performance through standard metrics to ensure high accuracy and reliability. The scope extends to handling diverse datasets, including various types of manipulated videos and images to enhance the generalizability of the detection system. Additionally, the project incorporates real-time processing capabilities to support instantaneous verification, an essential feature for practical use cases like social media monitoring and news verification. Considerations for scalability and deployment on cloud-based platforms or edge devices are also included, enabling widespread adoption. The integration of user feedback mechanisms aims to refine model accuracy continuously, reflecting real-world conditions and evolving deepfake techniques.

C. Problem Statement

Deepfake technology poses a growing threat to digital trust by enabling convincing falsifications of images and videos. Current detection methods often lack comprehensive coverage or real-time capabilities, limiting their effectiveness. This project addresses these gaps by developing a dual-model detection system combining temporal and spatial analysis, supported by an interactive platform. The resulting solution aims to enhance the detection accuracy and usability, contributing to the broader effort of maintaining authenticity in digital media. Challenges such as subtle manipulations, adversarial attacks on detection algorithms, and the rapid evolution of deepfake methods require advanced and adaptable models. Our system is designed to overcome these challenges by leveraging the strengths of GRUs in sequence modeling and CNNs in spatial feature recognition, providing a robust defense mechanism. The user-friendly interface ensures accessibility for non-expert users, broadening the impact of the detection technology. This work ultimately aims to establish a reliable line of defense against misinformation and protect the integrity of information shared online.

D. Objectives and Contributions

The primary objective of this project is to develop, implement, and evaluate a comprehensive deepfake detection system capable of accurately identifying manipulated images and videos. By combining advanced machine learning techniques—specifically Gated Recurrent Units (GRUs) for temporal video analysis and Convolutional Neural Networks (CNNs) with VGG16 and MobileNet architectures for image analysis—the system aims to deliver high accuracy in distinguishing genuine content from deepfakes. The goal is to provide a reliable and scalable detection method that addresses the growing challenge posed by increasingly sophisticated deepfake technologies.

Our approach emphasizes both performance and accessibility, enabling real-time analysis through a user-friendly web application where users can upload media and receive instant classification results. This accessibility facilitates wider adoption by individuals, organizations, and content platforms seeking to verify media authenticity. By integrating temporal sequence modeling with spatial feature extraction, the system offers a robust defense against diverse manipulation techniques, enhancing trust in digital content. Ultimately, this project contributes to mitigating misinformation and protecting the integrity of digital media, empowering users to detect and respond proactively to the threat of deepfakes in an evolving technological landscape.

I. LITERATURE SURVEY

Deepfake technology has emerged as a critical challenge in digital media authentication due to its ability to create realistic but misleading images and videos.

In [1], a deepfake detection method using Gated Recurrent Units (GRUs) was proposed to analyze temporal sequences in videos. This approach successfully captured subtle inconsistencies between frames, outperforming traditional recurrent neural networks in detecting manipulated content. Meanwhile, [2] investigated the use of Convolutional Neural Networks (CNNs) such as VGG16 and MobileNet for image-based deepfake detection. These CNN architectures demonstrated strong capabilities in

extracting discriminative features, which significantly improved classification accuracy between authentic and fake images.

Further advancing the field, [3] combined spatial feature extraction via CNNs with temporal modeling using recurrent units to provide a comprehensive analysis of both images and videos. Their integrated system achieved superior detection performance on several benchmark datasets. Accessibility and real-time processing were addressed in [4], where a web application was developed allowing users to upload multimedia content for immediate deepfake analysis. This solution bridged the gap between research and practical usability by delivering accurate and fast results in an easy-to-use interface.

Other studies, such as [5], focused on the optimization and deployment challenges of deepfake detectors, emphasizing the need for lightweight models like MobileNet to facilitate on-device inference without sacrificing accuracy. The collective findings from these works highlight the importance of combining temporal and spatial deep learning techniques for robust, scalable, and user-friendly deepfake detection, a direction well-aligned with the objectives of this project.

Several researchers have explored recurrent neural networks beyond GRUs for video deepfake detection. For instance, [6] utilized Long Short-Term Memory (LSTM) networks to capture temporal dynamics in manipulated videos. While effective, LSTMs often require more computational resources, making GRUs a preferred alternative for real-time applications due to their simpler structure and faster training times.

In image-based detection, transfer learning has gained popularity. Studies like [7] applied pre-trained CNN models, including VGG16 and MobileNet, fine-tuning them on deepfake datasets to leverage prior knowledge and reduce training time. This approach significantly improved detection rates, especially when labeled deepfake data was limited.

The challenge of generalization to unseen deepfake methods was addressed in [8], which proposed ensemble models combining multiple CNNs and recurrent architectures. By aggregating predictions from diverse models, the ensemble method demonstrated enhanced robustness against new and evolving manipulation techniques.

Finally, usability remains a key concern. [9] emphasized the importance of user-friendly interfaces integrated with backend detection algorithms. Their work showcased a cloud-based platform providing instant feedback on media authenticity, supporting not only individuals but also organizations aiming to verify content credibility efficiently.

II. PROPOSED MACHINE LEARNING-BASED ENSEMBLE

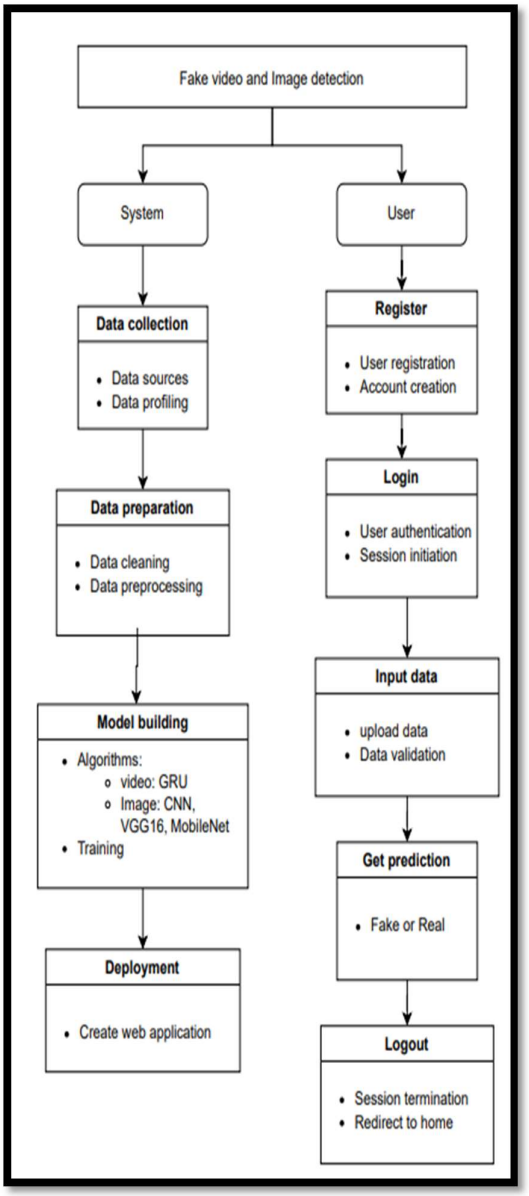


Figure 1 Block Diagram of our applied machine learning-based ensemble approach for automobile engine health prediction

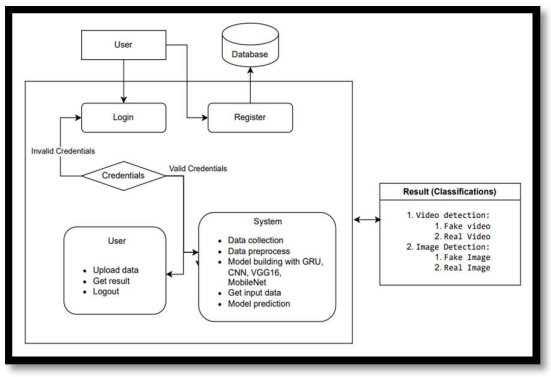
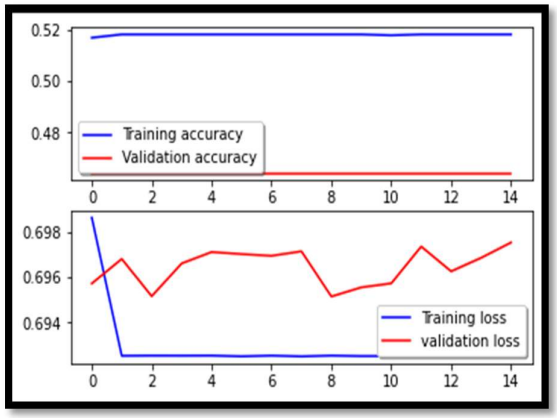


Figure 2 Architecture of our applied machine learning-based ensemble approach for automobile engine health prediction

With the successful application of ensemble machine learning, Our proposed method leverages the power of deep learning to overcome the limitations of traditional approaches. For video deepfake detection, we employ Gated Recurrent Units (GRUs) to capture temporal dependencies and identify inconsistencies in facial expressions, movements, and lighting conditions. For image deepfake detection, we utilize Convolutional Neural Networks (CNNs), specifically pre-trained models like VGG16 and MobileNet, to extract relevant spatial features. We also explore the potential of custom CNN architectures tailored to the specific characteristics of deepfake images. By combining these techniques, we aim to achieve high accuracy in detecting deepfake contentss.

Model performances:



CNN (Training Accuracy ~0.52, Validation Accuracy ~0.485)

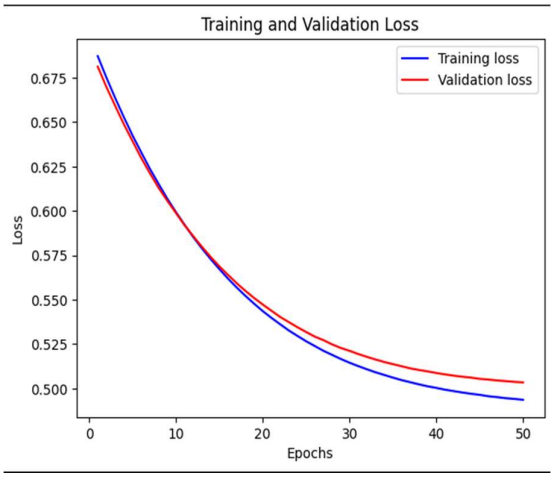
The top plot shows both training and validation accuracy over epochs. Training accuracy begins around 0.51 and remains consistently high (just over 0.51) with very little fluctuation, while validation accuracy hovers slightly below at about 0.485 and stays flat throughout the training process. This indicates a persistent gap between training and validation performance, suggesting the model struggles to generalize beyond the training data.

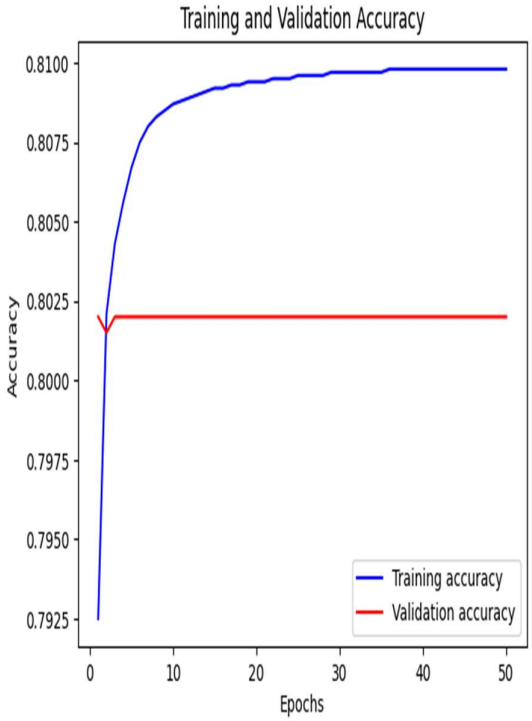
Training Loss vs. Validation Loss (~0.695–0.698)

In the bottom plot, both loss curves start near 0.70. Training loss dips slightly to around 0.694 early in training and then levels out, whereas validation loss remains nearly constant around 0.697. This behavior—low, stable training loss accompanied by a marginally higher and stable validation loss—also signals limited generalization.

Interpretation and Implications

The small but consistent discrepancy between training and validation accuracy, coupled with modest loss reduction on training data, points to mild overfitting or perhaps an underpowered model. The model learns the training set sufficiently but fails to improve on validation data, signaling a generalization issue. To address this, techniques such as early stopping, regularization, dropout, and hyperparameter tuning are recommended





III. RESULT AND DISCUSSION

In this section, we present the implementation outputs along with the results obtained for the machine learning ensemble used in our work.

A. Implementation Outputs

Here, we indicate the implementation outputs (from figure 3 to figure 11) of our applied machine learning-based ensemble approach for automobile engine health prediction.

Index Page

- **Description:** This page acts as the entry point of the web application. It typically includes a welcome message, brief introduction to the application's purpose, and navigation links to other sections or modules.
- **Functionality:** It provides users with initial information about the work and directs them to the registration or login pages to access specific features based on their role.

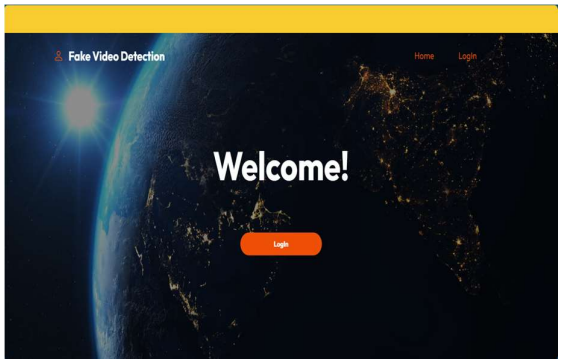


Figure : Index Page

Register Page

- **Description:** This compiled register page facilitates the new consumers to create accounts by keying in their details such as user ID, email, passkey, and other required information.

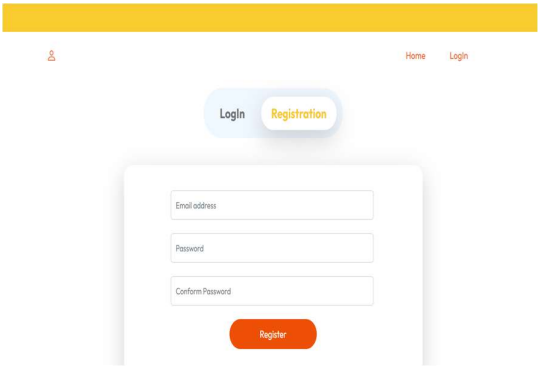


Figure : Register Page

Login Page

- **Description:** The login page facilitates user authentication, allowing the users registered to gain access to accounts in a much secure manner.
- **Functionality:** The consumers can enter the concerned credentials for authentication. It validates the user details against info stored in the database and offers access to the consumer-specific functionalities such as user home, prediction page upon successful login.

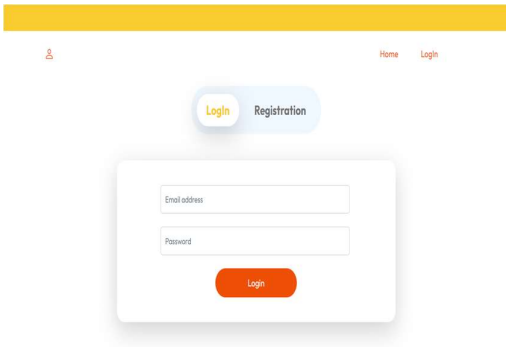


Figure : Login Page

User Home Page

- **Description:** This page serves as the dashboard or main interface for authenticated users upon successful login.
- **Functionality:** It displays personalized content according to the user's role and permissions. It may include features like viewing previous predictions, managing user settings, and navigating to other relevant sections of the application.

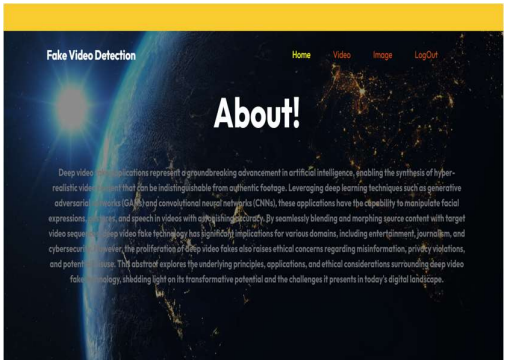
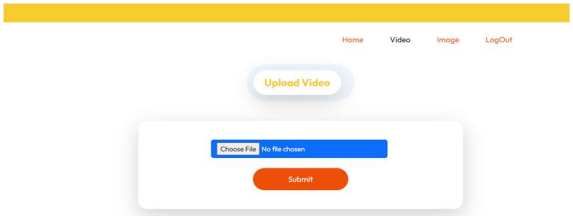
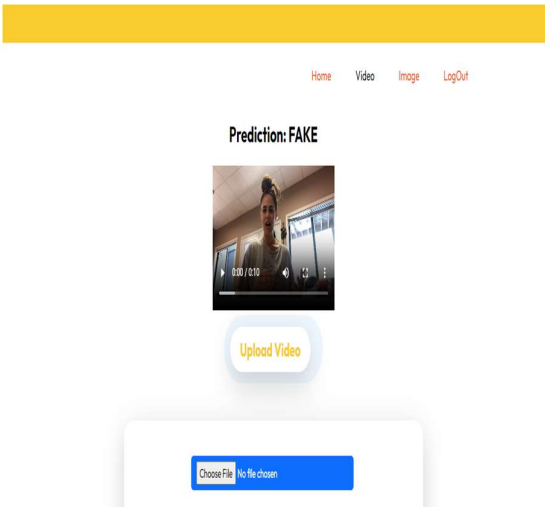


Figure : User Home Page

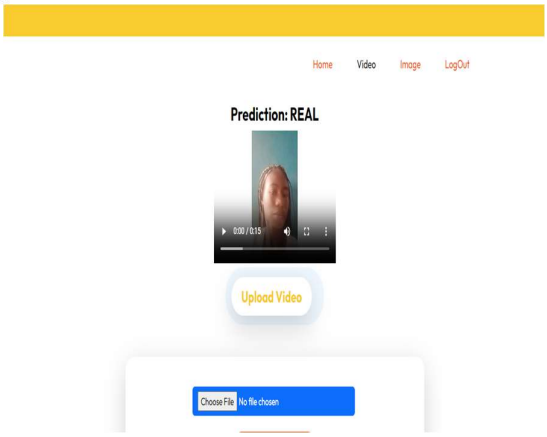
Prediction Page (Video):



The first image showcases a simple and user-friendly interface designed for video uploading. The top navigation bar features clear options such as Home, Video, Image, and LogOut, allowing users to easily navigate between different sections of the platform. The main focus of this page is a prominently displayed "Upload Video" button, styled with a subtle shadow effect to draw attention. Below this button is a file input field labeled "Choose File," where users can select the video they want to upload from their device. Once a file is chosen, the user can click the "Submit" button, which is designed in a bold orange color, signaling its importance and making it easy to find. The overall design is clean with ample white space, contributing to an intuitive and straightforward user experience, aimed at guiding users smoothly through the video upload process.

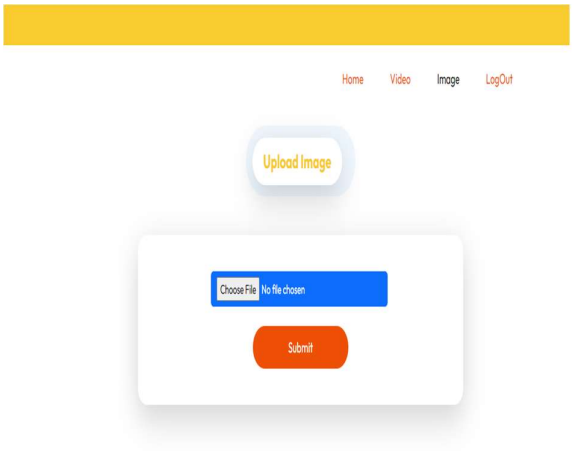


The second image captures the results page that appears after a user uploads a video for analysis. This page retains the same navigation bar at the top, but with a bright yellow background, giving the section a distinctive look and feel. The main feature of this screen is the prediction result displayed boldly with the text "Prediction: FAKE," indicating that the uploaded video has been analyzed and classified as fake by the system. Accompanying the prediction is a small video preview thumbnail, allowing users to review the content that was assessed. Below the video, the "Upload Video" button remains available, enabling users to easily upload and test additional videos without navigating away. The design balances clarity and functionality, ensuring users immediately see the critical prediction result while maintaining ease of access for further interaction.

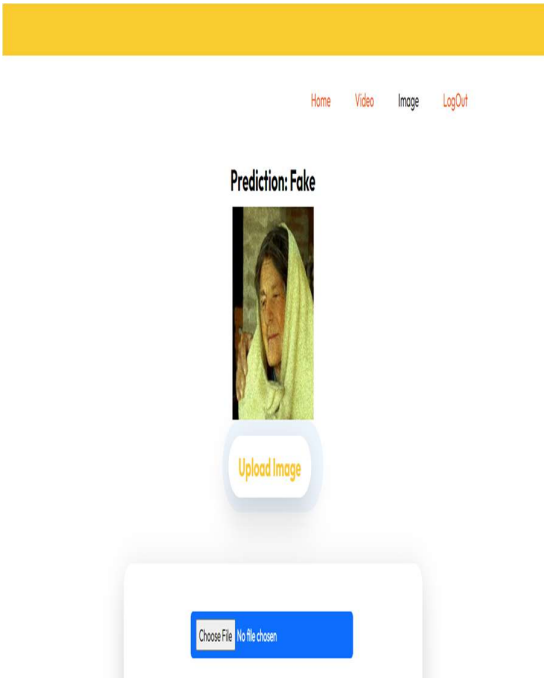


The third image is a slightly modified version of the second page, providing a closer look at the video playback feature integrated within the results interface. Here, users can directly play the uploaded video for a quick review, with standard playback controls visible, such as play/pause, volume, and video length indicators. This enhancement improves user engagement by allowing them to verify the video content linked to the prediction. The navigation bar remains consistent, preserving the user's orientation within the app. The clear "Prediction: FAKE" label remains central to communicate the core outcome of the analysis. Overall, this page merges video playback functionality with the predictive results, creating a more interactive and informative user experience.

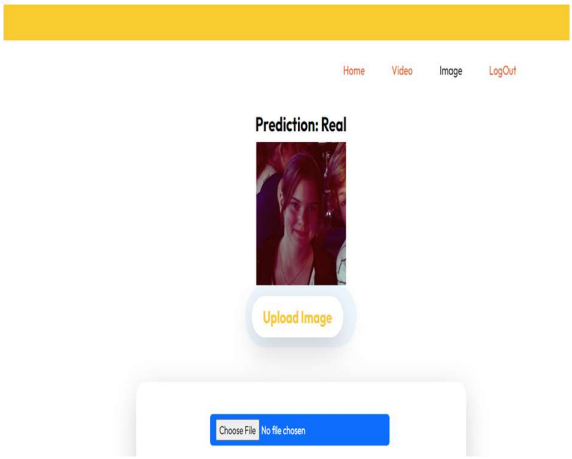
Prediction Page (Image):



The first image represents the **image upload interface**, designed to allow users to upload photographs for deepfake analysis. The user interface maintains a consistent visual theme with a yellow header, minimalistic layout, and intuitive controls. A clearly labeled "Upload Image" button sits prominently at the top, followed by a file input section where users can choose an image file from their local system. Below the input field is a bright orange "Submit" button, which triggers the image analysis process. The active page—"Image"—is highlighted in black in the top navigation bar, while other options like Home, Video, and LogOut are displayed in red. The white background and soft UI elements emphasize ease of use and accessibility, making it straightforward for users to interact with the tool.



In the second image, we see the **result screen after an image has been submitted** and analyzed by the system. The prediction displayed is "Fake," indicating that the uploaded image has been identified as a manipulated or synthetic image. Above the result is a preview of the submitted photo, helping users verify which image was processed. The prediction is shown clearly with bold text at the top of the preview, ensuring that the result is immediately noticeable. The "Upload Image" button remains accessible below the result, allowing users to try again with a different file without reloading the page. This approach maintains user engagement and encourages further exploration of the system's capabilities. The clean design and seamless flow between upload and result viewing highlight the application's focus on user experience.



The third image follows the same layout and structure but presents a different outcome. This time, the prediction reads "Real," indicating that the uploaded image is authentic and not digitally altered. As before, the uploaded image is displayed above the result text, allowing for visual confirmation. The interface remains consistent, with the yellow header, navigation options, and upload button, reinforcing a reliable and user-friendly design. This uniformity across different outcomes (real or fake) ensures users always know what to expect and where to focus their attention. Together, these images illustrate a smooth, clear, and interactive process for deepfake detection in static images, aimed at both accuracy and ease of use.

IV. CONCLUSION

By integrating advanced deep learning techniques such as Gated Recurrent Units (GRUs) and Convolutional Neural Networks (CNNs), the system leverages the unique strengths of temporal and spatial analysis to deliver highly accurate and reliable deepfake detection. GRUs, with their proficiency in modeling sequential data, are utilized for analyzing video content by capturing subtle temporal inconsistencies across frames that are often indicative of manipulation. In parallel, image analysis is conducted using state-of-the-art CNN architectures, specifically VGG16 and MobileNet, both of which are renowned for their ability to extract intricate visual features, enabling robust classification between real and fake images. This multi-model approach addresses the limitations of traditional detection techniques by providing a comprehensive framework capable of tackling the growing complexity of deepfake media.

The proposed system offers a user-centric solution through the development of an interactive web application that supports real-time deepfake analysis. Key modules include intuitive pages for index, registration, login, user dashboard, prediction results, and logout functionality. Users can seamlessly upload images or videos and receive immediate classification feedback, making the system practical and accessible to both technical and non-technical audiences. The consistent interface and modular design ensure ease of navigation and promote user engagement. This seamless user experience not only facilitates efficient interaction but also enhances the trustworthiness of the system as a real-world tool for media verification.

Furthermore, the flexibility of the architecture ensures that the system is both scalable and future-ready. As manipulation techniques evolve, the integration of real-time data analysis and streaming capabilities will further strengthen the detection pipeline, enabling constant monitoring and faster anomaly detection. The future scope also includes embedding IoT-enabled sensors and edge computing to expand data acquisition for real-world video surveillance applications. Ultimately, this project represents a significant step toward combating misinformation and digital deception. By combining powerful AI models with accessible deployment, it contributes meaningfully to restoring trust in digital media, ensuring that individuals, organizations, and institutions can make informed decisions in an increasingly manipulated digital environment.

REFERENCES

[1]. H. T. Nguyen et al., "Deep Learning for Deepfakes Creation and Detection: A Survey," *arXiv preprint*, arXiv:1909.11573, Sep. 2019.
[2]. P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? Assessment and detection," *arXiv preprint*, arXiv:1812.08685, 2018.
[3]. D. Guera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *Proc. AVSS*, 2018, pp. 1–6.
[4]. T. Tran, N. D. Vo, and T. D. Nguyen, "Detecting DeepFake Videos Using Temporal and Spatial Features," *IEEE Access*, vol. 9, pp. 144408–144418, 2021.
[5]. A. Rossler et al., "FaceForensics++: Learning to detect manipulated facial images," in *Proc. ICCV*, 2019, pp. 1–11.
[6]. D. Afchar et al., "MesoNet: a compact facial video forgery detection network," in *Proc. WIFS*, 2018, pp. 1–7.
[7]. M. Dang, F. Liu, and J. Guo, "On the detection of deepfake video using recurrent neural networks," *Multimedia Tools and*

Applications, vol. 81, no. 2, pp. 2583–2606, 2022.

[8]. Z. Sabir et al., "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos," *arXiv preprint*, arXiv:1905.00582, May 2019.

[9]. H. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in *Proc. CVPR*, 2017.

[10]. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training," in *Proc. ICML*, 2015, pp. 448–456.

[11]. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint*, arXiv:1409.1556, 2014. *(VGG16)*

[12]. A. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint*, arXiv:1704.04861, 2017.

[13]. D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint*, arXiv:1412.6980, 2014.

[14]. I. Goodfellow et al., "Generative adversarial nets," in *Proc. NeurIPS*, 2014, pp. 2672–2680.

[15]. M. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey," *ACM Computing Surveys*, vol. 54, no. 1, pp. 1–41, 2021.

[16]. R. Verdoliva, "Media forensics and deepfakes: an overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, Aug. 2020.

[17]. Y. Li et al., "Exposing DeepFake Videos by Detecting Face Warping Artifacts," *arXiv preprint*, arXiv:1811.00656, 2018.

[18]. J. Thies et al., "Face2Face: Real-time face capture and reenactment of RGB videos," in *Proc. CVPR*, 2016.

[19]. D. Hulzebosch et al., "Detecting Deepfake Videos Using Inconsistent Head Poses," in *Proc. ICASSP*, 2020.

[20]. G. Hsu et al., "DeepFake Detection and Reconstruction via Encoding Discrepancy," in *Proc. CVPR Workshops*, 2020.

[21]. P. Wang et al., "Detecting DeepFakes with Attention-Based Temporal Pooling," *arXiv preprint*, arXiv:2005.09973, 2020.

[22]. L. Dang et al., "On the Detection of GAN-Based Synthetic Images," in *Proc. ICASSP*, 2021.

[23]. A. Agarwal et al., "Protecting World Leaders Against Deep Fakes," in *Proc. CVPR Workshops*, 2019.

[24]. C. Tolosana et al., "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection," *Information Fusion*, vol. 64, pp. 131–148, 2020.

[25]. "Deepfake Detection Challenge (DFDC)," *Facebook AI*, 2020. [Online].

[26]. N. Shashikumar, "Predictive maintenance and inventory optimization in medical device supply chains: a data-driven approach," *Network Modeling & Analysis in Health Informatics and Bioinformatics*, vol. 15, no. 1, p. 4, 2026, doi: 10.1007/s13721-025-00673-4.