International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Data-Driven Regression Framework for Reliable Fish Growth Prediction in Aquaculture Farms

B. Maheshwari[1], Gaddam Rathan Kumar[2], Bandari ShivaKumar[2], Erraboina Shruthi[2], Allanki Srija[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

Aquaculture output has grown by over 35% in the past decade, yet field reports indicate that nearly 20% of expected growth is lost each cycle due to insufficient monitoring and inaccurate estimation. Despite steady production increases, the absence of reliable automated prediction systems limits farm efficiency. Traditional methods, which involve manually capturing fish to estimate length and weight, are slow, inconsistent, and highly sensitive to operator skill. Frequent handling also disturbs the pond environment, stressing fish and affecting growth behavior. These challenges make continuous monitoring difficult, especially under rapidly changing environmental conditions. Real-world applications such as adaptive feeding, early detection of growth deviations, and dynamic environmental adjustments require real-time, accurate prediction support. Current systems largely rely on linear regression models, basic regression techniques, or manually engineered features from simple sensors, which struggle to capture the nonlinear patterns inherent in aquaculture. To address this, we propose ConvETR, a regression framework combining a Convolutional Neural Network (CNN) with an Extra Trees Regressor (ETR), designed to process continuous visual sensor streams rather than static images. The CNN extracts meaningful temporal–spatial features, capturing density variations, movement behavior, and subtle water-column cues related to growth, while the ETR stabilizes predictions by handling noise and nonlinear relationships. This integrated, non-intrusive approach improves accuracy, enables real-time decisions, and offers a scalable solution for modern IoT-enabled aquaculture systems requiring reliable fish-growth regression.

Keywords: Fish growth, aquaculture, spatial features, convolutional neural network, early identification.

Received: 06-02-2026 Accepted: 13-03-2026 Published: 20-03-2026

1. INTRODUCTION

The global demand for aquatic products has positioned aquaculture as a cornerstone of food security and environmental sustainability [4]. Among the various methodologies, Aquaponics the symbiotic integration of aquaculture (raising aquatic animals) and hydroponics (cultivating plants in water) has emerged as a highly efficient, closed-loop system. However, the viability of these systems depends on a delicate biological equilibrium between fish, plants, and nitrifying bacteria. As established in contemporary research, the quality of water directly affects the rate of growth, the efficiency of feed, and the overall health of the entire ecosystem [10]. Minor fluctuations in parameters such as Dissolved Oxygen (DO), pH levels, and nitrogenous compounds (Ammonia, Nitrate, and Nitrite) can lead to catastrophic system failure. Despite the benefits of aquaponics, many practitioners face a "knowledge gap" regarding species selection, as different fish and plant pairings have highly specific, and often conflicting, water quality requirements.

Figure 1 shows the fish growth (fish production) across all Indian states in 2025, measured in million tonnes. It highlights that Andhra Pradesh is the largest contributor with 5 million tonnes, shown in the darkest shade. Other major fish-producing states include West Bengal (1 million tonnes), Odisha (0.921 million tonnes), Gujarat (0.883 million tonnes), and Tamil Nadu (0.883 million tonnes). States like Uttar Pradesh (0.55), Assam (0.45), Bihar (0.5), and Karnataka (0.685) also make moderate contributions. Meanwhile, smaller and hilly states such as Himachal Pradesh, Uttarakhand, Sikkim, Mizoram, and Nagaland report very low production ranging between 0.001 and 0.04 million tonnes.

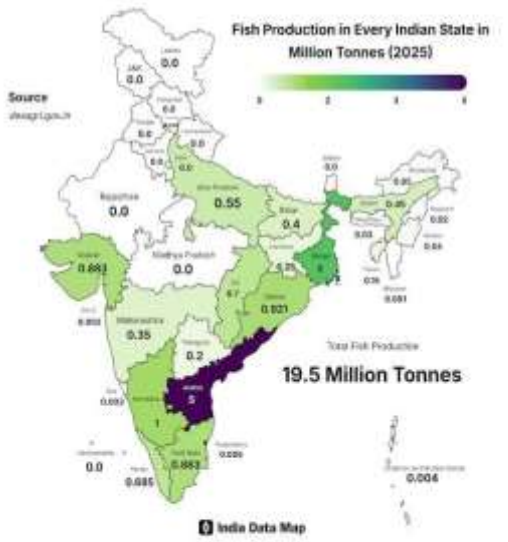

Figure 1. Fish growth in every Indian state in million tonnes.

India's total fish production in 2025 is 19.5 million tonnes, as indicated at the bottom of the map. Many central and northern states, such as Rajasthan (0), Haryana (0), and Delhi (0), show negligible or no fish production. Coastal regions and river-rich eastern states dominate the map due to favorable aquaculture environments. By improving infrastructure, promoting IoT-based monitoring, and providing training to farmers, India can further strengthen its fish production and move toward self-sufficiency while creating more economic opportunities in rural areas.

2. LITERATURE SURVEY

Recent advancements in aquaculture have shifted from manual oversight to data-driven management. Research in this domain primarily focuses on integrating Internet of Things (IoT) sensors with Deep Learning (DL) architectures to enhance productivity and environmental sustainability.

2.1 Fish Growth and Biomass Estimation

Accurate biomass estimation is critical for feeding optimization and harvest planning. Shaikh et al. [1] developed a Convolutional Neural Network (CNN) model to estimate fish length and weight by analyzing real-time IoT data, including water temperature and dissolved oxygen. This approach addresses the limitations of manual sampling, which is often stressful for the fish. Similarly, An-Qi et al. [4] identified growth prediction as a cornerstone of food security, noting that while DL models show promise, challenges remain in model generalization across different aquatic species and environments.

2.2 Intelligent Water Quality Management

Maintaining optimal water parameters is the primary challenge in intensive farming. Alghamdi et al. [2] explored the transformation of Biofloc Technology (BFT) into "smart" systems using AI analytics to manage microbial communities and stabilize water fluctuations. For specific parameter prediction, Shete et al. [8] proposed a hybrid CNN-Self Attention (SA)-Bidirectional Simple Recurrent Unit (BISRU) model, achieving an R2 of 0.9765 for dissolved oxygen (DO) estimation.

Other researchers have focused on ensemble methods; Mousa et al. [7] compared various regression techniques, finding that ensemble approaches (Random Forest and XGBoost) provided the most proactive decision-making capabilities. Furthermore, Palanikkumar et al. [10] emphasized that IoT-based smart water monitoring is essential for managing the complex symbiotic relationships in aquaponics systems, where water quality directly impacts fish, plants, and bacterial health.

2.3 Health, Behavior, and Disease Diagnosis

The transition from hand-crafted features to automated health monitoring has been a major trend. Kheriji et al. [3] conducted a comprehensive survey on DL-based multimodal health diagnosis, arguing that fusing visual imagery with behavioral signals provides a more robust framework than traditional manual inspections. In terms of behavioral classification, Janani et al. [11] introduced a region-based CNN optimized by Fish Swarm Optimization (FSO) to recognize "behavioral states" like aggression or resting, achieving a 94% accuracy rate. This real-time monitoring is further supported by AI-enabled decision-support systems that identify deviations from optimal conditions and alert farmers instantly [13].

2.4 Sustainability, Waste, and Ecological Impact

Beyond farm management, AI is being applied to the broader value chain and ecological conservation. Huang et al. [6] demonstrated the use of a 12-layer Deep Neural Network (DNN) to optimize the drying of fish by-products, reducing energy consumption by 35% and supporting waste valorization. However, technological optimism is balanced by research into environmental stressors. Wang et al. [14] utilized meta-analysis to show that microplastic (MP) exposure significantly inhibits growth and survival, while Islam et al. [15] highlighted the urgent need for better monitoring of hormonal residues in fish flesh, which pose significant risks to human consumers. These studies underscore the necessity of integrating ecological monitoring such as the Hybrid Machine Learning Strategy (HMLS) proposed by Xu et al. [9] to ensure the long-term sustainability of aquatic ecosystems.

3. PROPOSED SYSTEM

The proposed fish growth deep regression system as shown in Fig. 2 integrates IoT-based water quality data with a hybrid CNN–ETR model to accurately predict fish length and fish weight. Unlike existing linear, ridge, and lasso regression methods that assume simple linear relationships, the proposed approach captures complex non-linear interactions between environmental parameters and biological growth. Sensor data is processed, analyzed, and modeled through deep feature extraction using a CNN, while the ETR performs robust regression on the extracted features. The final system is deployed as a desktop application using Tkinter, enabling users to input real-time data and obtain reliable fish growth predictions.

The dataset contains time-based aquaculture sensor data such as temperature, turbidity, dissolved oxygen, ph, ammonia, nitrate, and population details, along with fish growth indicators like fish length (cm) and fish weight (g) as target variables. The data is first pre-processed by removing duplicates, handling missing values, and scaling numerical features to ensure consistent model learning. Exploratory Data Analysis (EDA) is then performed to understand data distribution and identify relationships between environmental factors and fish growth. Initially, Linear Regression is used as a baseline model but struggles with complex patterns. Ridge Regression improves stability by reducing multicollinearity through L2 regularization, while Lasso Regression performs feature selection using L1 regularization but still has limitations in capturing non-linear relationships.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

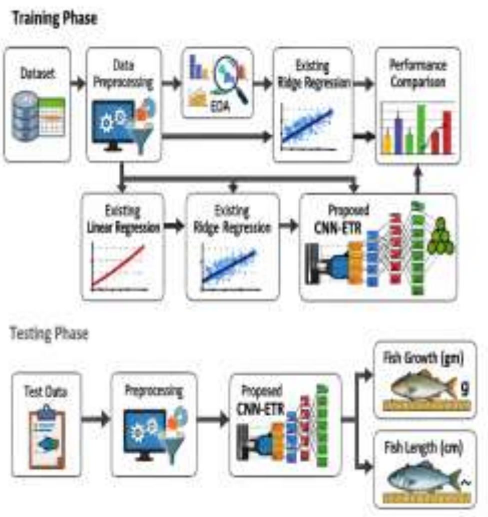

Figure 2. Proposed system architecture.

To overcome these issues, the proposed ConvETR model combines CNN for deep feature extraction with Extra Trees Regressor for ensemble-based prediction, enabling better learning of complex growth patterns. The models are compared using regression metrics such as MAE, MSE, RMSE, and R^2 score, where ConvETR achieves superior accuracy and robustness. Finally, the trained model is tested on unseen data to predict fish length and weight, and the complete system is integrated into a Tkinter-based desktop application that allows users to input water quality parameters and obtain real-time fish growth predictions.

4. RESULTS ANALYSIS

Figure 3 presents a combined visualization of feature relationships and data variability in the aquaponics dataset, where the correlation heatmap on the left illustrates the strength and direction of relationships among environmental and biological parameters, and the box plot on the right highlights the distribution of fish weight along with potential outliers. From the heatmap, fish length and fish weight exhibit a very strong positive correlation (0.91), indicating that increases in fish length are strongly associated with increases in fish weight, while nitrate concentration also shows a strong positive correlation with fish length (0.74) and fish weight (0.53), suggesting its importance in fish growth.

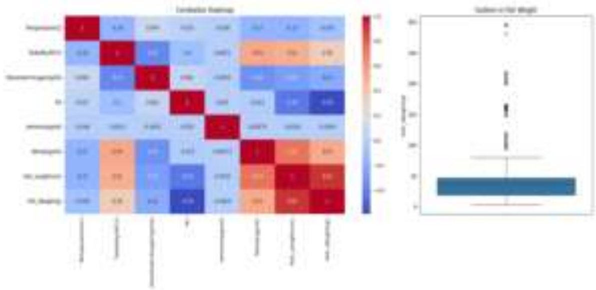

Figure 3. Visualization of feature relationships and data variability.

Turbidity demonstrates moderate positive correlations with fish length (0.51) and nitrate (0.54), whereas pH shows a moderate negative relationship with fish weight (−0.59), indicating that higher pH levels may adversely affect weight gain. Other parameters such as temperature and ammonia show weak correlations with growth metrics, implying a relatively smaller influence under the observed conditions. The box plot reveals that while most fish weight values are concentrated within a moderate range, several high-value outliers are present, indicating variability in growth and the occurrence of unusually large fish, which may result from biological diversity or favorable localized conditions within the system.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

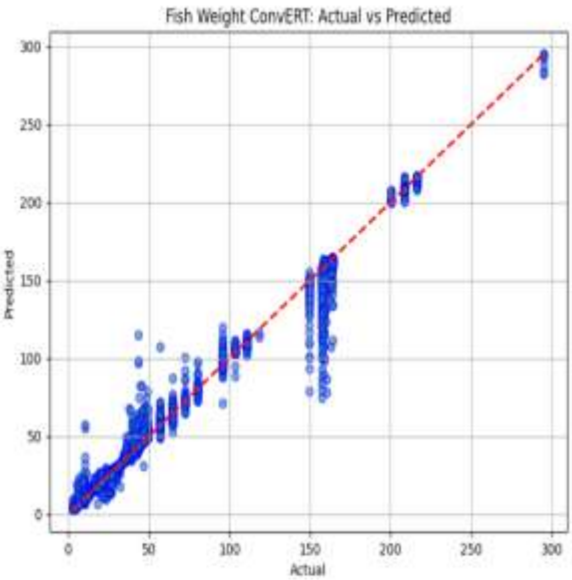

Figure 4. Fish weight scatter plot using the ConvETR

Figure 4 shows the actual vs. predicted fish weight results using the ConvETR model, where blue dots represent individual predictions and the red dashed line indicates the ideal one-to-one prediction line. The points are tightly clustered along the reference line across the entire weight range, demonstrating an excellent agreement between actual and predicted values. Compared to linear, Lasso, and Ridge regression plots, this figure exhibits significantly less dispersion and minimal deviation from the ideal line, indicating very low prediction error and strong generalization capability. Even at higher weight values, the predictions remain highly accurate with only minor variance, highlighting the model's ability to capture complex and nonlinear relationships in the data.

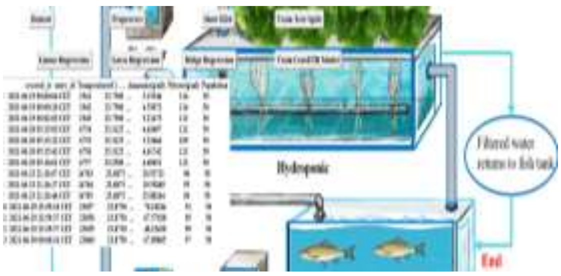

Figure 5. Prediction on sample test data.

Overall, the visualization confirms the superior performance of the ConvETR model for fish weight prediction, consistent with its high R^2 score and low error metrics. Figure 5 illustrates the prediction framework for sustainable aquaponics fish farming using real-time environmental and population data collected at regular intervals. The dataset includes parameters such as water temperature, ammonia concentration, nitrate levels, and fish population, which together influence fish growth prediction. As shown in Figure 5, the water temperature remains relatively stable in the range of 23.75°C to 25.31°C, creating suitable thermal conditions for fish health. Ammonia levels vary from approximately 4.55 g/ml to 20.54 g/ml, while nitrate concentrations show wider fluctuations, increasing from about 96 g/ml to nearly 114 g/ml across different timestamps. The fish population is consistently maintained at 50, indicating controlled stocking density. These continuously logged parameters are fed into the prediction framework to estimate growth-related outcomes, enabling timely monitoring of water quality, early detection of stress conditions, and data-driven decision-making for sustainable aquaponics management.

5. CONCLUSION

The proposed ConvETR-driven aquaponics monitoring framework demonstrates a highly effective and intelligent approach for predicting fish growth while maintaining sustainable water quality management. Experimental results show that traditional regression models such as linear, Lasso, and ridge achieve moderate performance, with fish length prediction MAE values around 2.09–2.10 cm and R^2 scores near 0.72, while fish weight prediction shows higher errors with MAE values of approximately 18.26 g and R^2 scores around 0.62. In contrast, the proposed ConvETR model significantly outperforms these methods, achieving an exceptionally low MAE of 0.0739 cm and RMSE of 0.2498 cm for fish length prediction, along with a high R^2 score of 0.9975. Similarly, for fish weight prediction, the ConvETR model achieves an MAE of 0.6748 g, an RMSE of 2.9725 g, and an R^2 score of 0.9957, demonstrating near-perfect prediction accuracy. These results confirm that the ConvETR model effectively captures complex, non-linear relationships between water quality parameters and fish growth, which are not adequately modeled by traditional regression techniques. By integrating real-time sensor data with advanced deep learning, the system supports accurate growth forecasting, efficient nutrient utilization, and continuous water recirculation in aquaponics systems. So, the framework enhances decision-making for fish farmers, reduces manual intervention, and promotes environmentally sustainable aquaculture practices through data-driven intelligence.

REFERENCES

[1] Shaikh, Asadullah, Wahidur Rahman, Kaniz Roksana, Tarequl Islam, Mohammad Motiur Rahman, Hani Alshahrani, Adel Sulaiman, and Mana Saleh Al Reshan. "An improved deep CNN-based freshwater fish classification with cascaded bio-inspired networks." *Automatika* 66, no. 2 (2025): 249-280.

[2] Alghamdi, Mansoor, and Yasmeen G. Haraz. "Smart Biofloc Systems: Leveraging Artificial Intelligence (AI) and Internet of Things (IoT) for Sustainable Aquaculture Practices." *Processes* 13, no. 7 (2025): 2204.

[3] Kheriji, Lazhar, Abdelmalek Kouadri, and Majdi Mansouri. "Deep Learning based Fish Health Monitoring and Diagnosis: A Review." *IEEE Access* (2025).

[4] Wu, An-Qi, Ke-Lei Li, Zi-Yu Song, Xiuhua Lou, Pingfan Hu, Weijun Yang, and Rui-Feng Wang. "Deep Learning for Sustainable Aquaculture: Opportunities and Challenges." *Sustainability* 17, no. 11 (2025): 5084.

[5] Arabelli, Rajeshwarrao, T. Bernatin, Venkataramana Veeramsetty, and CH Hussaian Basha. "Deep learning based optimal fish species identification to maximize production in fish ponds." *Scientific reports* 15, no. 1 (2025): 37688.

[6] Huang, Yo-Ping, and Simon Peter Khabusi. "Artificial intelligence of things (AIoT) advances in aquaculture: a review." *Processes* 13, no. 1 (2025): 73.

[7] Ali, Khaled Abdeen Mousa, Chagyou Li, Mohamed Fawzi Abdelshafie Abuhussein, Elwan Ali Darwish, and Gomaa Galal Abd El-wahhab. "Advanced Chain Regression and Deep Learning Models for Fish By-Product Drying Optimization: An Intelligent Conveyor System for Sustainable Waste Valorization." *Journal of Food Process Engineering* 48, no. 6 (2025): e70148.

[8] Shete, Rupali P., Amith Shekhar C, Yogeshwari V. Mahajan, Anupkumar M. Bongale, and Deepak Dharrao. "IoT-driven ensemble machine learning model for accurate dissolved oxygen prediction in aquaculture." *Discover Internet of Things* 5, no. 1 (2025): 94.

[9] Xu, Longqin, Wenjun Liu, Cai Chengqing, Tonglai Liu, Xuekai Gao, Ferdous Sohel, Murtaza Hasan, Mansour Ghorbanpour, Shahbaz Gul Hassan, and Shuangyin Liu. "Hybrid deep learning framework for real-time DO prediction in aquaculture."

Scientific reports 15, no. 1 (2025): 24643.

[10] Palanikkumar, D., T. Anuradha, Jothi Ramalingam, and S. S. Sivaraju. "A hybrid machine learning strategy for aquatic plant surveillance in sustainable aqua-ecosystems using IoT attributes." *Aquaculture* 609 (2025): 742779.

[11] Janani, V., S. Kamalakkannan, and P. Kavitha. "IoT-based Animal Species Recognition and Behavior Analysis using Deep Learning Technique." In *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, pp. 905-912. IEEE, 2025.

[12] Gupta, Manya, Gargi Mishra, Supriya Bajpai, Abhinav Bhardwaj, and Milind Gautam. "AI in Agriculture: A Comprehensive Exploration of Technological Transformation." *Emerging Smart Agricultural Practices Using Artificial Intelligence* (2025): 105-132.

[13] Abid, Muhammad Adeel, Madiha Amjad, Kashif Munir, Hafeez Ur Rehman Siddique, and Anca Delia Jurcut. "IoT-based smart Biofloc monitoring system for fish farming using machine learning." *IEEE Access* 12 (2024): 86333-86345.

[14] Wang, Jiaqi, Fan Wu, Shunqi Dong, Xusheng Wang, Shunhao Ai, Zhengtao Liu, and Xiaonan Wang. "Meta-analysis of the effects of microplastic on fish: Insights into growth, survival, reproduction, oxidative stress, and gut microbiota diversity." *Water Research* 267 (2024): 122493.

[15] Islam, Purba, Md Imran Hossain, Popy Khatun, Rony Ibne Masud, Shadia Tasnim, Mahir Anjum, Md Zahorul Islam, Salman Shahriar Nibir, Kazi Rafiq, and Md Anwarul Islam. "Steroid hormones in fish, caution for present and future: A review." *Toxicology Reports* 13 (2024): 101733.

[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] 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

[18] 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

[19] Todupunuri, A. . (2024). Artificial Intelligence Ethics: Investigating Ethical Frameworks, Bias Mitigation, and Transparency in AI Systems to Ensure Responsible Deployment and Use of AI Technologies. International Journal of Innovative Research in Science,Engineering and Technology, 13(09), 1–14. https://doi.org/10.15680/ijirset.2024.1309002

[20] 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

[21] 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

[22] Gaddam, S. INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING.

[23] Bajarang Bhagwat, V. (2023). Optimizing Payroll to General Ledger Reconciliation: Identifying Discrepancies and Enhancing Financial Accuracy. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 1(4). https://doi.org/10.56975/jaafr.v1i4.501636

[24] Srinivasa Kalyan Immadi. (2025). Harnessing Artificial Intelligence In Oracle Hcm: Revolutionising Workforce Management With Automation And Predictive Analytics. International Journal of Data Science and IoT Management System, 4(4), 7–13. https://doi.org/10.64751/ijdim.2025.v4.n4.pp7-13

[25] S. M. K. P. (2025). Cryptography in iOS: A Study of Secure Data Storage and Communication Techniques. International Journal on Science and Technology, 16(1). https://doi.org/10.71097/ijsat.v16.i1.1403

[26] Suhasnadh Reddy Veluru, Sai Teja Erukude, and Viswa Chaitanya Marella. 2025. Multimodal Detection of Fake Reviews using BERT and ResNet-50. In 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 877–882.

[27] Cyril, H. P. (2025). Event-Driven Provisioning Architectures For Modern Telecom Networks: Overcoming Legacy Limitations And Enabling Autonomous 6g Operations. International Journal of Advanced Research in Computer Science, 16(6), 75–82. https://doi.org/10.26483/ijarcs.v16i6.7389

[28] 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

[29] Reddy, S. K. (2025). Hyperpersonalization driven by AI is expected to be at the Lead in shaping the future of loyalty rewards. Journal of Emerging Technologies and Innovative Research.

[30] Reddy, S. K. R. (2021). Strengthening the Security of Loyalty Reward Systems: An In-Depth Analysis of Emerging Cyber Threats and Protection Mechanisms. Journal of Computational Analysis and Applications, 29(6).

[31] 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.

[32] Uday Kumar Kalae. (2025). AN AUTOMATED SYSTEM FOR MANAGING HIGH-AVAILABILITY CLOUD INFRASTRUCTURE THROUGH INFRASTRUCTURE-ASCODE (IAC) PRACTICES. American Journal of AI Cyber Computing Management, 5(2), 42–50. https://doi.org/10.64751/ajaccm.2025.v5.n2.pp42-50

[33] 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

[34] 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