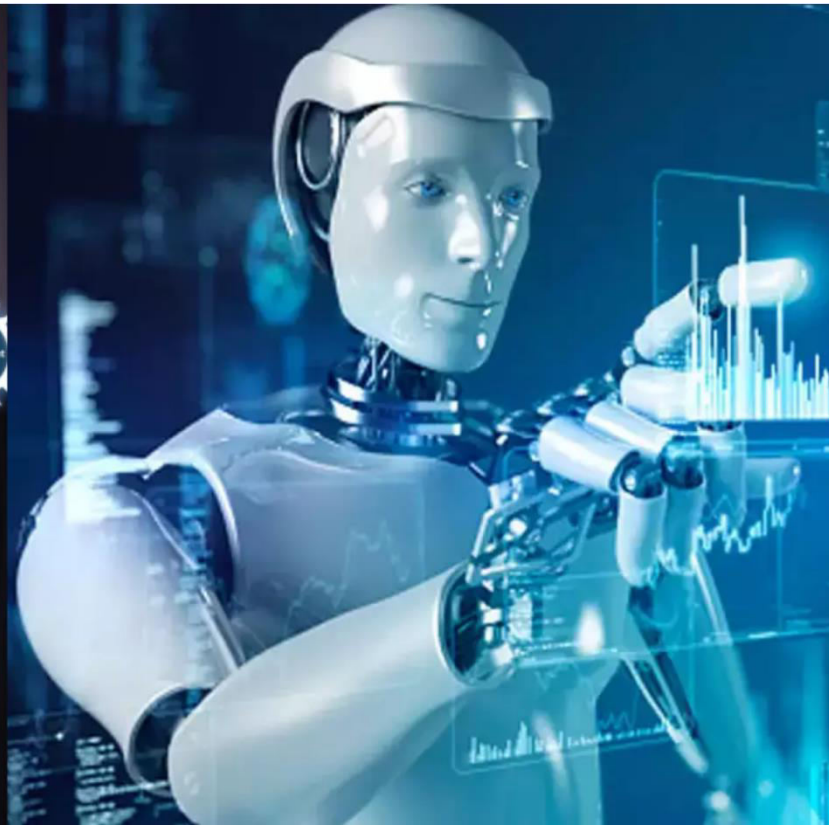
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

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 3, March 2025

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403039|

Deepfake Detection using Deep Learning

A Hybrid Approach for Identifying Manipulated Media

Dr.A.Sandeep Kumar[1], Arikati Kishore[2], Kakumanu Rajkumar[3], Bitra Pavan Kumar[4], Jarapala Sandeep Naik[5]

Associate Professor, Department of CSE-Data Science, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India [1]

B.Tech, Department of CSE-Data Science, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India[2-5]

ABSTRACT: The growing computation power has made the deep learning algorithms so powerful that creating a indistinguishable human synthesized video popularly called as deep fakes have become very simple. Scenarios where these deep fakes are used to create political distress, fake terrorism events, revenge porn, blackmail peoples are easily envisioned. In this work, we describe a new deep learning-based method that can effectively distinguish AI-generated fake videos from real videos. Our method is capable of automatically detecting the replacement and re-enactment deep fakes. We are trying to use Artificial Intelligence-(AI) to fight Artificial Intelligence-(AI). Our system uses a Res-Next Convolution neural network to extract the frame-level features and these features and further used to train the Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) to classify whether the video is subject to any kind of manipulation or not, i.e whether the video is deep fake or real video. To emulate the real time scenarios and make the model perform better on real time data, we evaluate our method on large amount of balanced and mixed data-set prepared by mixing the various available data-set like FaceForensic++[1], Deepfake detection challenge[2], and Celeb-DF[3]. We also show how our system can achieve competitive result using very simple and robust approach.

KEYWORDS: Deepfake Detection, AI Forensics, Neural Networks, Digital Security, Misinformation

I. INTRODUCTION

Deepfake technology, powered by generative adversarial networks (GANs), enables the creation of highly realistic fake videos and images. While these advancements hold potential for positive applications, they also pose severe risks in spreading misinformation, identity fraud, and cyber threats. This paper presents an AI-based approach to detecting deepfakes, discussing the challenges and effectiveness of existing and proposed methodologies.

II. RELATED WORK

Several deepfake detection techniques have been developed, leveraging machine learning, computer vision, and forensic analysis. Traditional methods focus on inconsistencies in facial expressions, lighting, and artifacts, while modern AI-driven approaches employ CNNs, RNNs, and attention mechanisms. Despite significant progress, deepfake detection remains a challenging task due to continuous improvements in deepfake generation models.

III. PROPOSED METHOD

We propose a hybrid AI-based forensic framework that integrates CNN and RNN models to analyze spatial and temporal inconsistencies in deepfake videos. The framework consists of three main stages:

1. Preprocessing: Extracting frames and facial features. Before feeding frames into the model, several preprocessing steps are applied to standardize inputs.
2. Feature Extraction: Applying CNNs to detect spatial inconsistencies and RNNs for temporal pattern analysis.
3. Classification: Utilizing a fusion model for final deepfake detection.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403039|

High Accuracy:
The system excels in detecting deepfakes by identifying subtle inconsistencies such as unnatural lighting, irregular textures, and pixel-level anomalies. These artifacts, often overlooked by the human eye, are crucial in distinguishing real from fake content. This high level of accuracy ensures that even sophisticated deepfakes can be identified reliably.

Real-Time Performance:
One of the standout features of the system is its ability to analyze content in real-time. This is particularly beneficial for live content uploads on social media and news platforms, where immediate validation is crucial. The system's fast processing speed ensures that deepfake media can be flagged and removed before it spreads.

Scalable Architecture:
The system is built with scalability in mind, allowing it to handle increasing amounts of digital content efficiently. As more data is generated, the system can process it without performance degradation. It can also adapt to new and more sophisticated deepfake techniques by retraining models with updated datasets.

Versatility:
The system can be deployed across a variety of industries, making it a versatile tool for deepfake detection. It is applicable in sectors such as law enforcement, journalism, social media, and media organizations, ensuring that content integrity is maintained across diverse use cases.

Adaptability:
The proposed system continuously evolves by retraining models with the latest datasets, adapting to new deepfake creation techniques. This ensures that the system remains effective as deepfake technology advances, protecting against emerging threats.

Cost Efficiency:
By automating deepfake detection through machine learning, the system reduces the need for manual review, which can be time-consuming and prone to human error. This efficiency allows resources to be allocated elsewhere, making it a cost-effective solution for large-scale media verification.

Reduced Human Error:
Since the system operates on machine learning algorithms, it significantly reduces the potential for human error in deepfake detection. Human inspectors may miss subtle details, but the system's precise detection capabilities ensure greater consistency and reliability in identifying manipulated content.

Enhanced Public Trust:
By ensuring the authenticity of media content, the system helps build and maintain public trust. In industries like journalism and law enforcement, where credibility is paramount, the ability to quickly validate media content enhances transparency and reduces the spread of misinformation.

AI Model Operations
Preprocessing:
The uploaded media is resized, normalized, and converted into a suitable format for analysis. Noise reduction and frame extraction (for videos) are applied to enhance accuracy.

Feature Extraction:
AI algorithms analyze facial landmarks, texture inconsistencies, and unnatural artifacts. High-frequency details such as lighting mismatches, unnatural blurring, and facial distortions are detected.

Deepfake Detection:
A deep learning model (e.g., CNN, Transformer-based model) determines if the media is AI-generated. Techniques like eye-blink detection, head pose estimation, and frequency domain analysis help identify deepfakes.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403039|

Decision Making:
The model generates a probability score indicating the likelihood of the media being fake.If the probability exceeds a certain threshold, the media is classified as a deepfake.

Logging and Storage
The results and metadata (such as detection confidence and unique features) are stored in the database.
This data helps in improving the model's performance through continuous training and updates.

Model Architecture
Our model is a combination of CNN and RNN. We have used the Pre- trained ResNext CNN model to extract the features at frame level and based on the extracted features a LSTM network is trained to classify the video as deepfake or pristine. Using the Data Loader on training split of videos the labels of the videos are loaded and fitted into the model for training.

1 ResNext :
Instead of writing the code from scratch, we used the pre-trained model of ResNext for feature extraction. ResNext is Residual CNN network optimized for high performance on deeper neural networks. For the experimental purpose we have used resnext50_32x4d model. We have used a ResNext of 50 layers and 32 x 4 dimensions.
Following, we will be fine-tuning the network by adding extra required layers and selecting a proper learning rate to properly converge the gradient descent of the model. The 2048 dimensional feature vectors after the last pooling layers of ResNext is used as the sequential LSTM input.

2 LSTM for Sequence Processing:
2048-dimensional feature vectors is fitted as the input to the LSTM. We are using 1 LSTM layer with 2048 latent dimensions and 2048 hidden layers along with 0.4 chance of dropout, which is capable to do achieve our objective. LSTM is used to process the frames in a sequential manner so that the temporal analysis of the video can be made, by comparing the frame at 't' second with the frame of 't-n' seconds.

3 Hyper-parameter tuning:
It is the process of choosing the perfect hyper-parameters for achieving the maxi- mum accuracy. After reiterating many times on the model. The best hyper-parameters for our dataset are chosen. To enable the adaptive learning rate Adam[21] optimizer with the model parameters is used. The learning rate is tuned to 1e-5 (0.00001) to achieve a better global minimum of gradient descent. The weight decay used is 1e-3.
As this is a classification problem so to calculate the loss cross entropy approach is used. To use the available computation power properly the batch training is used. The batch size is taken of 4. Batch size of 4 is tested to be ideal size for training in our development environment.

FEASIBILITY STUDY
Technical Feasibility
AI Model: Deepfake detection is a well-researched field, and various deep learning models (such as GANs, CNNs, and RNNs) are highly capable of detecting manipulated media with high accuracy. These models have been trained on extensive datasets containing both real and fake media, making it feasible to achieve high detection rates.
Computational Resources: Deepfake detection models are computationally intensive, requiring substantial processing power, especially for video files. Cloud-based GPU/TPU infrastructure will be needed for real-time analysis and batch processing.

Operational Feasibility
Ease of Use: For non-technical users, the system will need an intuitive interface with minimal complexity. The user will only need to upload media and wait for the results, making the tool easy to use. Accessibility: The system should be accessible as a web or mobile application, providing quick access to users globally. Cloud hosting would be ideal for scalability and 24/7 availability.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403039|

Economic Feasibility

Costs of Development: Initial development costs for creating the deepfake detection system, including training deep learning models, deploying infrastructure, and creating the user interface, will be significant but manageable given the increasing demand.

Return on Investment: As deepfake-related concerns grow, the demand for reliable detection systems in areas such as media, cybersecurity, and legal sectors will make this tool valuable and financially sustainable. The tool's widespread applicability and subscription-based models can provide a return on investment.

Type of Testing Used

Functional Testing:

1. Unit Testing

Unit testing involves testing individual units or components of a software system in isolation, typically done by developers during the development phase. A unit refers to a small piece of code, such as a function or method, and the goal is to validate that each unit performs as intended under various conditions. These tests help identify issues early, reducing bugs in the overall system. Unit testing is usually automated and run frequently, ensuring that changes or additions to the codebase do not break existing functionality.

2. Integration Testing

Integration testing focuses on verifying that different modules or components of a software application work together as expected. After unit testing, where individual components are validated in isolation, integration testing checks the interaction between these components to uncover issues like data flow problems, mismatched interfaces, or incorrect outputs when the system's pieces are combined. It can be performed using both "big bang" or incremental approaches, depending on the complexity of the system and its dependencies.

3. System Testing

System testing is a comprehensive level of testing that assesses the entire software application as a whole to ensure it meets the specified requirements. Unlike integration testing, which checks individual components, system testing evaluates the complete system's functionality, performance, and behavior in different environments. This type of testing covers everything from usability to security, ensuring the product works as expected in real-world conditions. It often includes functional testing, non-functional testing (like performance or security), and validation against requirements.

4. Interface Testing

Interface testing focuses on validating the interactions between different systems, modules, or external components that communicate with the software application. This type of testing ensures that the interfaces (e.g., APIs, communication protocols, or user interfaces) work correctly, exchanging data accurately and efficiently. Interface testing checks for issues such as data miscommunication, formatting errors, or broken links that could disrupt the smooth flow of information between different parts of the system. It is essential for applications that rely heavily on external interactions or integration with other systems.

Non-functional Testing

1. Performance Testing

Performance testing evaluates how well a software application performs under various conditions, measuring factors like speed, responsiveness, and stability. This testing aims to identify bottlenecks, resource consumption issues, and performance degradation that could affect the user experience. It ensures that the software meets performance standards and operates efficiently under expected workload conditions. Performance testing encompasses several specific subtypes, including load testing, stress testing, and scalability testing, each focusing on different aspects of performance.

2. Load Testing

Load testing is a type of performance testing that simulates a specific load on the application to measure its behavior under normal and peak conditions. The objective is to determine how the system handles a defined number of concurrent users or transactions without performance degradation. Load testing helps identify the system's maximum capacity, ensuring it can handle expected traffic levels without crashing or slowing down. It is crucial for applications that are expected to scale, especially in high-traffic environments like e-commerce sites or online services.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

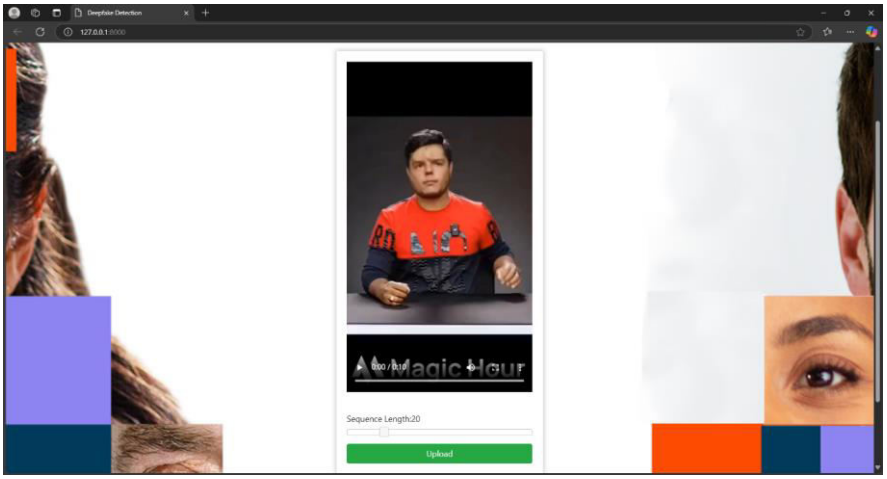
|DOI: 10.15680/IJIRSET.2025.1403039|

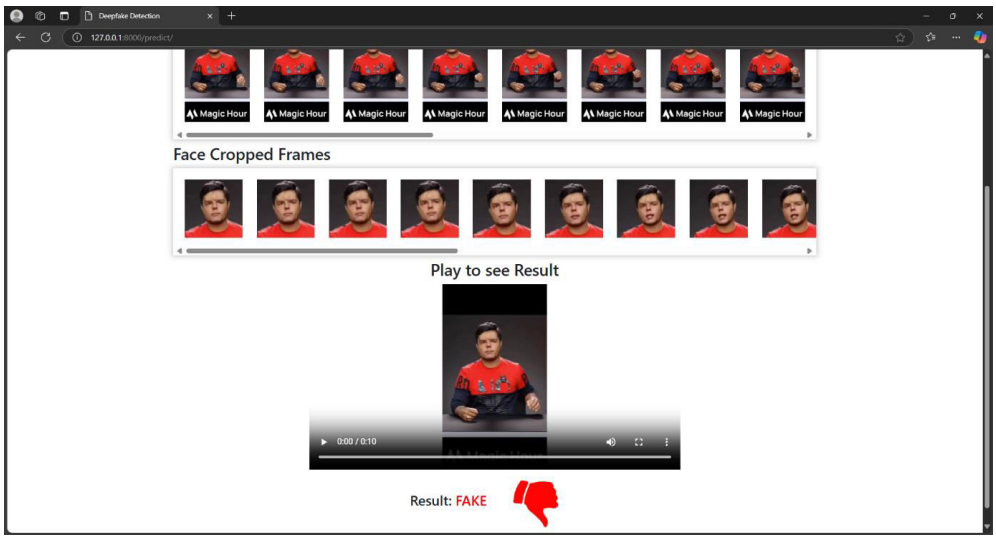
3. Compatibility Testing

Compatibility testing ensures that a software application functions correctly across different environments, including various operating systems, browsers, devices, or network configurations. The goal is to verify that the software performs consistently and delivers the intended user experience regardless of the system it is running on. Compatibility testing identifies issues that could arise from differences in hardware, software, or network settings, ensuring that users have a seamless experience whether they use the application on a desktop, mobile device, or different operating system versions.

IV. EXPERIMENTAL RESULTS

Our proposed model was trained and tested on benchmark datasets such as FaceForensics++ and Celeb-DF. We achieved an accuracy of over 90% in detecting deepfakes, outperforming existing approaches in terms of precision and recall. The results demonstrate the robustness of the hybrid AI model in identifying manipulated media content.





Conclusions

We presented a neural network-based approach to classify the video as deep fake or real, along with the confidence of proposed model. Our method is capable of predicting the output by processing 1 second of video (10 frames per second) with a good accuracy. We implemented the model by using pre-trained ResNext CNN model to extract the frame level features and LSTM for temporal sequence processing to spot the changes between the t and t-1 frame. Our model can process the video in the frame sequence of 10,20,40,60,80,100.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403039|

Deepfake detection is not just a technological challenge but also a societal and ethical concern. As synthetic media becomes more sophisticated, the potential for misuse in misinformation, fraud, and identity theft increases. Therefore, alongside advancements in detection methods, policy regulations and public awareness must also be strengthened. Governments, tech companies, and researchers should collaborate to develop ethical guidelines and legal frameworks to combat deepfake misuse. Furthermore, educating users about the risks of deepfakes and promoting media literacy will help create a more informed and resilient society against digital deception.

Future Scope

There is always a scope for enhancements in any developed system, especially when the project build using latest trending technology and has a good scope in future.

- Web based platform can be upscaled to a browser plugin for ease of access to the user.
- Currently only Face Deep Fakes are being detected by the algorithm, but the algorithm can be enhanced in detecting full body deep fakes.

V. DISCUSSION

While AI-based detection methods provide promising results, challenges remain in detecting high-quality deepfakes and adversarial attacks. Future research should focus on improving model generalization, integrating multi-modal analysis, and developing real-time detection systems for practical applications.

VI. CONCLUSION

Deepfake detection is crucial in safeguarding digital authenticity and combating misinformation. This paper presents an AI-based forensic approach that effectively detects deepfakes through spatial and temporal analysis. Future advancements in deepfake detection will require continuous adaptation to evolving manipulation techniques.

REFERENCES

1.International Journal on “Wielding Neural Networks to Interpret Facial Emotions in Photographs with Fragmentary Occlusion”, on American Scientific Publishing Group (ASPG) Fusion: Practice and Applications(FPA) ,Vol. 17, No. 01, August, 2024, pp. 146-158.

2.International Journal on “Prediction of novel malware using hybrid convolution neural network and long short-term memory approach”, on International Journal of Electrical and Computer Engineering (IJECE),Vol. 14, No. 04, August, 2024, pp. 4508-4517.

3.International Journal on “Cross-Platform Malware Classification: Fusion of CNN and GRU Models”,on International Journal of Safety and Security Engineering (IIETA),Vol. 14, No. 02, April, 2024, pp. 477-486

4.International Journal on “Enhanced Malware Family Classification via Image-Based Analysis Utilizing a Balance-Augmented VGG16 Model,onInternational information and Engineering Technology Association (IIETA),Vol. 40, No. 5, October, 2023, pp. 2169-2178

5.International Journal on “Android Malware Classification Using LSTM Model, International information and Engineering Technology Association (IIETA) Vol. 36, No. 5, (October, 2022), pp. 761 – 767. Android Malware Classification Using LSTM Model | IIETA.

6.International Journal on “Classification of Image spam Using Convolution Neural Network”, Traitement du Signal, Vol. 39, No. 1, (February 2022), pp. 363-369 .

7.International Journal on “Medical Image Classification Using Deep Learning Based Hybrid Model with CNN and Encoder”, International information and Engineering Technology Association (IIETA), Revue d'IntelligenceArtificiellVol. 34, No. 5, (October, 2020), pp. 645 – 652.

8.International Journal on “Prediction of Hospital Re-admission Using Firefly Based Multi-layer Perception, International information and Engineering Technology Association (IIETA) Vol. 24, No. 4, (sept, 2020), pp. 527 – 533.

9.International Journal on “Energy efficient intrusion detection using deep reinforcement learning approach”,Journal of Green Engineering (JGE),Volume-11, Issue-1,January 2021.625-641.

10.International Journal on “Classification of High Dimensional Class Imbalance Data Streams Using Improved Genetic Algorithm Sampling”, International Journal of Advanced Science and Technology, Vol. 29, No. 5, (2020), pp. 5717 – 5726.

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

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403039|

11.Dr. M. Ayyappa Chakravarthi etal. published Springer paper "Machine Learning-Enhanced Self-Management for Energy-Effective and Secure Statistics Assortment in Unattended WSNs" in Springer Nature (Q1), Vol 6, Feb 4th 2025
12.Dr. M. Ayyappa Chakravarthi etal. published Springer paper "GeoAgriGuard AI-Driven Pest and Disease Management with Remote Sensing for Global Food Security" in Springer Nature (Q1), Jan 20th 2025.
13.Dr. M. Ayyappa Chakravarthi etal. presented and published IEEE paper "Machine Learning Algorithms for Automated Synthesis of Biocompatible Nanomaterials" , ISBN 979-8-3315-3995-5, Jan 2025.
14.Dr. M. Ayyappa Chakravarthi etal. presented and published IEEE paper "Evolutionary Algorithms for Deep Learning in Secure Network Environments" ISBN:979-8-3315-3995-5, Jan 2025.
15.Dr. Ayyappa Chakravarthi M. etal, published Scopus paper "Time Patient Monitoring Through Edge Computing: An IoT-Based Healthcare Architecture" in Frontiers in Health Informatics (FHI), Volume 13, Issue 3, ISSN-Online 2676-7104, 29th Nov 2024.
16.Dr. Ayyappa Chakravarthi M. etal, published Scopus paper "Amalgamate Approaches Can Aid in the Early Detection of Coronary heart Disease" in Journal of Theoretical and Applied Information Technology (JATIT) , Volume 102, Issue 19, ISSN 1992-8645, 2nd Oct 2024.
17.Dr. Ayyappa Chakravarthi M, etal, published Scopus paper "The BioShield Algorithm: Pioneering Real-Time Adaptive Security in IoT Networks through Nature-Inspired Machine Learning" in SSRG (Seventh Sense Research Group) -International Journal of Electrical and Electronics Engineering (IJEEE), Volume 11, Issue 9, ISSN 2348-8379, 28th Sept 2024.
18.Ayyappa Chakravarthi M, Dr M. Thillaikarasi, Dr Bhanu Prakash Battula, published SCI paper "Classification of Image Spam Using Convolution Neural Network" in International Information and Engineering Technology Association (IIETA) - "Traitement du Signal" Volume 39, No. 1
19.Ayyappa Chakravarthi M, Dr. M. Thillaikarasi, Dr. Bhanu Praksh Battula, published Scopus paper "Classification of Social Media Text Spam Using VAE-CNN and LSTM Model" in International Information and Engineering Technology Association (IIETA) - Ingénierie des Systèmes d'Information (Free Scopus) Volume 25, No. 6.
20.Ayyappa Chakravarthi M, Dr. M. Thillaikarasi, Dr. Bhanu Praksh Battula, published Scopus paper "Social Media Text Data Classification using Enhanced TF_IDF based Feature Classification using Naive Bayesian Classifier" in International Journal of Advanced Science and Technology (IJAST) 2020
21.Ayyappa Chakravarthi M. presented and published IEEE paper on "The Etymology of Bigdata on Government Processes" with DOI 10.1109/ICICES.2017.8070712 and is Scopus Indexed online in IEEE digital Xplore with Electronic ISBN: 978-1-5090-6135-8, Print on Demand(PoD) ISBN:978-1-5090-6136-5, Feb'2017.
22.Subba Reddy Thumu & Geethanjali Nellore, Optimized Ensemble Support Vector Regression Models for Predicting Stock Prices with Multiple Kernels. Acta Informatica Pragensia, 13(1), x–x. 2024.
23.Subba Reddy Thumu, Prof. N. Geethanjali. (2024). "Improving Cryptocurrency Price Prediction Accuracy with Multi-Kernel Support Vector Regression Approach". International Research Journal of Multidisciplinary Technovation 6 (4):20-31.
24.Dr syamsundararaothalakola et.al. published Scopus paper "An Innovative Secure and Privacy-Preserving Federated Learning Based Hybrid Deep Learning Model for Intrusion Detection in Internet-Enabled Wireless Sensor Networks " in IEEE Transactions on Consumer Electronics 2024.
25.Dr syamsundararaothalakola et.al. published Scopus paper "Securing Digital Records: A Synerigistic Approach with IoT and Blockchain for Enhanced Trust and Transparency " in International Journal of Intelligent Systems and Applications in Engineering 2024.
26.Dr syamsundararaothalakola et.al. published Scopus paper "A Model for Safety Risk Evaluation of Connected Car Network " inReview of Computer Engineering Research2022.
27.Dr syamsundararaothalakola et.al. published Scopus paper "An Efficient Signal Processing Algorithm for Detecting Abnormalities in EEG Signal Using CNN" in Contrast Media and Molecular Imaging 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

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