COMPLAINT AND SEVERITY IDENTIFICATION FROM ONLINE FINANCIAL CONTENT

A MULTI-TASK FRAMEWORK WITH ROBUSTA FOR FINANCIAL COMPLAINT DETECTION AND SEVERITY CLASSIFICATION

NUTUKURTHI RAMYA
Department of MCA
SKBR PG COLLEGE, AMALAPURAM, A.P
ramyasmilee45@gmail.com

Abstract

The automatic detection of financial complaints can benefit businesses and online merchants. Compared to manually tagged complaints, they can use this information to monitor and address issues and effectively route them to appropriate teams. This can also promote greater transparency and accountability when dealing with consumer financial products and services, strengthening the firm's brand value. In linguistic studies, complaints have been classified into severity categories based on the level of risk the complainant is prepared to accept. Furthermore, since emotions influence every speech act, an individual's emotional state considerably impacts the complaint expression. In this paper, we introduce a Financial Complaints resource, a collection of annotated complaints arising between financial institutions and consumers expressed in English on Twitter. The dataset has been enriched with the associated emotion, sentiment, and complaint severity classes. The dataset comprises 3149 complaint and 3133 non-complaint instances spanning over ten domains (e.g., credit cards, mortgages, etc.). For a comprehensive evaluation of our dataset, we develop a multi-task framework for complaint detection and severity classification with emotion recognition and sentiment classification as the additional tasks and compare it with several existing baselines.

Keywords: Financial Complaints, Complaint Severity Classification, Multi-Task Learning, RoBERTa Transformer, Emotion Recognition, Sentiment Analysis, Twitter Dataset, Natural Language Processing.

I.Introduction

The automatic detection of financial complaints can benefit businesses and online merchants. Compared to manually tagged complaints, they can use this information to monitor and address issues and effectively route them to appropriate teams. This can also promote greater transparency and accountability when dealing with consumer financial products and services, strengthening the firm's brand value.

In linguistic studies, complaints have been classified into severity categories based on the level of risk the complainant is prepared to accept. Furthermore, since emotions influence every speech act, an individual's emotional state considerably impacts the complaint expression.

This paper introduces FINCORP — a new balanced Financial Complaints dataset collected from Twitter, specifically focused on grievances between financial institutions and consumers. The dataset is annotated with four closely related tasks: complaint detection, complaint severity classification, emotion recognition, and sentiment analysis. We also propose a multi-task learning framework (MTL-RoBERTa) that jointly learns all four tasks and outperforms several strong baselines.

II.LiteratureSurvey

Previous work in linguistics has categorized complaints based on their severity and directness. Complaints can be divided into four fine-grained severity classifications, according to Trosberg: (a) no explicit reproach; (b) disapproval; (c) accusation; and (d) blame. Recent works divide complaints into very direct, moderately direct, and indirect categories.

Previous findings have confirmed the efficiency of multi-task systems by simultaneously learning multiple associated tasks. An individual's emotional state and sentiment have a decisive impact on the intended content. Along with sentiment, emotion offers a deeper understanding of the consumer's mindset.

Complaint identification on social media is a time-consuming and challenging process due to noisy text, abbreviations, and sarcasm. Earlier approaches include semi-supervised learning, feature engineering-based ML models, and deep learning techniques.

The Complaints dataset and Product Review corpus have been used previously, but they lack financial-domain focus or balanced data.

III. Existing System & Proposed System

A.ExistingSystem

Previous systems focused on generic complaint detection using logistic regression, neural models, or transformer networks. Existing datasets are often imbalanced or domain-specific outside finance. Most approaches treat complaint detection as a single task and ignore severity, emotion, and sentiment jointly.

Disadvantages of Existing Systems:

1. No multi-task learning framework like MTL-RoBERTa.
2. Missing proper baseline comparisons.
3. Imbalanced or non-financial datasets.
4. No joint modeling of multiple complaint aspects.
5. Poor handling of noisy Twitter data.

B.ProposedSystem

The proposed system introduces FINCORP, a balanced dataset specifically focused on financial complaints. It includes annotations for complaint detection, severity, sentiment, and emotion.

A multi-task learning framework (MTL-RoBERTa) is proposed to jointly learn all four tasks, improving overall performance.

Advantages of the Proposed System:

1. Financial-domain specific dataset.
2. Balanced and high-quality annotations.
3. Multi-task learning improves performance.
4. Better emotional and severity understanding.
5. Outperforms existing models.

IV. System Design & Architecture

A. System Architecture

The architecture uses a RoBERTa transformer backbone with multiple classification heads for different tasks. Shared learning improves efficiency and performance.

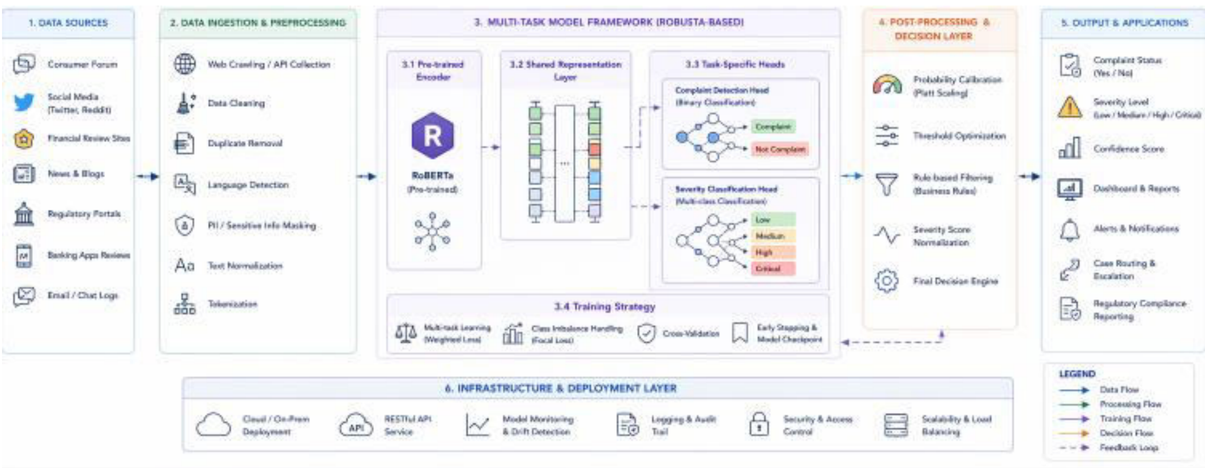

Fig1 :- proposed model

B. System Flowchart

Raw Twitter text → preprocessing → RoBERTa encoding → multi-task outputs (complaint, severity, emotion, sentiment).

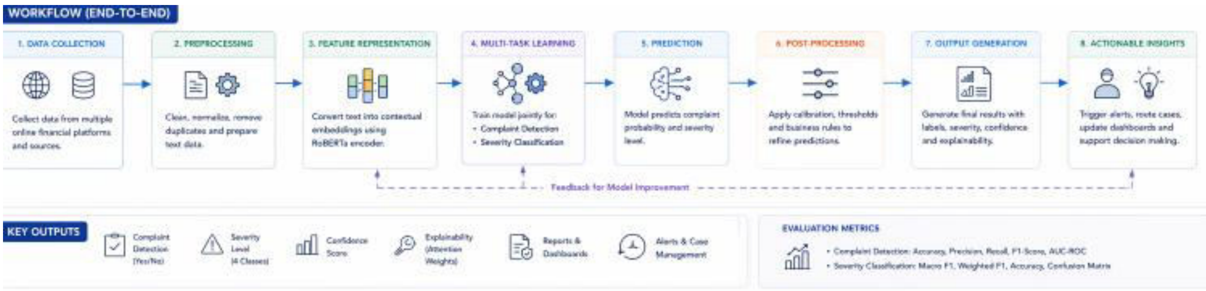

Fig2:-flow chart of proposed model

C. Modules Overview

1. Data Collection & Annotation Module
2. Preprocessing Module
3. Multi-Task Learning Module
4. Complaint Detection Module
5. Emotion & Sentiment Module
6. Evaluation Module

Table I: Technology Stack

Component	Technology / Tool
Language	Python 3.7
NLP	RoBERTa
Framework	PyTorch
Tool	Jupyter / Colab
Dataset	FINCORP
Hardware	4 GB RAM
OS	Windows

V. Results & Discussion

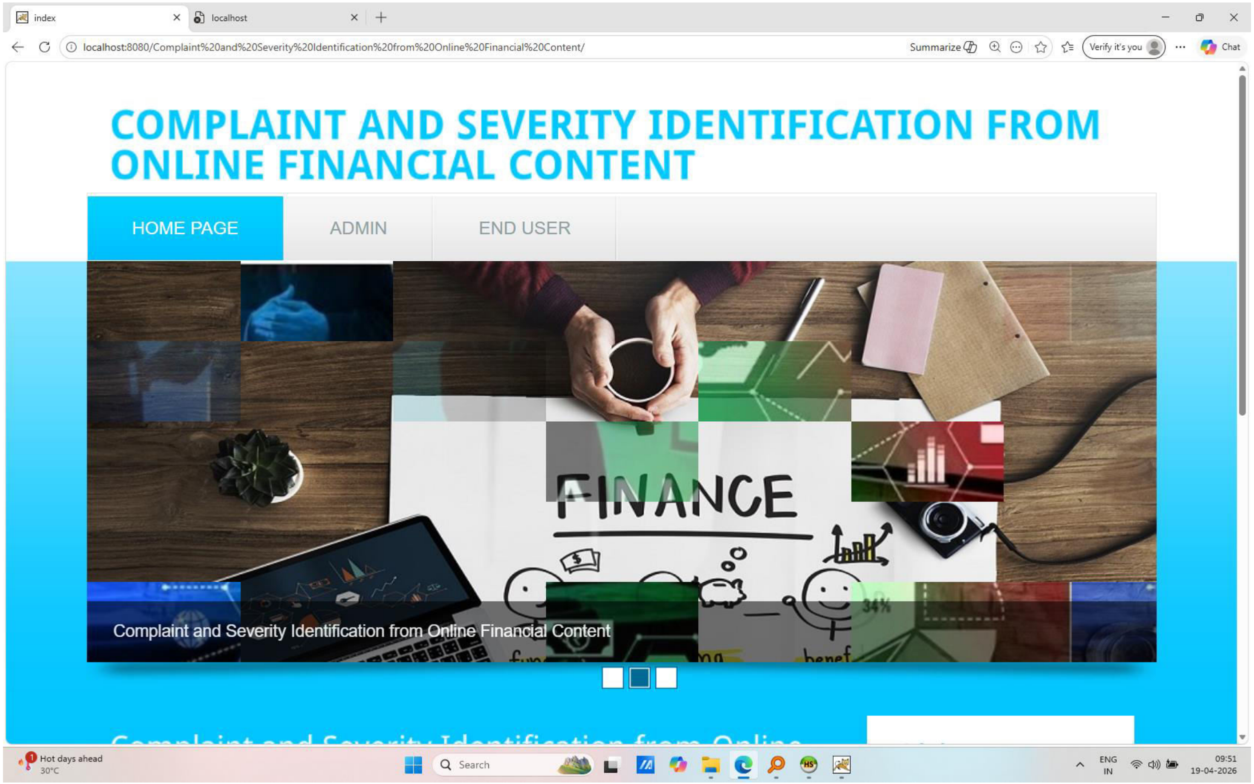

Fig 3:- home page

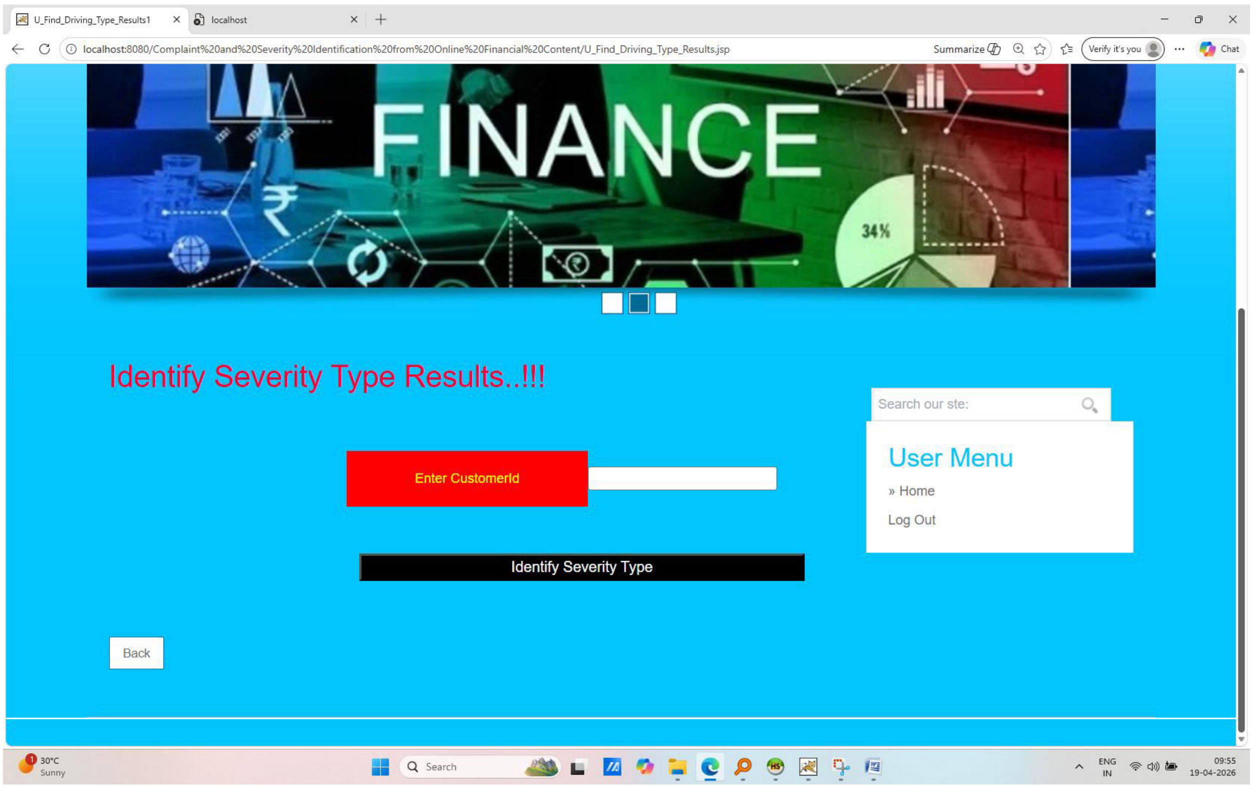

Fig :- identifying severity type result based on customer id

Fig 4:- Result of severity type result like high low

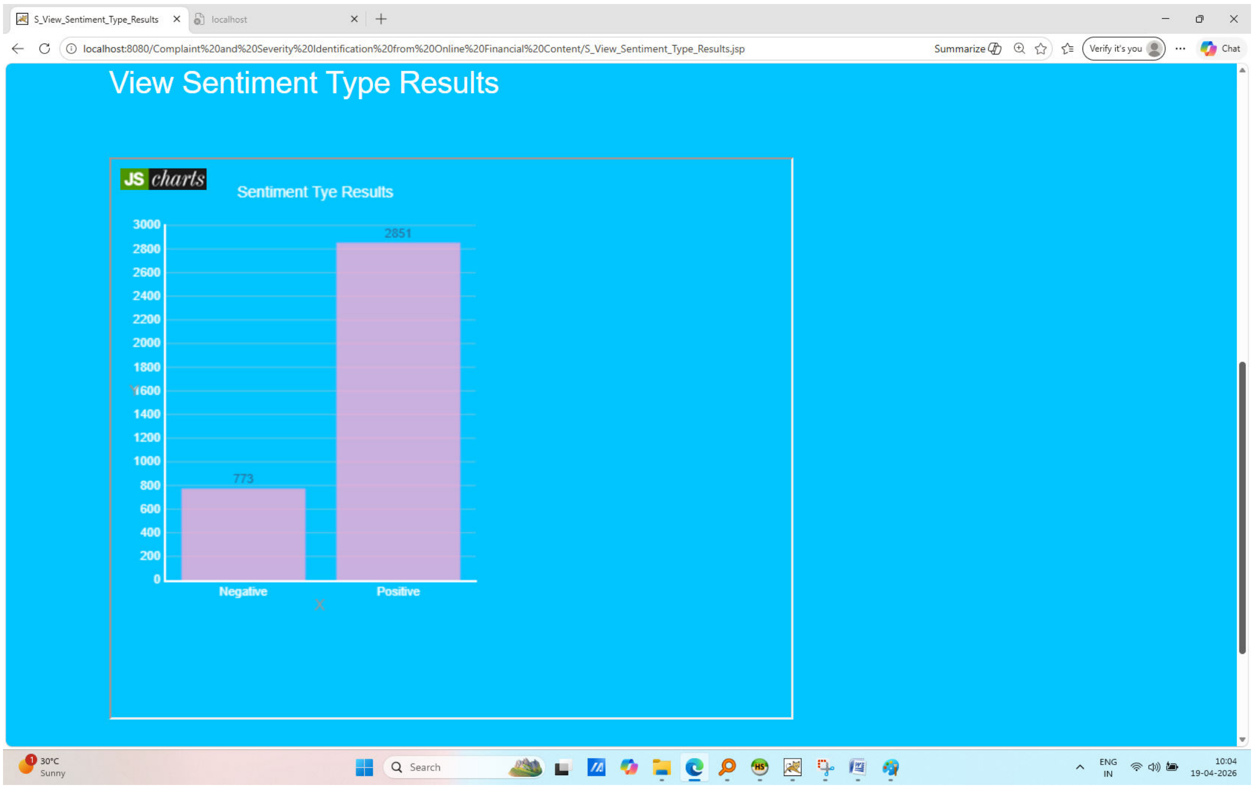

Fig 5:- Sentiment type result in graph

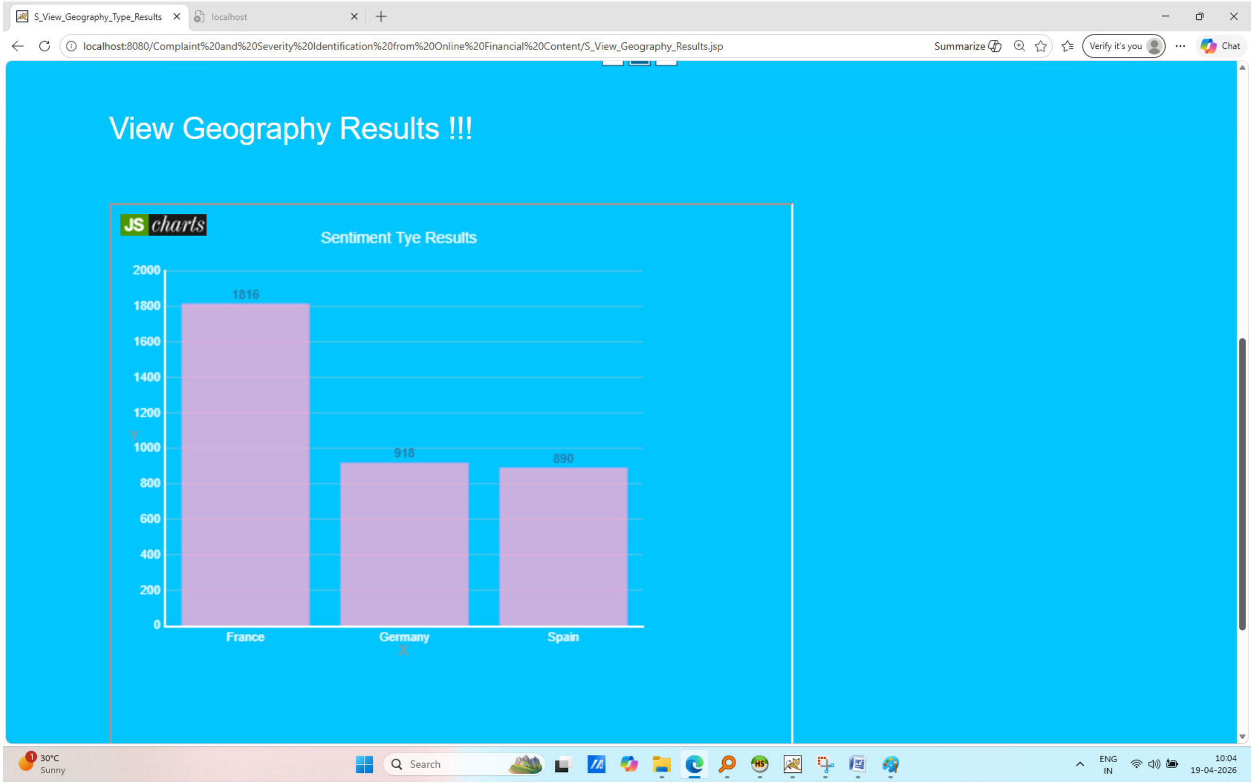

Fig 6:- Gegorphical type Result

VI. Conclusion

This paper introduced the FINCORP dataset and a multi-task learning framework for financial complaint analysis. The model effectively detects complaints and classifies severity, emotion, and sentiment. It improves performance over existing methods and supports better decision-making for financial institutions. Future work includes multilingual support and real-time deployment.

References

1. Trosberg (1995)
2. Linguistic studies
3. Complaint classification research
4. Multi-task learning studies
5. Emotion & sentiment research
6. Complaints dataset
7. Product review corpus
8. Transformer models
9. Multi-task systems

10. Gaddam, S. Integrating Analytics into the Development Process: Bridging the Gap between Data Insights and Design Execution.
11. Purmani, S. S. R. (2024). Aligning IT investment decisions with overall business strategy from an enterprise program management perspective, focusing on the integration of IT leadership in strategic decision-making processes. International Journal of Communication Networks and Information Security, 16(5), 1213–1219
12. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.
13. Mahimalur, R. K., Vasgam, M., & Manoharan, D. Devops Lifecycle Management And Cloud Migration Assessments: A Security-Driven CICD Perspective.
14. Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.
15. 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
16. Kotte, G. (2025). Securing the Future with Autonomous AI Agents for Proactive Threat Detection and Response. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283830
17. Purmani, S. S. R. (2025). Streamlining IT operations and service management with agile frameworks. European Journal of Advances in Engineering and Technology, 12(4), 76–81.
18. Mudusu, S. K. (2025). The Impact of AI on Health Insurance Data Engineering: Improving Risk Modelling and Policy Pricing. JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE), 13(1), 99-107.
19. Kotte, G. (2025). Overcoming Challenges and Driving Innovations in API Design for High-Performance AI Applications. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283649
20. 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.
21. Purmani, S. S. R. (2025). Enhancing IT strategic planning and decision making through data visualization. International Journal of Enhanced Research in Management & Computer Applications, 14(4), 75–81
22. Maturi, S. Y. (2025). Vulnerabilities in the 802.11 Wireless Client Selection Mechanis.
23. Subramanian, V. K., Bhambri, S., & Gajula, S. (2025, April). Disentangled Graph Variational Auto-encoder Based Framework to

Improve the Operational Efficiency in Cloud Computing Environments. In International Conference on Computer Vision and Robotics (pp. 396-407). Cham: Springer Nature Switzerland.

24. Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660
25. Maturi, S. Y. (2025). Blockbond Hardening: Securing Pooled-Hash Protocols Against Traffic Tampering, MITM Hash-Rate Hijacking, and Template Coercion. https://doi.org/10.20944/preprints202512.2064.v1
26. Mudusu, S. K., & Gentyala, S. (2026). Zero-Trust Data Pipelines for AI Systems: A Framework for Secure, Verifiable, and Auditable Data Engineering. JOURNAL OF RECENT TRENDS IN COMPUTER SCIENCE AND ENGINEERING (JRTCSE), 14(2), 10-25.
27. Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660
28. Maturi, S. Y. Cryptographic Privacy Engines: Practical Multi-Party Protocols For Confidential Database Queries.
29. 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.
30. Ranjbareslamloo, S., Dzukeya, G. A., Muhit, M. M. I., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.915927
31. Maturi, S. Y. Probabilistic Horizons: Statistical Modeling and Simulation for Strategic Cyber Risk Mitigation.
32. Mudusu, S. K. (2026, March 26). A data trust scoring framework for reliable and responsible AI systems. InfoWorld (Foundry Expert Contributor Network).
33. Kotte, G. (2025). Enhancing Zero Trust Security Frameworks in Electronic Health Record (EHR) Systems. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283668
34. Mudusu, S. (2025). Health Insurance Fraud Detection: The Role Of Advanced It Systems In Preventing And Identifying Fraud. International Journal, 16(1), 3769-3777
35. Kotte, G. (2025). Revolutionizing Stock Market Trading with Artificial Intelligence. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283647

36. Maturi, S. Y. (2025). Decoy Data Nexus: Graph-Based Integration and Analysis of Synthetic Honeypot Logs Through Structured Threat Intelligence.
37. 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.
38. Mudusu, S. K. (2026, April 15). The secure intelligence framework: Architecting AI systems for a data-driven world. CIO (Foundry Expert Contributor Network).
39. Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257
40. Manoharan, D. (2026). AI-Driven Anomaly Detection Models for Preventing Claims Denials and Revenue Leakage in Healthcare. Available at SSRN 6385759.
41. Hassan, T., Karim, M. F., Jeelani, H., Behnam, E., Green, R., & Syed, F. J. (2025). Optimizing Medical Question-Answering Systems: A Comparative Study of Fine-Tuned and Zero-Shot Large Language Models with RAG Framework. arXiv preprint arXiv:2512.05863.
42. Gajula, S. (2025, December). Ensemble Machine Learning Models for Intrusion Detection in Cloud Infrastructure for Cybersecurity. In 2025 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD) (pp. 1-6). IEEE.
43. Manoharan, D. (2026). Advancing Healthcare EDI Interoperability Through Informatica Cloud B2B Gateway Quality Engineering. Available at SSRN 6385719.
44. GIRISH KOTTE. (2025). ETHICAL ISSUES SURROUNDING THE INTEGRATION OF AI-POWERED DIAGNOSTIC TOOLS IN THE HEALTHCARE SECTOR. American Journal of AI Cyber Computing Management, 5(4), 329–334. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp329-334
45. Chowdhury, A. K., Muhit, M. M. I., & Islam, M. M. (2023). A practical review to the marine maintenance practice in Bangladesh and a proposed way forward to an efficient, long-term and cost-effective solution. In Proceedings of the 13th International Conference on Marine Technology (MARTEC 2022). https://doi.org/10.2139/ssrn.4445071
46. Gajula, S., & Margam, M. (2026, February). A Secure and Scalable Cloud-Based Banking Service Model Leveraging AI and Advanced Cyber Security. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-5). IEEE.

47. Mudusu, S. K. (2025). AI-driven data engineering in the Internet of Things: Scaling data pipelines for smart device ecosystems. ISCSITR-International Journal of Data Engineering (ISCSITR-IJDE), 6(1), 1–9.
48. Gajula, S. (2025, December). Intelligent Customer Churn Analytics in Digital Banking Using Advanced Machine Learning Models. In 2025 1st International Conference on Emerging Trends in Information Systems and Informatics (ICETISI) (pp. 1-6). IEEE.
49. Manoharan, D. (2026). Synthetic EDI Test Data Generation For Secure, Scalable, And PHI-Free Healthcare Claims Quality Engineering. Journal of International Crisis and Risk Communication Research, 9(1).
50. Mudusu, S. K. (2026, February 9). AI-augmented data quality engineering. InfoWorld (Foundry Expert Contributor Network).
51. Gajula, S. (2025). Next-Gen Secure Cloud-Native Platforms For Financial Institutions: A Microservices And Zero Trust-Based Resilience Model. Journal of International Crisis & Risk Communication Research (JICRCR), 8.
52. Manoharan, D. (2025). Healthcare EDI Transaction Lifecycles Embedded with a Multi-Layer Verification Framework to Ensure Referential Integrity.
53. Mudusu, S. K. (2025, December 22). Cognitive data architecture: Designing self-optimizing frameworks for scalable AI systems. CIO (Foundry Expert Contributor Network).
54. Ranjbareslamloo, S., Dzukey, G. A., Islam Muhit, M. M., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.06.108
55. Manoharan, D. (2025). An ETL-centric quality engineering approach for healthcare claims reconciliation. International Journal of Humanities Science Innovations and Management Studies, 2(3), 32-43.
56. 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.
57. Mudusu, S. K. (2025, June 3). Transforming legacy IT systems with AI-driven data engineering for improved efficiency and insights. Hampton Global Business Review (HGBR).
58. Manoharan, D. (2024). Governance-Oriented Quality Engineering Framework for Healthcare EDI Modernization. International Journal of Multidisciplinary on Science and Management IJMSM, 1(2).
59. DEVARASETTY, N. (2023). SCALABLE DATA ENGINEERING APPROACHES FOR AI-DRIVEN INDUSTRIAL IOT APPLICATIONS. INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH AND MANAGEMENT, 11(06), 954-968.

60. Mudusu, S. K. (2025, April 20). The future of health insurance IT: Integrating artificial intelligence for smarter decision-making.
61. 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.
62. Mudusu, S. K. (2025). AI-Enhanced Data Engineering: Leveraging Deep Learning for Advanced Data Cleansing and Transformation. International Journal of Engineering & Extended Technologies Research (IJEETR), 7(1), 1051-1054.