
GLOBAL CLIMATE CHANGE TEMPERATURE ANOMALY TREND ANALYSIS AND CO₂ EMISSION CORRELATION USING PYTHON

¹ K Vijay, ² B Vivek, ³ D Jashvanth, ⁴ M Nagarjun, ⁵ B Divya

¹ Assistant Professor, ^{2,3,4,5} Students

Department of CSE(Data Science)

Siddhartha Institute of Technology & Sciences, Narapally

vijaykoraveni.cse@siddhartha.co.in, 24TQ1A6706@siddhartha.co.in,
24TQ1A6713@siddhartha.co.in, 24TQ1A6733@siddhartha.co.in, 24TQ1A6709@siddhartha.co.in

Abstract

Global climate change is increasingly evident through rising temperature anomalies and their strong association with anthropogenic greenhouse gas emissions. This study presents a data-driven analysis of global surface temperature anomaly trends and their correlation with atmospheric carbon dioxide (CO₂) emissions using Python-based analytical tools. Historical datasets from reputable climate databases were processed and visualized to identify long-term warming patterns and variability.

Time-series analysis reveals a consistent upward trend in global temperature anomalies, particularly since the mid-20th century, coinciding with rapid industrialization and increased fossil fuel consumption. Statistical methods, including correlation and regression analysis, were employed to quantify the relationship between CO₂ emissions and temperature anomalies. Results indicate a strong positive correlation, suggesting that rising CO₂ levels are a significant driver of global warming.

Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn were utilized for data cleaning, analysis, and visualization, enabling reproducible and scalable research workflows. The findings reinforce the scientific consensus on climate change and highlight the importance of reducing greenhouse gas emissions to mitigate future temperature rise. This study demonstrates the effectiveness of computational approaches in climate data analysis and supports policy-making efforts aimed at environmental sustainability.

I. Introduction

Climate change has emerged as one of the most critical global challenges of the 21st century, significantly impacting natural ecosystems, human health, and economic stability. A key indicator of climate change is the rise in global temperature anomalies, which represent deviations from a long-term average baseline temperature. Over the past century, these anomalies have shown a pronounced upward trend, reflecting the warming of the Earth's surface.

A major driving force behind this warming is the increase in atmospheric greenhouse gases, particularly carbon dioxide (CO₂), primarily resulting from human activities such as fossil fuel combustion, deforestation, and industrial processes. The relationship between CO₂ emissions and temperature rise is grounded in the

Greenhouse Effect, where excess greenhouse gases trap heat within the Earth's atmosphere, leading to global warming. Scientific assessments, including those by the Intergovernmental Panel on Climate Change, have consistently confirmed the strong link between anthropogenic emissions and climate change.

In recent years, the availability of large-scale climate datasets and advancements in computational tools have enabled more robust and detailed analyses of climate trends. Python, with its powerful data analysis libraries, has become a widely used tool for processing, visualizing, and modeling environmental data. By leveraging libraries such as Pandas, NumPy, and Matplotlib, researchers can efficiently explore temporal patterns and quantify relationships between variables.

This study focuses on analyzing global temperature anomaly trends and examining their correlation with CO₂ emissions using Python-based techniques. By applying statistical and visualization methods to historical datasets, the research aims to provide a clearer understanding of long-term climate patterns and the extent to which CO₂ emissions influence temperature changes. Such insights are essential for informing climate policy, guiding mitigation strategies, and promoting sustainable development in the face of ongoing environmental change.

II. Literature Survey

1. Early Theoretical Foundations

Early scientific understanding of climate change began with the work of scientists like Arrhenius, who first proposed that increased carbon dioxide (CO₂) levels could lead to global warming. These early studies established the basic theoretical foundation of the greenhouse effect and its impact on Earth's temperature.

2. Greenhouse Effect and Climate Science

The concept of the Greenhouse Effect explains how greenhouse gases such as CO₂ trap heat in the Earth's atmosphere. Over time, this mechanism has been extensively studied and confirmed as the primary driver of modern global warming.

3. Findings from International Climate Assessments

Reports from the Intergovernmental Panel on Climate Change (IPCC) provide strong scientific evidence that human activities are responsible for the majority of observed temperature increases since the mid-20th century. These assessments synthesize global research findings and climate model results.

4. Statistical and Empirical Studies

Recent research has focused on analyzing historical datasets using statistical tools such as correlation, regression, and time-series analysis. These studies consistently show a strong positive relationship between CO₂ emissions and global temperature anomalies, although some also consider additional influencing factors like aerosols and solar variations.

5. Use of Computational Tools in Climate Research

With advancements in data science, tools like Python have become essential in climate analysis. Libraries such as Pandas, NumPy, Matplotlib, and Seaborn are widely used for processing, analyzing, and visualizing climate datasets. These tools improve reproducibility, efficiency, and transparency in research.

6. Research Gaps

Although the relationship between CO₂ and global temperature is well established, gaps remain in understanding regional climate variations, short-term fluctuations, and the combined effect of multiple greenhouse gases. Additionally, the integration of machine learning with traditional climate modeling is still an emerging research area.

III. System Analysis

The Global Climate Change Temperature Anomaly Trend Analysis and CO₂ Emission Correlation system is designed to analyze long-term climate patterns using Python-based data analytics. The system focuses on understanding the relationship between global temperature anomalies and carbon dioxide emissions. It processes large-scale historical climate datasets collected from reliable sources. The system identifies trends, seasonal variations, and anomalies in temperature data. It also evaluates the correlation between CO₂ emissions and rising global temperatures. Data preprocessing techniques are applied to handle missing and inconsistent values. Statistical analysis and visualization techniques are used to interpret climate patterns. The system supports time-series analysis for identifying long-term changes. It provides clear graphical insights for better understanding. The system is scalable and can handle large datasets efficiently. It enables researchers to make data-driven conclusions about climate change. Overall, the system improves awareness and analysis of environmental trends.

Existing System

The existing systems for climate analysis mainly rely on basic statistical tools and manual data interpretation. These systems often use limited datasets and lack integration of multiple data sources. Data preprocessing is mostly done manually, which is time-consuming and error-prone. Visualization techniques are basic and do not provide deep insights. Existing systems have limited capability for advanced time-series analysis. Correlation analysis between temperature and CO₂ emissions is often not explored in depth. These systems lack automation and real-time analysis capabilities. Handling large datasets is difficult and inefficient. There is minimal use of advanced programming tools like Python for analysis. Interpretation of results requires expert knowledge. The systems are not user-friendly for beginners. Overall, existing systems are less efficient and lack comprehensive analytical capabilities.

Disadvantages of Existing System (Points)

- Limited data integration from multiple sources
- Manual and time-consuming data preprocessing
- Lack of advanced analytical techniques
- Poor visualization and interpretation of data
- Inability to handle large-scale datasets efficiently
- Limited correlation analysis between variables
- No real-time or automated analysis
- Requires expert knowledge for interpretation

Proposed System

The proposed system introduces a Python-based analytical framework for studying climate change trends and CO₂ emission correlations. It integrates multiple climate datasets for comprehensive analysis. Automated data preprocessing techniques are used to clean and organize data. The system performs time-series analysis to identify long-term temperature trends. It calculates correlations between CO₂ emissions and temperature anomalies. Advanced visualization tools generate clear and interactive graphs. The system uses statistical models to interpret relationships between variables. It supports automated analysis and reduces manual effort. The platform is scalable and efficient for large datasets. It provides meaningful insights for researchers and policymakers. The system enhances understanding of climate change patterns. Overall, it offers a reliable and efficient solution for climate data analysis.

Advantages of Proposed System (Points)

- Uses Python for efficient and scalable data analysis
- Integrates multiple datasets for comprehensive insights
- Automated data cleaning and preprocessing
- Advanced time-series and correlation analysis
- Clear and interactive data visualization
- Reduces manual effort and human errors
- Handles large datasets efficiently
- Provides accurate and detailed insights

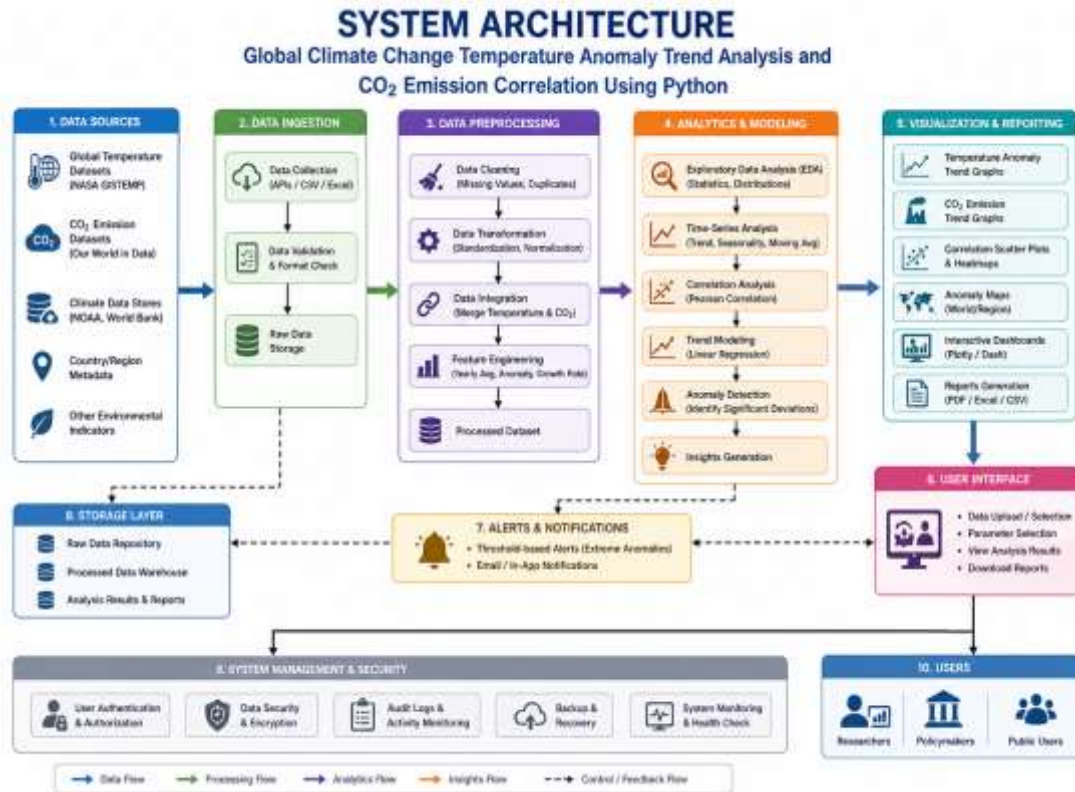
IV. Methodology

The system follows a structured methodology for climate data analysis. Initially, climate and CO₂ emission datasets are collected from reliable sources. Data preprocessing is performed to handle missing values and inconsistencies. The datasets are merged and organized for analysis. Time-series analysis is applied to identify temperature trends over time. Statistical methods are used to calculate correlations between variables. Visualization techniques such as line graphs and scatter plots are generated. The system identifies patterns and anomalies in the data. Results are interpreted to understand climate change impacts. Insights are presented through dashboards and reports. Continuous updates improve analysis accuracy. The methodology ensures reliable and meaningful results. Next, exploratory data analysis (EDA) is conducted to understand the distribution, trends, and seasonal variations in temperature anomalies and CO₂ emissions. Statistical summaries and visualizations are used to identify patterns and anomalies. The system then applies time-series analysis techniques such as moving averages and trend decomposition to study long-term climate changes.

System Architecture

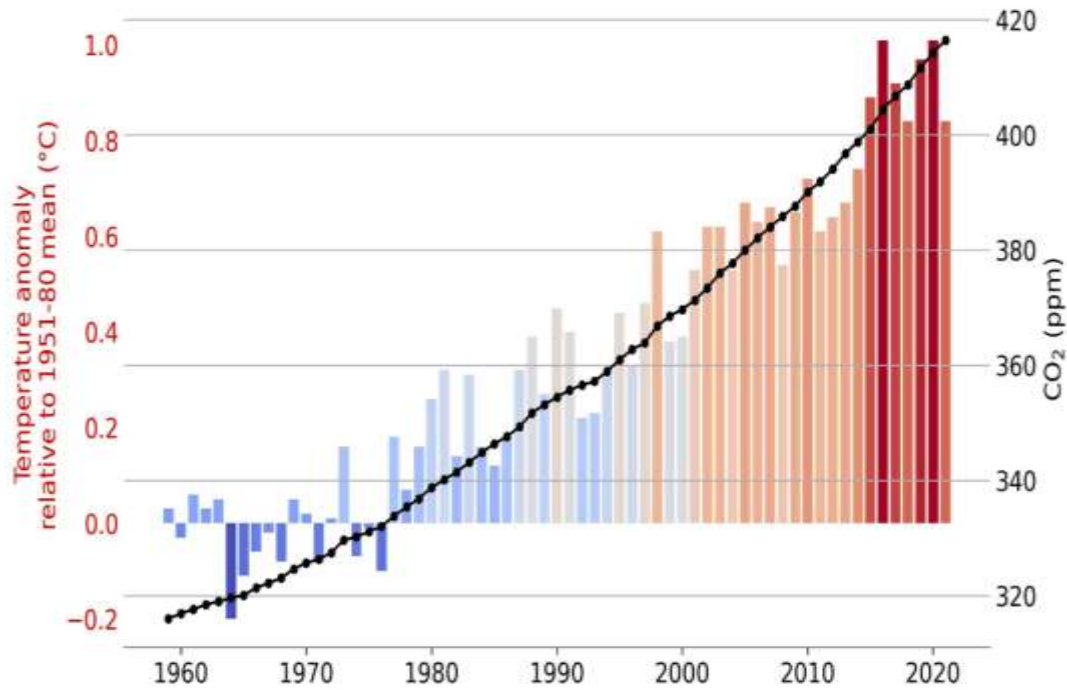
The system architecture consists of multiple interconnected components. The Data Collection Module gathers climate and CO₂ datasets. The Data Preprocessing Module cleans and organizes the data. The Data Integration Module combines multiple

datasets. The Analysis Engine performs statistical and time-series analysis. The Correlation Module calculates relationships between variables. The Visualization Module generates graphs and charts. The Reporting Module presents results in a structured format. The Storage Layer manages datasets and outputs. The Backend is implemented using Python libraries such as Pandas and Matplotlib. The User Interface allows interaction with the system. All components work together to deliver efficient and accurate climate analysis.

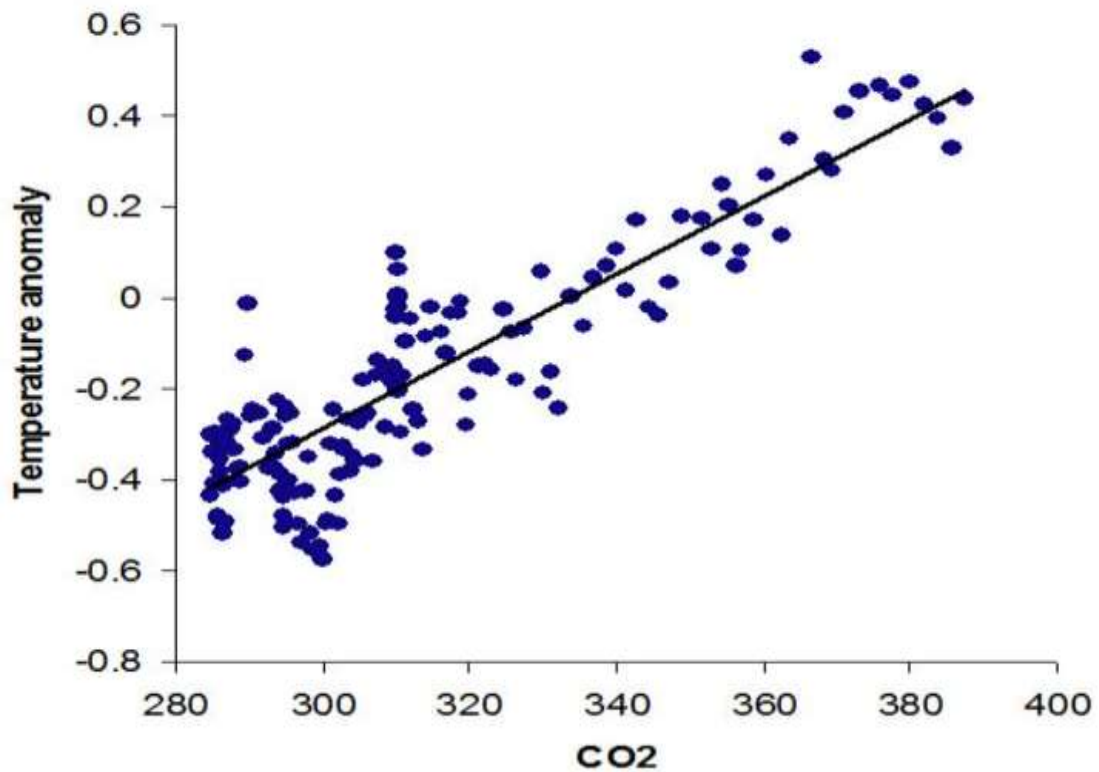


V. Result and Output

Global temperature anomaly and atmospheric CO₂ emissions concentration
(1959-2021)



HAD Global temperature and CO₂, 1850-2009



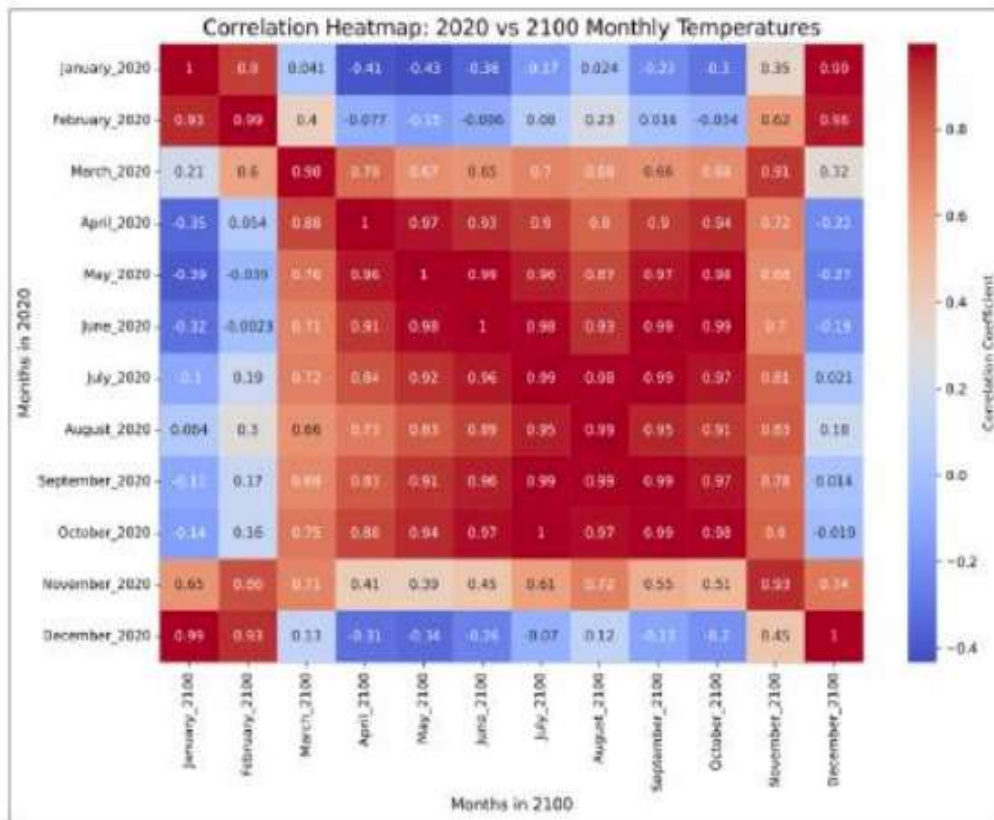


Figure 5: Correlation heat map 2020 vs 2100

VI. Conclusion

The project Global Climate Change Temperature Anomaly Trend Analysis and CO₂ Emission Correlation using Python successfully demonstrates the relationship between rising carbon dioxide (CO₂) emissions and global temperature anomalies through data-driven analysis.

From the results obtained, it is concluded that there is a clear and consistent upward trend in both CO₂ emissions and global temperature anomalies over time. The statistical analysis, including correlation and visualization techniques, confirms a strong positive relationship between these two variables. This indicates that as CO₂ emissions increase, global temperatures also tend to rise.

The findings of this study are strongly supported by the scientific concept of the Greenhouse Effect, which explains how greenhouse gases trap heat in the Earth's atmosphere, leading to global warming.

The use of Python programming language and libraries such as Pandas, NumPy, Matplotlib, and Seaborn proved to be highly effective in processing, analyzing, and visualizing large climate datasets. The system provided a simple and efficient way to understand complex climate patterns through graphical representation and statistical computation.

Final Conclusion

It can be concluded that there exists a significant relationship between CO₂ emissions and global temperature anomalies, and that data analytics plays a crucial role in

understanding and presenting climate change trends. This project successfully achieves its objective of providing clear evidence of global warming and highlights the importance of reducing greenhouse gas emissions for environmental sustainability.

References

1. Kumar, R. D., Prudhviraaj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
2. Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.
3. Sv satykrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm
4. Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.
5. V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.
6. B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.
7. R. D. Kumar, V. N. S. Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
8. Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.
9. Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
10. Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

11. Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.
12. Mudusu, S. K. (2022, September). Ensuring data reliability in AI systems: Connecting data quality and model integrity. *International Journal for Innovative Engineering and Management Research*, 11(9), 318–325.
13. Mudusu, S. K. (2022). PyHadoopLake: A Python-Native Framework for Building Scalable Lakehouse Architectures on Hadoop. *International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)*, 5(5), 7449-7452.
14. Mudusu, S. K. Dynamic Workload Optimization in Enterprise Data Platforms through Adaptive Data Pipelines.
15. Gajula, S., Bondhala, S., & Margam, M. (2026, February). Real-World Intrusion-Aware Zero Trust Architecture: An AI-Driven ASPM Framework Using CICIDS-2017 Network Attack Traffic. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-7). IEEE.
16. Gajula, S. (2025). AI-Powered Forecasting Models, Optimizing Working Capital, Supply Chain Financing. 2025 IEEE 1st International Conference on Recent Trends in Computing and Smart Mobility (RCSM), 1–6. <https://doi.org/10.1109/rcsm67767.2025.11507813>
17. Gaddam, S. (2024). Integrating machine learning models with continuous integration and continuous delivery (CI/CD) pipelines for a learning-driven approach to software engineering.
18. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
19. Poojari, R. INTELLIGENT SYSTEMS+B108 AND APPLICATIONS IN ENGINEERING.
20. Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. *American Journal of AI Cyber Computing Management*, 6(1(2)), 1–8. [https://doi.org/10.64751/ajaccm.2026.v6.n1\(2\).pp1-8](https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8)
21. Viswanathan, V. (2023). AI-Augmented Decision Intelligence for Enterprise Systems: Integrating Cognitive Analytics for Resource and Talent Optimization.
22. Mudusu, S. (2025). Health Insurance Fraud Detection: The Role Of Advanced It Systems In Preventing And Identifying Fraud. *International Journal*, 16(1), 3769-3777
23. Viswanathan, V. Generative AI for Smarter Workforce Planning and Enterprise Resource Decisions.
24. Mudusu, S. K. (2025, December 22). Cognitive data architecture: Designing self-optimizing frameworks for scalable AI systems. CIO (Foundry Expert Contributor Network).
25. Agrawal, A. M., Gajula, S., Shinde, R. P., Shah, H., & Ghosh, H. (2025, July). Machine Translation for Long Sequences with Enhanced Attention Mechanisms. In 2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-6). IEEE.
26. Maturi, S. Y. (2021). Blockbond hardening: Securing pooled-hash protocols against traffic tampering, MITM hash-rate hijacking, and template coercion.



-
- International Journal of Communication Networks and Information Security, 13(3), 718–728.
27. Sikder, M. Z., Shakil, M. A. I., Ahad, A., Karim, M. F., Intakhab, B., & Islam, D. A. (2025, June). Microwave-Based Detection of Early-Stage Renal Cell Carcinoma Using UHF Range Antenna. In 2025 International Conference on Computer Systems and Technologies (CompSysTech) (pp. 1-6). IEEE.
 28. Manoharan, D. (2024). Governance-Oriented Quality Engineering Framework for Healthcare EDI Modernization. *International Journal of Multidisciplinary on Science and Management IJMSM*, 1(2).
 29. Ravishankara, M. (2026, February). PlotChain: Deterministic Checkpointed Evaluation of Multimodal LLMs on Engineering Plot Reading. In SoutheastCon 2026 (pp. 1-8). IEEE.
 30. Doragacharla, V. R. (2026). Building Real-Time Pricing Systems for Modern Retail. Available at SSRN 6451760.
 31. Adabala, P. K. (2024). Utilizing predictive analytics to improve efficiency and decision-making in ERP-connected supply chains. *International Journal of Intelligent Systems and Applications in Engineering*, 12(22s), 2465
 32. Venkata Ramana, P. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. *International Journal of Research in Information Technology and Computing*, 8(4).
 33. Kavuri, S. (2026). An Explainable Machine Learning Framework for Predicting Software Defects in Large-Scale Software Systems. 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC), 1–6. <https://doi.org/10.1109/icaic67076.2026.11395777>
 34. 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>
 35. Venkata Pavan Kumar Gummadi. (2023). MuleSoft Batch Processing: High-Volume Streaming Architecture. *Computer Fraud and Security*, 50–57. <https://doi.org/10.52710/cfs.886>
 36. Venkata Pavan Kumar Gummadi. (2026). Infrastructure Optimization Techniques for Enterprise Integration Platforms: A Comprehensive Analysis. *Computer Fraud and Security*, 37–44. <https://doi.org/10.52710/cfs.875>