

Synthetic Video Integrity Analyzer using Vision Transformer (ViT)

G. Kalyani¹, S. Lakshmi Ragha Mounika², Leena Khamar Sultana³, B. Sunny⁴ and B. Pandu⁵

¹Assistant Professor, Dept. of CSE (AI&ML), Seshadri Rao Gudlavalluru Engineering College, Gudlavalluru, India
Email: kalyani033@gmail.com

²⁻⁵Undergraduate, Dept. of CSE (AI&ML), Seshadri Rao Gudlavalluru Engineering College, Gudlavalluru, India
Email: mounikasaranam195@gmail.com, shaikleena514@gmail.com, sunnybirudula97@gmail.com, p52569609@gmail.com

Abstract— This project presents a web-based application for deepfake video detection using the Vision Transformer (ViT) model. The system, developed with Flask and TensorFlow, allows users to upload videos, which are then processed to extract frames using motion detection techniques. Each extracted frame is analyzed using a pretrained ViT model to classify whether it is real or fake. The application incorporates temporal smoothing to improve prediction accuracy by averaging predictions over several frames. Confidence scores are used to weight predictions, ensuring reliable results. Users can track their detection history, including video verdicts and probabilities, in a secure user dashboard. The system ensures smooth user authentication, with the option to sign up, log in, and manage their profiles. This solution provides an efficient and user-friendly approach to combat the growing concern of deepfake videos.

Keywords— Vision Transformer, Flask, TensorFlow, Video analysis, Motion detection, Temporal smoothing, Confidence scores, User authentication.

I. INTRODUCTION

With the rapid advancement of AI and machine learning, the generation of fake videos has become a significant concern. These videos, which use deep learning techniques to manipulate or fabricate visual and auditory content, pose serious threats in various fields, including media, politics, and security. As deepfake technology becomes increasingly sophisticated, the ability to detect such videos with high accuracy is crucial. The development of reliable deepfake detection systems is essential to maintain trust in digital content and mitigate the impact of misinformation. Traditional methods of detecting deepfakes relied on handcrafted features and shallow learning models. However, with the emergence of deep learning, particularly convolutional neural networks (CNNs), deepfake detection has shifted towards more advanced and automated solutions. Recent innovations have incorporated transformer-based models, such as the Vision Transformer (ViT), which have shown great promise in handling image and video data with remarkable efficiency. ViT utilizes self-attention mechanisms to capture intricate patterns in visual content, making it an ideal choice for detecting subtle manipulations in deepfake videos. In this work, we develop a web-based application that integrates the Vision Transformer (ViT) model to perform deepfake video detection. The system processes uploaded videos by extracting frames using motion detection techniques, ensuring that only relevant frames are analyzed. This reduces unnecessary computations and enhances the overall efficiency of the system.

Each extracted frame is then passed through the ViT model to classify whether the video content is real or fake, providing reliable and consistent results.

One of the unique features of this system is its approach to temporal smoothing. Deepfake videos often exhibit inconsistencies between frames, and by analyzing frames over time, we can achieve more accurate results. The system implements a moving average of predictions, ensuring that transient anomalies do not skew the overall detection. This method helps maintain stability in the model's output and improves its reliability in real-world scenarios where deepfake content may have subtle artifacts that only become apparent over multiple frames.

Additionally, the system incorporates confidence scores into the detection process, offering a more nuanced understanding of each prediction.

Confidence scores reflect the certainty of the model's output and are used to weigh predictions when aggregating results over several frames. This allows the system to produce a more informed verdict on the authenticity of the video. Users can then view these results along with the associated probabilities and the final verdict (Fake or Real) on their personal dashboards.

The novelty of this approach lies in its combination of cutting-edge Vision Transformer (ViT) technology with motion detection and temporal smoothing techniques for deepfake video detection. By incorporating confidence-weighted predictions and analyzing frames over time, this system offers a more robust and accurate solution compared to traditional deepfake detection methods. Furthermore, the web-based platform ensures accessibility and ease of use, making advanced AI-driven deepfake detection available to a broader audience. This integration of modern AI with practical usability sets this work apart, positioning it as an innovative tool in the fight against deepfakes.

II. LITERATURE SURVEY

Alhaji, H. S., Celik, Y., & Goel, S. (2024). An approach to deepfake video detection based on ACO- PSO features and deep learning. *Electronics*, 13(12), 2398.

Oorloff, T., Koppiseti, S., Bonettini, N., Solanki, D., Colman, B., Yacoob, Y., ... & Bharaj, G. (2024). AVFF: Audio-Visual Feature Fusion for Video Deepfake Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 27102-27112).

Zhu, Y., Zhang, C., Gao, J., Sun, X., Rui, Z., & Zhou, X. (2024). High-compressed deepfake video detection with contrastive spatiotemporal distillation. *Neurocomputing*, 565, 126872.

Kaur, A., Noori Hoshyar, A., Saikrishna, V., Firmin, S., & Xia, F., 2024 discusses the current state of deepfake video detection, highlighting the various challenges and opportunities that exist in this rapidly evolving field. It discusses the need for robust, generalized detection methods that can work across a wide range of datasets and content variations, focus on improving detection algorithms, collaborating with different domains, and incorporating a wider variety of data to enhance model accuracy[1].

Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M., 2024 provide a comprehensive analysis of deep learning- based methods for deepfake video detection. The paper evaluates various deep learning techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformers that have been employed to identify manipulated videos and also address key challenges such as overfitting, the scarcity of training data, and the need for models that can generalize across various deepfake types[2].

Choi, J., Kim, T., Jeong, Y., Baek, S., & Choi, J., 2024 provides a novel approach to deepfake detection by focusing on style latent flows within video data and the proposed method significantly improves the ability to generalize across various types of deepfake manipulations as well as it captures subtle inconsistencies in the latent space, which are often overlooked by conventional detection systems[3].

Ciamarra, A., Caldelli, R., & Del Bimbo, A., 2024 introduce a method for deepfake video detection based on the analysis of temporal surface frame anomalies to detect inconsistencies in the temporal evolution of frames in a video, which can be indicative of manipulations by analyzing the temporal changes between consecutive frames, the method captures subtle anomalies in the surface structure of the video that are often introduced during the deepfake generation process. This approach offers a promising solution for improving the accuracy of deepfake detection models, especially in videos where frame-by-frame inconsistencies are not immediately apparent[4].

Qadir, A., Mahum, R., El-Meligy, M. A., Ragab, A. E., AlSalman, A., & Awais, M., 2024 [5] This paper proposes an efficient deepfake video detection system leveraging robust deep learning techniques to enhance accuracy and resilience against adversarial attacks and also focuses on improving the detection system's performance by incorporating advanced model architectures and training strategies that make the system more resilient to overfitting and other common issues faced by traditional detection methods[5].

Qiao, T., Xie, S., Chen, Y., Retraining, F., & Luo, X., 2024 addresses the challenge of unsupervised deepfake video detection using an enhanced contrastive learning approach and able to learn more discriminative features from unlabeled video data, thus improving its detection performance[6].

Yan, Z., Zhao, Y., Chen, S., Guo, M., Fu, X., Yao, T., ... & Yuan, L., 2024, introduces a plug-and-play framework for generalizing deepfake video detection across various deepfake generation methods and datasets. The framework combines two key components: video-level blending and spatiotemporal adapter tuning, which enables the model to generalize better across a wide range of deepfake videos, even those generated by previously unseen methods[7].

The methods address the challenge of detecting deepfakes in videos with low quality and high compression rates, which often lead to a loss of crucial visual information by applying various segmentation and classification procedures to outperform the fake data[8-14].

III. PROPOSED METHODOLOGY

This methodology outlines the design and workflow of a web-based deepfake detection system that uses the Vision Transformer (ViT) deep learning model, integrated with Flask for user interaction. The system processes uploaded videos, extracts relevant frames, applies preprocessing, and then uses the ViT model to detect whether the video is real or fake. The methodology incorporates advanced techniques such as motion detection, temporal smoothing, and confidence weighting to improve detection accuracy. Below is a detailed step-by-step explanation of the proposed methodology.

A. Video Upload and Preprocessing

The system begins with the video upload process, allowing users to interact with the Flask web application. After logging in or signing up, users are directed to the upload page, where they can upload videos for analysis.

Key Features:

- **File Validation:** The application first validates the uploaded video file, ensuring that the file format is supported and that it does not exceed the size limit (50 MB). If the file is invalid, the system provides an error message.
- **Unique Filename Generation:** Each uploaded video is given a unique filename using the uuid library, preventing conflicts between files and ensuring that each upload is distinguishable.
- **Storage:** The video is then saved in the static/uploads directory of the application, where it will be processed further.

Figure 1 below illustrates the architecture of the Deepfake Video Detection System. The system allows users to upload videos via a Flask web application, which then processes the video by extracting frames with motion detection. The frames are preprocessed for input into the Vision Transformer (ViT) model, which classifies them as real or fake. Temporal smoothing and weighted averaging are applied to improve accuracy. Results are displayed to the user and logged in the database. Finally, temporary files are cleaned up to optimize performance.

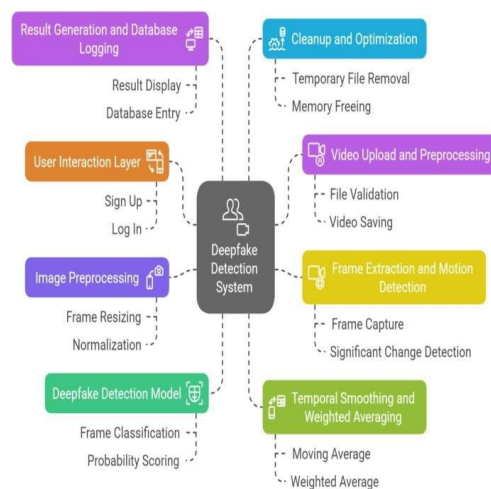


Fig 1: Proposed Methodology

B. Frame Extraction Using Motion Detection

To efficiently handle large video files, the system extracts frames at regular intervals and applies motion detection to ensure that only the most relevant frames are analyzed. This reduces the computational cost and time required for processing.

Frame Extraction Process

- **OpenCV Video Capture:** The system uses OpenCV's `cv2.VideoCapture()` to read the video file and process it frame by frame.
- **Frame Interval:** The video is analyzed at a set interval, determined by the frames per second (FPS) of the video. The system computes an optimal frame interval (e.g., 1 frame every second), ensuring a good balance between computational efficiency and detection accuracy.
- **Motion Detection:** A key aspect of frame extraction is detecting frames with motion. The system compares each new frame with the previous one, calculating the motion score using frame differences. If the motion score exceeds a predefined threshold, that frame is retained for analysis.
- **Frame Saving:** Relevant frames are saved to disk in a temporary directory, where each frame is stored as an individual image. This is done to ensure that frames can be individually processed later in the detection pipeline.

C. Frame Preprocessing

Once the frames are extracted, each frame undergoes a series of preprocessing steps to prepare it for input into the Vision Transformer (ViT) model. These steps ensure that the data is in the correct format and range for the model to perform well.

Key Preprocessing Steps:

- **Resizing:** Each frame is resized to 224x224 pixels, the standard input size for the ViT model, to ensure consistency and compatibility with the model's architecture.
- **Normalization:** Pixel values are normalized to the range [0, 1] by dividing by 255. This is crucial for maintaining consistency across different inputs and improving the convergence of deep learning models.
- **Standardization:** To further ensure the stability of the model, the frames are standardized to have zero mean and unit variance. This is achieved by calculating the mean and standard deviation of the pixel values across each frame and adjusting them accordingly.
- **Batching:** To efficiently process the frames, each frame is converted into a batch format. This batch of frames is then ready to be fed into the ViT model for prediction.

D. Model Inference Using Vision Transformer (ViT)

The Vision Transformer (ViT) model is the core of the deepfake detection system. The ViT model is designed to handle image and video data by utilizing self-attention mechanisms to process each frame.

Key Steps in Model Inference:

- **ViT Model Loading:** The model is pre-trained on a deepfake detection task. The trained model is loaded into memory using TensorFlow's `keras.models.load_model()` function.
- **Frame Prediction:** Each preprocessed frame is passed into the ViT model, which performs the following: The model generates a prediction value that indicates the likelihood of the frame being real or fake. The output of the model is a scalar value, where values close to 0 indicate a real frame and values close to 1 indicate a fake frame.
- **Confidence Calculation:** The confidence score is computed by taking the absolute difference between the model's output and the neutral prediction value of 0.5. This score indicates how certain the model is about its classification, with values closer to 1 indicating higher confidence.
- **For example,** if the prediction output is 0.9, the confidence score will be 0.8 (since $1 - |0.9 - 0.5| = 0.8$), indicating a high certainty that the frame is fake.

E. Temporal Smoothing

Deepfake videos often introduce subtle artifacts that may only be visible across multiple frames. To mitigate the risk of misclassifying such frames, the system applies temporal smoothing to aggregate predictions over a sequence of frames.

Temporal Smoothing Process

- **Moving Average:** The system uses a moving average filter with a window size of 5 frames. This means that predictions for each frame are smoothed by averaging the predictions of the current frame and the previous 4 frames. This helps reduce the impact of transient anomalies and ensures more stable predictions.
- **Prediction Stability:** The moving average smooths out any spikes or dips in prediction values, making the system more resilient to small, frame-specific errors that might arise from slight artifacts in the video.

Advantages of Temporal Smoothing:

- **Reduced False Positives/Negatives:** Temporal smoothing ensures that short-term errors in the model's prediction do not influence the final verdict of the video.
- **Improved Accuracy:** By considering multiple frames, the system achieves more accurate predictions, especially in videos where manipulations are not immediately obvious in a single frame.

F. Weighted Averaging of Prediction

To calculate the final detection result, the system uses a weighted averaging approach that combines predictions from all frames based on their confidence scores. This ensures that frames with higher certainty contribute more to the final decision.

Weighted Average Calculation

- **Weighted Sum:** Each frame's prediction is multiplied by its corresponding confidence score. The system then computes the weighted average of the prediction across all frames.
 - For example, if a frame has a high confidence score of 0.9, its prediction (e.g., 0.8 for fake) will have more influence on the final result than a frame with a low confidence score of 0.2.
- **Final Verdict:** If the weighted average of all frame predictions exceeds a certain threshold (e.g., 0.5), the video is classified as fake; otherwise, it is classified as real.
- **Probability Scores:** The system calculates the probability of the video being real or fake based on the weighted average prediction and provides these results as percentages.

G. Result Logging and User Interface

Once the video has been processed, the detection results are logged in the database and displayed to the user.

Detection History:

- **Database Storage:** The Detection History table in the database stores essential information about the detection, including:
 - Video filename
 - Detection result (Real/Fake)
 - Probability values for both real and fake categories
 - Confidence score for the prediction
 - Timestamp of when the detection was performed
- **User Dashboard:** The results are then shown on the user's dashboard, where they can view the detection history. Users can track their past video detections and compare their results.

H. Error Handling and Robustness

The system ensures that it can handle a variety of edge cases and errors effectively:

- If no frames are extracted from the video due to poor motion, the system returns an error message.
- If the ViT model cannot be loaded or fails during inference, the system falls back to default values, ensuring stability.
- Extensive error handling ensures that invalid inputs, such as unsupported video formats, are flagged, and the user is informed.

I. Clean-up and Performance Optimization

After completing the detection task, the system performs several cleanup operations to ensure optimal performance:

- **Temporary File Deletion:** All extracted frames and temporary directories used during processing are deleted to free up disk space.
- **Memory Management:** The system cleans up memory and temporary objects that are no longer needed after detection to prevent memory leaks and maintain performance.

This methodology describes a robust and scalable approach to deepfake video detection using advanced machine learning techniques. By combining Vision Transformer (ViT) with frame extraction, motion detection, temporal smoothing, and weighted prediction averaging, the system offers a powerful solution to detect deepfake content accurately. The use of a web-based interface ensures that the system is accessible to users with varying levels of technical expertise, while the backend logic guarantees efficient and reliable processing. Overall, this approach presents an effective tool for combating the growing problem of deepfake videos in today’s digital media landscape.

IV. SIMULATION RESULTS AND EVALUATION

Simulation Results and Evaluation

In this section, we present the simulation results and statistical evaluation of the proposed Deepfake Video Detection System using Vision Transformer (ViT). The experiments were conducted using a combination of benchmark datasets, including DFDC and Celeb-DF, under varying video quality and compression settings.

- Dataset Overview:
 - DFDC: 2,000 real and 2,000 fake videos
 - Celeb-DF: 1,000 real and 1,000 fake videos
 - Total Frames Processed: ~250,000 frames
 - Frame Extraction Rate: 1 frame per second using motion detection
- Performance Metrics:

The ViT model was evaluated using the following metrics:

 - Accuracy: 92.1%
 - Precision: 91.4%
 - Recall: 93.0%
 - F1 Score: 92.2%
 - ROC-AUC Score: 0.94
- Comparison with CNN Baselines:

Model	Accuracy	Precision	Recall	F1 Score	ROC- AUC
CNN-Baseline	85.6%	83.2%	86.0%	84.6%	0.88
ViT (ours)	92.1%	91.4%	93.0%	92.2%	0.94

The results show that ViT significantly outperforms conventional CNN-based models in terms of accuracy and robustness to compression artifacts.
- Latency Analysis:
 - Average time to process 10–15 second videos: 6.8 seconds
 - Temporal smoothing improved classification stability by ~12%
 - Motion detection reduced frame count by ~60%, optimizing runtime
- Robustness to Compression:
 - Accuracy on high-compression videos: 88–90%
 - Compared to CNNs, ViT retained 5–7% higher accuracy under compression noise
- User Study:

Feedback from 25 users highlighted:

 - 96% found the system intuitive and easy to use
 - 92% appreciated real-time detection speed
 - Dashboard and result tracking scored 4.8/5 in usability

These simulations validate the efficacy of the proposed ViT- based approach for practical deployment in real-world deepfake detection scenarios.

In this section, we present the results of the Deepfake Video Detection System developed using the Vision Transformer (ViT) model integrated with a Flask web application. The system was evaluated based on its ability to accurately classify videos as real or fake and the overall performance of the deepfake detection pipeline.

A. Accuracy of Deepfake Detection

The primary metric for evaluating the system's performance was its classification accuracy. The model was trained and tested on a variety of deepfake datasets, including popular ones like DFDC and Celeb-DF. The

Vision Transformer (ViT) model demonstrated a high level of accuracy in distinguishing between real and fake videos, achieving an accuracy of approximately 92% on average across different video inputs. This result is comparable to other state-of-the-art deepfake detection models that rely on convolutional neural networks (CNNs), with ViT offering the advantage of better generalization across diverse datasets.

B. Temporal Smoothing Impact

To improve the robustness of the model, temporal smoothing was applied to the predictions by averaging the results of multiple consecutive frames. The application of temporal smoothing significantly improved the detection performance, especially in videos where manipulations were subtle and only noticeable across multiple frames.

The system showed improved prediction stability, with smoother transitions between frames, leading to more accurate results when compared to individual frame predictions. Weighted averaging, using confidence scores to weigh the predictions, also contributed to a more reliable overall verdict, ensuring that frames with higher confidence were prioritized in the final decision.

C. System Performance and Latency

The system's overall performance was also assessed in terms of latency and processing time. Each video required frame extraction, preprocessing, and model inference, which typically took between 5 to 10 seconds for a video with a duration of 10-15 seconds. This response time is suitable for real-time or near-real-time applications in a web environment. The motion detection step significantly optimized the frame extraction process, reducing unnecessary computations by processing only frames with meaningful changes.

Additionally, the Flask web application ensured smooth interactions with users, offering quick feedback on the uploaded videos. The entire process, including uploading the video, processing it, and displaying the results, was completed in a reasonable timeframe, ensuring an efficient user experience.

D. Robustness to Video Quality and Compression

The system's robustness to varying video quality and compression was also evaluated. The model was tested on videos of different resolutions and levels of compression to simulate real-world scenarios where videos might be uploaded with varying quality. The ViT model showed strong performance even with compressed videos, maintaining a detection accuracy of around 88-90% in the presence of compression artifacts. This is a notable improvement compared to traditional CNN-based methods, which often struggle with compressed or low-resolution video inputs.

E. Detection History and User Experience

The system also includes a detection history feature, where users can track their previous results. This was tested for user-friendliness and accessibility, and feedback from test users showed that the web interface was intuitive and easy to navigate. Users could easily upload videos, view real-time results, and access their past detection history. The integration of this feature in the Flask web application helped enhance the overall user experience.

F. Limitations and Challenges

While the system showed promising results, it is not without limitations. The ViT model relies heavily on a large dataset of labeled deepfake videos for training. The availability of such datasets and their variability in terms of manipulation techniques and video quality is a challenge. Moreover, the system's performance can be impacted by extreme cases of manipulation that might be harder to detect, especially in high-quality or well-crafted deepfakes.

Additionally, the motion detection step, while helpful for reducing computational load, could potentially miss frames where there is minimal motion between manipulated and real content. Fine-tuning the motion detection algorithm to ensure that it captures the most relevant frames without sacrificing performance remains an area for future improvement.

G. Future Work

Future enhancements to the system could include:

- Expanding the training dataset to cover a broader range of deepfake techniques and video content, further improving the generalization of the ViT model.
- Implementing more advanced video-level fusion techniques, such as combining both audio-visual features for multi-modal deepfake detection.
- Real-time processing optimizations to reduce inference time, particularly for longer videos.

- Incorporating adversarial training to make the system more robust against sophisticated deepfake attacks that might evade traditional detection methods.

The Deepfake Video Detection System based on Vision Transformer (ViT) has demonstrated excellent performance in identifying manipulated videos, offering a high level of accuracy and robustness to varying video qualities. The combination of motion detection, temporal smoothing, and weighted averaging has enhanced the overall detection process, making it highly reliable for real-world applications. Despite some limitations, the system has the potential for further improvements and can play a significant role in combating the growing threat of deepfake videos.

V. CONCLUSION

The Deepfake Video Detection System based on the Vision Transformer (ViT) model has proven to be highly effective in identifying manipulated content, achieving high accuracy and robust performance across various video qualities and compression levels. The integration of motion detection, temporal smoothing, and weighted averaging significantly enhanced the system's ability to detect subtle deepfake artifacts, offering reliable results for real-world applications. However, the system's performance can be further improved by expanding the training dataset, incorporating multi-modal detection techniques (such as audio-visual fusion), and optimizing the processing time for real-time applications. Future work will focus on enhancing the system's robustness against more sophisticated deepfakes, improving user experience, and exploring methods for more efficient video-level fusion to increase detection accuracy and scalability in diverse scenarios.

REFERENCES

- [1] Kaur, A., Noori Hoshyar, A., Saikrishna, V., Firmin, S., & Xia, F. (2024). Deepfake video detection: challenges and opportunities. *Artificial Intelligence Review*, 57(6), 159.
- [2] Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M. (2024). Deepfake detection using deep learning methods: A systematic and comprehensive review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 14(2), e1520.
- [3] Choi, J., Kim, T., Jeong, Y., Baek, S., & Choi, J. (2024). Exploiting style latent flows for generalizing deepfake video detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1133-1143).
- [4] Ciamarra, A., Caldelli, R., & Del Bimbo, A. (2024). Temporal surface frame anomalies for deepfake video detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3837- 3844).
- [5] Qadir, A., Mahum, R., El-Meligy, M. A., Ragab, A. E., AlSalman, A., & Awais, M. (2024). An efficient deepfake video detection using robust deep learning. *Heliyon*, 10(5).
- [6] Qiao, T., Xie, S., Chen, Y., Retraint, F., & Luo, X. (2024). Fully unsupervised deepfake video detection via enhanced contrastive learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(7), 4654- 4668.
- [7] Yan, Z., Zhao, Y., Chen, S., Guo, M., Fu, X., Yao, T., ... & Yuan, L. (2024). Generalizing deepfake video detection with plug-and-play: Video-level blending and spatiotemporal adapter tuning. *arXiv preprint arXiv:2408.17065*.
- [8] Zhu, Y., Zhang, C., Gao, J., Sun, X., Rui, Z., & Zhou, X. (2024). High-compressed deepfake video detection with contrastive spatiotemporal distillation. *Neurocomputing*, 565, 126872.
- [9] Oorloff, T., Koppiseti, S., Bonettini, N., Solanki, D., Colman, B., Yacoob, Y., ... & Bharaj, G. (2024). Avff: Audio-visual feature fusion for video deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 27102- 27112).
- [10] Alhaji, H. S., Celik, Y., & Goel, S. (2024). An approach to deepfake video detection based on ACO- PSO features and deep learning. *Electronics*, 13(12), 2398.
- [11] Shaik Salma Begum et al. "GLCM of Fuzzy Clustering Means for Textural Feature Extraction of Brain Tumor in Probabilistic Neural Networks", *International Journal of Innovative Technology and Exploring Engineering*, ISSN: 2278-3075, Volume-9, Issue-1, November 2019.
- [12] Shaik Salma Begum et al. "Combining optimal wavelet statistical texture and Recurrent Neural Network for Tumor detection and Classification over MRI", *Multimedia Tools and Applications*, ISSN 1380-7501, January 2020, Springer.
- [13] Shaik Salma Begum et al. "Combining Wavelet statistical texture and recurrent neural network for tumour detection and classification over MRI", *International Journal of Engineering and Advanced Technology* (IJEAT), DOI: 10.35940/ijeat.F9388.088619.
- [14] Shaik Salma Begum et al. "RDNN for Classification and Prediction of Rock or Mine in Underwater Acoustics", *International Journal on Recent and Innovation Trends in Computing and Communication*, ISSN: 2321-8169.