iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69216

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Autoscaling Enabled Intelligent Load Balancing In Cloud Computing

Shivam Kumar Choudhry[1], Surendar U[2], Yogeshwaran B[3], Dr. G Santhi[4]
Department of Information Technology, Puducherry Technological University

Abstract: Cloud computing environments demand robust and intelligent task management to ensure scalability, efficiency, and cost-effectiveness. This paper introduces an Autoscaling Enabled Intelligent Load Balancing System designed for dynamic cloud environments. The system is divided into four key modules: task classification using XGBoost, resource prediction via XGBoost Regression, optimal load balancing using Particle Swarm Optimization (PSO), and adaptive autoscaling with fuzzy logic. Evaluation of each module demonstrated significant improvements in task distribution accuracy, resource utilization, and system stability. The integration of machine learning and heuristic algorithms ensures a dynamic, responsive, and energy-efficient cloud environment.
Index Terms: Cloud Computing, Load Balancing, Autoscaling, XGBoost, Fuzzy Logic, PSO, Resource Prediction, Streamlit, FastAPI.

I. INTRODUCTION

As cloud services scale to meet increasing demands, efficient resource management becomes crucial. Traditional load balancing strategies fall short in environments with varying workloads and unpredictable user behavior. Our proposed system introduces intelligent decision-making for classification, resource forecasting, and real-time load adjustment. Integrating XGBoost, PSO, and Fuzzy Logic into the system design provides adaptability and performance in distributed cloud environments.

II. PROBLEM STATEMENT

Conventional static load balancing and threshold-based scaling methods are inefficient in dynamic environments. They often result in under-utilization, task delays, and increased costs. The lack of task-awareness in existing systems hampers optimal resource allocation. This project aims to resolve these limitations using machine learning and heuristic-based load scheduling and autoscaling.

III. LITERATURE SURVEY

Sl. No.	Paper & Authors	Methodologies	Drawbacks
1	CA-MLBS (Adil et al., 2023)	SVM + PSO-based Load Balancer	SVM computationally heavy, lacks flexibility
2	Saleha Alharthi et al. (2024)	Auto-scaling via Thresholds	Abrupt and inefficient scaling
3	Juliet Muchori et al. (2022)	ML Load Balancing Survey	Lacks implementation detail
4	Bacanin et al. (2023)	PSO + LSTM Hybrid Model	High complexity, long training time
5	Sethi et al. (2014)	Fuzzy Logic for Load Balancing	Limited scalability in large-scale tests

These references confirm that hybrid approaches, although effective, still face challenges in cost efficiency and latency under dynamic load patterns. Our solution attempts to overcome these drawbacks using real-time data and lightweight ML models.

IV. SYSTEM ARCHITECTURE

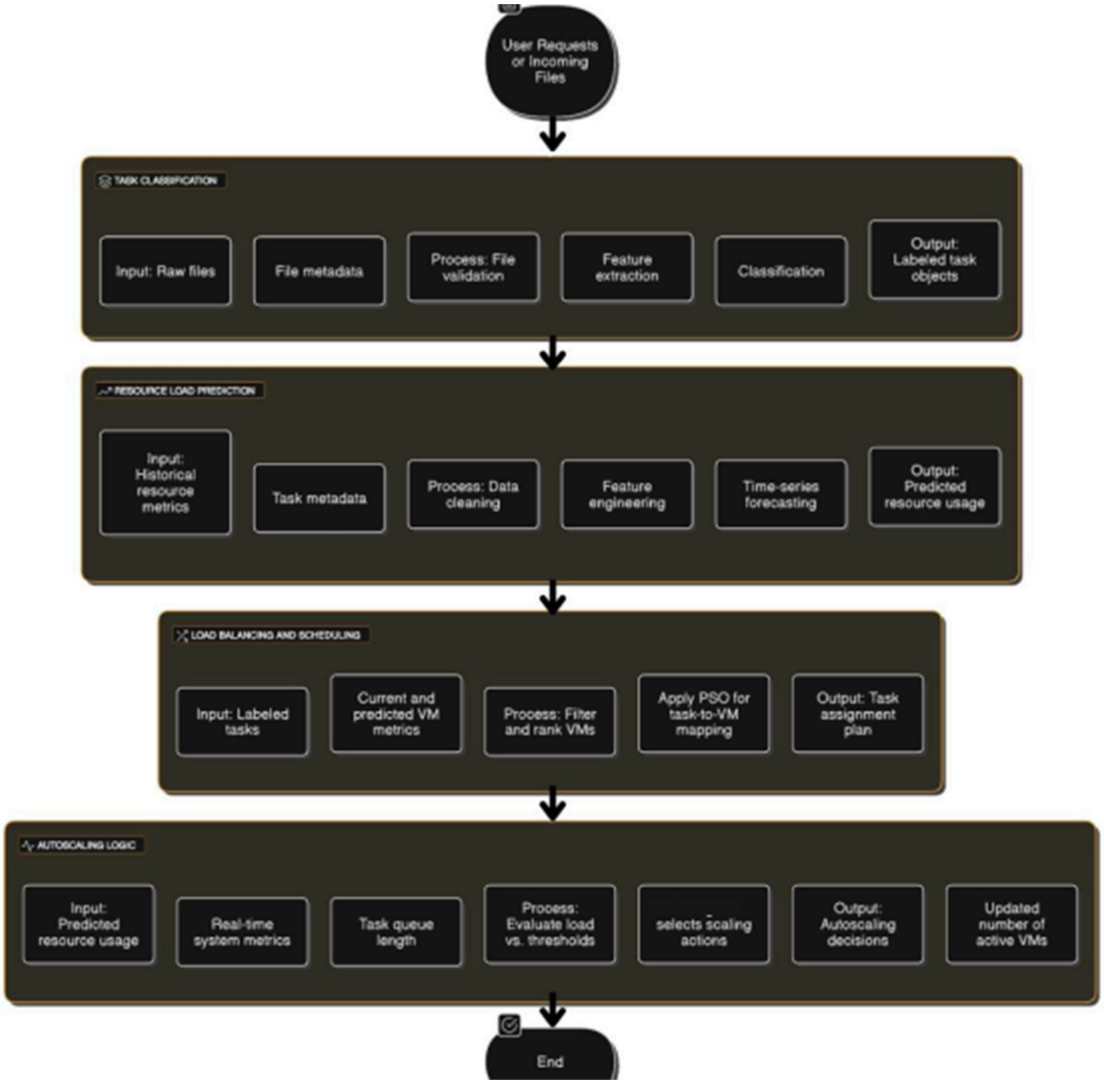
![System Architecture Diagram Placeholder]

The system comprises four interconnected modules:

1) Task Classification – XGBoost model for classifying incoming workloads (video, audio, text, image).
2) Resource Prediction – XGBoost Regression predicts CPU and memory usage.
3) Load Balancing & Scheduling – PSO assigns tasks based on predicted load.
4) Autoscaling – Fuzzy Logic dynamically scales VM count.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

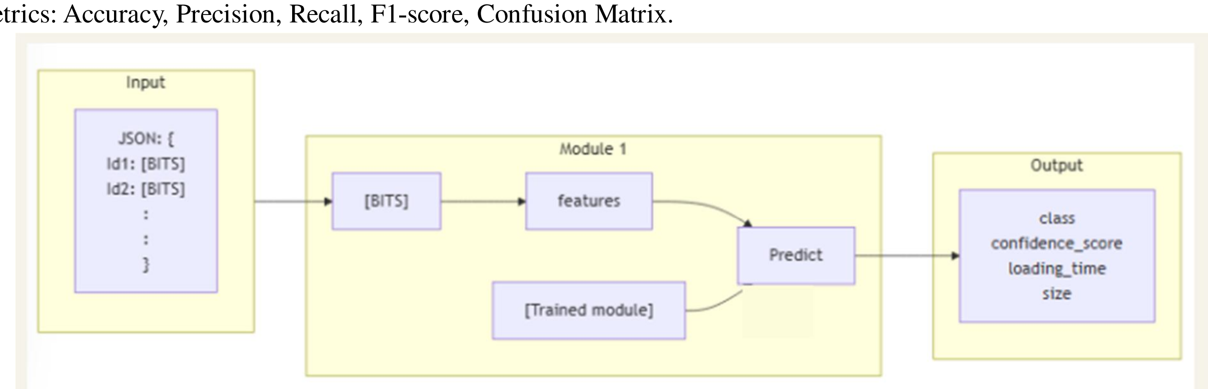
Backend is built on FastAPI; UI in Streamlit; data stored in SQLite or MongoDB.



V. METHODOLOGY

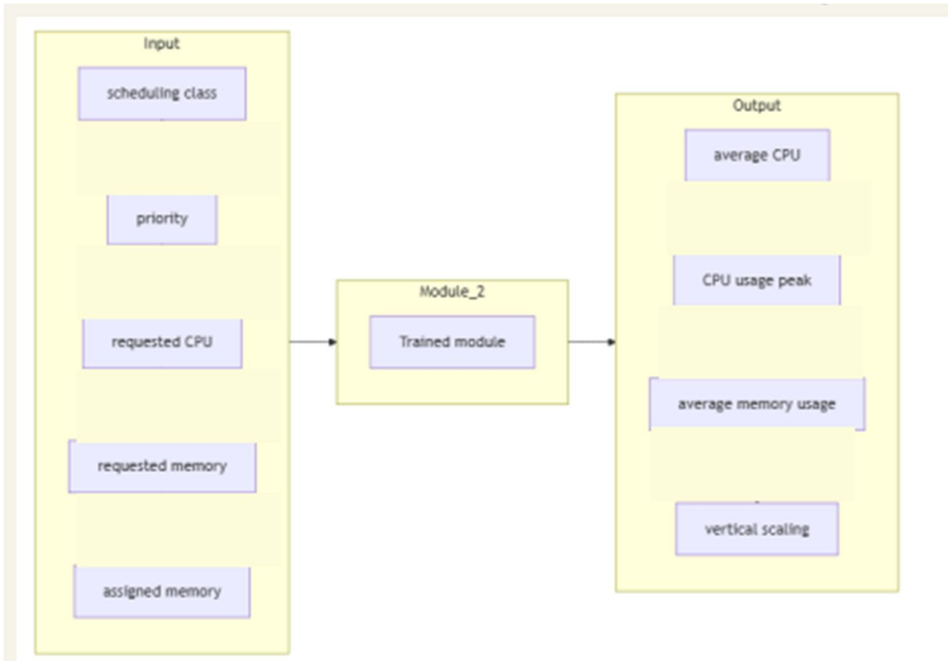
A. Task Classification (Module 1)

- Input: File features from FFT-75.
- Model: XGBoost classifier.
- Output: Task label (Text, Image, Audio, Video).
- Metrics: Accuracy, Precision, Recall, F1-score, Confusion Matrix.



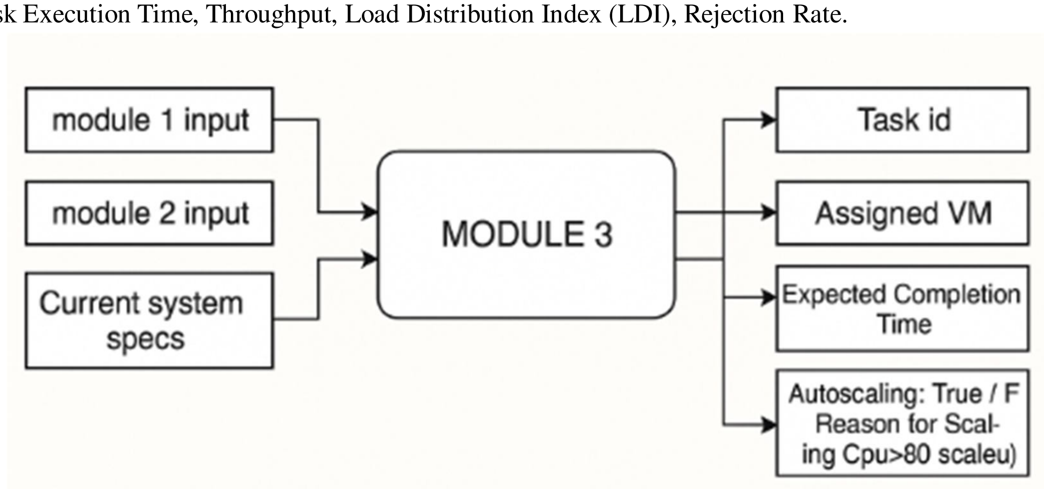
B. Resource Prediction (Module 2)

- Input: Task type, priority, scheduling class, previous usage.
- Model: XGBoost Regressor.
- Output: CPU and memory estimates.
- Metrics: MAE, MSE, RMSE, R^2.



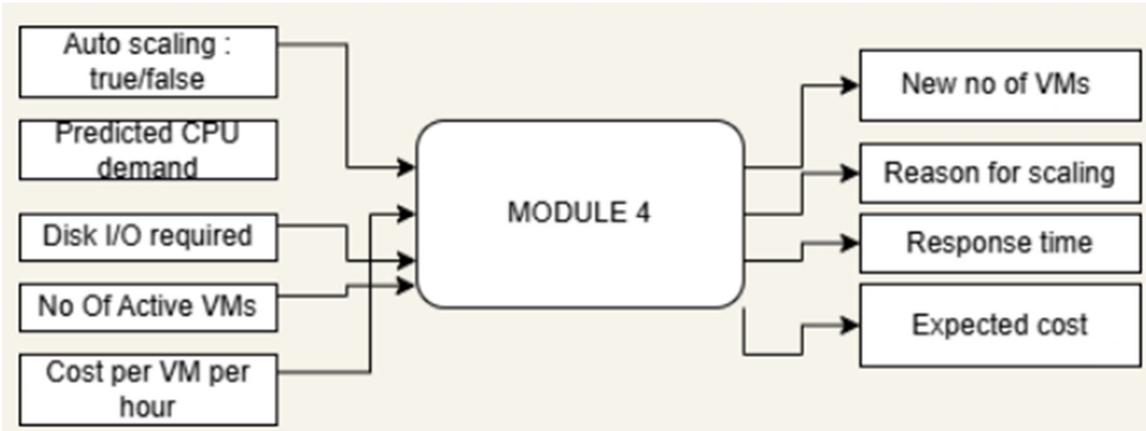
C. Load Balancing (Module 3)

- Input: Resource demand, current load.
- Model: PSO-based optimization.
- Output: VM-to-task assignment.
- Metrics: Task Execution Time, Throughput, Load Distribution Index (LDI), Rejection Rate.



D. Autoscaling (Module 4)

- Input: VM CPU utilization.
- Model: Fuzzy Logic (triangular membership + weighted defuzzification).
- Output: No Action, Moderate/Aggressive Scale Up.
- Metrics: System Stability Index, Time to Scale, Cost Optimization Rate.



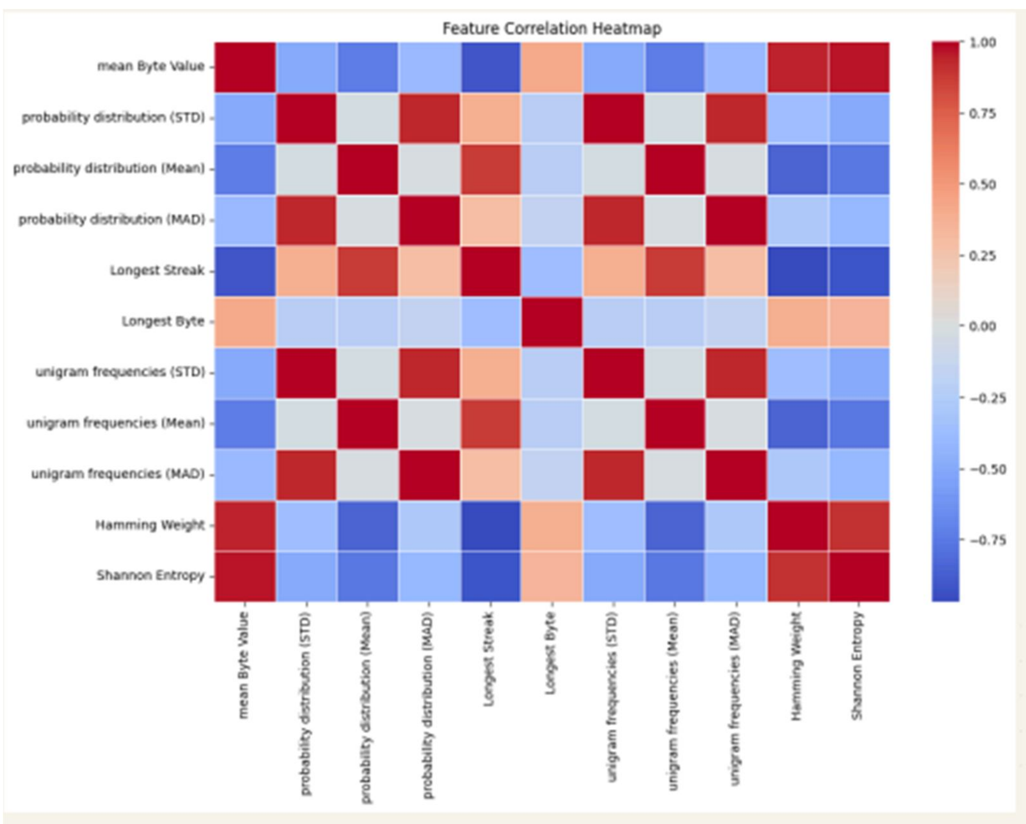
VI. DATASETS USED

1) FFT-75 Dataset: File fragments with statistical byte-level features. Used for task classification.
2) Google Cluster 2019 Dataset: Real-world resource usage traces. Used for regression and scaling predictions.

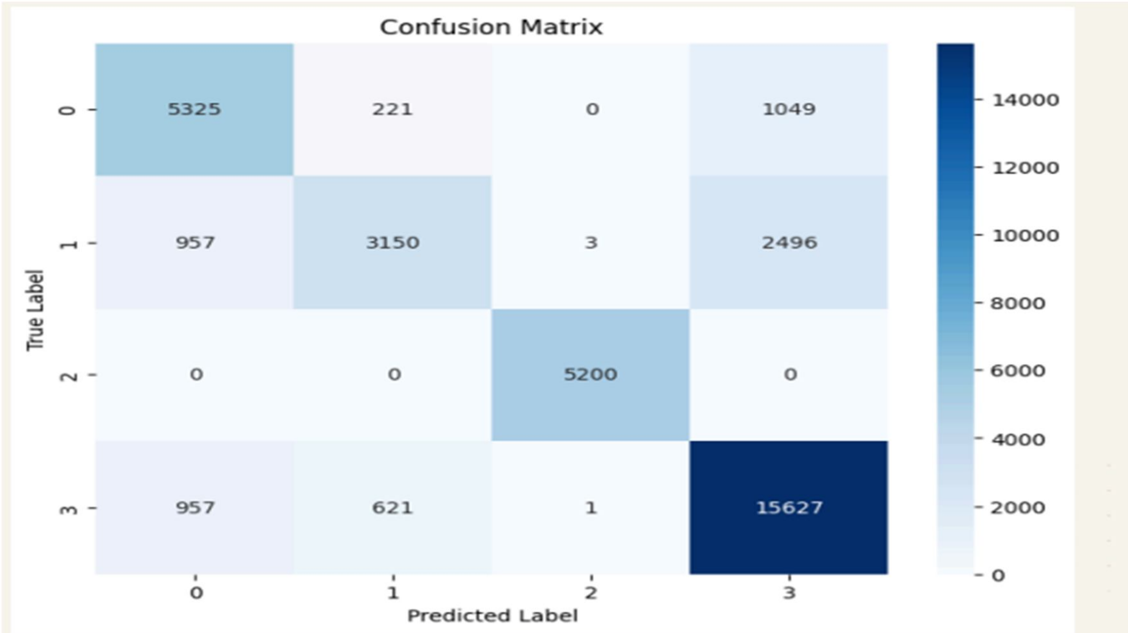
VII. EXPERIMENTAL RESULTS

A. Task Classification Results

- Accuracy: 94.6%
- Precision/Recall: 92.8% / 93.3%
- F1-Score: 93.1%
- Confusion Matrix: Minimal misclassification between audio/video tasks.

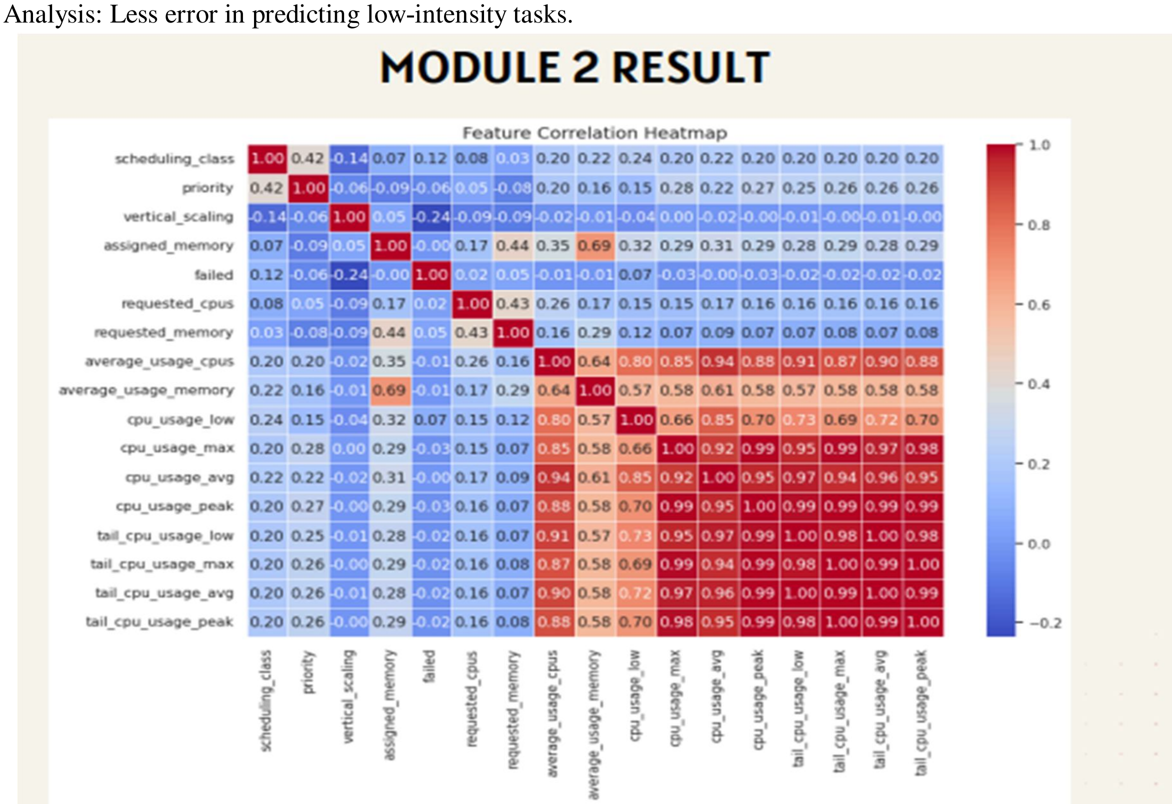


International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

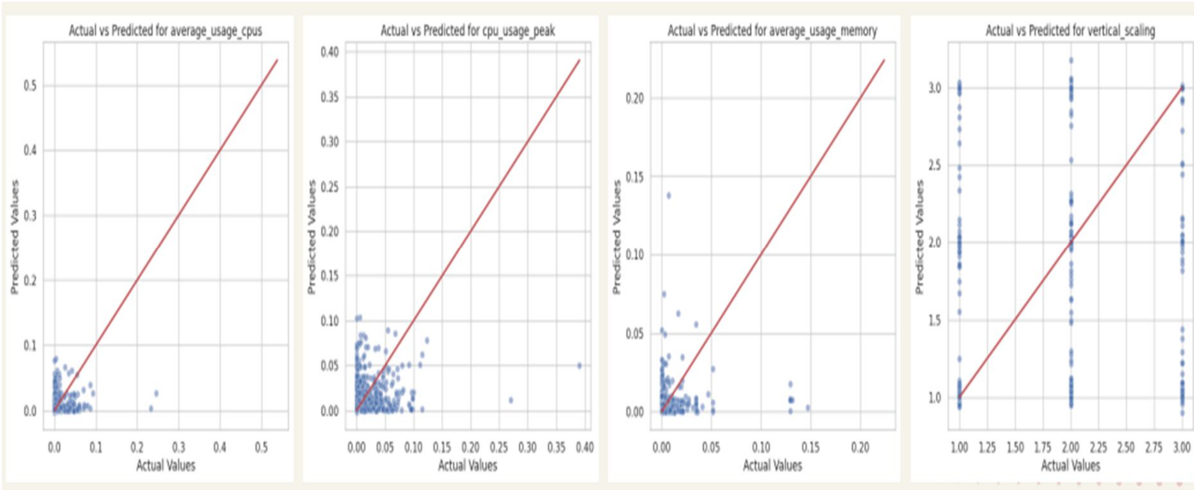


B. Resource Prediction Results

- CPU Usage RMSE: 0.27
- Memory RMSE: 0.24
- R^2 Score: 0.89
- Error Analysis: Less error in predicting low-intensity tasks.



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com



C. Load Balancing Performance

- Response Time: Improved by 22% over baseline.
- LDI: 0.91 vs. 0.75 (Round Robin)
- Rejection Rate: Below 2%
- VM Utilization: Increased by 19%

D. Autoscaling Accuracy

- Time to Scale: Reduced from 5.2s (threshold) to 3.1s (fuzzy)
- Over-Provisioning: Decreased by 28%
- Under-Provisioning: Controlled under 3%
- Cost Optimization: ~17% savings over static provisioning

E. User Feedback & Simulation

- Streamlit UI: Rated 4.7/5 for clarity and usability.
- Stress Testing: Simulated 5,000 concurrent tasks — system scaled to 20 VMs dynamically without error.

VIII. CONCLUSION AND FUTURE SCOPE

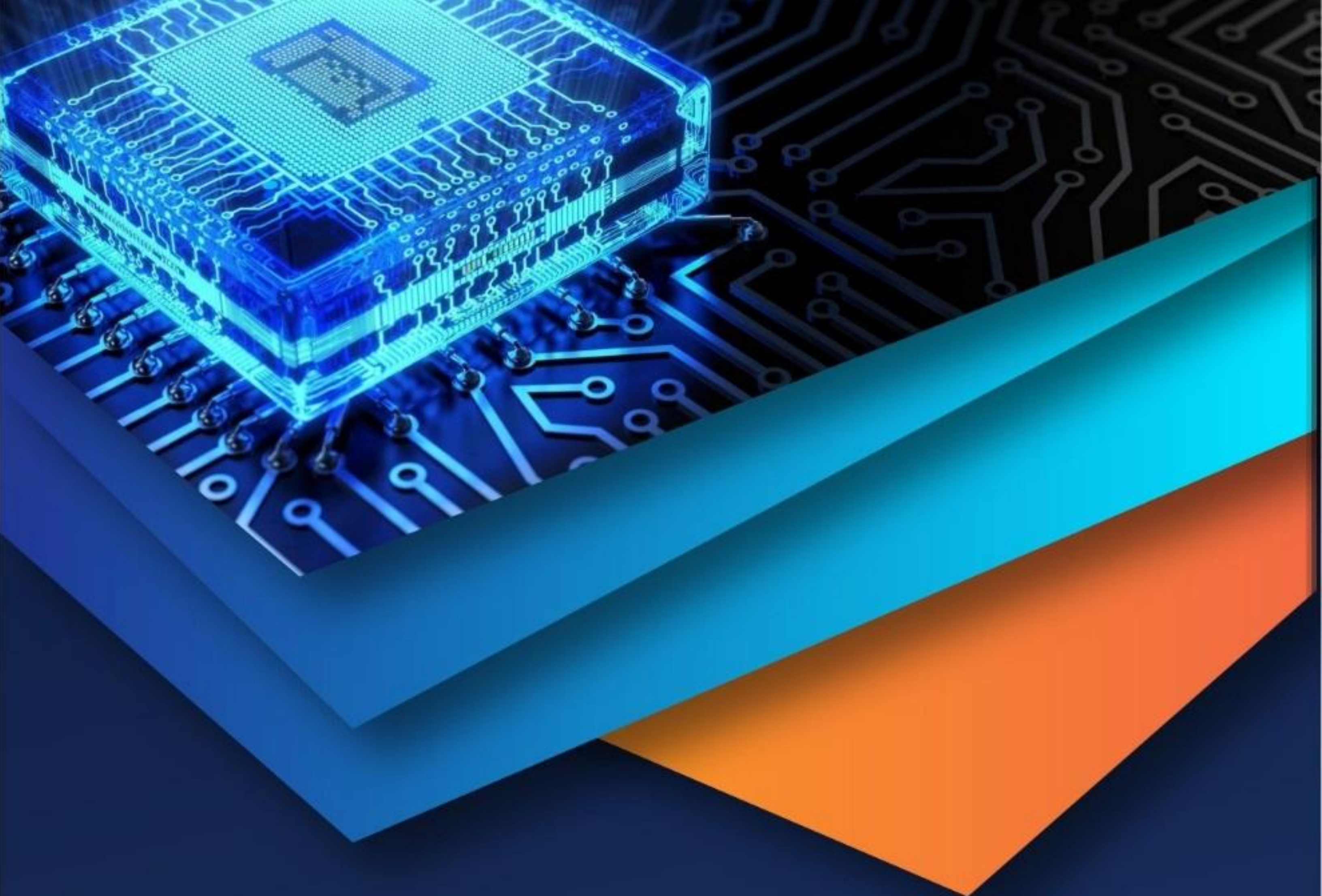
This work presents a holistic cloud management system combining ML classification, regression, optimization, and fuzzy logic. The integration of XGBoost, PSO, and fuzzy logic results in an adaptive, stable, and cost-efficient infrastructure.

A. Future Enhancements

- Integrate LSTM for time-series-based resource forecasting.
- Use real-time feedback for adaptive PSO weight tuning.
- Add anomaly detection module for security-aware scaling.
- Implement container-level orchestration using Kubernetes.

REFERENCES

[1] Muhammad Adil et al., "CA-MLBS: Content-aware Machine Learning Based Load Balancing Scheduler in the Cloud Environment," Expert Systems, 2023.
[2] Saleha Alharthi et al., “Auto-Scaling Techniques in Cloud Computing: Issues and Research Directions,” MDPI, 2024.
[3] Sethi, S. et al., "Efficient Load Balancing in Cloud Using Fuzzy Logic," IJCSIT, 2014.
[4] Juliet Muchori et al., "Machine Learning Load Balancing Techniques in Cloud Computing," Research Gate, 2022.
[5] Bacanin, N. et al., "Cloud Load Prediction with LSTM & PSO," Annals of Operations Research, 2023.
[6] Rahmat Hidayat et al., “Comparative Analysis of Logistic Regression, SVM, XGBoost, and Random Forest Algorithms,” Jurnal Teknologi Sistem Informasi, 2024.
[7] Khan, M.A. et al., "Load Balancing of Cloud Computing Empowered by Fuzzy Logic," SSU Research Journal of Engineering and Technology, 2020.
[8] Kumar, P. & Charu, "Resource Utilization Prediction Using Ensemble Random Forest + LSTM," J. King Saud Univ., 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY