Proceedings of the Third International Conference on Augmented Intelligence and Sustainable Systems (ICAISS-2025)
IEEE Xplore Part Number: CFP25CB2-ART; ISBN: 979-8-3315-0724-4

Unified E-Commerce Platform Aggregator using NLP and Machine Learning Techniques

Ramalakshmi. P
Assistant Professor
Department of Information Technology
Agni College of Technology
Chennai, Tamil Nadu
ramalakshmi.it@act.edu.in

Senthil G. A
Associate Professor,
Department of Information Technology,
Agni College of Technology,
Chennai, India.
senthilga@gmail.com

Geerthik S
Associate Professor,
Department of Information Technology,
Agni College of Technology,
Chennai, India
geerthiks@gmail.com

Nanthini Priya. R
Department of Information Technology
Agni College Of Technology
Chennai, Tamil Nadu
nanthinir0110@gmail.com

Praveen. M
Department of Information Technology
Agni College Of Technology
Chennai, Tamil Nadu
praveen9080208131@gmail.com

2025 Third International Conference on Augmented Intelligence and Sustainable Systems (ICAISS) | 979-8-3315-0724-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAISS61471.2025.11042138

Abstract **- The study proposed the integration of Natural Language Processing (NLP) with Machine Learning (ML) techniques like ARIMA (Autoregressive Integrated Moving Average) and LSTM (Long Short-Term Memory) to develop a smart predictive system for e-commerce platforms. NLP extracts insights from customer feedback, product descriptions, and social media sentiments to understand customer preferences and market trends. ARIMA, a time-series forecasting model, predicts demand fluctuations based on historical sales data, enabling better stock management, while LSTM captures long-term dependencies in sequential data for more accurate forecasting and product recommendations. By combining these techniques, the system enhances decision-making, optimizes pricing strategies, and improves customer engagement through personalized recommendations. Additionally, it helps businesses maintain an efficient inventory system by reducing overstocking and stockouts, ultimately increasing profitability. This approach not only benefits businesses but also enhances the shopping experience by making e-commerce platforms more adaptive and responsive to changing market demands. The integration of NLP, ARIMA, and LSTM ensures a data-driven strategy that leverages AI and ML to transform the e-commerce industry through predictive analytics and intelligent decision-making.**

Keywords— Unified E-Commerce Platform, Natural Language Processing (NLP), Product Comparison, Data Aggregation, ARIMA (Autoregressive Integrated moving average), LSTM (Long Short-Term Memory), E-Commerce Aggregator, Sales Forecasting, Price Comparison.

I. INTRODUCTION

In the rapidly evolving e-commerce landscape, businesses are inundated with vast amounts of data from customer interactions, product reviews, and sales transactions. effectively analyzing this data is crucial for informed decision-making and maintaining a competitive edge. recent studies have explored the integration of NLP with ML models to enhance predictive analytics in e-commerce. for instance, Vavliakis et al. (2021) proposed a hybrid approach combining ARIMA and LSTM models to optimize sales forecasting, demonstrating improved accuracy by capturing both linear and nonlinear sales trends [1]. Similarly, Li et al. (2024) developed a combinatorial optimization model that leverages ARIMA and LSTM for demand forecasting, aiming to enhance inventory management and supply chain efficiency [2]. Building upon these methodologies, this project seeks to integrate NLP with ARIMA and LSTM models to develop a predictive system tailored for e-commerce platforms. The NLP component will process unstructured textual data, which includes customer reviews and social media posts, to extract sentiments and emerging trends. These insights will inform the ARIMA model for time-series forecasting, capturing linear patterns in sales data, while the LSTM model will address complex, nonlinear relationships and long-term dependencies. By combining these techniques, the system aims to provide accurate sales forecasts, optimize inventory levels, and enhance customer engagement through personalized recommendations [3]. This integrated approach not only aligns with existing research but also endeavors to advance predictive analytics in e-commerce by offering a comprehensive, data-driven solution [4].

This research seeks to enhance the net shopping experience by means of offering a consolidated interface in which users can find product details from diverse e-commerce websites, including Amazon, Flipkart, and so on. within the gift e-trade panorama, customers stumble upon the difficulty of navigating between multiple apps to evaluate charges, opinions, and product attributes. This task addresses that problem by merging more than one e-commerce platform into an unmarried entity, simplifying the shopping enjoyment and holding customers' time, attempt, and tool garage. This project addresses that issue by merging multiple e-commerce platforms into a single entity, simplifying the shopping experience and conserving users' time, effort, and device storage. This integrated platform not only enhances the convenience of online shopping but also assists users in making more informed purchasing choices [5]. establishing a centralized solution promotes knowledgeable buying, improves user satisfaction, and has the potential to boost customer loyalty. This project illustrates how advanced

technology can revolutionize the e-commerce experience, rendering it more accessible and enjoyable for all [6].

II. METHODOLOGY

A. Application of NLP

In this research text data from several e-commerce websites, including product descriptions and customer evaluations, is analyzed and interpreted using NLP [7]. The platform uses NLP algorithms to classify items, comprehend user queries, and rank search results according to relevancy. NLP provides precise product matching across several e-commerce platforms by processing text data, giving customers access to consolidated information such as reviews, specs, and price comparisons for comparable products. In addition to improving search accuracy, this intelligent text analysis also increases customization, giving consumers the most pertinent information in a timely and effective manner. In current competitive times, seamless customer service is crucial for e-commerce companies. By classifying the queries and assigning priority according to the content, natural language processing allows them to expedite the query processing [8].

Fig. 1. NLP in Retail and E-commerce

B. Process of NLP in Unified E-commerce platform

A methodical approach to problem-solving is employed in tests and research. Finding a query or issue is the first step in the process, which is then followed by background research to comprehend what is already known [9]. A hypothesis is developed in light of this study and subsequently put to the test in an experiment. Troubleshooting is necessary to improve the process if the experiment does not go as planned. Following a successful operation, data is evaluated and a conclusion is drawn. The results are recorded as final results if they support the hypothesis. Even if the results only partially match, they nevertheless help guide future research since they could raise a new query or idea that starts a new round of testing. In the end, this iterative process advances knowledge by ensuring thorough testing and improvement. Finally, NLP facilitates cross-platform integration, enabling seamless comparison of products, reviews, and pricing across multiple e-commerce sites within a single interface. This improves decision-making for both businesses and customers, making the unified e-commerce platform more intelligent, efficient, and user-friendly.

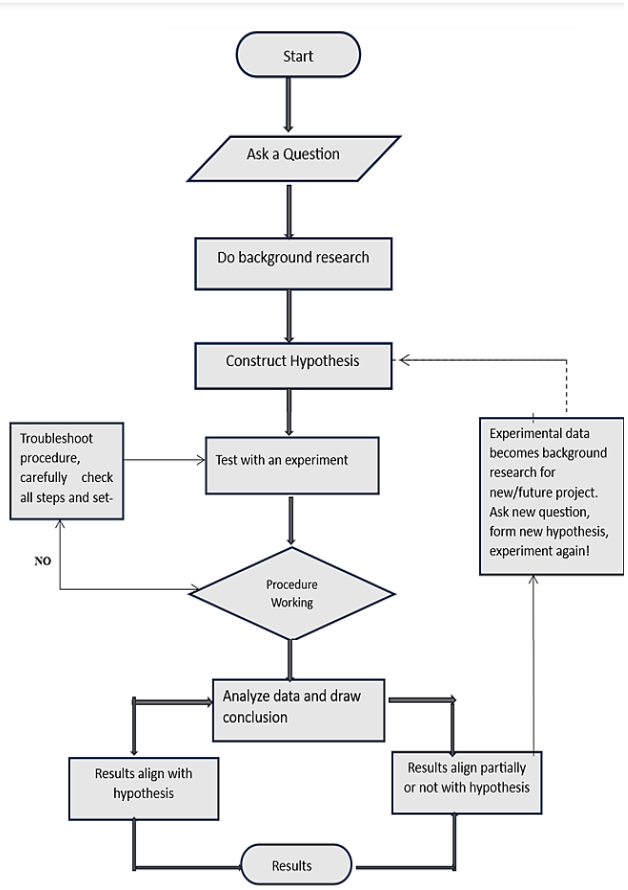


Fig. 2. Scientific method Workflow of NLP

C. Product classification using ML Techniques

NLP is integrated with Machine Learning techniques such as ARIMA and LSTM to enhance predictive modeling and decision-making [10]. The NLP component is used to procedure and extract meaningful insights from unstructured information, inclusive of purchaser evaluations, product descriptions, and social media sentiment, that are then fed into the ML models. ARIMA is employed for time-series forecasting, particularly useful in predicting trends and demand patterns based on historical data, while LSTM is applied for sequence prediction, capturing long-term dependencies in data sequences for more accurate outcomes [11]. Together, these techniques provide a comprehensive framework for analyzing and predicting customer behavior, optimizing inventory management, and improving overall business strategies.

(i) ARIMA for Sales Forecasting

The Autoregressive Integrated Moving Average (ARIMA) model is employed to predict future sales trends by analyzing historical sales data. ARIMA is effective in capturing linear dependencies in time-series data and is defined by three main parameters:

i. p (Autoregressive Order): Number of past values considered for forecasting.
ii. d (Differencing Order): Number of times the data needs to be differenced to make it stationary.
iii. q (Moving Average Order): Number of past forecast errors used in the prediction model.

$$Y_t'' = Y_t' - Y_{t-1}' = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \quad (1)$$

where Yt″ represents second-order differencing. This helps in removing trends and making the series stationary, which is essential for applying ARIMA effectively.

(ii) ***LSTM For Demand Prediction***

LSTM is a sort of recurrent neural network (RNN), that is utilized for demand forecasting via getting to know styles from sequential information. Unlike ARIMA, LSTM can capture nonlinear dependencies and long-term relationships in sales data, making it suitable for complex time-series predictions [12]. The LSTM model updates its cell state using the following equations:

Candidate Cell State Updating Equation

$$\tilde{C}_t = \tanh(W_c . [h_{t-1}, x_t] + b_c) \quad (2)$$

where:

i. Ct~ is the candidate cell state.
ii. W_c and b_c are weight and bias parameters.
iii. ht−1 is the previous hidden state.
iv. xt is the current input.

Hidden State Updating Equation

$$ht = ot * \tanh(Ct) \quad (3)$$

where,

i. ht is the current hidden state.
ii. Ot is the output gate value.
iii. Ct is the updated cell state.

These formulations make certain that the LSTM keeps lengthy-time period dependencies even when selectively forgetting beside the point statistics, making it highly effective for sequential data prediction.

D. Utilized tools and Technologies

The following table represents several technologies and techniques to provide a potent e-commerce product recommendation system. Text-based data is processed and analyzed using NLP libraries like NLTK, SpaCy, or BERT, and pertinent information is classified and predicted with the use of machine learning techniques like Linear SVM and Gaussian Naive Bayes [15]. RESTful APIs with JSON-based communication make API integration easier and enable smooth communication between many services. Web scraping technologies like Scrapy and Selenium are used to collect pertinent data from e-commerce websites to extract data [13].

Table 1. Utilized Tools, Algorithms and Technologies

Core Utilization	Tools and Technologies
Machine Learning Models	ARIMA and LSTM
API Integration	RESTful APIs, JSON-based API communication
Web Scraping Tools	Scrapy, Selenium
Backend Framework	Flask or Django
Frontend Framework	Html, CSS, React
Cloud Hosting and Deployment	AWS, Google Cloud, or Azure
NoSQL Database	MongoDB or Firebase
NLP	Libraries (e.g., NLTK, SpaCy, or BERT)

III. PROPOSED WORK

A unified e-commerce platform that leverages NLP and ML models like ARIMA and LSTM for product classification, demand forecasting, and enhanced user experience. NLP processes data from multiple platforms, extracting insights from product descriptions, reviews, and pricing. Sentiment analysis and named entity recognition help understand customer preferences and market trends. ARIMA predicts sales trends by analyzing historical data, identifying seasonal patterns, and optimizing inventory management. LSTM enhances demand forecasting by capturing complex, long-term dependencies in sales data [16]. Combining ARIMA and LSTM ensures accurate predictions for stock replenishment and pricing strategies. Additionally, NLP-powered product classification improves search accuracy and personalized recommendations, simplifying product discovery. This integration creates an efficient, intelligent e-commerce platform that saves time, optimizes business operations, and enhances the shopping experience for users and businesses [17].

The system collects information from several sources, including product names, pricing, reviews, and availability, and compiles it into a single, user-friendly interface using API integration and web scraping techniques. Tokenization, stop word removal, and word embeddings are important NLP techniques that aid in search query analysis, keyword recognition, and more accurate product matching [18]. A recommendation engine also improves the user experience by offering tailored product recommendations. A central database for storing aggregated data is also part of this system; it is updated often to ensure correctness. Users may locate the greatest items in a single, effective app and save time and effort by using this unified platform. By doing away with the need for several shopping applications, the entire procedure not only increases accessibility but also maximizes storage space on customers' smartphones [19].

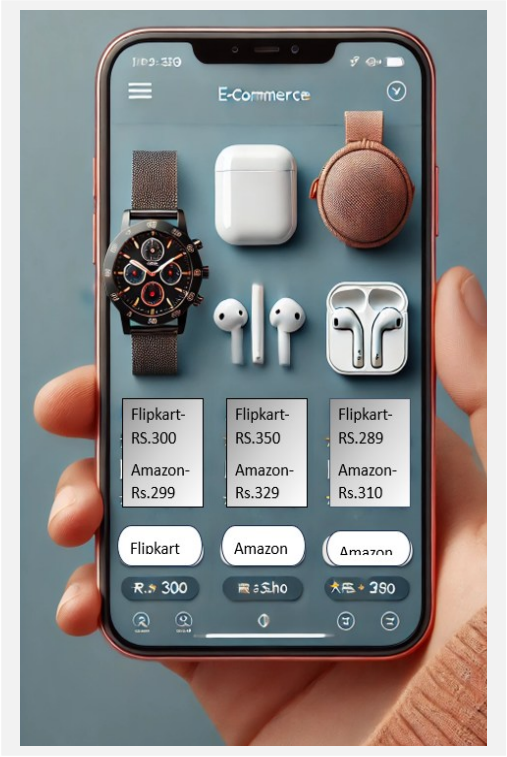


Fig. 3. Sample interface of Unified Platform

The suggested solution for the Unified E-Commerce Platform Aggregator Using NLP is to establish a single interface that consolidates items from numerous e-commerce sites, such as Amazon, Flipkart, etc, into a single app. This technology interprets user search queries using Natural Language Processing (NLP), guaranteeing accurate and pertinent results. The system gathers product information from several sources, such as pricing, reviews, and availability, by combining APIs and web scraping techniques. This information is then compiled and arranged for convenient comparison. This unified approach saves users time and storage space, simplifies the online shopping process, and offers them a comprehensive view of products across multiple platforms in one place [22]. Figure 3 shows sample interface of unified platform.

A. Effects and Advantages of this experiment

a. The research improves the shopping experience by streamlining product searches and assisting customers in finding pertinent products with accurate recommendations.
b. Product tags that are precise and pertinent reduce false information, enhancing user confidence and search engine dependability.
c. Automating tagging and suggestions frees up sellers' time so they can concentrate on other important aspects of their business.
d. Customer confidence is increased by accurate and well-structured product information, which promotes recurring business and loyalty. By showing customers products that closely match their tastes, personalized suggestions increase conversion rates [20].
e. The project helps firms make data-driven strategic decisions by producing insightful data about user behaviour and preferences [21].

Table 2. Benefits of a Proposed Unified Application

IMPACTS	ADVANTAGES
Increased precision of the tags	Eliminates false information
Speeds up product discovery	Improves user experience
Seller automation	Saves Time
Better suggestions	Promotes Sales
increases client trust	Promotes recurring business
Data-driven insights	Strengthens business strategy

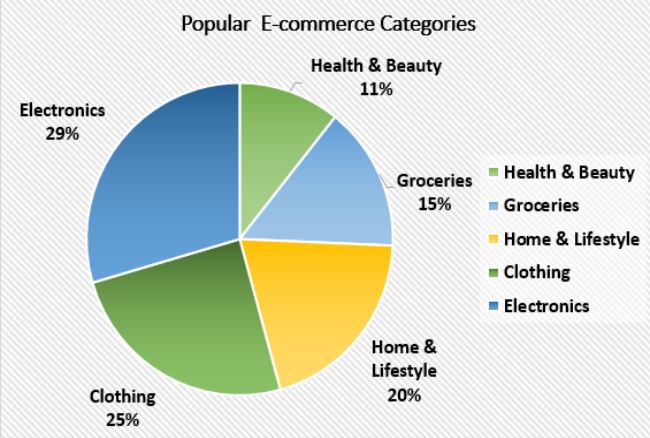
IV. RESULT AND DISCUSSION

The unified e-commerce platform successfully integrates data from multiple sources, allowing users to compare product prices, reviews, and features in a single interface. The use of NLP for product classification improves search accuracy and personalized recommendations, helping users find the best deals. ARIMA-based forecasting provides insights into sales trends, enabling businesses to manage inventory efficiently and reduce stock shortages. LSTM enhances demand prediction by capturing complex patterns in customer behavior, leading to more accurate sales forecasts. The comparison of different e-commerce platforms shows significant price variations, helping users make informed purchasing decisions.

Table 3. Share of online retail transactions over the years

Year	Share of Online Retail Sales
2021	**18.8%**
2022	**18.70%**
2023	**19.40%**
2024*	**20.10%**
2025*	**21%**
2026*	**21.80%**
2027*	**22.60%**

Performance analysis indicates that the integration of NLP with ML models improves overall platform efficiency, reducing search time and enhancing user experience. Businesses can optimize pricing strategies and inventory management, while customers benefit from a seamless and cost-effective shopping experience. The results are validated using standard performance metrics like RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error).



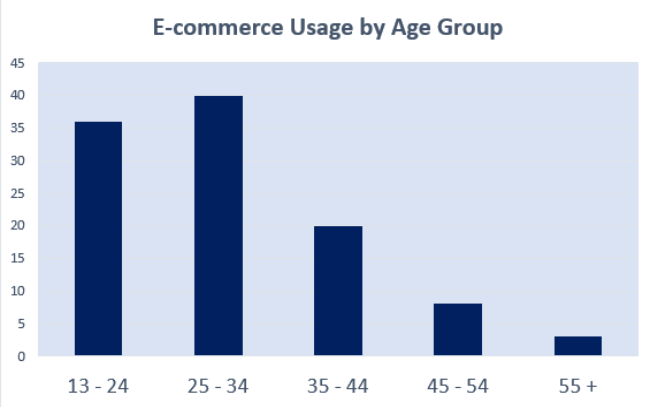


Fig. 4. Analyzation of E-commerce Engagement

The above graphs show the popularity of product categories and the use of e-commerce across different age groups. According to the bar graph on the left, users between the ages of 18 and 24 and 25 and 34, who together make up 75% of e-commerce users, participate the most, with usage sharply declining as users get older. This trend suggests that younger populations are more likely to shop online. According to the pie chart on the right, the most sought-after categories are electronics (30%), apparel (25%), and home & lifestyle [8]. These findings highlight the necessity for e-commerce platforms to prioritize popular product categories while adjusting their products and marketing tactics to appeal to younger customers.

V. CONCLUSION

This research successfully develops a unified e-commerce platform that integrates NLP and ML techniques like ARIMA and LSTM for product classification, price comparison, and demand forecasting. By gathering data from multiple platforms, users can compare product prices, reviews, and trends in one place, enabling better purchasing decisions. NLP improves search accuracy, sentiment analysis, and personalized recommendations, enhancing user experience. ARIMA predicts sales trends using historical data, while LSTM captures long-term patterns for better demand forecasting. This integration optimizes inventory management and pricing strategies, benefiting businesses and consumers alike. The results show improved platform efficiency, reduced search time, and accurate demand predictions. Users save time by accessing multiple e-commerce sites in a single interface, while businesses minimize overstock and shortages. This approach promotes a transparent and competitive market, ensuring consumers get the best deals. Future improvements could include expanding platform support, real-time price tracking, and AI-driven recommendations. Overall, this platform bridges multiple e-commerce services, providing a seamless and cost-effective shopping experience.

In the future, this research can be improved by integrating advanced AI technologies to enhance product recommendations, demand forecasting, and user experience. Using deep learning models like Transformers for NLP can improve product classification, sentiment analysis, and personalized search results. AI-powered chatbots and voice assistants can be introduced to help users find the best deals and assist with queries in present. Reinforcement learning can be used to develop smart pricing strategies that adjust based on demand, competitor prices, and customer preferences. Additionally, AI can be implemented for fraud detection, ensuring secure transactions and preventing fake product listings. The platform can also be expanded to support more e-commerce websites, allowing for a wider range of product comparisons. Real-time AI-driven price tracking can help users get instant alerts on price drops, ensuring they make purchases at the best possible rates. By continuously improving with AI, this unified e-commerce platform can become more efficient, intelligent, and user-friendly, providing both consumers and businesses with a seamless and smarter shopping experience.

REFERENCE

[1] Senthil, G.A., Prabha, R., Suganthi, S., Sridevi, S., Shanthi, N. (2023). An IoT-Enabled Smart Network Traffic Signal Assistant System for Emergency Vehicles Using Computer Vision. In: Raj, J.S., Perikos, I., Balas, V.E. (eds) Intelligent Sustainable Systems. ICoISS 2023. Lecture Notes in Networks and Systems, vol 665. Springer, Singapore.

[2] Shitian Li, Junzhe Zhang, Ziyu Zhang, Xv Chu, Lili tune, Xiaojun Wang, Combinatorial Optimisation model for E-commerce Retail merchant call for Forecasting based on ARIMA and LSTM. Facts Structures and Economics (2024) Vol. 5: 91-99. DOI: http://dx.doi.org/10.23977/infse.2024.050513.

[3] Khurana, D., Koli, A., Khatter, ok. et al. Natural language processing: The nation of the artwork, modern trends and challenges. Multimed gear Appl eighty two, 3713–3744 (2023). https://doi.org/10.1007/s11042-022-13428-four.

[4] Acosta Gutiérrez, Gina. (2020). Comparative observation of NLP and device mastering techniques for Sentiment analysis and topic Modeling on Amazon evaluations. worldwide journal of computer science Engineering. nine. 159-one hundred seventy. 10.21817/ijcsenet/2020/v9i2/200902007.

[5] Aljbour, M., Avcı, İ. (2024). sales Prediction in E-commerce platforms the use of gadget mastering. In: Rasheed, J., Abu-Mahfouz, A.M., Fahim, M. (eds) drawing close Networks and Sustainability in the AIoT generation. FoNeS-AIoT 2024. Lecture Notes in Networks and Structures, vol 1036. Springer, Cham. https://doi.org/10.1007/978-three-031-62881-8_17

[6] J. Lv, "Research on inventory management and call for Forecasting of E-commerce Platform primarily based on ARIMA and LSTM fashions," 2024 IEEE 3rd Global Conference on Electrical Engineering, Massive Statistics and Algorithms (EEBDA), Changchun, China, 2024, pp. 494-499, doi: 10.1109/EEBDA60612.2024.10485781.

[7] Qi-Qiao He, Cuiyu Wu, Yain-Whar Si, LSTM with particle Swam optimization for income forecasting, digital trade research and packages, volume 51,2022,101118, ISSN 1567-4223,https://doi.org/10.1016/j.elerap.2022.101118.

[8] Senthil, G. A., R. Prabha, A. Pomalar, P. Leela Jancy, and M. Rinthya. "Convergence of Cloud and Fog Computing for Security Enhancement." In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 1-6. IEEE, 2021. DOI: 10.1109/I-SMAC52330.2021.9640872.

[9] R. Prabha, M. Razmah, G. Saritha, R. Asha, S. G. A and R. Gayathiri, "Vivoice - Reading Assistant for the Blind using OCR and TTS," 2022 International Conference on Computer Communication and Informatics (ICCCI), 2022, pp. 01-07, doi: 10.1109/ICCCI54379.2022.9740877.

[10] D. Roopa, M. Subashini, S. Mathupriya, R. Prabha and G. A. Senthil, "A Novel Approach for Oil Spill Detection from SAR Images," *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, Chennai, India, 2023, pp. 1-8, doi: 10.1109/ICCEBS58601.2023.10448846.

[11] R. H, T. A, G. S, S. G. A and A. J, "MindSerenity: VR Ascents to Serenity - A Personalized Journey for Anxiety Reduction, Fostering Mental Well-Being Through Immersive and Tailored Experiences," 2024 7th International Conference on Devices, Circuits and Systems (ICDCS), Coimbatore, India, 2024, pp. 16-21, doi: 10.1109/ICDCS59278.2024.10560973.

[12] G. A. Senthil, S. Geerthik, R. Karthikeyan and G. Keerthana, "Face Recognition based Automated Smart Attendance using Hybrid Machine Learning Algorithms and Computer Vision," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024, pp. 606-611, doi: 10.1109/ICAAIC60222.2024.10574896.

[13] Senthil, G. A. & Prabha, R. (2025). Predictive Analysis-Based AI-Driven Data Security Authentication and Authorization for Medical Warehousing Mechanisms. In G. Kaur, J. Arora, V. Jain, & A. Shaikh (Eds.), Strategic Innovations of AI and ML for E-Commerce Data Security (pp. 371-396). IGI Global. https://doi.org/10.4018/979-8-3693-5718-7.ch015.

[14] Senthil G. A, R. Prabha, Arun M, A. Sahaya Anselin Nisha, S. Prabu. (2024). A Novel Predictive Analysis and Classification of Land Subsidence Vulnerability Mapping based on GIS using Hybrid Optimized Machine Learning Techniques and Computer Vision, Procedia Computer Science, Volume 233, pp 343-352,ISSN 1877-0509, https://doi.org/10.1016/j.procs.2024.03.224.

[15] Senthil, G.A., Abinaya, J., Oliviya, K.J., Keerthana, R. (2025). A Novel Revolutionizing Medical Surgery Procedures Using Mixed Reality. In: Geetha, R., Dao, NN., Khalid, S. (eds) Advances in Artificial Intelligence and Machine Learning in Big Data Processing. AAIMB 2023. Communications in Computer and Information Science, vol 2202. Springer, Cham. https://doi.org/10.1007/978-3-031-73065-8_1.

[16] G. S, S. G. A, J. D and A. J, "Deepfake Video Prediction Using Attention-Based CNN and Mel-Frequency Cepstral Coefficients," *2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, Trichirappalli, India, 2024, pp. 1-6, doi: 10.1109/ICEEICT61591.2024.10718393.

[17] S. G. A, R. Prabha, I. Thamarai, D. Roopa, R. K. R and C. Srimathi, "CardioNet: An Integrative AI-Machine Learning Framework for Enhanced Prediction and Management of Cardiovascular Diseases Using Deep Data Analytics and Clinical Insights," 2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ICPECTS62210.2024.10780317.

[18] Senthil, G.A., Prabha, R., Monica, K.M., Bindu, G., Kumar, M.J. (2024). A Novel Analysis and Classification of Pressure Ulcer Prediction for Coma Patients in ICU Using Deep Learning Techniques. In: Saha, A.K., Sharma, H., Prasad, M. (eds) Proceedings of International Conference on Intelligent Vision and Computing (ICIVC 2023). ICIVC 2023. Proceedings in Adaptation, Learning and Optimization, vol 20. Springer, Cham. https://doi.org/10.1007/978-3-031-71391-0_10.

[19] Senthil, G.A., Mathumitha, V., Prabha, R., Suganthi, S., Alagarsamy, M. (2023). Simulation on Natural Disaster Fire Accident Evacuation Using Augmented Virtual Reality. Inventive Communication and Computational Technologies. ICICCT 2023. Lecture Notes in Networks and Systems, vol 757. Springer, Singapore. https://doi.org/10.1007/978-981-99-5166-6_23.

[20] R. Prabha, G.A. Senthil, S. K. B. Sangeetha, S.U. Suganthi, D. Roopa. (2023). Network Routing and Its Real-Time Practice in Broadband Wireless Networks. Wiley online library, pp:99-112, https://doi.org/10.1002/9781119827603.ch5.

[21] Prabha, R., G. A. Senthil, S. K. B. Sangeetha, S. U. Suganthi, and D. Roopa. "Broadband Wireless Network Era in Wireless Communication–Routing Theory and Practices." Modeling and Optimization of Optical Communication Networks (2023): 267.

[22] R.M. Asha, P. Pondeepak, R. Prabha, G.A. Senthil, A. Padma Bharrathi, A novel approach effect of ocean acidification on oysters, Materials Today: Proceedings,2023,ISSN 2214-7853,https://doi.org/10.1016/j.matpr.2023.01.194.