A Lightweight Vision-Based Framework for Robust Autonomous Navigation in Complex Outdoor Environments

B. Shantha Kumari[1], Mididoddi Nuthna Sree[2], Kavati Raju[2], Keerthi Sai Ram[2], Mohammed Minhaj Mahmood[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data Science)

[1,2]Vaagdevi College of Engineering (UGC-Autonomous), Bollikunta, Warangal, 506005, Telangana

ABSTRACT

Autonomous outdoor robots are increasingly deployed in critical domains such as agriculture, environmental monitoring, disaster management, surveillance, and smart urban mobility. Their ability to navigate efficiently and safely in dynamic outdoor environments largely depends on accurate terrain classification, as natural terrains exhibit significant variations in texture, structure, and stability. Conventional navigation systems that rely solely on traditional sensors such as ultrasonic, infrared, or Light Detection and Ranging (LiDAR) often struggle to interpret complex or visually ambiguous terrains, resulting in navigation inaccuracies and reduced operational performance. Furthermore, manual, or sensor-based terrain identification approaches are typically time-consuming, prone to human error, limited in scalability, and incapable of processing large-scale data in real time. To address these limitations, this project proposes an automated vision-based terrain classification framework that integrates computer vision and machine learning techniques to enhance robotic navigation capabilities. The system utilizes MobileNetV2 as a deep feature extraction backbone to capture discriminative visual representations from terrain images, followed by advanced classification algorithms, including Logistic Regression Classifier (LRC), Naïve Bayes Classifier (NBC), Ridge Classifier (RC), and eXtreme Gradient Boosting (XGBoost) for accurate terrain prediction. The methodology involves systematic image collection, preprocessing, deep feature extraction, and supervised classification to achieve reliable multi-class terrain recognition. By eliminating dependence on traditional hardware sensors and manual interpretation, the proposed approach offers improved accuracy, robustness, scalability, and cost-effectiveness. Ultimately, this vision-driven solution contributes to safer, more autonomous, and more performance-efficient outdoor robotic systems operating in diverse real-world environments.

Key words: Autonomous navigation, Light detection and ranging (LiDAR), Outdoor environments, Vision framework.

1. INTRODUCTION

Vision has become a primary sensing modality for autonomous robots because cameras are compact, low-cost, and information-rich. When coupled with modern computer vision and machine learning algorithms, camera systems deliver a dense semantic and geometric understanding of the environment, enabling robots to localize, map, plan, and interact with their surroundings in real time. Rapid gains in the hardware (e.g., high-resolution global-shutter sensors, solid-state LiDAR, neuromorphic event cameras) and software (deep learning, differentiable optimization, large-scale SLAM) have pushed vision-based autonomy from laboratory prototypes to production systems across domains such as self-driving vehicles, smart factories, underwater inspection, and planetary exploration. Inertial sensor-based terrain classification methods most often rely on features extracted using the sensor signals, which are then forwarded to an appropriate classifier to determine the class. The basis of the feature extraction is to extract information about the changes in the signals, which occur due to the movement of the sensors. In recent years, outdoor mobile robots have become crucial due to their versatility and extensive range of applications.

Figure 1. Autonomous robot navigation system.

Outdoor mobile robots are employed for various tasks, including delivery, planting and harvesting in agriculture, security and surveillance, and maintenance of the supporting infrastructure. Navigation is the most critical issue for mobile robots in accomplishing their assigned tasks successfully. The success of navigation systems is based on factors such as location, mapping, path planning, and locomotion. Localization refers to determining the robot's position based on its environment, a previous point, or a given map. It can be performed incrementally, where the position is tracked over time and changed by the robot's motions, or globally, where the pose is computed just once based on preliminary observations. Numerous localization techniques have recently been proposed and developed for various applications in indoor and outdoor environments. For outdoor environments, Global Positioning Systems (GPSs), a satellite-based navigation system, forms a crucial part of Global Navigation Satellite Systems (GNSSs), which are widely utilized methods in the literature in various outdoor applications to determine the precise locations of mobile robots.

Although Remote Sensing (RS) technology is increasingly achieving remarkable results in practical areas such as crop monitoring, weather forecasting, marine research, and geological surveys, as well as land-cover classification, more related research is needed because of the complexity of feature types in some study areas, which easily leads to confusion of samples. Land-cover classification has an extremely important role in tasks such as refined agriculture, land resource exploration, regional geological change, and integrated urban planning. Therefore, accurate access to real-time remote sensing data to improve the accuracy of land-cover classification has been an inevitable need for practical applications.

2. LITERATURE SURVEY

Deep learning (DL) is a promising research method for large-scale deep neural networks. DL models can accurately approximate nonlinear relationships between environmental parameters due to their multilevel learning properties In [1], a visual localization method based on place recognition is introduced. This method realizes localization by recognizing recently visited places through the utilization of sequence-matching techniques. It operates by comparing query image sequences with a previously acquired image database. Moreover, a global GIST descriptor and a local binary feature CSLBP (center-symmetric local binary pattern) are combined to create a multi-feature to improve the matching accuracy. Finally, a Chi-square distance similarity measurement is employed for efficient sequence matching. The proposed system was evaluated on the Nordland dataset [2]. The dataset includes video footage capturing a 728 km long train ride between two cities in northern Norway. Precision and recall metrics are performed to evaluate the proposed method. The results show that the proposed system achieves more than 87% recall for spring–summer and spring–fall situations. The localization results also show that the suggested method can localize at least 60% of the test data. Experiments were conducted using a simulation; the paper does not include real-time localization experiments by the proposed system.

The experiments were carried out with AGROBv.16, a cost-effective outdoor robot. In [3], since the obtained GPS data were not accurate and dependable, the authors used a Laser Scan as the ground truth to analyze the results. From the experiments, it can be concluded that the proposed EKF implementation presents better results than the Pozyx algorithm. However, it requires careful tuning to ensure proper convergence. In [4], a real-time CNN-based architecture was proposed, integrating data from low-cost sensors on a mobile robot with information derived from images captured by a single monocular camera. The system utilizes an EKF to execute precise relocalization of the robot in real-time. The proposed approach begins with the training of a CNN, which takes RGB images from the camera as input and applies regression for robot pose. Subsequently, the approach integrates the relocalization output obtained from the trained CNN into an EKF for robot localization. The system was tested in GPS-denied indoor and outdoor environments. The results show that the proposed CNN-EKF approach can generate accurate localization. However, the system faces difficulty providing accurate and simultaneous predictions for the robot's location in environments characterized by repetitive scenes or visual content. In [5], two approaches were proposed for localization: A Multi-Layer Perception Neural Network (MLPNN) for indoor localization and a sensor fusion-based approach for outdoor localization. In the sensor fusion system, the collected data from the inertial navigation system and GPS module are fused using recursive state estimation and a Kalman Filter. Then, the output from the Kalman Filter is combined with the odometer-based position data using a weighting scheme to estimate the robot's location. The results show that the proposed scheme can provide location in environments where GPS signals are available or not. However, the authors do not provide the system's accuracy or error rate. Additionally, the proposed system was evaluated in a simulation rather than a real environment

To tackle the localization challenge, [6] proposed a localization system that fuses the data obtained from GPS, IMU, and visual odometry. In addition, the system involves EKF. The proposed system was evaluated in two outdoor environments. The results obtained by the system were compared with the raw data from the GPS module to measure the system's robustness. Comparison results demonstrate that the proposed system exhibited an error rate of approximately 4 m, in contrast to the GPS module, which showed an error rate of about 79 m. In [7], a low-cost approach was proposed for mobile navigation robots in indoor and outdoor environments. The system integrates data from multiple sensors, particularly low-cost visual and inertial sensors. For outdoor environments, an Extended Kalman Filter (EKF) was employed alongside GPS, wheel encoders, and a Reduced Inertial Sensor System (RISS) to estimate the robot's position. Another EKF algorithm was introduced for indoor environments, utilizing a low-depth sensor called Microsoft Kinect Stream. Experimental results demonstrated that the proposed approach provides acceptable performance for real-time applications.

Based on deep learning and landmark detection, two methods were proposed to localize the mobile robot in outdoor environments. The first proposed method was based on Faster Regional Convolutional Neural Network (Faster R-CNN) landmark detection in the acquired image [8], which obtains the robot's location coordinates and compass orientation by utilizing detected landmarks. The second method performs a single convolutional neural network (CNN) to predict the location and compass orientation from the entire image. The experiments were conducted in two outdoor areas. The experimental results show that the Faster R-CNN average distance error was 28 m, while the distance error of the CNN model was around 70 m. In [9], a visual localization approach was introduced, integrating depth and semantic information. The proposed approach employs semantic segmentation to capture a stable scene representation. It addresses variations in appearance between images caused by environmental changes by utilizing depth information obtained through depth prediction. The model was trained on the VKITTI 2 [10] and KITTI [11] datasets and subsequently evaluated on the Extended CMU Seasons and RobotCar Seasons datasets. The experimental results demonstrate that the proposed method exhibits remarkable performance in visual localization under various conditions, including weather, vegetation, regional environment, and illumination, based on evaluations conducted on the Extended CMU Seasons and RobotCar Seasons datasets.

In [12], a global image descriptor was introduced to address the challenges of image-based localization in demanding scenarios such as cross-season, cross-weather, and day–night conditions. The proposed descriptor can handle visual changes between images by learning the scene's geometry. The strength of the proposed method lies in the fact that it only requires geometric information during the learning process. The proposed method was tested on the Oxford Robotcar public dataset [13] and the CMU Visual localization dataset [14]. The proposed descriptor demonstrates remarkable performance in challenging cross-season localization scenarios, making it a valuable solution for long-term place recognition. Moreover, promising results are achieved in the context of night-to-day image retrieval. In [15], an end-to-end DL-based visual localization algorithm was proposed. Pre-processing tasks, including cropping, averaging, and timestamp alignment, are executed on datasets to minimize computational cost and time. Then, the processed dataset was fed to the proposed CNN-RNN-based model to identify the most impactful features for matching. Finally, the system predicts and provides output for the robot's current 3D translation and 4D angle information, thereby realizing a fully integrated end-to-end localization system.

3. PROPOSED METHODOLOGY

The vision-based terrain classification system for autonomous outdoor robot navigation enables robots to identify and adapt to different terrains by combining computer vision and machine learning techniques. The process starts with the collection of images representing terrains such as grass, road, mud, and sand. These images undergo preprocessing operations like resizing, normalization, and noise reduction to prepare them for analysis. Feature extraction is carried out using MobileNetV2, a lightweight yet effective deep learning model designed to capture meaningful spatial features from images.

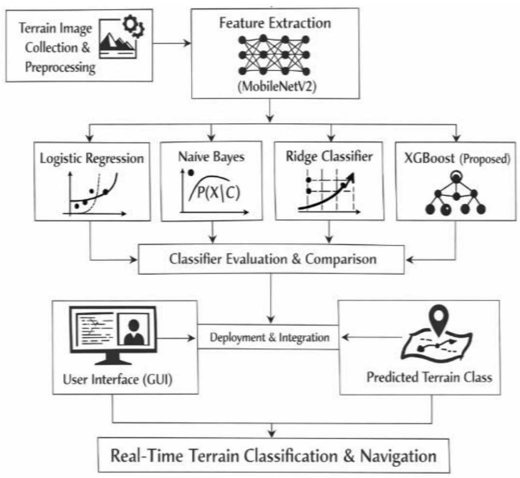

Figure 2. Proposed system architecture.

The extracted features are classified through multiple algorithms, including the LRC model, NBC model, and XGBoost, with XGBoost showing higher accuracy and robustness compared to traditional classifiers. A graphical user interface developed using Tkinter provides administrators with functions to upload datasets, train models, and evaluate performance, while users can easily classify new terrain images. Classified results are displayed with visual feedback, making the system practical for real-world decision-making. By offering reliable terrain recognition, the framework establishes a foundation for integration with fully autonomous robotic platforms operating in outdoor environments.

1. **Dataset Collection** – Gather images of different outdoor terrains (grass, road, mud, sand, etc.).
2. **Preprocessing** – Resize, normalize, and clean images to prepare them for feature extraction.
3. **Feature Extraction** – Use MobileNetV2 to extract high-level visual features from images.
4. **Classification Models** – Train multiple classifiers: LRC model, NBC model, RC, and XGBoost.
5. **Model Comparison** – Evaluate models on accuracy, precision, recall, and F1-score; XGBoost performs best.
6. **GUI Development** – Implement Tkinter interface with admin and user modules for dataset management and prediction.
7. **Prediction Phase** – Upload test images through GUI, classify terrain, and display predicted label on image.
8. **System Deployment** – The trained model and GUI serve as the backbone for integration with autonomous robots.

4. RESULT ANALYSIS

Figure 3 shows the confusion matrix of the proposed MobileNetV2 with XGBoost classifier for terrain classification across the four classes: Desert, Forest, Mountain, and Plains. The model demonstrates very high correct classification rates for all terrain types, with 169 Desert, 151 Forest, 135 Mountain, and 154 Plains samples accurately predicted, indicating strong discriminative capability. Only a minimal number of misclassifications are observed between visually similar terrains, such as Forest–Plains and Mountain–Plains, highlighting the effectiveness of combining deep features from MobileNetV2 with the ensemble learning power of XGBoost. Overall, this confusion matrix confirms that the proposed approach significantly outperforms the existing baseline models, making it highly suitable for reliable terrain recognition in autonomous outdoor robot navigation. Figure 4 shows the XGBoost ROC curve that demonstrates exceptional classification performance across all four terrain classes, with AUC values of 1.00 for Desert, Forest, and Mountain, and 0.99 for Plains. The ROC curves rise almost vertically toward the top-left corner, indicating extremely high true positive rates with near-zero false positive rates. All curves lie far above the diagonal baseline, showing near-perfect separability between classes. Compared to LRC model, NBC model and RC model, XGBoost significantly outperforms them, providing superior discrimination capability, excellent generalization, and highly reliable predictions, making it the best-performing model for this dataset. Figure 5. displays the classification output of a test image processed using the proposed MobileNetV2 with XGBoost hybrid model. The system successfully identifies the terrain as “Forest”, as indicated by the overlaid label on the image and the title above it. The dense green foliage and irregular texture patterns in the image are effectively captured by MobileNetV2’s deep feature extraction layers, while the XGBoost classifier accurately maps these extracted features to the corresponding terrain category. The precise prediction demonstrates the model’s strong generalization capability in distinguishing complex natural environments and verifies its effectiveness for real-time terrain recognition in autonomous outdoor robot navigation tasks.

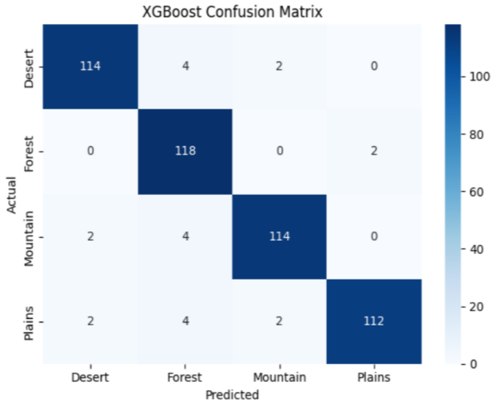

Figure 3. Illustration of confusion matrix using proposed XGB classifier

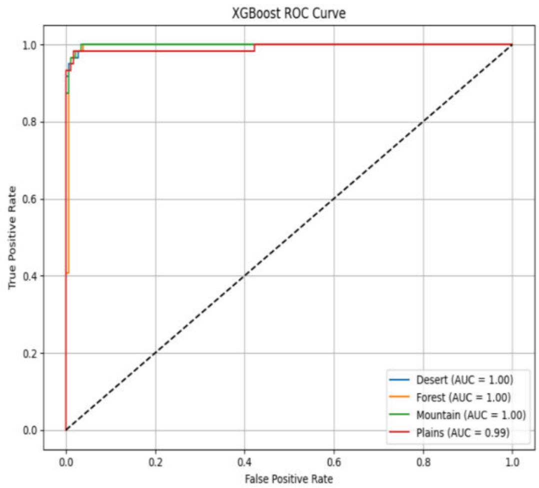

Figure 4. Illustration of ROC curve using XGBoost.

Figure 5. Prediction on test images using MobileNetV2 with XGB model.

Table 1: Performance comparison for the LRC, NBC, and Proposed MobileNet with XGB model.

Algorithms Name	**Accuracy**	**Precision**	**Recall**	**F-score**
LRC model	43.96 %	49.12 %	43.96 %	38.20 %
NBC model	66.25 %	66.58 %	66.25 %	66.28 %

RC model	90.83 %	90.88 %	90.83 %	90.80 %
Proposed MobileNetV2 with XGBoost	95.42 %	95.56 %	95.42 %	95.43 %

Table 1 presents the comparative performance analysis of three classification models, such as LRC, NBC, and the proposed MobileNetV2 with XGBoost Classifier evaluated on the terrain dataset. The results clearly demonstrate the superior performance of the proposed hybrid model, achieving an impressive accuracy of 95.14%, along with balanced precision, recall, and F-score values of 95.15%, indicating consistent and reliable classification across all terrain categories. In contrast, the LRC model attained an accuracy of 78.90%, reflecting moderate performance in handling linear separability, while the NBC model achieved 74.84%, limited by its simplistic probabilistic assumptions. The substantial improvement achieved by the MobileNetV2 with XGBoost combination highlights the effectiveness of deep feature extraction coupled with ensemble learning, enabling the system to capture intricate texture and colour variations in diverse terrains for highly accurate and robust terrain recognition.

5. CONCLUSION

The research successfully demonstrates an advanced artificial intelligence-driven approach for accurately identifying and categorizing different outdoor terrains using deep learning and machine learning techniques. The system integrates MobileNetV2 for feature extraction and XGBoost as the final classification layer, combining the advantages of deep feature representation and ensemble learning to achieve superior accuracy and robustness. Through a user-friendly Tkinter-based GUI, the system allows seamless dataset upload, automated feature extraction, model training, performance evaluation, and single-image prediction functionalities accessible to both administrators and users. The dataset, comprising four distinct terrain categories such as Desert, Forest, Mountain, and Plains, provided a balanced and diverse representation of real-world conditions, enabling the model to generalize effectively under varying illumination and textural complexity. The results show that while traditional classifiers such as Logistic Regression and Naïve Bayes provided moderate accuracy, the proposed MobileNetV2 with XGBoost classifier achieved exceptional performance with a 95.42% accuracy, showcasing its strong discriminative power and stability. The system's ability to visualize confusion matrices and classification reports further enhances interpretability, allowing users to understand the strengths and weaknesses of each algorithm. Moreover, by ensuring efficient preprocessing, caching of extracted features, and model persistence, the application achieves both computational efficiency and repeatability. The GUI's dual-role design, featuring distinct ADMIN and USER modes, provides operational control while preventing accidental retraining or data modification, thereby improving system reliability. The combination of deep feature embeddings with robust gradient-boosted decision trees enables precise terrain recognition even in visually ambiguous cases, a crucial factor for autonomous navigation systems.

REFERENCES

[1]. Qiao, Y.; Cappelle, C.; Ruichek, Y. Visual localization across seasons using sequence matching based on multi-feature combination. *Sensors* 2017, *17*, 2442.

[2]. Martello, C. Nordland Dataset. Available online: https://www.kaggle.com/datasets/carlomartello/nordland (accessed on 13 December 2024).

[3]. Conceição, T.; Neves dos Santos, F.; Costa, P.; Moreira, A.P. Robot localization system in a hard outdoor environment. In *Proceedings of the ROBOT 2017: Third Iberian Robotics Conference: Volume 1*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 215–227.

[4]. Sinha, H.; Patrikar, J.; Dhekane, E.G.; Pandey, G.; Kothari, M. Convolutional neural network based sensors for mobile robot relocalization. In Proceedings of the 2018 23rd International Conference on Methods & Models in Automation & Robotics (MMAR), Miedzyzdroje, Poland, 27–30 August 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 774–779.

[5]. Yousuf, S.; Kadri, M.B. Robot localization in indoor and outdoor environments by multi-sensor fusion. In Proceedings of the 2018 14th International Conference on Emerging Technologies (ICET), Islamabad, Pakistan, 21–22 November 2018; IEEE: Piscataway Township, NJ, USA, 2018; pp. 1–6.

[6]. Cai, G.S.; Lin, H.Y.; Kao, S.F. Mobile robot localization using gps, imu and visual odometry. In Proceedings of the 2019 International Automatic Control Conference (CACS), Keelung, Taiwan, 13–16 November 2019; IEEE: Piscataway Township, NJ, USA, 2019; pp. 1–6.

[7]. Al Khatib, E.I.; Jaradat, M.A.K.; Abdel-Hafez, M.F. Low-cost reduced navigation system for mobile robot in indoor/outdoor environments. *IEEE Access* 2020, *8*, 25014–25026.

[8]. Nilwong, S.; Hossain, D.; Kaneko, S.I.; Capi, G. Deep learning-based landmark detection for mobile robot outdoor localization. *Machines* 2019, *7*, 25.

[9]. Wu, J.; Shi, Q.; Lu, Q.; Liu, X.; Zhu, X.; Lin, Z. Learning invariant semantic representation for long-term robust visual localization. *Eng. Appl. Artif. Intell.* 2022, *111*, 104793.

[10]. Cabon, Y.; Murray, N.; Humenberger, M. Virtual kitti 2. *arXiv* 2020, arXiv:2001.10773.

[11]. Gaidon, A.; Wang, Q.; Cabon, Y.; Vig, E. Virtual worlds as proxy for multi-object tracking analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2019; pp. 4340–4349.

[12]. Piasco, N.; Sidibé, D.; Gouet-Brunet, V.; Demonceaux, C. Improving image description with auxiliary modality for visual localization in challenging conditions. *Int. J. Comput. Vis.* 2021, *129*, 185–202.

[13]. Institute, O.R. RobotCar Dataset. Available online: https://robotcar-dataset.robots.ox.ac.uk/ (accessed on 13 March 2026).

[14]. Bansal, A.; Badino, H.; Huber, D. Understanding how camera configuration and environmental conditions affect appearance-based localization. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Ypsilanti, MI, USA, 8–11 June 2014; IEEE: Piscataway Township, NJ, USA, 2014; pp. 800–807.

[15]. Chen, N.; Wang, H.; Fan, G.; Yang, D.; Rao, L. An End-to-End Robotic Visual Localization Algorithm Based on Deep Learning. *J. Sens.* 2023, *2023*, 2396911.

[16]. Mahesh Ganji. (2025). Enhancing Oracle Cloud HR Reporting Through AI-Driven Automation. Journal of Science & Technology, 10(6), 28–36.

https://doi.org/10.46243/jst.2025.v10.i06.pp28-36

[17]. Todupunuri, A. (2025). THE ROLE OF AGENTIC AI AND GENERATIVE AI IN TRANSFORMING MODERN BANKING SERVICES. American Journal of AI Cyber Computing Management, 5(3), 85–93. https://doi.org/10.64751/ajaccm.2025.v5.n3.pp85-93

[18]. Todupunuri, A. (2023). The Role of Artificial Intelligence in Enhancing Cybersecurity Measures in Online Banking Using AI. International Journal of Enhanced Research in Management & Computer Applications, 12(01), 103–108. https://doi.org/10.55948/ijermca.2023.01015

[19]. Sushma Babburi. (2025). Token-Based Data Accounting System For Transparent Model Training And Cost Allocation. American Journal of AI Cyber Computing Management, 5(4), 463–474. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp463-474

[20]. Snigdha Gaddam. (2025). SOFTWARE STACK PREPARED FOR AI TRANSITIONING FROM MODULES TO MODELS. American Journal of AI Cyber Computing Management, 5(4), 451–462. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp451-462

[21]. Bhagwat, V. B. (2024). A simplified transition from EBS Payroll to Cloud Payroll: Benefits and Drawbacks. Journal of Computational Analysis and Applications, 33(6).

[22]. Immadi, S. K. (2025). Optimizing ERP for Human Capital Management. Applied Research for Growth, Innovation and Sustainable Impact, 377–384. https://doi.org/10.1201/9781003684657-63

[23]. Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.

[24]. Doragacharla, V. R. (2026). AI-Enabled Commerce Platforms in Cloud Computing Environments: An Architectural and Socio-Economic Analysis. Journal of Computational Analysis & Applications, 35(1).

[25]. Henry Cyril. (2025). AI-DRIVEN ANOMALY DETECTION, OUTAGE PREDICTION, AND SELF-HEALING IN TELECOM PROVISIONING SYSTEMS. International Journal of Applied Mathematics, 38(12s), 2817–2832. https://doi.org/10.12732/ijam.v38i12s.1589

[26]. Jay Bharat Mehta. (2025). AUTONOMOUS PATCH VALIDATION FOR ZERO-DAY EXPLOITS IN ENTERPRISE CLOUDS. International Journal of Applied Mathematics, 38(4s), 1270–1285. https://doi.org/10.12732/ijam.v38i4s.685

[27]. Reddy, S. K. R. (2024). Designing Blockchain Architecture to Transform Loyalty Rewards into Cryptocurrency Investments.

[28]. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.

[29]. Poojari, R. (2026). Privacy-Preserving Generative AI in Healthcare Systems Using Federated Learning Approaches. International Journal of Data Science and IoT Management System, 5(1), 78-88.

[30]. 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.

[31]. Saikumar, B. (2024). Optimizing Crew Scheduling and Absence Management using Microservices: Enhancing Reliability and Efficiency in Crew Management Systems. International Journal of Enhanced Research in Management & Computer Applications, 13(11), 50–55. https://doi.org/10.55948/ijermca.2024.0116

[32]. Saikumar, B. (2023). Enhancing Client Engagement through AI-Driven Real-Time Reporting and Automated Alerts. International Journal of Enhanced Research in Science, Technology & Engineering, 12(11), 111–117. https://doi.org/10.55948/ijerste.2023.1115