JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Raspberry Pi-Enabled Multi-Modal AI System for Deepfake Detection in Multimedia

Dr. E. Anant Shankar
Department of ECE,
Sri Venkasteswrara College of Engineering (Autonmous),
Tirupati, A.P.,India
anant.shankar16@gmail.com

P.Pavan Kalyan Reddy
Department of ECE,
Sri Venkasteswrara College of Engineering (Autonmous),
Tirupati, AP.India
pavankalyanreddy2005@gmail.com

P.Sravanthi
Department of ECE
Sri Venkasteswrara College of Engineering (Autonmous),
Tirupati, A.P., India
sravanthisravs3894@gmail.com

N. Bhavya
Department of ECE,
Sri Venkasteswrara College of Engineering (Autonmous),
Tirupati, A.P., India
bhavyabhavi0604@gmail.com

T. Baby Shalini
Department of ECE
Sri Venkasteswrara College of Engineering (Autonmous),
Tirupati, A.P., India
thadushalini@gmail.com

D. Ajay
Department of ECE
Sri Venkasteswrara College of Engineering (Autonmous),
Tirupati, A.P., India
devarakondaajay9876@gmail.com

Abstract— **This project presents a Raspberry Pi–based multi-modal AI edge system designed for real-time deepfake detection across images, videos, and audio. Leveraging a Raspberry Pi as the central controller, the system integrates a USB webcam to capture visual data and analyze facial authenticity, while audio inputs are monitored for signs of manipulation. A memory card stores system data and AI models for local processing at the edge, minimizing reliance on cloud computing and ensuring faster response times. The system provides immediate feedback via an LCD display, showing sensor status and detection results. An audible buzzer alerts users to suspicious or abnormal content, enhancing security and situational awareness. The setup is modular, with 20 connectors for flexible expansion, enabling the integration of additional sensors or AI modules. This prototype demonstrates the feasibility of deploying edge-based AI systems for detecting multimedia deepfakes in real-time, combining cost-effective hardware with intelligent software.**

***Keywords*: Raspberry Pi, Deep fake Detection, Multi-modal AI , Edge Computing, Real-time Monitoring**

INTRODUCTION

As the technologies of artificial intelligence and deep learning rapidly evolve, the production of rather realistic synthetic media is becoming more frequent. Deepfakes are one of such technologies that attract much focus because they can be used to alter images, videos, and audio in such a way that human viewers perceive it as authentic. Although this type of technologies may be utilized in the context of entertainment and creative usage, it also triggers some

serious concerns connected with misinformation, violation of privacy, and security in the digital realm. With the increased availability of deepfake generation tools, the problem of identifying manipulated multimedia content has become a major research problem. Conventional deepfake detectors commonly utilize high-performance computing infrastructure or cloud computing processing infrastructures. Even though these solutions are capable of high accuracy, they might be constrained by high latency; high cost of computation; and reliance on the internet connection. Lots of real-life applications, particularly remote monitoring or security application, in the past, there has been an increasing demand to have compact and efficient systems, which can easily handle detection duties on their own without being heavily dependent on external infrastructure. One of the potential solutions to reduce these challenges is edge computing, which allows performing data processing on embedded devices. The edge-based systems are able to minimize response time, maximize privacy and increase reliability of the system by conducting an analysis in the proximity of the data area. The Raspberry Pi and other single-board computers provide an affordable platform to deploy simple artificial intelligence models that can be used in real-time. In this regard, the given system proposes a multi-modal deepfake detection system deployed on a Raspberry Pi platform. The system is a visual and audio analysis integrated

to detect possible manipulations of multimedia content. To analyze facial authenticity with the help of a USB web camera, real-time visual data is captured and audio input is analyzed to identify the possible changes. The results of detection are presented in the form of an LCD interface, and a buzzer allows generating instant notifications in case any suspicious activity is detected. The combination of edge-based processing and modular hardware will enable the system to show a realistic and scalable method of real-time deepfake detection with low-cost embedded technology.

RELATED WORKS

Creation and Detection of Deepfakes: A Survey by Yisroel Mirsky and Wenke Lee is an extensive review of the deepfake technology. The paper explains the creation of deepfake content based on the methods of deep learning and summarizes some of the detection methods that have been created to detect manipulated media. [1].

The article Anticipating and Addressing the Ethical Implications of Deepfakes in the Context of Elections by Nicholas Diakopoulos and Deborah Johnson examines the ethical dangers of the deepfake technology in the political climate. The paper constitutes the significance of the manipulated media in shaping the opinion of people and interfering with the democratic procedures, especially in election periods[2].

The Distinct Wrong of Deepfakes by Alexander de Ruiter explores the ethical and philosophical issues of deepfakes. The study offers insight into how deepfake content may break trust, privacy, and identity as highlighted by the social and ethical impacts of synthetic media[3].

The article Deep Fakes: A Coming threat to privacy, democracy, and national security by Bobby Chesney and Danielle Citron is about the systemic dangers of the deepfake technology. The authors discuss the possible risks of personal privacy, political stability, and national security due to realistic synthetic media[4].

Making Deepfakes Gets Cheaper and Easier With the Help of A.I. by Stuart A. Thompson states that the progress in the field of artificial intelligence has made deepfakes creation more affordable and reachable. The article describes how the availability of AI tools has increased making it easier to develop highly realistic manipulated videos[5].

Face Forgery Detection by 3D Decomposition This article by Xiangyu Zhu et al. suggests a deepfake detector that works with 3D facial geometry. The method breaks down the facial parts to detect discrepancies between authentic and fake facial features[6].

Self-Supervised Learning of Adversarial Example: Towards Good Generalizations in Deepfake Detection by Lin Chen and others describes a self-supervised learning model to enhance deepfake detection. Critical examples allow the model to improve its capacity to identify manipulation methods that are not visible to the human eye[7].

The article Multi-Mode Multi-Scale Transformers to Deepfake Detection by Jingjing Wang and others demonstrates a transformer architecture to detect deepfakes. The model is a multi-modal and multi-scale feature that integrates the ability to detect subtle artifacts within manipulated videos[8].

Video Transformer: Deepfake Detection by Incremental Learning by S. A. Khan and Honggang Dai is an article that suggests a video analysis transformer model with incremental learning. The approach will enable the system of detection to evolve and become more effective as new patterns of deepfakes will emerge[9].

The article ISTVT: Interpretable Spatial-Temporal Video Transformer to Deepfake Detection by Chao Zhao and others presents a spatial-temporal transformer model that is meant to detect deepfakes and can be interpreted. The model examines the spatial image characteristics and the time movement patterns on videos[10].
DFMT: End to End Deepfake Detection Framework Vision Transformer by Amin Khormali and Jin-Su Yuan is a vision transformer-based detection framework of deep fakes. The mechanism automatically isolates visual characteristics of the visuals and video to detect manipulated visual content[11].

Like Towards Generalizable Deepfake Detection with Locality-Aware AutoEncoder by Mengnan Du and co-authors, an autoencoder-based model concentrating on local facial features is proposed. The model will enhance generalization of varied deepfake generation techniques[12].

Towards Solving the DeepFake Problem: Improving DeepFake Detection with Dynamic Face Augmentation by Subhankar Das and co-authors explores the ability of using dynamic face augmentation to make deepfake detection models stronger. The technique increases the variety of training data to make the detection systems more robust[13].

Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale by Alexey Dosovitskiy and others presents a model called Vision Transformer. It establishes that transformer models can be effectively used to process image patches to perform large-scale image recognition tasks, which shapes several of the current deepfake detector methods[14].

Proposed method

The proposed model introduces a Raspberry Pi–based multi-modal AI edge system capable of real-time deepfake detection across images, videos, and audio. By integrating a USB webcam and microphone, the system captures live multimedia data for local processing using lightweight AI models stored on a memory card. An LCD displays sensor and detection status, while a buzzer alerts users to suspicious or manipulated content. This edge-based approach reduces dependence on cloud computing, ensures low-latency detection, and supports simultaneous monitoring of multiple media types. The modular design with 20 connectors allows easy expansion for additional sensors or AI modules, providing a practical, cost-effective solution for real-time deepfake detection. The overall architecture of the proposed system is illustrated in Fig. 1.

Block Diagram

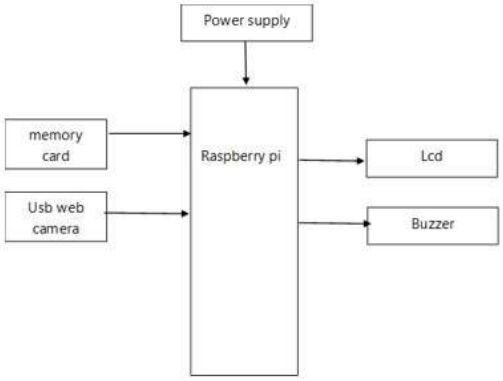

Fig. 1. Block diagram of the proposed system

Methodology

Principle of Functioning:

The Deepfake Detection System proposed is aimed at doing real-time analysis of multimedia information such as images, video and audio files to detect manipulated or fake information with the assistance of AI techniques. The system is designed to be based on Raspberry Pi, which serves as the central processing unit. The visual data is captured by a USB web camera to check the authenticity of the face, and audio streams are detected to detect any attempts of any manipulations or exploitation. All the gathered data are processed on-site with the help of pre-trained AI models on a memory card, without depending on cloud computing and assuring a quicker reaction time. The machine learning algorithms examine the characteristics obtained after processing visual and audio data to categorize the content as genuine or a deepfake with a great level of precision. The system status and detection results are shown on an LCD screen and can be seen in real-time. The system sends alerts to users in case of suspected deepfake content and this increases the security and situational awareness.

Hardware & Alerts

The hardware can be summarized as a Raspberry Pi, the USB web camera, a sound device (microphone), a memory card containing local models, LCD display, a buzzer and a modular interface that has a 20 connector board on which to use in the future. The Raspberry PI processes visual and audio information as it arrives through deepfake detection AI algorithms. The LCD display offers real-time information regarding the activities in the system such as detection results and sensor position. Once manipulated material is detected, a buzzer triggers a sound signal that users can use to be aware of possible threats to their security. The modular connector system enables the incorporation of other sensors or AI units, which makes the system rambling and adaptable to different deployment conditions. This is a compact design that is cost-effective and smart software-based to facilitate effective edge-based multimedia deepfake detection.

Power Requirements

It is necessary to have a constant controlled power supply to guarantee the constant functionality of the Raspberry Pi, USB web camera, microphone, LCD display, and other peripherals. The use of reliable power allows capturing data without delays, processing AI in real-time, and generating alerts immediately, as well as scaling the system with the help of adding new modules. The system is trained on publicly available datasets of both authentic and forged images, videos and audio samples. Video frames, image resizing, and audio signal transformation to spectrogram features are included in the preprocessing of the data. The visual and audio manipulation patterns are detected using a lightweight convolutional neural network (CNN). The optimized model is then deployed to the Raspberry Pi after being trained using a high-performance computer to run real-time deepfake detection at the edge. The performance evaluation of the implemented deepfake detection models is summarized in Table I. A comparison between the existing approaches and the proposed system is shown in Table II.

TABLE I. PERFORMANCE COMPARISON

Parameter	Specification / Metric	Description
Central Controller	Raspberry Pi	Serves as the main processing unit that collects visual and audio data, runs AI models for deepfake detection, and controls system operations.
Visual Input	USB Webcam	Captures real-time images and video streams to analyze facial authenticity and detect manipulated content.
Audio Input	Microphone	Monitors audio streams to detect tampering or deepfake audio signals.
Data Storage & Processing	Memory Card	Stores AI models and system data for local edge processing, minimizing reliance on cloud computing and enabling faster response times.
Detection Analysis	AI Models (Edge-based)	Processes collected multimedia data to classify content as authentic or deepfake with high accuracy.
Status Display	LCD Module (16×2)	Shows real-time system status, detection results, and alerts for local monitoring.

Expansion Interface	Modular Connectors (20 pins)	Allows integration of additional sensors or AI modules for system scalability and flexibility.
Audible Alert	Buzzer	Produces an alert sound to notify users when suspicious or manipulated content is detected.

TABLE II. PERFORMANCE COMPARISON OF DEEPFAKE DETECTION MODELS

Feature	Existing Model	Proposed Model
Processing	Cloud-based processing	Edge-based processing using Raspberry Pi
Media Detection	Mostly single media (image/video)	Multi-modal (image, video, audio)
Hardware	High-end computers or servers	Low-cost Raspberry Pi system
Response	Dependent on network	Faster local processing
User Alert	Software notifications	LCD display and buzzer alert
Deployment	Limited portability	Compact and portable system

Results

Fig. 2. Architecture of the proposed deepfake detection system
Fig. 2 shows the architecture of the multimodal AI edge system built on Raspberry Pi for detecting deepfake content in images, videos, and audio. The multimedia information is recorded by a USB webcam, and it is processed by the Raspberry Pi with the help of AI-based deep learning models to determine the manipulated data. The system examines both aesthetical and visual characteristics including the facial inconsistencies and unnatural patterns to categorize the media as a legitimate one or a fake one. A 16x2 LCD display avails a very user-friendly interface to the user to choose detection modes and see results. The system is fast with better privacy and less reliance on cloud services because it detects on the edge device, which makes it appropriate in digital media verification and security systems.

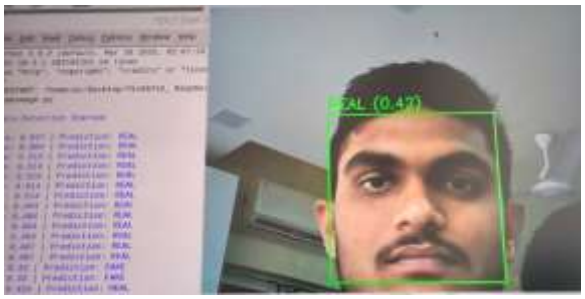

Fig. 3. System interface showing deepfake detection process
Fig. 3 illustrates the experimental design and real-time working of the proposed deepfake detection system for identifying real images. A webcam connected through a USB interface captures the facial image of a person. The captured image is then processed by the AI model, which analyzes facial features and patterns to verify authenticity. The system detects the face in the captured frame, extracts important visual features, and compares them with the trained model to determine whether the image is real. The classification result is immediately displayed on the monitor. This setup demonstrates a real-time face verification system using a webcam as the input source, enabling quick identification of genuine images within a few seconds.

Fig. 4. Deepfake image detection result using webcam input

Fig. 4 presents the deepfake detection result obtained using the webcam input. The captured facial image is analyzed by the AI model, and the detected face is highlighted with a bounding box on the screen. Based on feature analysis, the system classifies the image as fake, indicating that the content is manipulated. This demonstrates the system's ability to perform real-time detection of deepfake images.

Conclusion

In summary, the phenomenon of deepfakes is developing, with the rapid improvement of generative artificial intelligence, thus, it is becoming more and more realistic and hard to detect when images, videos, and sound are being manipulated. It has been emphasized in the literature that unlike the earlier methods of detection which mainly employed the visual artifacts,

contemporary techniques focus on multimodal techniques, that is, combining the facial, temporal and audio features to enhance robustness and generalization. They are constantly revealed by surveys and recent studies that multimodal fusion systems are more effective in comparison to single-modality systems, especially in response to high quality and cross-domain deepfakes. Nonetheless, issues of diversity of databases, real-world generalization, adversarial resistance and edge device lightweight implementation are present. Given the potential to improve security, trust and authenticity in digital media ecosystems, future studies ought to consider a more effective, explainable and resource efficient model that can detect in real time, particularly edge-based systems like Raspberry PI to improve future studies.

References

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

[2] N. Diakopoulos and D. Johnson, "Anticipating and addressing the ethical implications of deepfakes in the context of elections," New Media & Society, vol. 23, no. 7, pp. 2072–2098, Jul. 2021.

[3] A. de Ruiter, "The distinct wrong of deepfakes," Philosophy & Technology, vol. 34, no. 4, pp. 1311–1332, Dec. 2021.

[4] B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," California Law Review, vol. 107, pp. 1753–1819, 2019.

[5] S. A. Thompson, "Making deepfakes gets cheaper and easier thanks to AI," The New York Times, Mar. 12, 2023.

[6] X. Zhu, H. Wang, H. Fei, Z. Lei, and S. Z. Li, "Face forgery detection by 3D decomposition," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), Jun. 2021, pp. 2928–2938.

[7] L. Chen, Y. Zhang, Y. Song, L. Liu, and J. Wang, "Self-supervised learning of adversarial examples: Towards better generalization for deepfake detection," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), Jun. 2022, pp. 18689–18698.

[8] J. Wang, Z. Wu, W. Ouyang, X. Han, J. Chen, Y. G. Jiang, and S. N. Li, "M2TR: Multi-modal multi-scale transformers for deepfake detection," in Proc. Int. Conf. Multimedia Retrieval (ICMR), Jun. 2022, pp. 615–623.

[9] S. A. Khan and H. Dai, "Video transformer for deepfake detection with incremental learning," in Proc. 29th ACM Int. Conf. Multimedia, Oct. 2021, pp. 1821–1828.

[10] C. Zhao, C. Wang, G. Hu, H. Chen, C. Liu, and J. Tang, "ISTVT: Interpretable spatial–temporal video transformer for deepfake detection," IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1335–1348, 2023.

[11] A. Khormali and J. S. Yuan, "DFDT: An end-to-end deepfake detection framework using vision transformer," Applied Sciences, vol. 12, no. 6, p. 2953, Mar. 2022.

[12] M. Du, S. Pentyala, Y. Li, and X. Hu, "Towards generalizable deepfake detection with locality-aware autoencoder," in Proc. 29th ACM Int. Conf. Information and Knowledge Management (CIKM), Oct. 2020, pp. 325–334.

[13] S. Das, S. Seferbekov, A. Datta, M. S. Islam, and M. R. Amin, "Towards solving the deepfake problem: Improving deepfake detection using dynamic face augmentation," in Proc. IEEE/CVF Int. Conf. Computer Vision Workshops (ICCVW), Oct. 2021, pp. 3769–3778.

[14] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learning Representations (ICLR), 2021.