



# **Integrated News Summarization, Classification and Recommendation System**

**Dr P Ashok Kumar<sup>1</sup>, S.Vikas<sup>2</sup>, P.Dhanush<sup>3</sup>, G.Varun<sup>4</sup>, B.Bhargav Sai<sup>5</sup>**

Associate Professor, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India<sup>1</sup>

IV B. Tech, Department of CSE (Data Science), ACE Engineering College, Hyderabad, India<sup>2,3,4,5</sup>

**ABSTRACT:** The “Integrated News Summarization, Classification and Recommendation System” is an intelligent platform designed to mitigate the challenges of information overload and the absence of personalization commonly found in conventional news delivery systems. The proposed system integrates external news APIs with advanced Natural Language Processing (NLP) and Machine Learning (ML) techniques to perform topic classification, personalized recommendation generation, and automatic article summarization, thereby enhancing the efficiency and quality of news consumption. The front-end application, developed using React, ensures cross-platform compatibility and an interactive user interface, while Python-based backend modules manage data processing, model execution, and analytical operations. Incorporating secure user authentication and storage mechanisms, the system ensures data integrity and privacy. Overall, it delivers an informed, time-efficient, and personalized news reading experience, significantly improving user engagement and content relevance.

**KEYWORDS:** News Summarization, NLP, Machine Learning, TF-IDF, Cosine Similarity, T5-SMALL, Recommendation System, React, Flask, MongoDB.

## **I. INTRODUCTION**

Traditional news applications primarily focus on aggregating content from various online sources, offering users large volumes of information but failing to cater to contextual and personalized needs [1]. Most existing platforms rely on basic topic-based filtering such as sports, technology, or business, without providing deeper semantic understanding of the content. As a result, users are often overwhelmed by excessive information and forced to manually read lengthy articles to extract key insights. The primary objective of the Integrated News Summarization and Recommendation System is to focus the challenges by leveraging advanced Natural Language Processing (NLP) and Machine Learning (ML) techniques [4][7]. The system automatically generates concise summaries of long news articles, organizes content into meaningful topic-based categories, and provides personalized news recommendations based on user preferences. By combining summarization and personalization, the solution aims to deliver a context-aware, time-efficient, and intelligent news consumption experience [2].

## **II. EXISTING SYSTEM**

Current news apps (e.g., major aggregators) mainly function as large-scale content aggregation platforms that collect real-time articles from multiple publishers using APIs or feeds. They help users access diverse news in one place and organize it into broad categories like politics, sports, and technology. They also provide trending sections based on engagement metrics and allow limited personalization through selected topics, language, or region preferences. Some platforms enhance engagement using multimedia content such as images and short videos. However, these systems rely mostly on publisher metadata and basic filtering rather than true contextual analysis. They lack automated summarization, deep user behaviour analysis, and advanced personalized recommendations. As a result, users often face information overload and must read lengthy or irrelevant articles.

### **A. Drawbacks of Existing System**

1. No automatic summarization.
2. Broad Classification.
3. Weak Personalization.
4. Data is scattered and not integrated from multiple sources.

### III. PROPOSED MODEL

The proposed model is an AI-driven news application designed to enhance the digital news reading experience by reducing information overload and delivering concise, personalized content [21]. It fetches real-time news from reliable APIs and processes articles using Natural Language Processing (NLP) and Machine Learning (ML) techniques to present structured and relevant information to users. News articles are organized into topic-based categories using metadata from the external API, ensuring accurate classification without additional models. The system also uses the T5-Small transformer model to generate short summaries of lengthy articles, helping users quickly understand key information [3]. Additionally, personalized recommendations are provided using content-based filtering techniques like TF-IDF and cosine similarity based on user preferences and reading behaviour. The application is built with a React-based frontend and a scalable backend to handle API communication, NLP processing, and recommendation functionality efficiently.

#### A. Advantages of the Proposed System

- 1. Reduces Information Overload:** The system provides concise summaries of long news articles using transformer-based models, helping users quickly understand important information without reading full articles.
- 2. Personalized News Recommendations:** By analysing user preferences and reading behaviour using techniques like TF-IDF and cosine similarity, the system suggests news articles that match the user's interests.
- 3. Time-Efficient News Consumption:** Summarized content allows users to stay updated with current events in a shorter time compared to traditional news reading.
- 4. Real-time News Updates:** The application fetches the latest news from reliable APIs, ensuring that users always receive up-to-date information [19].
- 5. Improved User Experience:** A modern React-based frontend with a scalable backend ensures smooth navigation, fast loading, and efficient processing of news data.

### IV. METHODOLOGY

#### A. Data Gathering and Cleaning Module:

The Data Gathering and Cleaning Module serves as the foundation of the system. It fetches and prepares real-time news data for machine learning tasks such as summarization, and analysis, and classification [20].

#### B. News Summarization Module (T5-SMALL):

T5-SMALL generates short, meaningful summaries to reduce reading time while preserving key information. Uses deep attention mechanisms in T5-SMALL to capture the relationships between words for improved semantic understanding [6].

#### C. News Classification Module:

The News Classification Module organizes articles into relevant categories using automatic topic labelling provided by the external news API. Instead of training a supervised machine learning classifier, the system relies on source-level metadata that assigns each article to predefined topics such as Politics, Sports, Technology, Business, and Health. This approach ensures reliable categorization while maintaining real-time performance and low computational overhead [12].

#### D. Personalized Recommendation Engine (TF-IDF + Cosine Similarity):

Computes content similarity between articles using TF-IDF vectors and Cosine Similarity. Tracks user reading patterns and preferences to personalize recommendations. Continuously updates suggestions based on the latest user interactions and trending topics [9].

**E. Application Integration Module (Flask + React PWA):**

This module integrates all machine learning and backend functionalities with the React frontend to deliver a seamless mobile user experience.

**V. SYSTEM ARCHITECTURE**

The News System follows a modular client-server architecture that integrates a React PWA frontend, Flask backend, machine learning components, Mongo-DB database, and external news APIs to deliver real-time summarized and personalized news. The frontend handles user interaction such as login, article viewing, and displaying summaries and recommendations, while the Flask backend manages API requests, fetches news data, processes text, runs the T5-Small summarization model, and generates recommendations using TF-IDF with cosine similarity. Mongo-DB stores user profiles, article metadata, and interaction data, while external news APIs provide real-time articles and topic information.

**Fig 1: System Architecture**

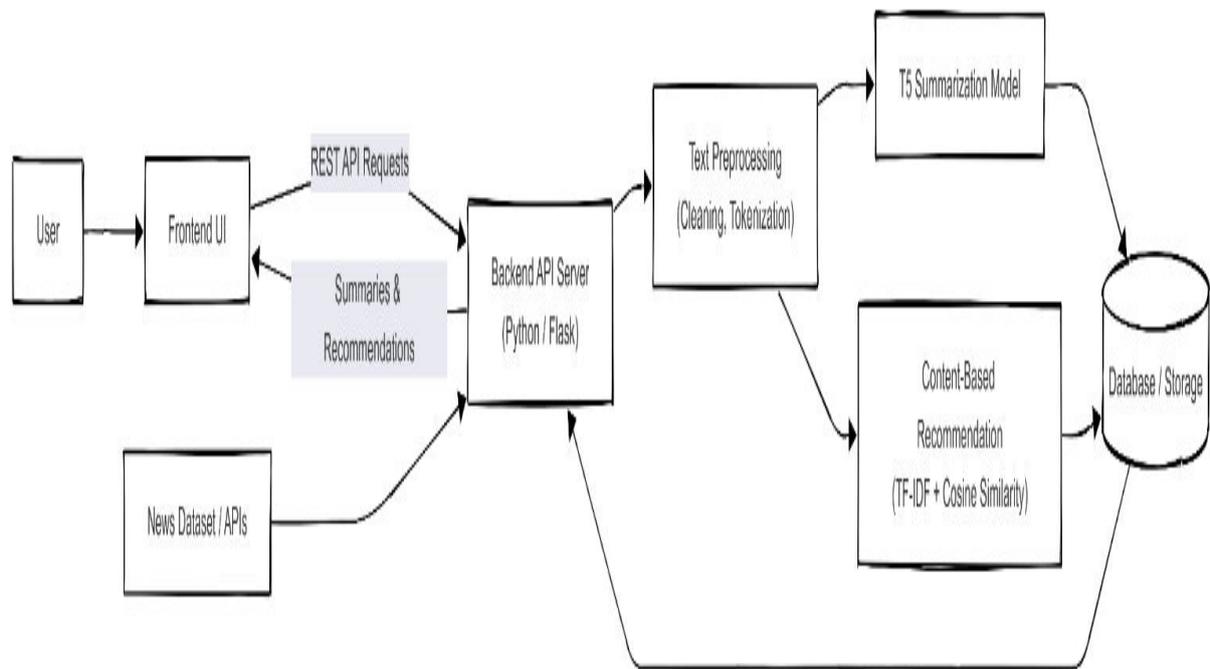
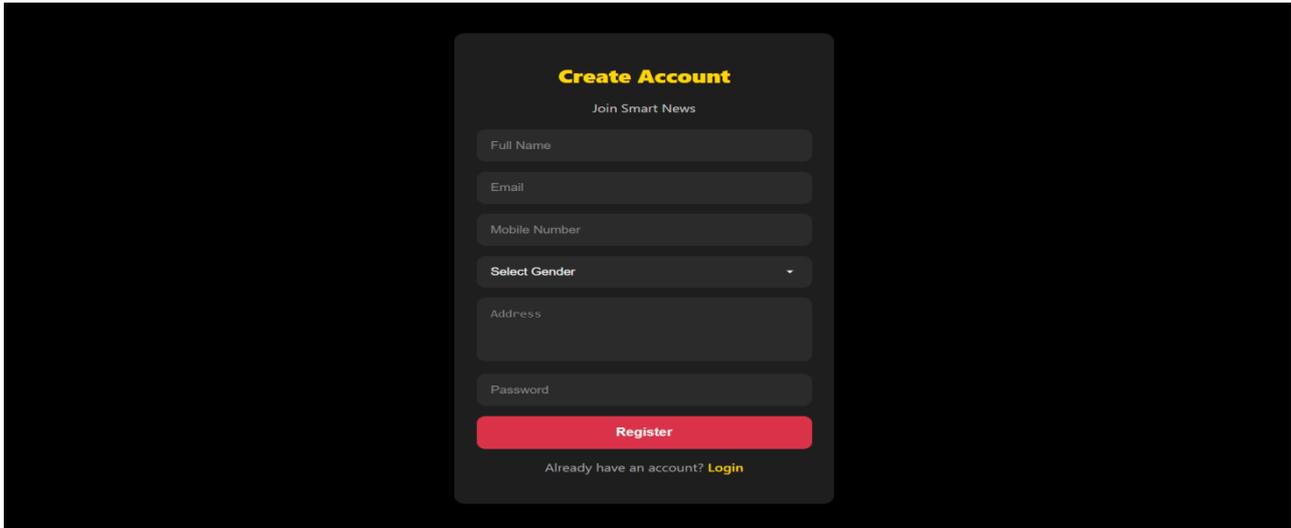


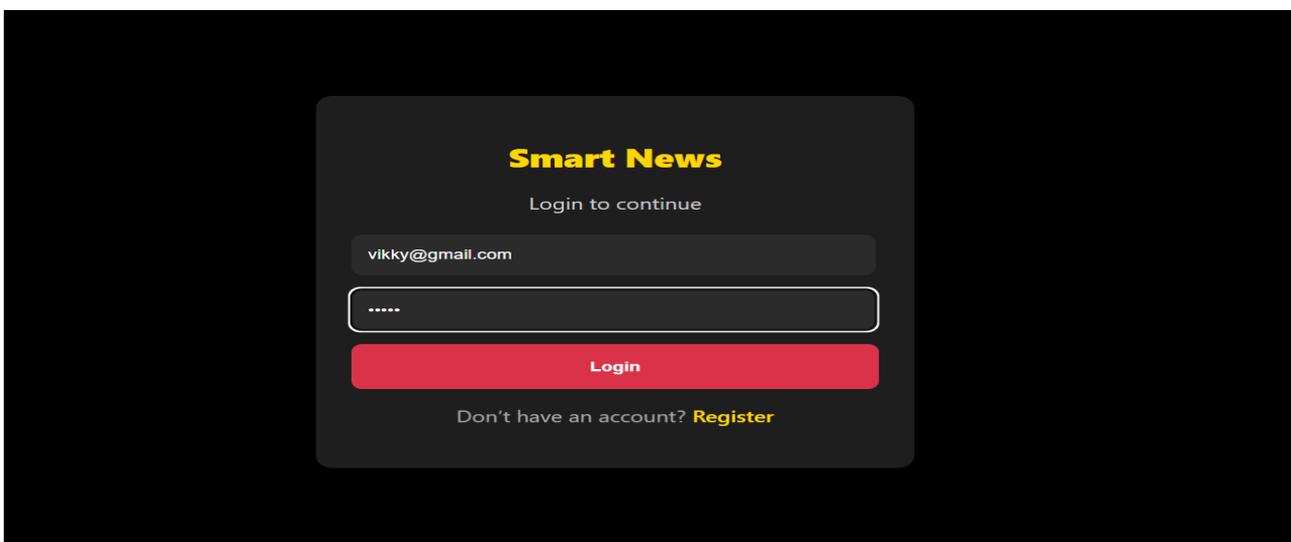
Figure 1 shows System architecture which ensures efficient data flow, scalability, and a seamless user experience by separating presentation, processing, and data management components.

**VI. RESULTS AND OUTPUT****Step 1: User Registration****Fig 2: Register Page**

The Figure-2, system allows new users to create an account through the registration interface by entering details such as name, email, mobile number, gender, address, and password. Once the information is submitted, the system securely stores the user details in the database and creates a new account for accessing the application.

**Step 2: User Login**

The system allows registered users to log in by entering their email and password through the login interface. After successful authentication, the user is granted access to the Smart News application to view personalized news summaries and recommendations for details refer to Figure-3.

**Fig 3: Login Page**

**Step 3: News Dashboard Display**

In Figure-4, After logging in, the system displays the main Smart News dashboard where users can view trending and recent news articles. The interface organizes news into different categories such as Health, Tech, Business, Sports, and Recommended to help users easily browse and read relevant content.

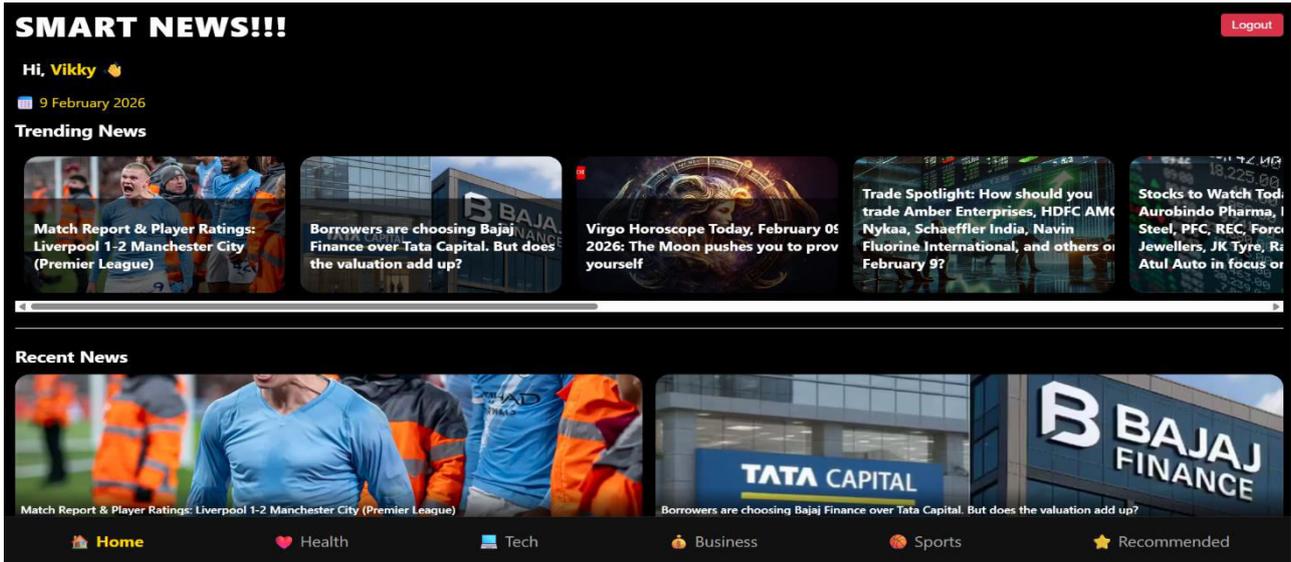


Fig 4: News Dashboard Page

**Step 4: Personalized News Recommendation**

In Figure-5, The system analyses the user's reading preferences and behaviour to generate personalized news recommendations. Using content-based filtering techniques such as TF-IDF and cosine similarity, the application suggests relevant articles tailored to the user's interests.

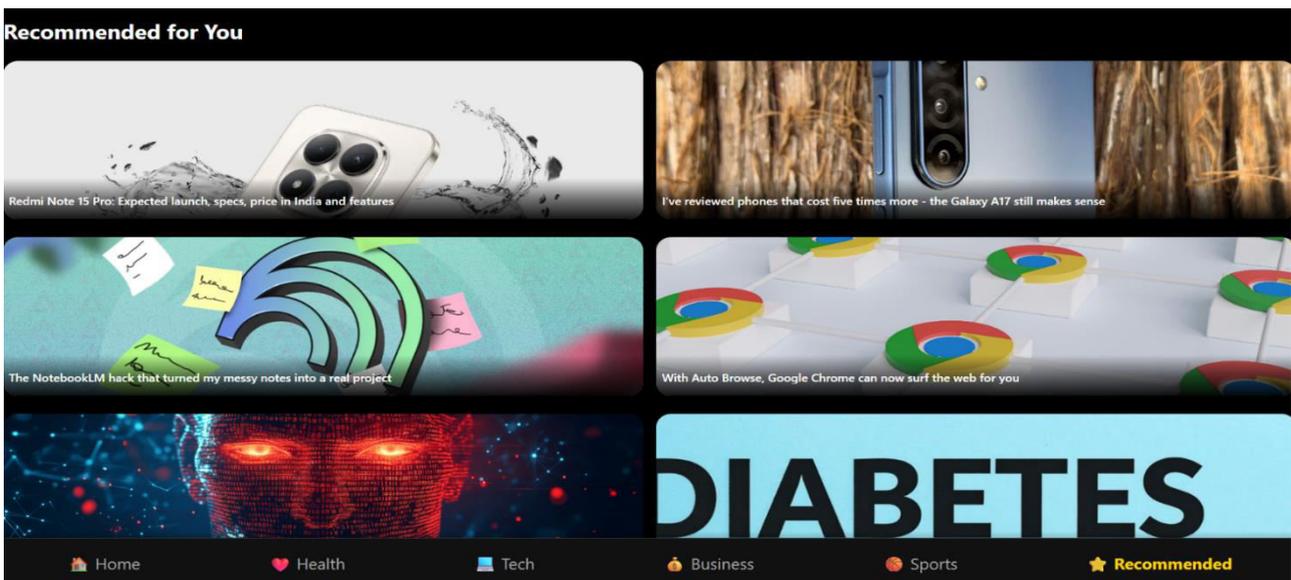


Fig 5: Recommendation Page

### Step 5: AI-Based News Summarization

When a user selects a news article, the system generates a concise summary using the T5-based NLP summarization model. This allows users to quickly understand the key points of the article without reading the entire content, with an option to view the full article if needed, refer Figure-6 for details.

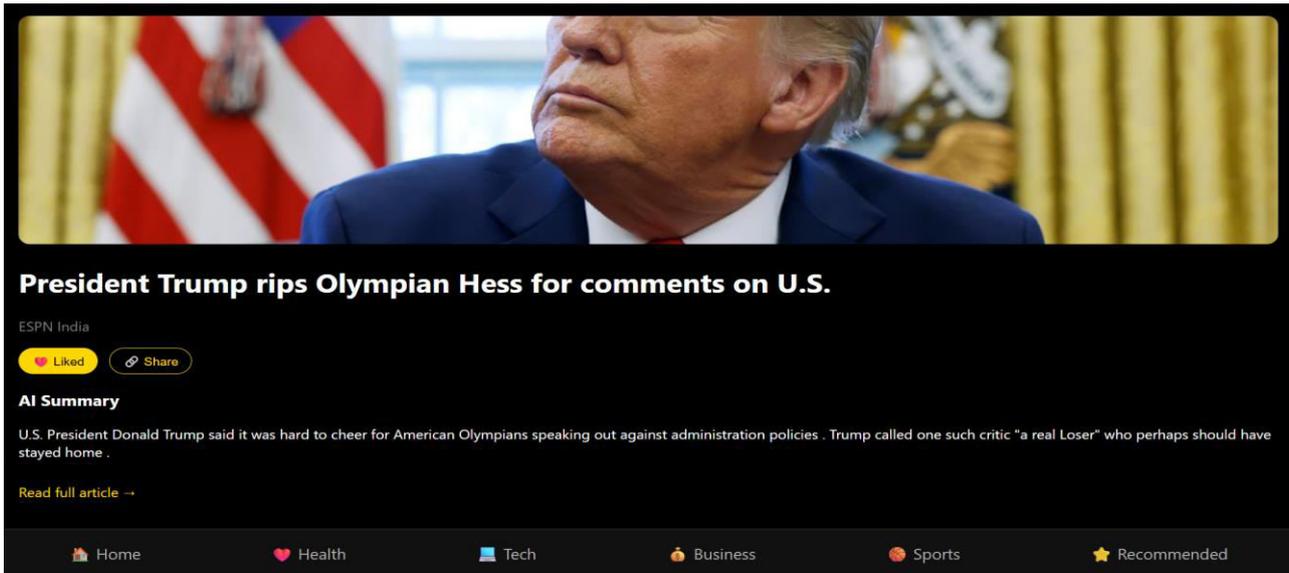


Fig 6: News Details Page

## VII. CONCLUSION AND FUTURE SCOPE

The Integrated News Summarization, Classification, and Recommendation System was successfully designed and implemented as an intelligent news application. The system addresses major challenges of traditional news platforms such as information overload and lack of personalization by delivering concise, categorized, and personalized news content in real time.

The application uses a modern technology stack with a React-based Progressive Web Application (PWA) frontend, Flask backend, and Mongo-DB for data storage. Advanced NLP and ML techniques are integrated, including the T5-SMALL transformer model for news summarization and TF-IDF with cosine similarity for personalized recommendations, while topic classification is handled using reliable API-provided metadata.

Overall, the system provides a scalable, efficient, and user-friendly platform that improves news reading efficiency and engagement. The modular architecture also allows future enhancements such as multilingual support, push notifications, and advanced recommendation models.

The system can be improved by adding support for multiple languages and regional news, allowing users to receive news in their preferred language and location. Advanced deep learning or hybrid recommendation models can be used to provide more accurate and personalized news suggestions. The application can also include push notifications to instantly inform users about breaking news and important updates. Additionally, deploying the system on cloud platforms like AWS, Azure, or Google Cloud can improve scalability, reliability, and support for a large number of users.

**REFERENCES**

- [1] Zhou, X., Li, J., & Ren, X. (2022). Segmentation-Based End-to-End Summarization for Long News Articles. <https://arxiv.org/abs/2205.07847>
- [2] Nayeem, M., & Chali, Y. (2022). NEWTS: A Topic-Focused Summarization Corpus for Topic-Aware Text Summarization. <https://arxiv.org/abs/2201.09987>
- [3] Poudel, P., et al. (2022). Multilingual News Classification using Fine-Tuned T5-SMALL and SVM Models. <https://arxiv.org/abs/2204.12811>
- [4] Kumar, A., & Gupta, V. (2023). A Comparative Study of News Categorization using Machine Learning and Transformer Models. <https://arxiv.org/abs/2302.06712>
- [5] El Mahdaoui, A., et al. (2023). News Genre Classification using Ensemble Transformers (RoT5-SMALLa + Longformer). <https://arxiv.org/abs/2310.03455>
- [6] Liu, Y., et al. (2019). RoT5-SMALLa: A Robustly Optimized T5-SMALL Pretraining Approach. <https://arxiv.org/abs/1907.11692>
- [7] Lewis, M., et al. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation and Summarization. <https://arxiv.org/abs/1910.13461>
- [8] Raffel, C., et al. (2020). T5: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. <https://arxiv.org/abs/1910.10683>
- [9] Devlin, J., et al. (2019). T5-SMALL: Pre-Training of Deep Bidirectional Transformers for Language Understanding. <https://arxiv.org/abs/1810.04805>
- [10] KUMAR, P. ASHOK, GSatish KUMAR, and SETTI NARESH KUMAR. "Improve the Capacity of Uniform Embedding for Efficient JPEG Steganography Based on DCT." (2015).
- [11] Kumarr, P. Ashok, et al. "In-vestigating Context-Aware Sentiment Classification Using Machine Learning Algorithms." Accelerating Discoveries in Data Science and Artificial Intelligence 1. ICDSAI 2023. Springer Proceedings in Mathematics & Statistics. Vol. 421. 2024.
- [12] Sunkavalli, Jayaprakash, et al. "Experimentation Analysis of VQC and QSVM on Sentence Classification in Quantum Paradigm." In 2024 International Conference on Computing, Sciences and Communications (ICCSC), pp. 1-5. IEEE, 2024.
- [13] Kumar, P. Ashok. "Event Based Time Series Sentiment Trend Analysis." ResearchGate.
- [14] Pandi, Chiranjeevi, et al. "A Survey: Recommender System For Trustworthy."
- [15] Kumar, P. Ashok, B. Vishnu Vardhan, and Pandi Chiranjeevi. "Correction to: Investigating Context-Aware Sentiment Classification Using Machine Learning Algorithms." In XVIII International Conference on Data Science and Intelligent Analysis of Information, pp. C1-C1. Cham: Springer Nature Switzerland, 2023.
- [16] Supraja, Thatikonda, et al. "Lstm-based cloud workload prediction using pso-ga hybrid algorithm." Security Issues in Communication Devices, Networks and Computing Models. CRC Press, 2025. 44-52.
- [17] Sunkavalli, Jayaprakash, et al. "Experimentation Analysis of VQC and QSVM on Sentence Classification in Quantum Paradigm." 2024 International Conference on Computing, Sciences and Communications (ICCSC). IEEE, 2024.
- [18] Supraja, Thatikonda, et al. "Exploring IoT and 5G in Smart Healthcare: Applications, Challenges and Future Directions." CVR Journal of Science and Technology 28.1 (2025): 12-21.
- [19] Supraja, T., Chiranjeevi, P., Kumar, P. A., & Swapna, Y. (2025). Exploring IoT and 5G in Smart Healthcare: Applications, Challenges and Future Directions. CVR Journal of Science and Technology, 28(1), 12-21.
- [20] Supraja, Thatikonda, Pandi Chiranjeevi, P. Ashok Kumar, and Yenugula Swapna. "Exploring IoT and 5G in Smart Healthcare: Applications, Challenges and Future Directions." CVR Journal of Science and Technology 28, no. 1 (2025): 12-21.
- [21] Supraja, T., Chiranjeevi, P., Kumar, P.A. and Swapna, Y., 2025. Exploring IoT and 5G in Smart Healthcare: Applications, Challenges and Future Directions. CVR Journal of Science and Technology, 28(1), pp.12-21.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details