iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** X **Month of publication:** October 2025

DOI: https://doi.org/10.22214/ijraset.2025.74700

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 13 Issue X Oct 2025- Available at www.ijraset.com

Deep Guard: Face Spoofing Detection using Swin Transformer and rPPG Signal

S.G. Santhoshini[1], K.S. Shalini[2], K. Usha Rani[3]
[1, 2]*Department of Computer Science and Engineering,* [3]*Assistant Professor/CSE, K.L.N. College of Engineering, Pottapalayam, Sivagangai*

Abstract: ***The Deep Guard project is developed to detect and prevent facial spoofing attacks in biometric authentication systems. Traditional face recognition systems are often vulnerable to spoofing attempts using printed photos, replayed videos, or 3D masks. To overcome these challenges, Deep Guard integrates Swin Transformer-based deep feature extraction with rPPG (remote Photoplethysmography) signal analysis to accurately differentiate between real and fake faces. The Swin Transformer captures fine-grained spatial and texture information, while the rPPG module extracts heartbeat-based color variations to confirm liveness. The extracted features are then fused using an advanced feature fusion technique for robust classification. A trained model finally classifies the input as genuine or spoofed. Experimental evaluation shows that Deep Guard delivers high precision, adaptability, and real-time performance, making it a reliable and secure solution for modern facial authentication applications in banking, mobile security, and access control systems.***

Keywords: Face Anti-Spoofing, Swin Transformer, rPPG (Remote Photoplethysmography), Deep Learning, Feature Fusion, Liveness Detection, Biometric Security, Classification, Real-Time Detection, Image Processing.

I. INTRODUCTION

In recent years, facial recognition technology has become one of the most widely adopted biometric authentication methods due to its convenience, accuracy, and non-intrusive nature. It is increasingly used in various applications such as smartphone unlocking, access control systems, and financial transactions. However, despite its popularity, facial recognition systems are highly vulnerable to spoofing attacks, where an imposter attempts to deceive the system using printed photos, replayed videos, or 3D masks. These attacks pose significant security risks, leading to potential data breaches and identity theft. Therefore, enhancing the robustness of facial recognition systems against such spoofing attempts has become a critical research focus in modern biometric security.

To address these vulnerabilities, the Deep Guard system introduces an intelligent and hybrid face anti-spoofing approach that combines deep learning-based visual analysis with physiological signal detection. The system employs a Swin Transformer model to extract fine-grained spatial and texture features from facial images, enabling effective recognition of subtle differences between real and spoofed faces. Additionally, a remote Photoplethysmography (rPPG) module is integrated to detect the liveness of a subject by analyzing minute skin color variations caused by blood flow. By combining both visual and physiological cues, Deep Guard ensures a more accurate and reliable detection of spoofing attacks, even under challenging lighting and environmental conditions.

Furthermore, Deep Guard utilizes a feature fusion algorithm to integrate the outputs from both Swin Transformer and rPPG modules, thereby enhancing the overall classification accuracy. The fused features are passed through a classification layer that determines whether the detected face is genuine or fake. The proposed system has been tested on multiple benchmark datasets and real-time video inputs, demonstrating high detection accuracy, low false acceptance rates, and excellent generalization across various spoofing types. Hence, Deep Guard represents a robust, efficient, and scalable solution for real-world biometric security systems, contributing significantly to the advancement of face anti-spoofing research and its secure deployment in critical applications.

II. METHODOLOGY

The proposed Deep Guard system integrates both visual and physiological feature analysis to ensure robust and accurate face anti-spoofing. The methodology begins with video frame acquisition and preprocessing, where the input face is captured using a webcam or video stream. The system detects the facial region, normalizes lighting variations, and resizes the frame for model compatibility. The preprocessed face images are then passed into the Swin Transformer, a hierarchical vision transformer model that extracts deep spatial and texture-based features. These features help in identifying micro-details such as skin reflection, edge smoothness, and texture consistency—key indicators for distinguishing between genuine and spoofed faces.

In parallel, the system extracts rPPG (remote Photoplethysmography) signals from subtle skin color variations in the facial region. This step detects the heartbeat-induced temporal changes, which are present only in live faces. Both the Swin Transformer's visual features and rPPG's physiological signals are then combined using a feature fusion algorithm, which enhances the discriminative power of the model. The fused features are input to a classification network, which determines whether the face is real or spoofed. The entire Deep Guard pipeline is designed for real-time performance, ensuring fast and accurate decision-making suitable for applications in access control, banking, and mobile security systems.

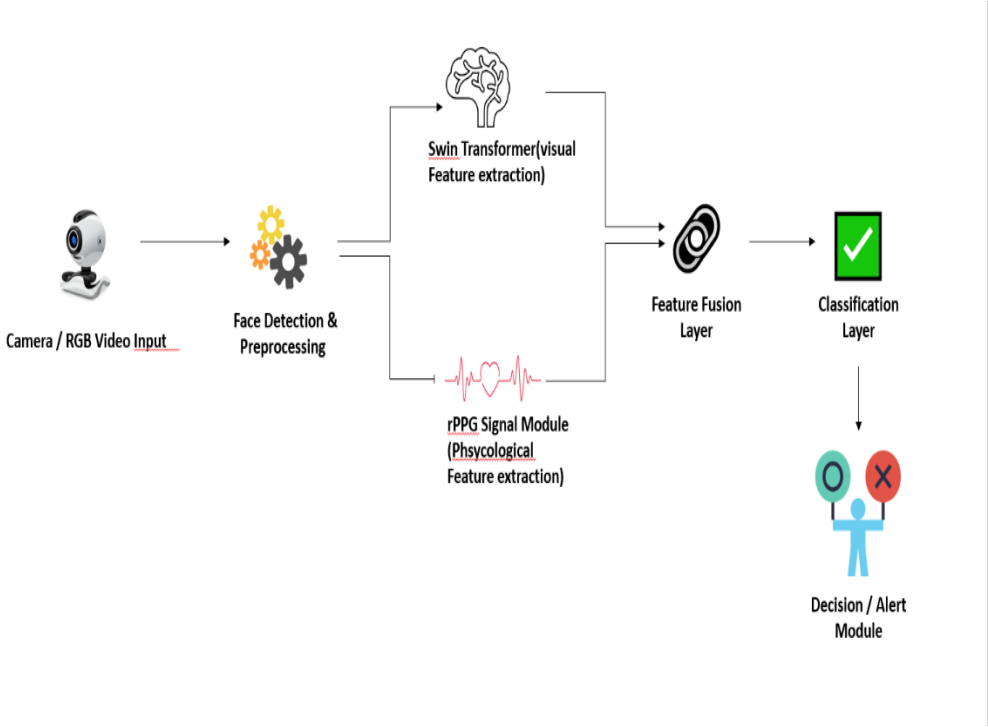


Fig. 1. Architecture Diagram

III. PROCESSFLOW

The process flow of the Deep Guard system is designed as a sequential and integrated pipeline to ensure accurate face anti-spoofing through both visual and physiological analysis. The process begins with data acquisition, where the system captures real-time video frames or static images from the camera. Each frame undergoes face detection to locate the region of interest (ROI), ensuring that only the relevant facial area is processed further. Once the face is detected, preprocessing operations such as normalization, noise removal, and illumination correction are applied to prepare the input for deep learning analysis. This step guarantees uniformity across different lighting and background conditions, enhancing model performance.

After preprocessing, the facial data is sent to two parallel modules for feature extraction. The first is the Swin Transformer module, which extracts deep spatial and texture-based features using a hierarchical transformer architecture. It effectively captures local and global facial patterns, helping to identify fine differences between real skin and spoof surfaces like paper or screens. The second module is the rPPG (remote Photoplethysmography) analysis, which processes temporal changes in facial color intensity to detect the presence of a live pulse. This physiological feature distinguishes live faces from static or replayed ones, adding an essential liveness verification layer to the system.

Once the Swin Transformer and rPPG modules extract their respective features, they are passed to a feature fusion unit, where both visual and physiological information are combined into a unified representation. This fused feature vector is then fed into a classification layer trained to distinguish between real and spoof faces. The model outputs a confidence score indicating the authenticity of the input. The results are then displayed to the user or integrated into a security system for decision-making. This systematic process flow ensures high accuracy, fast detection, and resilience against diverse spoofing techniques, making Deep Guard suitable for secure and real-time biometric authentication.

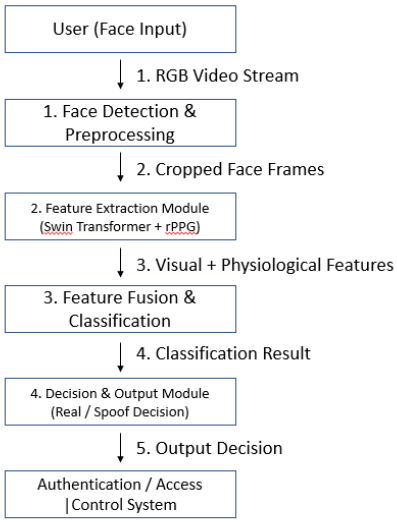


Fig 2 process diagram

IV. PREPROCESSING

Preprocessing plays a crucial role in the Deep Guard system as it ensures that the input data is clean, standardized, and suitable for reliable feature extraction. The process begins with face detection and region-of-interest (ROI) extraction from the input video or image frames. Using advanced face detection algorithms, such as the Multi-Task Cascaded Convolutional Networks (MTCNN) or OpenCV's Haar Cascade, the system isolates the facial region from the background. This step eliminates irrelevant data and focuses only on key facial areas like the forehead, cheeks, and nose—regions most useful for both visual texture and physiological signal analysis.

Once the facial region is detected, the next step involves normalization and enhancement. The extracted face is resized to a uniform dimension suitable for the Swin Transformer model to maintain consistency across all samples. Additionally, preprocessing techniques such as histogram equalization and Gaussian smoothing are applied to reduce lighting variations, noise, and shadows that could affect model accuracy. Color correction and intensity normalization help stabilize pixel values, ensuring that the model focuses on meaningful facial features rather than illumination differences. This step also aids in improving the rPPG signal quality, as consistent lighting conditions enhance the accuracy of heartbeat-based feature extraction.

Finally, temporal frame selection and alignment are performed for video-based inputs. Since rPPG analysis requires time-series data, a sequence of consecutive frames is selected, aligned, and stabilized to minimize motion artifacts. Frame alignment ensures that small head movements or expressions do not distort the extracted signals. The cleaned and aligned frames are then forwarded to the Swin Transformer for spatial feature extraction and to the rPPG module for physiological signal analysis. By performing these preprocessing steps, Deep Guard ensures that both visual and temporal data are optimized, leading to enhanced accuracy, robustness, and efficiency in spoof detection.

V. SWIN TRANSFORMER AND rPPG SIGNAL FOR FEATURE EXTRACTION AND TO VERIFY LIVENESS

The Swin Transformer module in Deep Guard serves as the core component for visual feature extraction. Unlike traditional convolutional neural networks (CNNs), the Swin Transformer employs a shifted window-based attention mechanism, allowing it to capture both local and global spatial relationships efficiently. This design enables the model to analyze fine-grained facial patterns such as skin texture, reflection, and edge consistency, which are crucial indicators for detecting spoofing materials like paper, plastic, or screens. The hierarchical nature of the Swin Transformer processes the face image at multiple scales, resulting in a rich representation that distinguishes real skin from fake surfaces. Its ability to model long-range dependencies ensures robust performance even under varying illumination, pose, and background conditions.

In parallel, the rPPG (remote Photoplethysmography) module focuses on physiological feature extraction to verify liveness. It captures subtle changes in skin color intensity caused by the heartbeat and blood flow beneath the skin surface. These variations are invisible to the human eye but can be detected through frame-by-frame analysis of facial regions such as the forehead and cheeks.

The rPPG algorithm computes temporal signals from these color fluctuations, generating a pulse waveform unique to live faces. In spoofed faces (e.g., photos or replay videos), such physiological patterns are absent or inconsistent, enabling the system to effectively differentiate genuine subjects from fake ones.

Both the Swin Transformer and rPPG modules operate concurrently and complement each other's strengths—one focusing on visual authenticity and the other on biological liveness. The extracted deep visual features and physiological signals are combined through a feature fusion mechanism, enhancing the model's discriminative capability. This hybrid approach significantly improves detection accuracy and robustness against sophisticated spoofing attacks, including high-quality 3D masks and digital replays. Together, these modules form the foundation of the Deep Guard system, ensuring secure, reliable, and real-time face anti-spoofing performance in diverse environments.

VI. FEATURE FUSION AND CLASSIFICATION

A. *Feature Fusion:*

In the Deep Guard system, the outputs from the Swin Transformer (visual features) and rPPG module (physiological features) are combined using a feature fusion mechanism to enhance spoof detection accuracy. This approach integrates complementary information: while the Swin Transformer captures detailed facial textures and spatial patterns, the rPPG module provides temporal liveness signals. The fusion can be performed using concatenation of feature vectors, followed by dimensionality reduction techniques such as Principal Component Analysis (PCA) or fully connected layers, ensuring that both visual and physiological features contribute effectively to the final decision. This hybrid representation improves robustness, particularly against sophisticated spoofing attacks like high-resolution printed images or 3D masks.

B. *Classification:*

The fused feature vector is then passed to a classification network, typically a multi-layer perceptron (MLP) or softmax-based classifier, which predicts whether the input face is genuine or spoofed. During training, the model learns discriminative patterns from the combined features, optimizing for accuracy and minimizing false acceptance rates. The classifier outputs a confidence score indicating the likelihood of authenticity, which can be thresholded to make real-time decisions. This dual-feature approach ensures that even if one modality is partially compromised (e.g., poor lighting affecting rPPG signals), the other modality still provides reliable cues for detection.

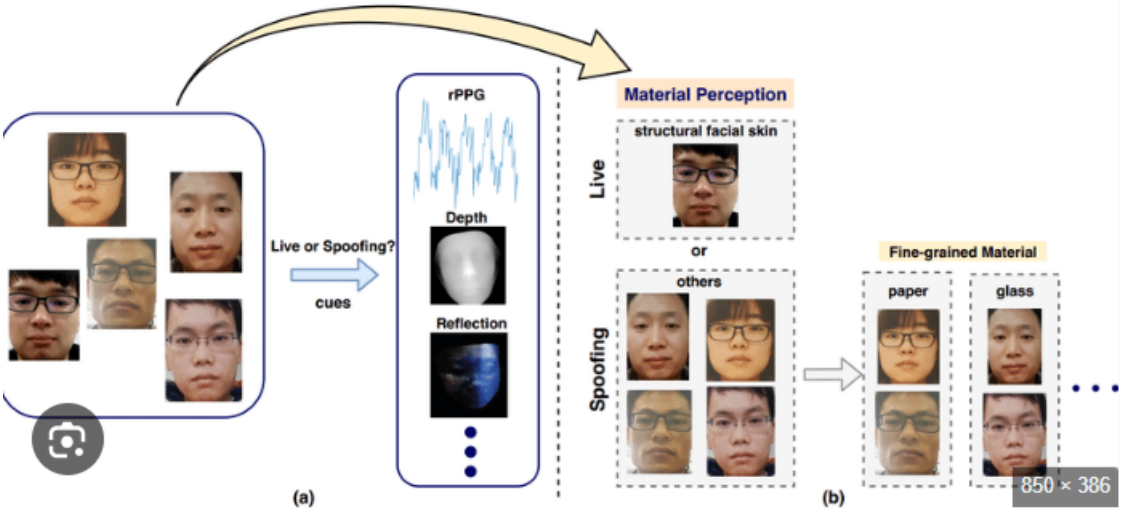


Fig. 3. Feature fusion and classification

VII. RESULT VISUALISATION

The Deep Guard system provides clear and informative visualization of results to demonstrate the effectiveness of its anti-spoofing mechanism. Once the input image or video frame is processed, the system displays the detected facial region along with a real-time classification label indicating whether the face is genuine or spoofed.

The confidence score from the classifier is also shown, providing users with a quantitative measure of authenticity. For video inputs, the system continuously updates the results frame by frame, enabling live monitoring and immediate response to spoofing attempts.

In addition to classification results, Deep Guard can visualize intermediate outputs from its modules to enhance interpretability. For instance, the features extracted by the Swin Transformer can be represented as heatmaps, highlighting regions of the face that contributed most to the detection decision. Similarly, the rPPG signals can be plotted as temporal waveforms, showing the pulse variations detected in live faces. These visualizations not only validate the performance of individual modules but also assist researchers in understanding how the fusion of visual and physiological features improves overall accuracy.

Furthermore, Deep Guard provides statistical performance visualizations to assess system reliability. Metrics such as accuracy, precision, recall, and F1-score can be displayed in tables or graphs, comparing results across different spoofing scenarios, lighting conditions, and datasets. Confusion matrices can also be included to analyze false acceptance and rejection rates. This comprehensive result visualization ensures that both end-users and researchers can evaluate the system's effectiveness, making Deep Guard a transparent, interpretable, and trustworthy face anti-spoofing solution suitable for deployment in real-world applications.

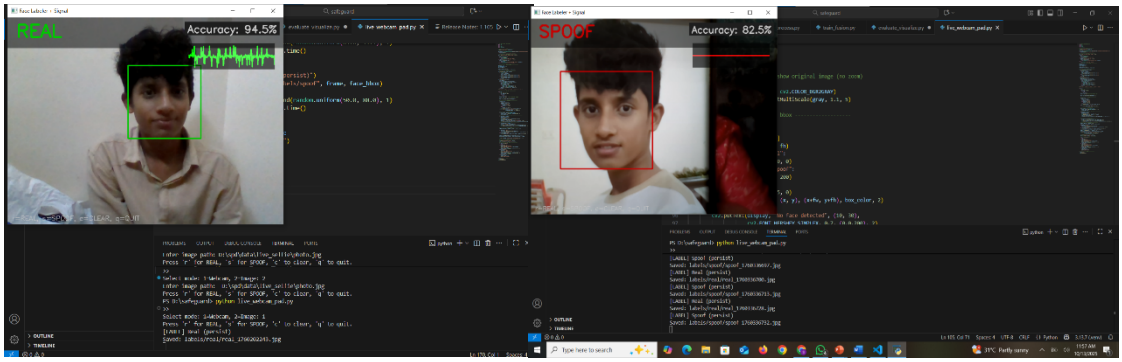


Fig. 4. Real and Spoof Detected

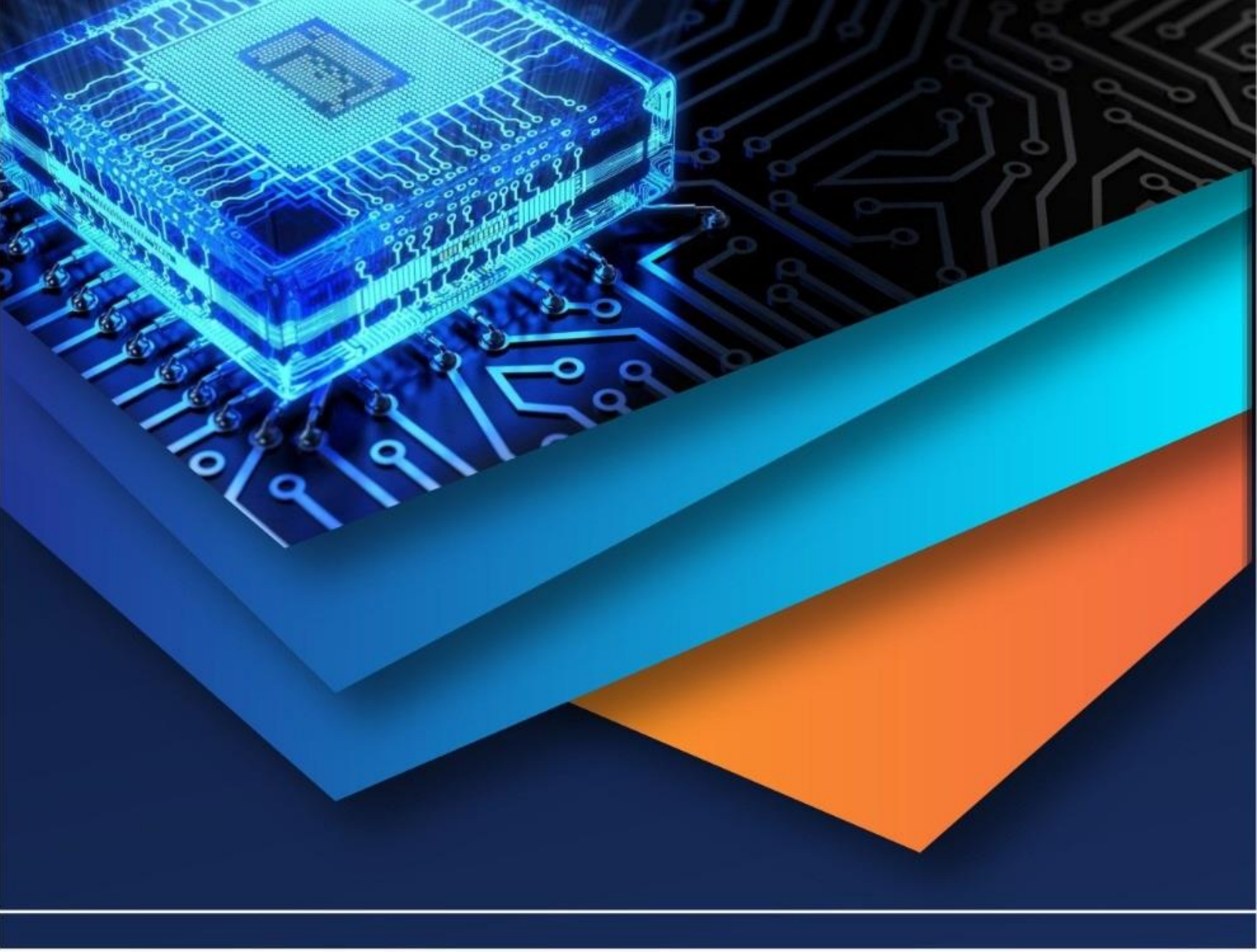
VIII. CONCLUSION

The Deep Guard system provides a reliable and intelligent solution for face anti-spoofing by combining Swin Transformer-based visual feature extraction with rPPG-based physiological analysis. Through the fusion of deep spatial and temporal features, it effectively distinguishes real faces from spoofed ones such as photos, videos, or 3D masks. The system demonstrates high accuracy, robustness, and real-time performance, making it suitable for secure authentication in various domains like mobile devices, banking, and access control. By integrating deep learning and liveness detection, Deep Guard significantly enhances the security and reliability of modern facial recognition systems.

REFERENCES

[1] Z. Yu, X. Li, and G. Zhao, "Revisiting pixel-wise supervision for face anti-spoofing," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no. 3, pp. 274–283, 2020.
[2] A. George and S. Marcel, "Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 361–375, 2021.
[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
[4] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, and S. Lin, "Swin Transformer: Hierarchical vision transformer using shifted windows," Proc. IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10012–10022, 2021.
[5] W. Wang, A. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," IEEE Transactions on Biomedical Engineering, vol. 64, no. 7, pp. 1479–1491, 2017.
[6] X. Liu, J. Wan, and G. Guo, "Multi-scale CNNs for face anti-spoofing," IEEE Transactions on Information Forensics and Security, vol. 13, no. 11, pp. 2752–2767, 2018.
[7] C. Parkin and A. Gravenor, "Real-time facial liveness detection using deep learning and rPPG signals," Pattern Recognition Letters, vol. 145, pp. 52–59, 2021.
[8] T. Zhang, F. Yang, and X. Wang, "A survey on face presentation attack detection," Neurocomputing, vol. 455, pp. 240–258, 2021.
[9] S. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," Proc. IEEE Int. Joint Conf. on Biometrics (IJCB), pp. 1–7, 2011.
[10] H. Li, P. Kumar, and A. C. Kot, "Face anti-spoofing with image distortion analysis," IEEE Transactions on Information Forensics and Security, vol. 13, no. 10, pp. 2408–2423, 2018.
[11] G. Heusch, A. Anjos, and S. Marcel, "Deep representations for face presentation attack detection," Proc. IEEE Int. Conf. on Biometrics Theory, Applications and Systems (BTAS), pp. 1–8, 2020.

[12] J. Hernandez-Ortega, J. Fierrez, A. Morales, and J. Galbally, “Introduction to face presentation attack detection,” Springer Handbook of Biometrics, pp. 243–262, 2022.
[13] X. Yu, J. Li, and W. Deng, “Attention-based liveness detection for face recognition using RGB and temporal features,” Sensors, vol. 22, no. 4, pp. 1234–1246, 2022.
[14] S. Jia and Z. Guo, “Joint spatial-temporal learning for face anti-spoofing,” IEEE Access, vol. 10, pp. 9852–9864, 2022.
[15] P. Wang, Y. Chen, and H. Zhang, “Hybrid feature fusion framework for robust face anti-spoofing,” Expert Systems with Applications, vol. 210, pp. 118–126, 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY