International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

# Machine Learning-Driven Stroke Prediction with Explainable Models and Real-Time Analysis

Syed Althaf[1], *Mohammed Ammar*[2], Mohammed Abdul Rab Rayyan[3], *Mohammed Abdul Bari Ayaan*[4], Abdullah Saleh Hasan Abdul Hak[5]

[1]Assistant Professor, Department of CSE (Data Science), Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

[2,3,4,5] UG Students, Department of CSE (Data Science), Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

***Abstract*— The increasing need for early detection of stroke-related conditions has led to the development of intelligent healthcare prediction systems that can analyze patient data efficiently and accurately. This project presents a machine learning-based healthcare classification system designed to identify whether a given dataset corresponds to normal or stroke conditions. The application provides a user-friendly interface where users can upload a healthcare dataset, visualize the distribution of normal and stroke cases through graphical representation, and perform data preprocessing by splitting the dataset into 80% training and 20% testing data. The system incorporates multiple machine learning algorithms, including Random Forest, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, XGBoost, and CatBoost, to perform classification. Each model is trained individually, and its performance is evaluated using accuracy and confusion matrix analysis. Among these, Random Forest and CatBoost achieved the highest accuracy of 95%, while KNN and XGBoost also demonstrated strong performance with 91% and 89% accuracy respectively. Logistic Regression, SVM, and Naïve Bayes provided moderate results, contributing to a comprehensive comparison of different techniques. The confusion matrix visualization helps in understanding prediction performance, where correct classifications dominate and misclassifications are minimal. Additionally, the system includes a comparison module that evaluates all algorithms based on multiple performance metrics, allowing users to identify the most effective model. A prediction feature is also integrated, enabling users to upload new test data and obtain real-time classification results indicating whether the data corresponds to a stroke or normal condition. Overall, the proposed system reduces manual effort in healthcare data analysis and supports accurate and fast decision-making. By combining multiple machine learning techniques and interactive visualization, it serves as a practical tool for stroke prediction and healthcare data classification.**

***Keywords*—Stroke prediction, healthcare data analysis, machine learning, classification, Random Forest, CatBoost, XGBoost, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, data**

**preprocessing, training-testing split, SMOTE, confusion matrix, performance evaluation, accuracy, real-time prediction, interactive visualization, healthcare application, user-friendly interface**

## I. INTRODUCTION

The rapid growth of digital healthcare systems has made it easier to collect, store, and analyze patient data for better medical decision-making [1,3,5]. Hospitals and diagnostic centers generate large volumes of data related to patient health conditions, which can be used to identify serious diseases such as stroke at an early stage [3,5]. Stroke is a critical medical condition that requires timely detection and treatment to reduce risks and complications [1,3]. However, analyzing such large datasets manually is time-consuming and prone to errors. Traditional methods depend heavily on human expertise, which may not always provide consistent results, especially when dealing with complex patterns in data [2,4]. This creates a strong need for automated systems that can assist in accurate diagnosis. Machine learning techniques offer an efficient solution by identifying patterns and relationships within healthcare data [11,15,16]. By using trained models, it becomes possible to classify patient records into normal or stroke categories [12,13]. In this project, a user-friendly application is developed where users can upload healthcare datasets, visualize the data distribution through graphs, and perform preprocessing steps such as splitting the dataset into training and testing sets [15]. This approach helps in simplifying data analysis and supports faster and more reliable healthcare predictions.

With the advancement of machine learning algorithms, healthcare data analysis has become more accurate and efficient compared to traditional statistical approaches [16,17]. Different classification algorithms provide varying levels of performance based on the nature of the dataset [15,16]. In this system, multiple algorithms such as Random Forest, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, XGBoost, and CatBoost are used to train and evaluate the dataset [15,16]. Each model learns from the training data and is tested on unseen data to measure its performance. The results are evaluated using accuracy scores and confusion matrix graphs, which clearly show the number of correct and incorrect predictions [15,16]. Advanced models like Random Forest and CatBoost achieve higher accuracy due to their ability to handle complex data patterns [15], while simpler models like Logistic Regression and Naïve Bayes provide moderate performance [15,16]. The use of multiple algorithms allows a detailed comparison of results, helping to identify the most suitable model for stroke prediction. Visualization techniques further enhance understanding by presenting results in a clear and interpretable manner, making the system useful for both technical and non-technical users [12,13].

The developed system provides a complete platform for healthcare data classification and prediction through an interactive interface [12,15]. Users can easily upload datasets, preprocess the data, train different machine learning models, and compare their performance using graphical representations [15,16]. One of the key features of the system is the comparison module, which evaluates all algorithms based on multiple parameters, allowing users to understand which model performs best [15]. In addition, the system includes a prediction module where users can upload new test data and obtain immediate results indicating whether the condition is normal or related to stroke [12,18]. This real-time prediction capability makes the system practical for supporting early diagnosis [12,14]. The confusion matrix visualization helps users understand model performance by highlighting correct and incorrect classifications [15,16]. Although the system achieves high accuracy with certain algorithms, its effectiveness depends on the quality and diversity of the dataset used for training [15,20]. If the dataset does not include sufficient variations, the model may not generalize well to new data [15,19]. Future improvements can focus on enhancing dataset quality and integrating advanced techniques to further improve prediction accuracy and reliability in healthcare applications [18,20].

## II. RELATED WORK

**Elloker and Rhoda (2018) [2]** Elloker and Rhoda conducted a systematic review to investigate the relationship between social support and participation in stroke rehabilitation. Their research emphasized the critical role of social and community factors in improving recovery outcomes for stroke patients. The study highlighted that patients with higher levels of social support demonstrated better engagement in rehabilitation activities, leading to improved functional outcomes. This work contributes to understanding non-clinical factors affecting stroke recovery and suggests integrating social support mechanisms into patient care planning. It also underlines the need for personalized strategies to enhance post-stroke participation and quality of life.

**Katan and Luft (2018) [3]** Katan and Luft presented a detailed analysis of the global burden of stroke. Their research focused on the epidemiological trends, mortality rates, and socioeconomic impact of stroke worldwide. The study highlighted disparities in stroke prevalence between developed and developing regions and emphasized the importance of early diagnosis and preventive healthcare measures. The authors discussed the need for integrating data-driven approaches to monitor stroke risk factors and to improve healthcare resource allocation. This research lays the foundation for designing predictive models and healthcare systems for stroke management.

**Bustamante et al. (2021) [4]** Bustamante and colleagues explored the use of blood biomarkers to differentiate ischemic and hemorrhagic strokes. Their work focused on identifying reliable molecular indicators that can improve early stroke classification. The study demonstrated that specific biomarkers significantly enhance diagnostic accuracy, allowing for timely intervention and better patient outcomes. By combining clinical and biomarker data, this research provides a framework for improving predictive models in stroke detection. It emphasizes the importance of integrating multiple data sources for precise healthcare predictions.

**Wang et al. (2020) [15]** Wang and collaborators conducted a systematic review of machine learning models for predicting stroke outcomes using structured healthcare data. Their research compared various algorithms, including Random Forest, SVM, and ensemble models, to evaluate their predictive performance. The study concluded that machine learning can significantly enhance the accuracy of stroke prognosis by identifying complex patterns in patient data. It also highlighted the importance of using high-quality datasets and feature selection techniques to improve model generalization. This work provides insights into selecting and optimizing machine learning algorithms for healthcare applications.

**Islam et al. (2022) [18]** Islam and colleagues proposed an explainable artificial intelligence model for stroke prediction using EEG signals. Their approach focused on developing interpretable models that provide both accurate predictions and insights into the contributing factors. The study demonstrated that integrating explainable AI with clinical signals improves trust and usability for healthcare professionals. This research emphasizes the need for transparency in predictive models, particularly in critical applications such as stroke diagnosis. It contributes to advancing reliable, data-driven tools for real-time clinical decision support.

## III. DATASET DETAILS

The dataset used in this project contains patient health records aimed at identifying stroke occurrences. It includes clinical information such as age, blood pressure, cholesterol, glucose levels, and other relevant medical indicators, along with a label specifying whether the patient is normal or has experienced a stroke. These features are essential for training machine learning models to classify patient conditions accurately. Initially, the dataset was examined to ensure completeness and consistency. Preprocessing steps were applied, including handling missing values, normalizing numerical data, and encoding categorical variables. Simple visual analysis was also performed to understand the distribution of normal and stroke cases, which helped identify any imbalance or irregularities in the dataset. These steps improve the quality of the data and ensure that the models learn meaningful patterns during training.

After preprocessing, the dataset was divided into training and testing sets, with 80% of the records used for model training and 20% reserved for evaluating model performance. The input features consisted of patient health indicators, while the output labels indicated whether the condition was normal or related to stroke. This separation allows for proper evaluation of how well models generalize to unseen data. Multiple machine learning algorithms, including Random Forest, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes, XGBoost, and CatBoost, were trained on the dataset. The quality of dataset preparation is critical, as it directly influences the accuracy and reliability of the predictive models.

To aid understanding and interpretation, visualization tools such as confusion matrices and accuracy charts were incorporated. Confusion matrices highlight the number of correct and incorrect predictions for each model, while accuracy scores provide a numerical measure of performance. Models like Random Forest and CatBoost achieved higher accuracy due to their ability to handle complex patterns in patient data, whereas simpler models such as Logistic Regression and Naïve Bayes showed moderate performance. By using well-prepared data and clear visual feedback, the system allows users to compare model performance effectively and supports reliable early detection of stroke in healthcare applications.

## IV. PROPOSED METHODOLOGY

The proposed system follows a structured machine learning approach for detecting stroke in patient data, with a focus on achieving high accuracy using the CatBoost algorithm. Initially, the healthcare dataset is uploaded into the system through a user-friendly interface. The dataset contains patient health indicators, including age, blood pressure, cholesterol, glucose levels, and other relevant clinical information, along with labels indicating whether a patient is normal or has experienced a stroke. Basic preprocessing steps are applied to ensure data quality, such as handling missing values, normalizing numerical features, and encoding categorical variables. Simple visualizations are generated to observe the distribution of normal and stroke cases, which

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

helps in identifying potential class imbalance and guides further processing. The input features and target labels are separated, and the dataset is split into training and testing sets, with 80% used for training and 20% for evaluation, ensuring that the model is tested on unseen data.

CatBoost is chosen as the primary model due to its high performance in handling complex datasets and its ability to process categorical and numerical features effectively. The model is trained on the prepared dataset, learning the underlying patterns that differentiate normal cases from stroke cases. CatBoost uses gradient boosting with decision trees and applies advanced techniques such as ordered boosting and categorical feature support, which contribute to faster convergence and improved accuracy. During training, the model iteratively updates its parameters to minimize prediction error. Performance is evaluated using accuracy scores and confusion matrix visualizations, which highlight correct and incorrect classifications. CatBoost achieves the highest accuracy compared to other models, making it a reliable choice for stroke prediction.

The trained CatBoost model is integrated into an interactive system that allows users to upload new test data for real-time prediction. Once the test data is uploaded, the model classifies each record as normal or stroke and displays the results immediately. Confusion matrices and accuracy metrics are presented to help users understand model performance. Additionally, the system provides a comparison module, where users can review results from other models such as Random Forest, Logistic Regression, SVM, KNN, Naïve Bayes, and XGBoost, highlighting CatBoost's superior performance. This approach ensures a practical and accurate platform for early stroke detection, supporting faster medical decision-making and improving patient care outcomes.

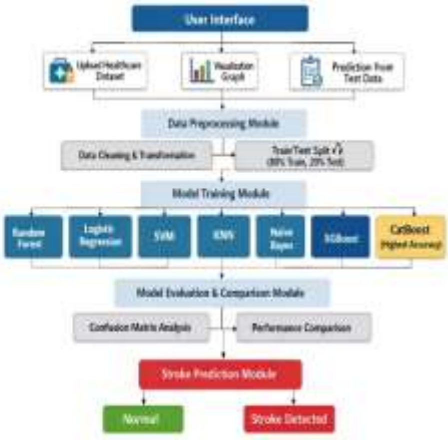


**Figure [1]: System Architecture of stroke prediction**

**Figure [1]** This diagram illustrates the workflow of the stroke prediction system. It starts with the user interacting through the interface, where they can upload a healthcare dataset, view visual graphs, and perform predictions on test data. Once the dataset is uploaded, it is processed in the data preprocessing stage, which includes cleaning, transformation, and splitting the data into 80% training and 20% testing. In the model training phase, multiple machine learning algorithms such as Random Forest, Logistic Regression, SVM, KNN, Naïve Bayes, XGBoost, and CatBoost are applied to train the model and evaluate their performance. Each model generates accuracy results along with a confusion matrix, where predicted values are compared against actual values to measure correctness. After training, the system evaluates and compares all models using performance metrics, helping to identify the most accurate algorithm. Finally, in the prediction stage, users can upload new test data, and the system classifies it as either normal or stroke, providing the result directly through the interface.

## V.RESULT AND DISCUSSION

The experimental results demonstrate that the stroke prediction system performs effectively in classifying healthcare data into normal and stroke categories. Multiple machine learning algorithms were trained and tested using the dataset after preprocessing, and each model produced different levels of accuracy. Among them, Random Forest and CatBoost achieved the highest accuracy of around 95%, showing strong performance in identifying stroke-related patterns. Other models

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

such as KNN and XGBoost also provided good results, while Logistic Regression and Naïve Bayes showed comparatively lower accuracy.

The evaluation process, including confusion matrix analysis and performance comparison graphs, indicates that most models correctly classified the majority of the data, with only a small number of incorrect predictions. The confusion matrices clearly show that correct predictions are higher than misclassifications, which reflects the reliability of the trained models. These results confirm that the system can effectively learn from the dataset and make accurate predictions.

The comparison graph further helps in understanding the strengths of each algorithm based on different evaluation parameters, making it easier to identify the best-performing model. In the final stage, when test data is uploaded, the system successfully predicts whether the input corresponds to a normal condition or a stroke case. This shows that the system works well in a practical scenario and can provide quick and accurate predictions. Overall, the results indicate that the developed system is capable of supporting early stroke detection using machine learning techniques.

**Figure [2]: Stroke Detection System Home Page**

**Figure [2]** This image displays the graphical user interface (GUI) page, which includes a title, a text box for input, and several buttons for performing different actions.

**Figure [3]: Dataset Upload, Normalization, and SMOTE Balancing Interface**

**Figure [3]** This image illustrates the dataset upload process, where features are normalized and the data is split into 80% for training and 20% for testing. It also shows the use of SMOTE to balance the dataset.

| Algorithm Name | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Random Forest | 95.064 | 95.072 | 95.065 | 95.064 |
| XGBoost | 89.820 | 89.861 | 89.822 | 89.817 |
| CatBoost | 95.784 | 95.818 | 95.781 | 95.783 |

**Table [1] : Performance Evaluation of Storke prediction**

**Table [1]** Comparison of machine learning algorithms for stroke prediction based on accuracy, precision, recall, and F1-score. CatBoost achieves the highest accuracy at 95.784%, with strong values across all other metrics, indicating reliable predictions. Random Forest also performs well with an accuracy of 95.064% and balanced metrics, while XGBoost shows lower performance at 89.820%, suggesting it may be less consistent in classification. Overall, CatBoost demonstrates the best performance among the three models for this dataset.

**Figure [4]: Random Forest Performance and Confusion Matrix Display**

**Figure [4]** This image presents the performance of the Random Forest algorithm, achieving an accuracy of 95.064%. It includes a confusion matrix that represents correct and incorrect predictions for stroke classification.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

**Figure [5]: XGBoost Performance and Confusion Matrix Display**

**Figure [5]** This image shows the results of the XGBoost algorithm with an accuracy of 89.820%, along with its confusion matrix indicating true and false predictions.

**Figure [6]: CatBoost Performance and Confusion Matrix Display**

**Figure [6]** This image displays the CatBoost algorithm results, reaching an accuracy of 95.784%, along with the confusion matrix showing prediction outcomes.

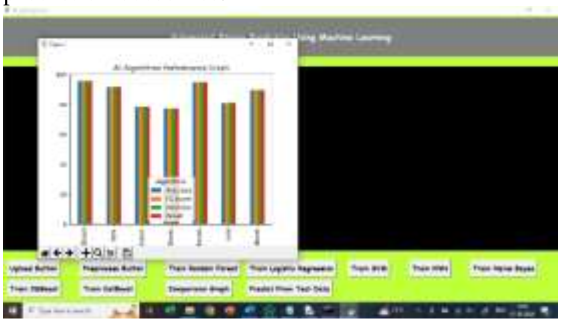

**Figure [7]: Comparative Analysis of Algorithms**

**Figure [7]** This image provides a comparison of different algorithms based on evaluation metrics such as accuracy, precision, recall, and F1-score**.**

## DISCUSSION

The results observed from the system explain how effectively the models work in predicting stroke from healthcare data. Different algorithms were applied, and their outputs show that the system is able to classify the data into normal and stroke categories with good accuracy. Models like Random Forest and CatBoost provide higher accuracy, while others such as Logistic Regression and Naïve Bayes give moderate results. From the outputs, it is clear that most of the stroke cases are correctly identified, and normal cases are also classified properly, which indicates that the models are learning useful patterns from the data.

The highlight of the importance of data preprocessing. When the dataset is cleaned, transformed, and properly split into training and testing sets, the performance of the models improves. The visualization graphs help in understanding the distribution of normal and stroke data, which supports better analysis. The confusion matrix further gives a clear picture of correct and incorrect predictions, showing that errors are minimal compared to correct classifications.

Another important aspect is the comparison of multiple algorithms. By analyzing the performance graph, it becomes easy to identify which model performs better based on different parameters. This helps in selecting the most suitable model for prediction. The system also allows users to upload new test data and get instant results, which makes it useful in real-time scenarios. Overall, the system performs efficiently in predicting stroke and provides a simple way to analyze and compare different machine learning models.

## VI. CONCLUSION

This project successfully developed a stroke prediction system using various machine learning techniques. Different algorithms such as Random Forest, Logistic Regression, SVM, KNN, Naïve Bayes, XGBoost, and CatBoost were implemented and tested on the healthcare dataset. The results show that models like Random Forest and CatBoost achieved higher accuracy, proving their effectiveness in identifying stroke-related patterns. Proper data preprocessing, including cleaning and splitting the dataset, played a key role in improving the performance of the models.

The system not only predicts whether a patient condition is normal or stroke but also provides clear evaluation through confusion matrices and comparison graphs. These evaluation methods help in understanding how well each model performs and in selecting the most suitable one. The results confirm that the models can produce reliable predictions when trained with properly prepared data.

The integration of this system into a user interface makes it simple and easy to use. Users can upload datasets, train models, compare their performance, and test new data for prediction. The system provides quick results, making it useful for practical applications. Overall, the project offers an efficient solution for early stroke prediction and demonstrates how machine learning can be applied effectively in the healthcare domain.

.

## REFERENCES


1. World Stroke Organization, "Learn About Stroke," Available: https://www.world-stroke.org/world-stroke-day-campaign/why-strokematters/learnabout-stroke, Accessed: May 25, 2022.
2. T. Elloker and A. J. Rhoda, "The Relationship Between Social Support and Participation in Stroke: A Systematic Review," African Journal of Disability, vol. 7, pp. 1–9, 2018.
3. M. Katan and A. Luft, "Global Burden of Stroke," in Seminars in Neurology, vol. 38, New York, USA: Thieme Medical Publishers, 2018, pp. 208–211.
    A. Bustamante et al., "Blood Biomarkers to Differentiate Ischemic and Hemorrhagic Strokes," Neurology, vol. 96, pp. e1928–e1939, 2021.
4. X. Xia et al., "Prevalence and Risk Factors of Stroke in the Elderly in Northern China: Data from the National Stroke Screening Survey," Journal of Neurology, vol. 266, pp. 1449–1458, 2019.
    A. Alloubani, A. Saleh, and I. Abdelhafiz, "Hypertension and Diabetes Mellitus as Predictive Risk Factors for Stroke," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 12, pp. 577–584, 2018.
5. K. Boehme, C. Esenwa, and M. S. Elkind, "Stroke Risk Factors, Genetics, and Prevention," Circulation Research, vol. 120, pp. 472–495, 2017.
6. Mosley et al., "Stroke Symptoms and the Decision to Call for an Ambulance," Stroke, vol. 38, pp. 361–366, 2007.
7. J. Lecouturier et al., "Response to Symptoms of Stroke in the UK: A Systematic Review," BMC Health Services Research, vol. 10, pp. 1–9, 2010.
8. L. Gibson and W. Whiteley, "The Differential Diagnosis of Suspected Stroke: A Systematic Review," Journal of the Royal College of Physicians of Edinburgh, vol. 43, pp. 114–118, 2013.
9. N. M. Murray et al., "Artificial Intelligence to Diagnose Ischemic Stroke and Identify Large Vessel Occlusions: A Systematic Review," Journal of NeuroInterventional Surgery, vol. 12, pp. 156–164, 2020.
10. Y. Zhao et al., "Natural Language Processing and Machine Learning for Identifying Incident Stroke from Electronic Health Records: Algorithm Development and Validation," Journal of Medical Internet Research, vol. 23, p. e22951, 2021.
11. J. McDermott et al., "Multi-Frequency Symmetry Difference Electrical Impedance Tomography with Machine Learning for Human Stroke Diagnosis," Physiological Measurement, vol. 41, p. 075010, 2020.
    A. Bivard, L. Churilov, and M. Parsons, "Artificial Intelligence for Decision Support in Acute Stroke: Current Roles and Potential," Nature Reviews Neurology, vol. 16, pp. 575–585, 2020.
12. W. Wang et al., "A Systematic Review of Machine Learning Models for Predicting Outcomes of Stroke with Structured Data," PLoS ONE, vol. 15, p. e0234722, 2020.
13. M. S. Sirsat, E. Fermé, and J. Câmara, "Machine Learning for Brain Stroke: A Review," Journal of Stroke and Cerebrovascular Diseases, vol. 29, p. 105162, 2020.
    A. K. Arslan, C. Colak, and M. E. Sarihan, "Prediction of Ischemic Stroke Using Different Medical Data Mining Approaches," Computer Methods and Programs in Biomedicine, vol. 130, pp. 87–92, 2016.

14. M. S. Islam et al., "Explainable Artificial Intelligence Model for Stroke Prediction Using EEG Signal," Sensors, vol. 22, no. 24, p. 9859, 2022.
15. E. Dritsas and M. Trigka, "Stroke Risk Prediction with Machine Learning Techniques," Sensors, vol. 22, no. 13, p. 4670, 2022.
16. Kokkotis et al., "An Explainable Machine Learning Pipeline for Stroke Prediction on Imbalanced Data," Diagnostics, vol. 12, no. 10, p. 2392, 2022.
17. Babburi, S. Lightweight Distributed Provenance Framework for Edge and IoT Data Systems.
18. Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.
19. 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
20. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
21. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
22. Mahimalur, R. K., Vasgam, M., & Manoharan, D. Devops Lifecycle Management And Cloud Migration Assessments: A Security-Driven CICD Perspective.
23. Viswanathan, V. Generative AI for Smarter Workforce Planning and Enterprise Resource Decisions.
24. Gajula, S. (2026, March). Two Pillars of Banking Intelligence: A Comparative Analysis of AI Techniques for Fraud Prevention and Churn Mitigation. In 2026 14th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-6). IEEE.
25. Maturi, S. Y. (2023). Crowdsourced frontier: Unveiling autonomous adversarial cybercapabilities via open AI competition. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 275–284.
26. Pavan Kumar Adabala. (2026). Best Practices for Enterprise System Integration in Modern Organizations. Journal of Information Systems Engineering and Management, 11(2s), 1137–1146. https://doi.org/10.52783/jisem.v11i2s.14558.
27. Srikanth Kavuri. (2025). AI-DRIVEN TEST AUTOMATION FRAMEWORKS: ENHANCING EFFICIENCY AND ACCURACY IN SOFTWARE QUALITY ASSURANCE. International Journal of Applied Mathematics, 38(10s), 699–710. https://doi.org/10.12732/ijam.v38i10s.990
28. Gummadi, V. P. K., Chilamkurthi, L. S., & Kavuri, S. (2026). Securing APIs in Government Clouds and Runtime Fabric Using FIPS-Enabled MuleSoft. 2026 International Conference on Artificial Intelligence, Systems, and Emerging Technologies (ICAISET), 1–6. https://doi.org/10.1109/icaiset66439.2026.11542099
29. Shashank, A. (2025). AI-Enhanced ETL Processes: Leveraging Artificial Intelligence for Optimized Data Integration Systems. Journal Of Multidisciplinary, 5(8), 219-225.
30. Kandula, S. T. R., Susarla, R. S., & Boyapati, P. K. (2025, July). Enhanced Cyber Security Using Global Local Artificial Neural Network Based Intrusion Detection in Big Data Environment. In 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC) (pp. 426-431). IEEE.
31. Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219