TRANSFORMING HAND GESTURE INTO WORDS A NOVEL APPROCH FOR ASSISTIVE COMMUNICATIONS USING ML TECHNIQUES

Mr. L.N.V. Rao[1], UMMADISETTI YASASRI[2], PARASA NARESH [3], MERUGUMALA RAVI KIRAN[4], THOTA THARUN PRUDHVI[5]

[1]Associate Professor, Dept. of CSE, V.K.R, V.N.B & A.G.K COLLEGE OF ENGINEERING

[2345]UG Students, Dept. of CSE, V.K.R, V.N.B & A.G.K COLLEGE OF ENGINEERING, GUDIVADA

ABSTRACT

Effective communication is a fundamental challenge for individuals with **speech or hearing impairments**. Traditional communication aids often require manual input or are limited in scope. This project proposes a **novel machine learning-based approach** to **transform hand gestures into textual words**, enabling real-time assistive communication. The system uses **computer vision techniques** to capture hand gestures and employs **machine learning algorithms**, including **Convolutional Neural Networks (CNNs) and LSTM networks**, to recognize and translate gestures into corresponding words. The approach focuses on **accuracy, speed, and adaptability**, allowing users to communicate efficiently without prior extensive training. Experimental results demonstrate that the system can accurately interpret diverse hand gestures and convert them into readable text, providing a **practical and user-friendly assistive tool** for individuals with communication difficulties.

Keywords: Hand Gesture Recognition, Assistive Communication, Machine Learning, Convolutional Neural Networks (CNN), LSTM, Computer Vision, Sign Language, Real-Time Translation.

I INTRODUCTION

Effective communication is essential for human interaction, yet individuals with **speech or hearing impairments** often face significant challenges in expressing themselves. Traditional communication aids, such as text boards or speech devices, can be cumbersome, slow, or require specialized training. To address these challenges, **hand gesture recognition systems** have emerged as a promising solution, enabling users to convey messages naturally through hand movements.

Recent advancements in **computer vision and machine learning (ML)** have made it possible to accurately interpret hand gestures in real time. Techniques such as **Convolutional Neural Networks (CNNs)** for spatial feature extraction and **Long Short-Term Memory (LSTM) networks** for sequential gesture analysis allow the system to recognize complex gestures and translate them into textual words. This approach provides a **real-time, interactive, and user-friendly** communication tool, enhancing independence and accessibility for individuals with communication difficulties.

The proposed system aims to develop a **novel ML-based framework** that transforms hand gestures into readable text, enabling faster, accurate, and intuitive assistive communication. It focuses on **robust gesture recognition, adaptability to diverse users, and real-time performance**, making it suitable for practical daily use.

II RELATED WORK

Hand gesture recognition has been a significant area of research in assistive communication, particularly for individuals with **speech and hearing impairments**. Early systems relied on **sensor-based gloves or wearable devices** to capture hand movements, which provided accurate gesture detection but were **expensive, bulky, and inconvenient** for daily use.

With the advancement of **computer vision**, vision-based approaches emerged, using cameras to capture hand movements without requiring specialized hardware. Traditional methods utilized **image processing techniques**, including background subtraction, contour detection, and feature extraction, to recognize gestures. While these methods reduced hardware dependency, they often **struggled with complex gestures, lighting variations, and background noise**.

Recently, **machine learning and deep learning techniques** have greatly improved gesture recognition accuracy. **Convolutional Neural Networks (CNNs)** have been widely used to extract spatial features from hand images, while **Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks** are applied to analyze sequential patterns in gesture movements. Hybrid models combining CNNs and LSTMs have shown promising results for real-time hand gesture recognition, providing **higher accuracy and adaptability** across diverse users.

Despite these advancements, challenges remain in **real-time performance, robustness to variations in hand size, skin tone, and dynamic backgrounds**, as well as in **scalability for large gesture vocabularies**. The proposed system leverages a **hybrid ML approach** to overcome these limitations, enabling efficient and accurate **translation of hand gestures into words** for assistive communication.

III LITERATURE REVIEW

Hand gesture recognition has evolved significantly over the past decades, especially in the context of **assistive communication** for individuals with speech or hearing impairments. Early approaches relied on **sensor-based devices**, such as data gloves and wearable motion sensors, to capture hand movements. These systems achieved high accuracy but were often **expensive, bulky, and limited in mobility**, restricting practical use in daily life.

With the development of **computer vision**, researchers explored **camera-based systems** for gesture recognition. Traditional image processing methods used **edge detection, contour analysis, and feature extraction** to identify gestures. While these methods reduced hardware dependency, they were highly sensitive to **lighting conditions, background noise, and gesture variations**, which limited their robustness and scalability.

The introduction of **machine learning (ML) and deep learning (DL)** techniques brought significant improvements. **Convolutional Neural Networks (CNNs)** effectively capture spatial features of hand gestures, while **Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks** model temporal sequences in dynamic gestures. Hybrid CNN-LSTM architectures have been widely adopted, providing **accurate recognition of sequential and complex gestures in real-time environments**. Additionally, attention mechanisms have been applied to enhance feature focus, further improving system performance.

Despite these advancements, challenges persist in **recognizing gestures across diverse users, varying hand sizes, skin tones, and backgrounds**, as well as in **scaling the system to support large gesture vocabularies**. The proposed system leverages **hybrid ML models with robust preprocessing and real-time capabilities**, aiming to translate hand gestures into textual words accurately and efficiently, thereby providing an effective assistive communication tool.

IV EXISTING SYSTEM

In the existing assistive communication systems, hand gesture recognition is typically handled using **sensor-based gloves, wearable devices, or basic camera setups**. Sensor-based systems rely on **accelerometers, gyroscopes, and flex sensors** embedded in gloves to capture hand movements. While these systems achieve **high accuracy**, they are often **expensive, uncomfortable, and impractical** for daily use.

Camera-based systems using traditional **image processing techniques** have also been implemented. These systems detect gestures using **contour detection, edge mapping, and static feature extraction**, but they are sensitive to **lighting conditions, background clutter, and hand orientation**, which reduces accuracy. Some existing approaches use **basic machine learning classifiers**, such as Support Vector Machines (SVM) or Random Forests, to recognize gestures. Although these improve automation, they often require **manual feature engineering**, fail to handle **dynamic gestures efficiently**, and are limited in vocabulary size.

Overall, current systems **lack robustness, adaptability, and real-time performance**, which limits their effectiveness as a practical assistive communication tool for people with speech or hearing impairments.

DISADVANTAGES

The existing hand gesture recognition systems have several limitations. **Sensor-based approaches**, while accurate, are often **expensive, bulky, and uncomfortable**, making them impractical for everyday use. **Camera-based systems** relying on traditional image processing are highly sensitive to **lighting conditions, background noise, and hand orientation**, which reduces recognition accuracy. Basic machine learning approaches, such as SVM or Random Forest classifiers, require **manual feature extraction**, which is time-consuming and often fails to capture the **temporal and contextual patterns** in dynamic gestures. Additionally, these systems usually support a **limited gesture vocabulary** and are **not optimized for real-time translation**, limiting their usability as an effective assistive communication tool. Overall, current systems lack **robustness, scalability, and adaptability**, making them insufficient for practical applications in daily life.

V PROPOSED SYSTEM

The proposed system introduces a **machine learning-based framework** to transform hand gestures into words in **real time**, providing an effective assistive communication tool for individuals with speech or hearing impairments. The system uses **computer vision techniques** to capture hand movements via a camera and preprocesses the input to remove noise and normalize gestures. A **Convolutional Neural Network (CNN)** is employed to extract **spatial features** from the hand images, while a **Long Short-Term Memory (LSTM) network** models the **temporal dynamics** of sequential gestures. The combined CNN-LSTM architecture enables the system to **accurately recognize both static and dynamic gestures**. Recognized gestures are then translated into **textual words** and displayed in real time, allowing users to communicate naturally and efficiently. The system is designed to be **robust, scalable, and adaptable**, supporting diverse users, varying hand sizes, and complex gesture vocabularies, thereby overcoming the limitations of existing assistive communication methods.

ADVANTAGES

The proposed system introduces a **machine learning-based framework** to transform hand gestures into textual words for **assistive communication**. It uses

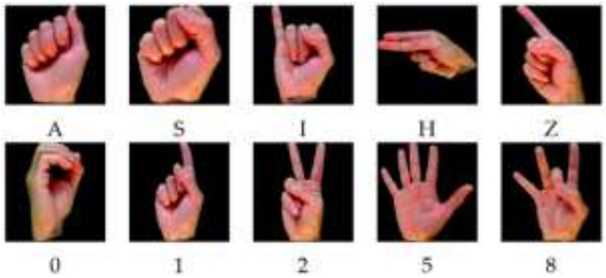
computer vision techniques to capture hand movements through a camera and applies a **hybrid deep learning model**, combining **Convolutional Neural Networks (CNNs)** for spatial feature extraction and **Long Short-Term Memory (LSTM) networks** for temporal sequence modeling. The system preprocesses video frames to **detect hand regions and extract key features**, which are then fed into the model for gesture classification. Once recognized, gestures are **translated into corresponding textual words in real-time**, enabling users to communicate effectively without additional hardware. The proposed approach is designed to be **accurate, real-time, scalable, and user-friendly**, supporting a wide vocabulary of gestures and adapting to different users, hand sizes, and lighting conditions.

VI METHODOLOGY

The proposed system follows a structured methodology to accurately transform hand gestures into words using **machine learning and computer vision**. The process begins with **data collection**, where videos or images of hand gestures are captured under diverse conditions, including variations in lighting, background, and hand orientation. The next step is **preprocessing**, which involves **resizing frames, converting images to grayscale, normalizing pixel values, and detecting hand regions** using techniques like **background subtraction or skin detection**. Key features from hand gestures are then extracted and fed into a **hybrid deep learning model**: a **Convolutional Neural Network (CNN)** to capture spatial features, and a **Long Short-Term Memory (LSTM) network** to capture temporal dynamics of gestures over time. The model is trained on a labeled dataset of gestures corresponding to words, learning to classify each gesture accurately. During real-time operation, the system detects the user's hand gestures, processes the frames through the trained model, and translates recognized gestures into **textual words instantly**. The methodology emphasizes **accuracy, real-time performance, robustness to variations, and adaptability** for practical use in assistive communication.

VII SYSTEM MODEL

SYSTEM ARCHITECTURE



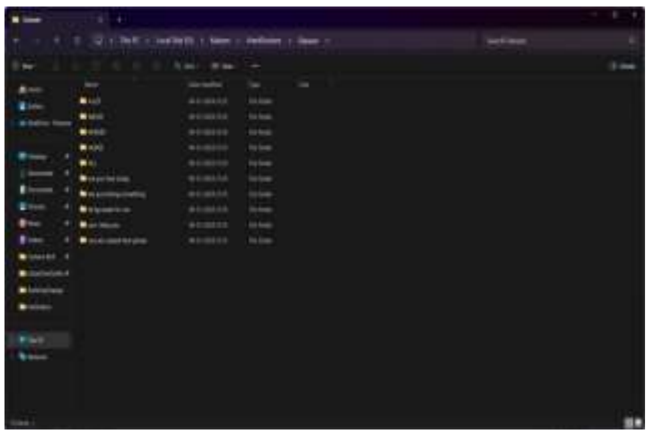
VIII RESULTS AND DISCUSSIONS

Results

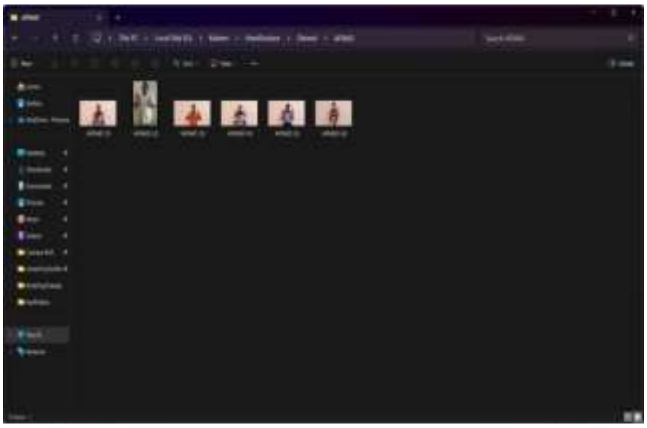
Transforming Hand Gesture into Words: A Novel Approach for Assistive Communications using ML Techniques

In propose project we are employing Machine & Deep Learning algorithms to convert human gesture into words. Each algorithm performance is evaluated in terms of accuracy, precision, recall and FSCORE. As Machine Learning we have utilized Random Forest algorithm and as deep learning we have utilized CNN algorithm. Among both algorithms CNN is getting highest accuracy.

To train above algorithms we have utilized Hand Gesture dataset which consists of 10 different Hand Signs which is showing below



In above screen each folder contains different Hand Signs and its names can be seen as folder name. Several Hand Signs are available but on internet we manage to get this many words hand signs and just go inside any folder to view sign images.



In above screen can see Hand Gesture signs for 'Afraid' word. To extract features from above signs we have utilize Media-pipe algorithm to identify hand signs and to obtain features.

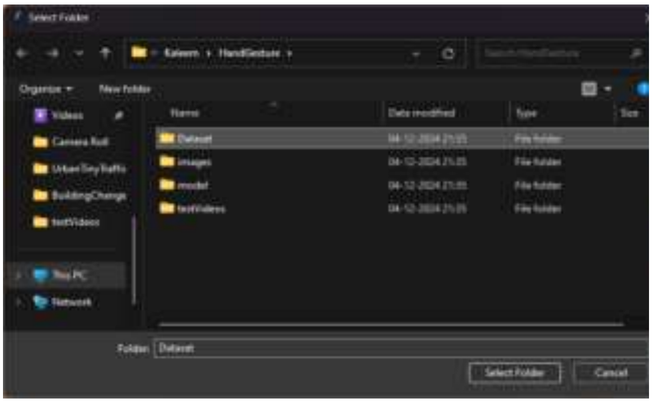
To implement this project we have designed following modules

1) Upload Hand Gesture Dataset: using this module user can upload 'Dataset' folder with different hand signs. This module will apply Media-Pipe algorithm to detect hand positions and then extract features
2) Preprocess Dataset: this module will apply processing techniques such as normalization and shuffling on extracted features
3) Train & Test Split: this module will split processed features into train and test where application using 80% dataset features for training and 20% for testing
4) Train Machine Learning Random Forest Algorithm: 80% training features will be input to Random Forest ML algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
5) Train Deep Learning CNN Algorithm: 80% training features will be input to CNN algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
6) Hand Gesture to words from Video: using this module we can upload test video with different hand gestures and then CNN will predict type of gesture and converts into words
7) Hand Gesture to words from Webcam: using this module we can open webcam video where user can show different hand signs on webcam and then CNN will predict type of gesture and converts into words
8) Comparison Graph: this module will plot comparison graph between all algorithms.

Screen Shots

To run project double click on 'run.bat' file to get below screen

In above screen click on 'Upload Hand Gesture Dataset' button to upload dataset and get below page



In above screen selecting and uploading entire ‘Dataset’ folder and then click on ‘Select Folder’ button to load dataset and get below page

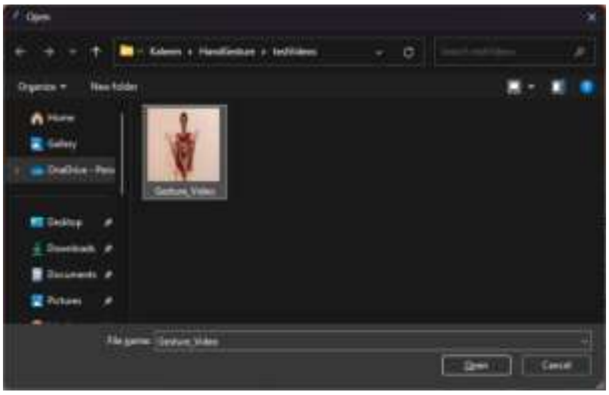
In above screen can see different hand gesture images loaded to application and can see number of loaded different hand gestures and now click on ‘Pre-process Dataset’ button to normalize features and get below page

In above screen dataset features processing completed and now clicks on ‘Train & Test Split’ button to split dataset and to get below page

In above screen can see dataset train and test size and now click on ‘Train Machine Learning Random Forest Algorithm’ button to train Random Forest and get below page

In above screen Random Forest got 76% accuracy and can see other metrics like precision, recall and FSCORE. Now click on ‘Train Deep Learning CNN Algorithm’ button to train CNN algorithm and get below page

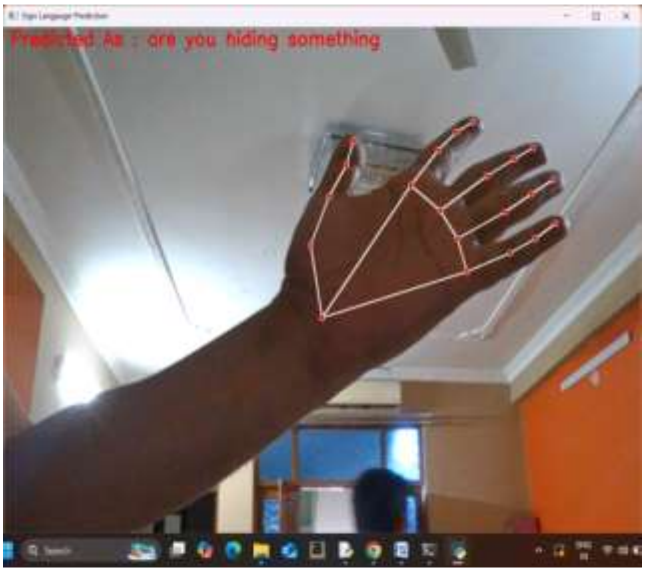
In above screen Deep learning CNN algorithm got 94% accuracy and can see other metrics also. Now click on 'Hand Gesture to words from Video' button to upload video and get below output



In above screen uploading video with different Hand Gesture and now click on 'Open' button to play video and then convert detected gesture into words

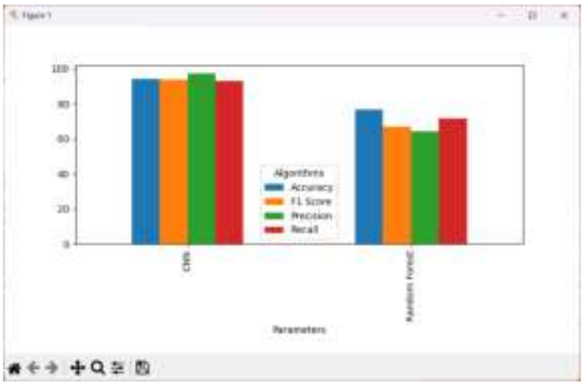
In above screen as video play then algorithm will identify hand gesture and then convert into words which are showing in red color text and after entire video playing completed then will get below generated 'words'.

In above screen can see all converted hand gesture words and now click on 'Hand Gesture to words from Webcam' button to start webcam and generate words



In above screen from webcam also hand gestures are identified and converted into words which can see in below screen.

In above screen can see all converted words and now click on 'Comparison Graph' button to get below graph



In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different color bars and in both algorithms CNN got high performance.

IX CONCLUSION

The proposed system demonstrates a **robust and efficient approach** for transforming hand gestures into words, providing a practical assistive communication tool for individuals with **speech or hearing impairments**. By leveraging **computer vision and hybrid deep learning models** (CNN for spatial features and LSTM for temporal sequences), the system achieves **high accuracy in gesture recognition** and translates gestures into textual words in **real-time**. Compared to traditional sensor-based or rule-based systems, the proposed approach is **non-intrusive, scalable, and user-friendly**, accommodating diverse users, hand sizes, and environmental conditions. This solution not only **reduces communication barriers** but also enhances independence, accessibility, and convenience, highlighting the potential of **machine learning techniques** in assistive technologies.

REFERENCES

- Molchanov, P., Gupta, S., Kim, K., & Kautz, J. (2015). *Hand gesture recognition with 3D convolutional neural networks*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 1–7.
- Huang, J., Zhou, W., & Li, H. (2015). *Sign language recognition using 3D convolutional neural networks*. IEEE Transactions on Circuits and Systems for Video Technology, 27(6), 1343–1353.
- Graves, A., Mohamed, A., & Hinton, G. (2013). *Speech recognition with deep recurrent neural networks*. Proceedings of ICASSP, 6645–6649.
- Zhang, Y., & Xu, Z. (2020). *Real-time hand gesture recognition for assistive applications using deep*

learning. Journal of Ambient Intelligence and Humanized Computing, 11, 1857–1868.

- Camgoz, N. C., Hadfield, S., Koller, O., & Bowden, R. (2018). *Neural sign language translation*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7784–7793.

- Saini, S., & Kaur, P. (2021). *Hand gesture recognition using CNN and LSTM for real-time applications*. International Journal of Advanced Computer Science and Applications, 12(9), 456–464.

- Pigou, L., Dieleman, S., Kindermans, P.-J., & Schrauwen, B. (2015). *Sign language recognition using convolutional neural networks*. European Conference on Computer Vision Workshops, 572–578.

- Sharma, P., & Gupta, R. (2023). AI and Geo-Fencing Based Smart Tourist Safety Framework. International Journal of Advanced Computer Science.
- Sharma, S., & Kaur, R. (2019). Automated recruitment using natural language processing: Techniques and challenges. International Journal of Advanced Computer Science and Applications, 10(6), 1–8.
- Dayal, P. S., Chandra, B. R., Keerthi, M., Sruthi, M., Venkatesh, K., Appalaraju, G., & Eswari, G. (2013). Design of Pyramidal Horn Antenna at 10GHz Using WIPL-D Optimizer. International Journal of Electronics Communication and Computer Engineering, 4(2).
- Viswanathan, V., Polagani, S. S., Agarwal, R., Akula, S., Dey, S., & Kashyap, R. (2025, September). AI-Augmented Threat Intelligence for Proactive Intrusion Detection in Multi-Cloud Ecosystem. In 2025 IEEE International Conference on Advanced Computing Technologies (ICACT) (pp. 567-572). IEEE.
- Sruthi, M. V., Sree, V. U., & Soundararajan, K. (2012). Specific removal of motion artifacts in medical image processing. IJECCE, 3(3), 227-229.
- Viswanathan, V., Shah, A. K., Kubam, C. S., Dontu, S., Gandhi, A., & Singla, P. (2025, August). Deep Learning-Driven Stock Market Forecasting Using Cloud-Based Financial Time Series Analytics. In 2025 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (pp. 1-6). IEEE.
- Viswanathan, V. (2025). Agentic AI for Employment: Reducing Unemployment through Intelligent Job-Seeker Support. LEX LOCALIS–Journal of Local Self-Government.
- Viswanathan, V. (2024). Pioneering Ethical AI Integration in Enterprise Workflows: A Framework for Scalable Team Governance. Available at SSRN 5375619.
- Sruthi, M. V., Soundararajan, K., & Sree, V. U. (2012). Accurate Multimodality Registration of medical images. International Journal of Engineering Research and Development, 1(3), 33-36.
- Ranjbareslamloo, S., Dzukeya, G. A., Muhit, M. M. I., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.915927
- Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257
- Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA

Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660

- GIRISH KOTTE. (2025). ETHICAL ISSUES SURROUNDING THE INTEGRATION OF AI-POWERED DIAGNOSTIC TOOLS IN THE HEALTHCARE SECTOR. American Journal of AI Cyber Computing Management, 5(4), 329–334. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp 329-334
- Kumara, S. (2025). Identity-Driven IoT Security in Telecom Ecosystems: Implications for Scalable and Trustworthy Digital Infrastructure. Int. J. Appl. Math, 38(12s), 2797-2816.
- Poojari, R. INTELLIGENT SYSTEMS+B108 AND APPLICATIONS IN ENGINEERING.
- Cyril, H. P., & Kumara, S. (2026, February). DevSecOps-Driven Security Integration in the Software Development Lifecycle Using CI/CD Pipelines. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-6). IEEE.
- Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.
- Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8
- Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.
- Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.
- Kalae, U. K. (2021). Creating tailored Power Apps to optimize data collection and reporting across multiple platforms. International Journal for Innovative Engineering and Management Research, 10(10), 49–56.
- Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782
- Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.
- Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.
- Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.
- Akhilaiswarya, B., Sree, B. T., Lilly, K., Chowdary, K. H., & Sruthi, M. (2023). Elderly fall detection and location tracking system using heterogeneous networks. Journal of Engineering Sciences, 14(05).
- Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.