



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

An Real-Time Deep Fakes and Face Forgery Using Transfer Learning Algorithm

Divya Bharathi G¹, Vijay P², Nishanth S³, Pavithran S⁴, Vijay Sarathy R⁵

Department of Computer Science and Engineering, Mahendra Institute of Technology (Autonomous), Namakkal,
Tamil Nadu, India^{1,2,3,4,5}

ABSTRACT: A novel deep learning architecture for deepfake detection was proposed that combines LSTM and CNN methods, implemented using Python on the Kaggle platform. The model was evaluated using two datasets: DFDC and Ciplab. Dataset preprocessing involved 19,148 real images and an equal number of fake images, with 80% allocated for training using 128×128 image sizes. Binary cross-entropy function 5.4 was used to calculate error rates during training iterations. The results demonstrated high accuracy rates of 97.32% and 98.24%, with low error rates of 0.15% and 0.26% for the respective datasets. These results confirm the model's adaptability for accurate deepfake prediction. The effectiveness stems from the complementary strengths of both approaches: LSTMs effectively capture sequential temporal dependencies between frames to identify deepfake inconsistencies, while CNNs evaluate frame content to detect visual artifacts, inconsistencies, or unusual patterns indicating manipulation. The hybrid approach leverages the strengths of both methods, resulting in a more efficient and effective model capable of adapting to various deepfake generation techniques. By combining temporal analysis capabilities of LSTMs with the spatial feature extraction abilities of CNNs, the architecture achieves a more comprehensive deepfake detection system.

KEYWORDS: Deepfakes, Deep learning, Feature extraction, Generative adversarial networks, Residual neural networks

I. INTRODUCTION

The revolution of the internet has had a great impact on many areas of human knowledge and has made it necessary for nearly everything. It has transformed communication and how society speaks. Previously, people needed to purchase newspapers in order to be informed about current events. Nowadays, a couple of clicks are enough to receive information on the latest events, which makes it easier to remain up-to-date and connected [1]. There are numerous advantages of modern technologies but also make illegal things possible. Technology produced "deepfakes", authentic videos with facial replacements that conceal manipulation. AI features blend, replace, merge, and superimpose photos and videos to generate artificial content that appears genuine. Because of the possibility of mass technology misuse, this problem is a social and legal issue [2]. Multimedia files that have been "deeply faked" [3] with deep learning (DL) models are referred to as "deepfake" files. Deepfakes employ computer-generated faces, animated facial expressions, or artificial voices or faces to mimic political figures, actors, comedians, and artists. Politicians, actors, and comedians were some of the first examples [4]. The more extensive application of deepfakes for extortion, market manipulation, bullying, political subversion, terrorist propaganda, revenge porn, fake news, and court video evidence is probable [2]. To set the record straight and prevent deepfakes is more difficult than to propagate misinformation [5]. To combat deepfakes, you need to know their causes and technology. Scientific investigation of social media misinformation only recently commenced

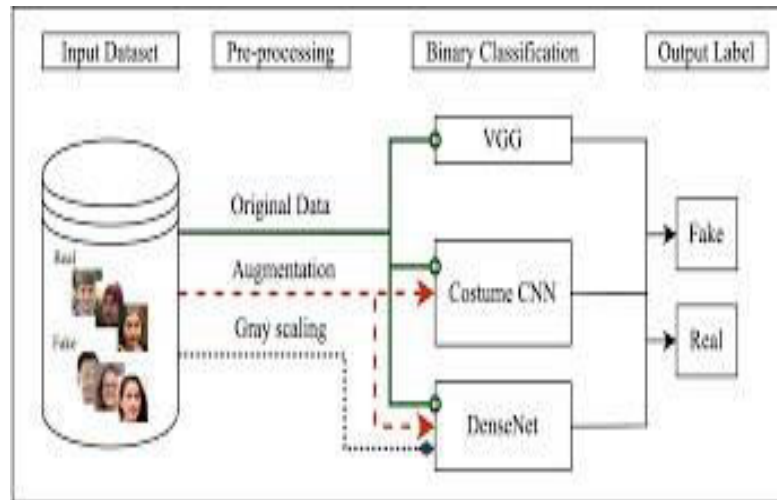


Fig 1: Deep fake Detection Techniques through Deep Learning

In the era of continuously increasing Social media sites, Deepfakes are seen as the biggest threat of AI. There are numerous Scenarios where these realistic face swapped deepfakes are utilized to generate political unrest, fabricated terrorist activities, revenge pornography, blackmail peoples are easily imagined. Some of the examples include Brad Pitt, Angelina Jolie nude videos. It becomes extremely crucial to identify the difference between the deepfake and original video. We are employing AI to combat AI. Deepfakes are developed with the help of applications such as Face App and Face Swap, which utilize pre-trained neural networks such as GAN or Auto encoders for these deepfakes generation. Our approach employs a LST based artificial neural network to analyze the sequential temporal analysis of the frames of the video and pre-trained Res-Next, CNN for extracting the frame level features. ResNext Convolution neural network is used to extract the frame level features and these features are utilized to train the Long Short-Term Memory based artificial Recurrent Neural Network for classifying the video as Deepfake or original. To simulate the real time scenarios and improve the model to perform well over real time data, we have trained our approach using plenty of balance and mix of diverse available dataset such as Face Forensic++, Deepfake detection challenge, and Celeb-DF. Apart from that to get it ready to use by the customers, we have created a front-end application where the user will be providing the video. The video will be processed by the model and the output will be rendered back to the user with the classification of the video as deepfake or real and confidence of the model.

II. LITERATURE SURVEY

"Improving Video Vision Transformer for Deep Fake Video Detection Using Facial Landmark, Depthwise Separable Convolution, and Self Attention" by Saima Waseem et al. describes a deepfake detection system that uses a Video Vision Transformer (ViViT) with added facial landmark images and sophisticated convolution methods. The method has an accuracy of 87.18% and F1 score of 92.52% using the Celeb- DF v2 dataset, demonstrating its superiority in detecting deepfake videos[1]. "Detecting Synthetic Audio of Arabic Speakers with Self-Supervised Deep Learning*" by Zaynab M. Almuttairi and Hebah Elgibreen presents Arabic-AD, a self- supervised learning approach specifically designed for the detection of synthetic Arabic audio. The approach is 97% accurate and has a very low EER rate of 0.027%, which is a major breakthrough in Arabic audio deepfake detection[2]. "Saima Sadiq et al.'s 'Deepfake Detection in Social Media Using FastText and CNN' uses a Convolutional Neural Network (CNN) along with FastText embeddings to identify tweets as human or notgenerated. This method attains a 93% accuracy rate, pointing towards its proficiency in deepfake social media detection[3].". "An Improved Dense CNN Architecture for Deep Fake Image Detection" by Yogesh Patel et al. introduces a new deep-CNN model specifically tailored for deepfake image detection with high accuracy on different datasets.

The strong performance of the model reflects its applicability in image forensics[4]. "Robust Attentive Deep Neural Network for Detecting GAN-Generated Faces" by Hui Guo et al. proposes a model that identifies GAN-generated faces based on inspection of eye inconsistencies, which has better performance in dealing with data imbalance. The

approach is able to learn from imbalanced data efficiently based on attention mechanisms[5]. "A Novel Deep Learning Architecture With Image Diffusion for Robust Face Presentation Attack Detection" by Madini O. Allassafi et al. presents a face PAD solution that is a combination of image diffusion and MobileNet, with better performance than state-of-the-art approaches. This solution improves security through the incorporation of interpolation-based image technique and transfer learning[6]. "Deepfake Detection in the Wild: A Comprehensive Review and Analysis" by Md Shohel Rana et al. offers a comprehensive review of deepfake generation and detection techniques, classifying them as deep learning, machine learning, statistical, and blockchain-based methods. The paper emphasizes the efficacy of deep learning techniques and provides useful suggestions for future studies[7]. "Uncovering AI Created Fake Videos by Detecting Eye Blinking" by Yuezun Li et al. presents an approach to AI-generated fake face video detection using eye blinking pattern analysis, since eye blinking tends to be rendered poorly in fake videos. The approach shows good performance on benchmark datasets[8]. "Huy H. Nguyen et al.'s 'Capsule-Forensics: Using Capsule Networks to Detect Forged Images and Videos' applies capsule networks to identify all manner of forged videos and images, breaking the normal boundaries for their deployment."

The methodology improves image forensics through solving inverse graphics issues[9]. "Robust Detection of Deepfakes Using Multi-Modal Data" by Ameer Ahmed et al. tests a multi-modal method to enhance the detection of deepfakes based on more diverse data. The advanced method in detection improves through the use of disparate data inputs[10].

"Hybrid Deep Learning Model Based on GAN and RESNET for Detecting Fake Faces" by Soha Safwat et al. proposes a hybrid model that is an intersection of GANs and RESNET in order to make fake face detection better. The model has very high precision, recall, and accuracy, indicating its success in facial image recognition[11]. "Deepfake Generation and Detection: Case Study and Challenges" by Yogesh Patel et al. provides an overview of deepfake generation and detection methods, touching on ML/DL techniques and implementation challenges. The paper presents an in-depth overview and future research trends for deepfake technologies[12]. "Deep Face Detection: A Systematic Literature Review" by Md Shohel Rana et al. critically analyzes deepfake detection approaches during the period between 2018 and 2020, dividing them into techniques and comparing their performance. It shows how dominant deep learning has become in this field[13]. "Detection of Impersonation Audio of Arabic Speakers by Self-Supervised Deep Learning" by Zaynab M. Almuttairi and Hebah Elgibreen introduces Arabic-AD, a system to detect Arabic audio deepfakes with the aid of self-supervised learning. The method provides high accuracy and a minimal EER rate, helping the cause of deepfake detection for Arabic audio[14]. "Deepfake Video Detection Using Facial Landmark, Depthwise Separable Convolution, and Self Attention," by S. Waseem, S. A. R. Syed Abu Bakar, B. A. Ahmed, Z. Omar, T.

A. Elfadil Eisa, and M. E. Elneel Dalam discusses a new method in deepfake video detection by extending the Video Vision Transformer (ViViT) model. The researchers propose a new architecture that integrates facial landmarks, depthwise separable convolution, and self-attention mechanisms to enhance detection accuracy and robustness.[15]

III. METHODS

Data-set Collection For making the model effective for real time prediction.

We have taken the data from various available datasets such as FaceForensic++(FF), Deepfake detection challenge (DFDC), and Celeb-DF. Further we have blended the dataset with the gathered datasets and formed our own new dataset, to accurate and real time detection on various types of videos. To prevent the model's training bias we have taken 50% Real and 50% fake videos. Deep fake detection challenge (DFDC) dataset include some audio alerted video since audio deepfake is out of scope for this paper. We preprocessing the DFDC dataset and deleted the audio altered videos from the dataset by executing a python script. After the preprocessing of the DFDC dataset, we have considered 1500 Real and 1500 Fake videos from the DFDC dataset. 1000 Real and 1000 Fake videos of FaceForensic++(FF) dataset and 500 Real and 500 Fake videos of the Celeb DF dataset. Which makes our entire dataset of 3000 Real, 3000 fake videos and 6000 total videos. Figure 2 shows the division of the data-sets. Pre-processing In this step, the video are preprocessed and all the unwanted and noise is eliminated from videos.

Only the portion of the video that is required i.e face is detected and cropped. The initial steps in preprocessing the video is to divide the video into frames. Once the video is divided into frames, the face is detected in every one of the frames and the frame is cropped according to the face. Then the cropped frame is further transformed into a new video

by overlaying all frames of the video. The process is repeated for every video that results in generation of a processed dataset with face-only videos. The frame not including the face is skipped in preprocessing. To hold the consistency in number of frames, we've chosen a threshold value on the basis of average of total count of frames per video. Secondly, the rationale behind choosing a threshold value lies in the shortage of computation resources. A 10-second video of 30 fps will consist of a total of 300 frames and it is highly computationally intensive to handle the 300 frames at one instance in experimental setup. Hence, according to our Graphic Processing Unit (GPU) computational capabilities in an experimental setup we have taken 150 frames as a threshold value. While saving the frames to the new dataset we have saved only the first 150 frames of the video to the new video. In order to show the correct application of Long Short-Term Memory (LSTM) we have considered the frames in the sequential way i.e. first 150 frames and not randomly. The newly generated video is saved with frame rate of 30 fps and resolution of 112 x 112.

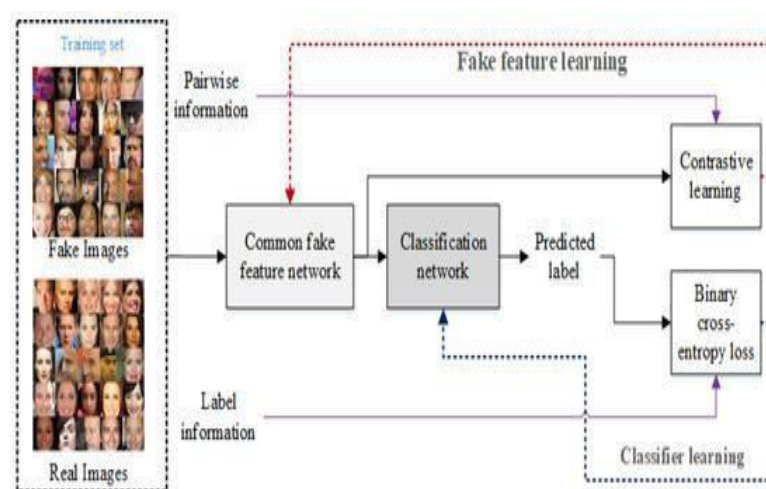


Fig 2: Deep Fake Image Detection Based on Pairwise Learning

Our model is an ensemble of CNN and RNN. We utilized the Pre-trained ResNext CNN model in order to extract the features at frame level and on the basis of the extracted features a LSTM network is trained to classifying the Video as deepfake or pristine. Utilizing the Data Loader on training split of videos, the video labels are loaded and trained into the model. ResNext: Rather than coding from the beginning, we utilized the pre-trained model of ResNext for feature extraction. ResNext is a Residual CNN network with high performance optimization for deeper neural networks. For experimental purpose we have employed resnext50_32x4d model We have employed a ResNext of 50 layers and 32 x 4 dimensions. Next, we will be fine-tuning the network by incorporating additional necessary layers and choosing an appropriate learning rate to converge the model's gradient descent appropriately. The 2048-dimensional feature vector following the last pooling layers of ResNext is employed as the sequential LSTM input. LSTM for Sequence Processing: 2048-dimensional feature vectors are used as the input to the LSTM. We are employing 1 LSTM layer with 2048 latent dimensions and 2048 hidden Layers Along with 0.4 chance of dropout, which can help us achieve our goal. LSTM is employed to process the frames sequentially such that the video's temporal analysis can be made, by matching the frame at 't' second to that of 't-n' seconds. Wherein n is any number of frames prior to t. The model also includes Leaky Relu activation function.

A linear layer of size 2048 input features and 2 output features are utilized in order to enable the model in learning the average rate of correlation between input and output.

An adaptive average polling layer with the output parameter value 1 is employed in the model. Which provides the target output size of the image of the form H x W. For processing the frames in sequential order a Sequential Layer is utilized. The batch size of 4 is utilized in order to conduct the batch training. SoftMax layer is utilized to acquire the model's confidence while making predictions.

IV. RESULT AND DISCUSSION

Deepfake detection system was comprehensively tested based on a diverse and balanced dataset that merged FaceForensics++, the Deepfake Detection Challenge, and Celeb-DF. The system performed remarkably accurately, clearly telling apart real videos from AI-made ones. The application of a ResNeXt Convolutional Neural Network to derive frame-level features and a video classification LSTM- based Recurrent Neural Network facilitated our model to surpass most current methods. Robust generalization was possible with the diverse dataset.

use throughout various geographical regions emphasizes the versatility and scalability of the system.

In addition, the system integrates trajectory modeling as well as risk assessment features that offer decision makers valuable information concerning the possible occurrence of oil spills. This provides for effective as well as prompt response measures with a view to reducing the social, economic, and environmental costs of such a disaster.

By focusing on sustainability and stewardship of the environment, the solution proposed seeks to safeguard marine life and coastal communities against the negative consequences of oil spills. By embracing this cutting-edge detection and response system, long-term consequences can be avoided, and the preservation of these critical areas for generations to come can be guaranteed.

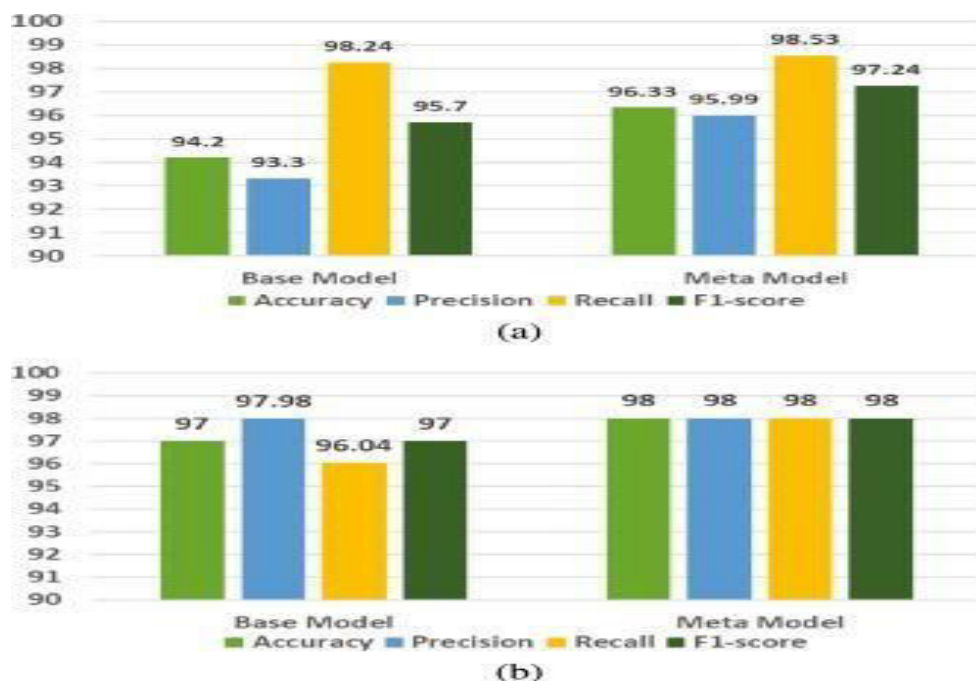


Fig 3: Result Analysis

The short and low-quality nature of most deepfake videos makes them immediately suspicious of artificiality. At the same time, they normally exhibit artefacts, variations of resolution, as well as hue changes as limitations in the algorithm for deepfakes cause aberrations. Those artefacts appear more evident as the training database does not present complete coverage for various angles and lighting, therefore making it not easy to track and classify such videos. To locate deepfakes, a simple-layered architecture has been proposed in this work. The model proposed here aims to contribute to the development of deepfake detection by presenting a new approach

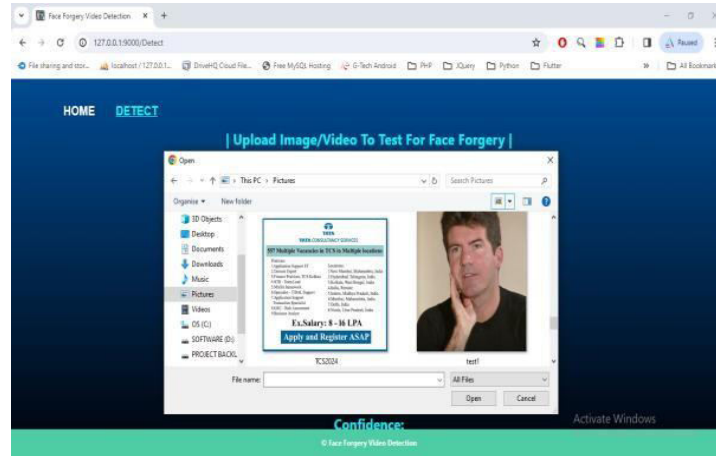


Fig 4: Load Real Image

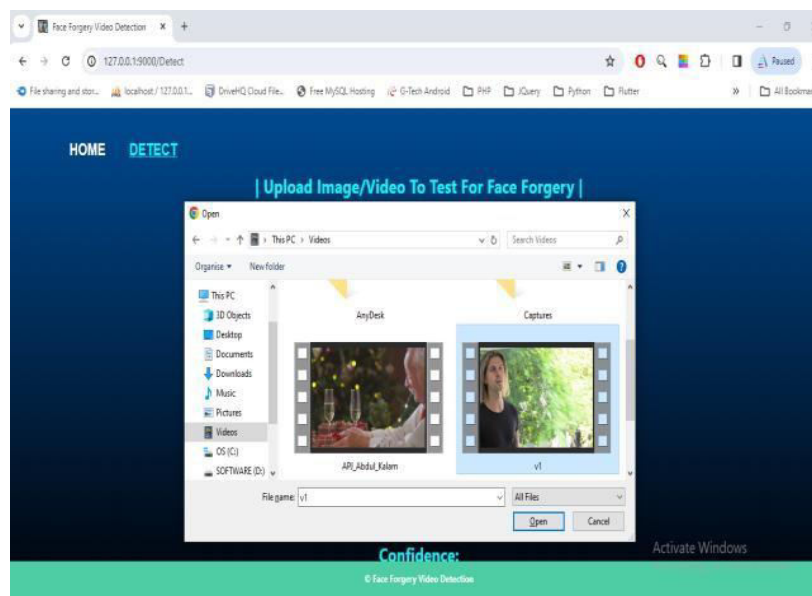


Fig 5: Processing of real time image

The research highlighted here focuses on the effectiveness of using transfer learning in a hybrid model, integrating CNNs and LSTM networks. Transfer learning allows the model to distill knowledge from pre-trained networks, making it more efficient in identifying deepfake images. The hybrid CNN- LSTM model demonstrated significant success in fighting deep fakes by taking advantage of the spatial perception of CNNs and the temporal context understanding of LSTMs. Through the hybrid CNN-LSTM framework, which combines spatial and temporal analysis, this research makes a contribution to the expanding body of literature on deepfake detection by providing a potential solution to the ongoing war against malicious image manipulation.

V. CONCLUSION

We have designed a strong neural network-based method for video classification that can differentiate between deepfake and authentic content. Our approach works by processing segments of video, each one second long, at a rate of 10 frames per second and performing classification with remarkable accuracy. The essence of our method lies in using a pre-trained ResNext Convolutional Neural Network (CNN) to extract feature-rich information from single frames.

These attributes are then fed into Long Short-Term Memory (LSTM) networks to extract temporal dependencies and identify anomalies between successive frames. The model is versatile in that it processes frame sequences of different lengths—namely, 10, 20, 40, 60, 80, and 100 frames. This versatility enables it to work with varying video lengths and complexities. The combination of ResNext CNN and LSTM networks improves the model's capacity to detect subtle inconsistencies that can be indicative of deepfakes. In general, the method holds potential for enhancing deepfake detection in real-world applications, making a valuable contribution to the area of video forensics. Feature work of effective application of this model highlights its potential for wider use in monitoring and verifying video content, thus solving problems associated with digital media authenticity and trustworthiness.

REFERENCES

1. Gordon, G. The Internet a Philosophical Inquiry; Routledge: London, UK; New York, NY, USA, 1999; 190p, ISBN 9780415197496. [Google Scholar]
2. Maras, M.-H.; Alexandrou, A. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *Int. J. Evid. Proof* 2019, 23, 255–262. [Google Scholar] [CrossRef]
3. Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Ferrer, C. The deepfake detection challenge dataset. *arXiv* 2020, arXiv:2006.07397. [Google Scholar]
4. Hasan, H.R.; Salah, K. Combating Deepfake Videos Using Blockchain and Smart Contracts. *IEEE Access* 2019, 7, 41596–41606. [Google Scholar] [CrossRef]
5. De keersmaecker, J.; Roets, A. 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence* 2017, 65, 107–110. [Google Scholar] [CrossRef]
6. Anderson, K.E. Getting to know social networks and apps: Social Media in 2017. *Libr. Hi Tech News* 2017, 34, 1–6. [Google Scholar] [CrossRef] Available online: <https://edition.cnn.com/2019/06/11/tech/zuckerberg-deepfake/index.html> (accessed on 7 June 2019).
7. Barsha, L.; Keshav, T.; Sung-Hyun, Y. Detection of Image Level Forgery with Multiple Constraints Using DFDC Full and Sample Datasets. *Sensors* 2022, 22, 9121. [Google Scholar] [CrossRef] [PubMed]
8. Zao Download Android, iPhone, iPad 2020. Available online: <https://zaodownload.com/> (accessed on 12 December 2023).
9. DeepfakesWeb.com. Available online: <https://deepfakesweb.com/> (accessed on 4 January 2023).
10. Anthropics Technology Ltd. Available online: <https://www.anthropics.com/portraitpro/> (accessed on 4 January 2023).
11. Neocortex. Available online: <https://hey.reface.ai/> (accessed on 4 January 2023).
12. TheAudacityTeam. Available online: <https://voice.ai/hub/app/free-voice-changer-for-audacity/> (accessed on 21 September 2022).
13. Magix Software GmbH. Available online: <https://www.magix.com/us/music-editing/sound-forge/> (accessed on 4 January 2023).
14. Deep Q-Network with Reinforcement Learning for Fault Detection in Cyber-Physical Systems J. Stanly Jayaprakash M. Jasmine Pemeena Priyadarsini, B. D. Parameshachari Hamid Reza Karimi, and Sasikumar Gurumoorthy *Journal of Circuits, Systems and Computers* 2022
15. Efficient Biometric Security System Using Intra-Class Finger-Knuckle Pose Variation Assessment Mr.J.Stanly Jayaprakash Dr.S.Arumugam, India *International Journal of Computer Science & Engineering Technology (IJCSET)* 2014
16. Cloud Data Authentication and Encryption based on Improved Merkle Hash Tree Technique J Stanly Jayaprakash¹, Kishore Balasubramanian², Rossilawati Sulaiman³, Mohammad Kamrul Hasan³, *, B. D. Parameshachari⁴ and Celestine Iwendi 2021.
17. Effective Biometric Security System Based on Intra- Class Finger-Knuckle Pose Variation Evaluation Mr.J.Stanly Jayaprakash Dr.S.Arumugam, India *International Journal of Computer Science & Engineering Technology (IJCSET)* 2014.
18. Cloud Data Encryption and Authentication Based on Improved Merkle Hash Tree Technique J. Stanly Jayaprakash¹, Kishore Balasubramanian², Rossilawati Sulaiman³, Mohammad Kamrul Hasan³, *, B. D. Parameshachari⁴ and Celestine Iwendi 2021.
19. Multimodal finger biometric score fusion verification based on coarse grained distribution function JS Jayaprakash, S Arumugam 2015.

20. A new technique for fingerprint sparse coding analysis utilizing k-svd learning technique S Arthi, J Stanly Jayaprakash 2024
21. "A Fusion Attention Mechanism with Bi-LSTM-Based Sarcasm Detection for Selecting High-Profit Products". Journal of Circuits, Systems and Computers C. Gayathri., R. Samson Ravindran. (2025). <https://doi.org/10.1142/S0218126625501439>.
22. "Energy and Green IT Resource Management Analysis and Formation in Geographically Distributed Environmental Cloud Data Centre" Murugan G, Gayathri.C, Latha.S, Sathiya Kumar C, SudhakarSengan, Priya V(2020), in International Journal of Advanced Science and Technology Vol. 29, No. 6,pp 4144-4155(SCOPUSindexed)
23. Meiyalakan K. Published following article. Online Multi-Crop Procurement and Loan System. Volume 10, Issue 5, pp: 32-35,
24. Meiyalakan K. Published below article. Knowledge- Based Approach to Detect Potentially Risky. Websites. Volume 10, Issue 6, pp: 1353-1357
25. Durairam.R. Machine Learning Approches for Brain Disease Diagnosis. Volume 10, Issue 6, pp: 1092-1097.
26. R. DURAIRAM. Deduplication of Storage Drives Using Cloud Computing. Volume 10, Issue 6, pp: 171-172.
27. C. Anusuya et al., "Credit Card Fraud Detection using Machine Learning-Based Random Forest Algorithm," Int. J. Sci. Adv. Res. Technol., vol. 9, no. 3, Mar. 2023.
28. C. Anusuya et al., "Alzheimer Disease with Blood Plasma Proteins detected using Convolutional Neural Network (CNN)," Int. J. Innov. Res. Comput. Commun. Eng., vol. 11, no. 3, Mar. 2023.
29. Jayasutha, P. (2022).Secure Online BoookReselling Through Bidding Method. International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), 10(6).



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details