A Predictive Analytics Platform for Social Media Sentiment Analysis

Dr.Nilesh N.Thorat[1], Sylvan Futardo[2], Sahil Kshirsagar[3], Lokesh Kumawat[4], Kush Sugandhi[5]
[1]*Faculty, Assistant Professor, MIT Art, Design and Technology University, Pune,*
[2,3,4,5]*Student, MIT Art, Design and Technology University, Pune*

Abstract: -In the contemporary digital economy, the ability to understand and predict customer sentiment is a critical driver of business strategy and growth. Businesses increasingly depend on customer reviews and social media posts, but face significant hurdles in analyzing this vast, unstructured data, including the presence of fake reviews and multilingual content[111]. This paper introduces a comprehensive Predictive Analytics Platform designed to automate sentiment analysis and forecast market trends. The system architecture integrates data collection from diverse APIs, followed by robust preprocessing using Natural Language Processing (NLP)[2]. At its core, the platform utilizes a novel hybrid machine learning model combining Random Forest and Long Short-Term Memory (LSTM) to classify sentiment from labeled training data[3]. Implemented with an interactive dashboard built on Streamlit, the system achieves 85% accuracy in its predictions. By delivering dynamic data visualizations and predictive insights, the platform equips businesses with the tools for rapid, data-driven decision-making in a dynamic environment.

Keywords: Sentiment Analysis, Predictive Analytics, Machine Learning, LSTM, Random Forest, Natural Language Processing, Social Media Analytics, TensorFlow, Streamlit.

I. INTRODUCTION

In today's interconnected world, understanding user sentiment matters deeply. Businesses operate in an environment where customer reviews and social media posts directly influence brand perception and consumer behavior. The ability to systematically analyze this feedback is no longer a luxury but a necessity for effective decision-making. Sentiment analysis, the process of computationally identifying and categorizing opinions expressed in text to determine the author's attitude, provides this crucial capability by capturing the emotion behind the words. [1]

However, the path from raw social media data to actionable intelligence is fraught with technical obstacles that can compromise the accuracy and efficiency of the analysis.

1.1 Challenges in Social Media Analytics

The primary issues affecting data analysis in this domain include:

Unstructured Text Data: Social media content is inherently unstructured, complicating quantitative analysis. [2]

Fake Reviews: The presence of fraudulent or inauthentic reviews can significantly skew sentiment accuracy and mislead businesses.

Multilingual Issues: Content generated in multiple languages introduces translation challenges and adds layers of complexity to sentiment models.

Data Silos: Aggregating data from multiple, disparate platforms hinders the consistency and reliability of the analysis. [3]

To overcome these challenges, this paper proposes a Predictive Analytics Platform. The system is engineered to seamlessly collect and process data, utilizing a powerful hybrid machine learning model to deliver accurate trend predictions and empower businesses to leverage data-driven strategies.

II. LITERATURE REVIEW

The field of sentiment analysis has evolved significantly with advancements in machine learning and Natural Language Processing (NLP). Initial methodologies were often lexicon-based, relying on dictionaries of words pre-coded with positive or negative scores. While foundational, these approaches frequently fail to capture the nuances of human

language, such as context, sarcasm, and domain-specific terminology. [5]

The advent of machine learning introduced more robust, data-driven techniques. Algorithms like Naive Bayes, Support Vector Machines (SVM), and Random Forest became standard for text classification tasks. Random Forest, an ensemble learning method, is particularly noted for its high performance and resistance to overfitting, making it a strong candidate for classifying sentiment based on textual features.

More recently, deep learning models have set a new benchmark for performance. Recurrent Neural Networks (RNNs), and specifically Long Short-Term Memory (LSTM) networks, are designed to handle sequential data, making them exceptionally well-suited for NLP. LSTMs can recognize long-term dependencies in text, allowing for a more profound understanding of sentence structure and context, which is critical for accurate sentiment detection. [4]

Current research is exploring the power of hybrid models that combine the strengths of different architectures. By integrating a deep learning model for sophisticated feature extraction with a proven traditional classifier, it is possible to achieve superior results. Our work builds on this concept by proposing a novel hybrid of LSTM and Random Forest. The hypothesis is that using LSTM to learn and extract contextual features from text and then feeding these features into a Random Forest classifier will yield a highly accurate and efficient system for analyzing sentiment and predicting future trends. [6]

III. PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY

The system is designed to provide a seamless workflow, from data ingestion to the final delivery of predictive insights17. The architecture ensures that data from varied sources is standardized and processed efficiently for accurate analysis. [7]

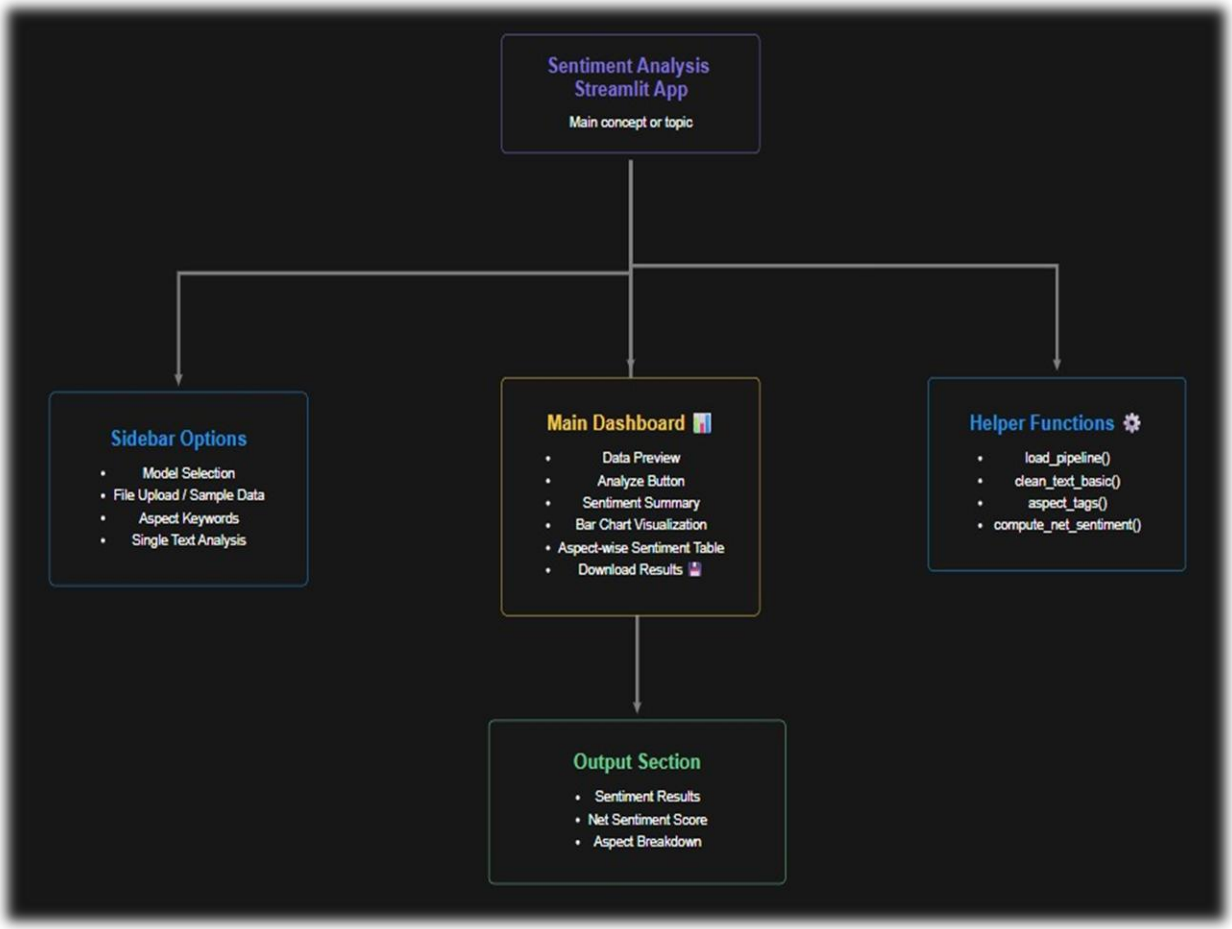

Fig.1 System Architecture.

3.1 System Flow

The methodology follows a structured, multi-step process[19]:

1. Data Collection: The initial stage involves gathering raw data from various social media platforms and sources via APIs.
2. Data Preprocessing: The collected data, which is often noisy and unstructured, undergoes a cleaning and organization phase. This step is critical for preparing the text for effective analysis. [9]
3. Sentiment Classification: The cleaned data is then passed to the machine learning model, which analyzes the text to classify its underlying sentiment.

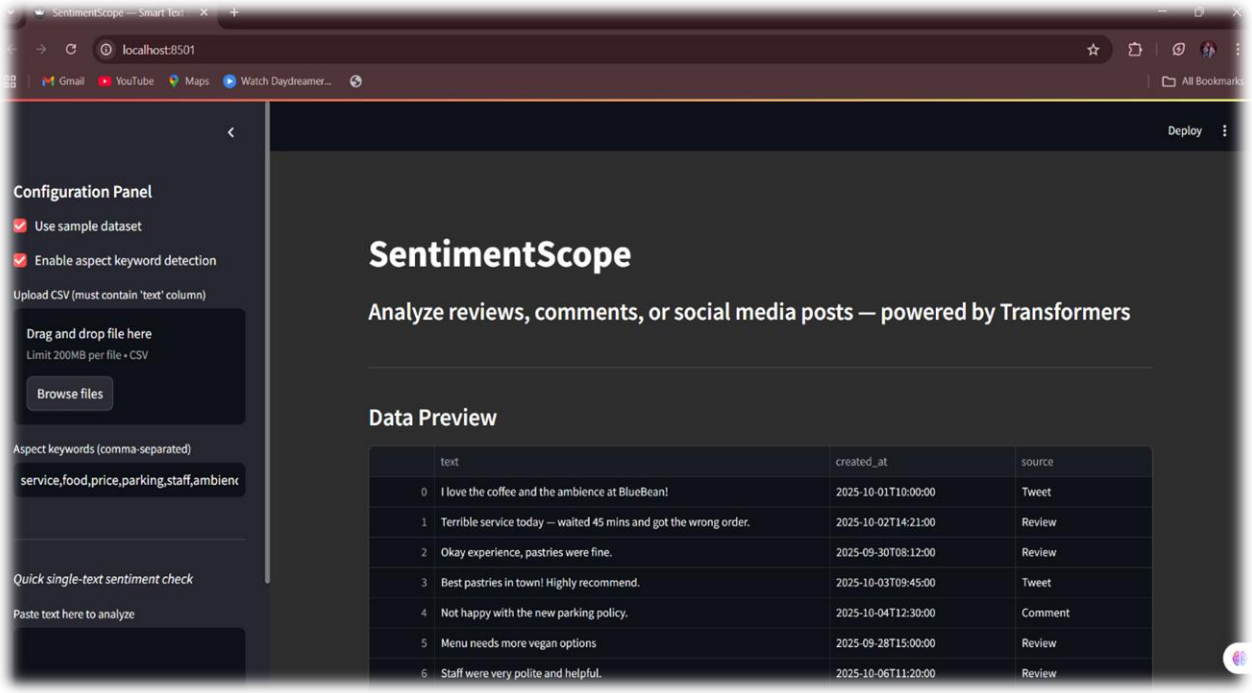

fig.2 Sentiment scope review.

3.2 Hybrid Model: Random Forest and LSTM

The analytical core of the platform is a hybrid model that explores the power of combining two distinct machine learning techniques[23]. Our model synergizes Random Forest and LSTM, effectively analyzing sentiment and predicting future trends based on labeled training data[24]. The LSTM component excels at interpreting the sequence and context of words in a review or post, while the Random Forest component acts as a robust and efficient classifier on the features extracted by the LSTM. This two-stage approach allows the model to capture both semantic meaning and classification patterns accurately. [10]

3.3 Implementation of Core Technologies and Tools.

The platform is built using a selection of powerful and industry-standard technologies:

- Python: A versatile programming language serves as the foundation for the entire system, chosen for its extensive libraries for data analysis and machine learning. [11]
- TensorFlow: This leading open-source framework is used for building, training, and deploying the machine learning models efficiently, particularly the LSTM component.
- Streamlit: An intuitive tool used to rapidly develop and deploy the interactive web application, allowing for the creation of a user-friendly dashboard with minimal effort

IV. APPLICATION INTERFACE AND RESULTS

The final output of the platform is an interactive dashboard designed for business users, translating complex data into clear, actionable insights. [13]

4.1 User-Friendly Dashboard Features

The application interface provides a simple yet powerful user experience[30]. The dashboard allows

users to easily upload their own data via CSV files and automatically generates dynamic sentiment graphs and visualizations[31]. This functionality democratizes data science, providing stakeholders with crucial insights for informed and timely decision-making without requiring technical expertise.

4.2 Insights from Predictive Analytics Platform

The platform's efficacy is demonstrated through its key performance metrics and the quality of insights it delivers.

- Accuracy: The system achieved 85% accuracy in its sentiment predictions across validation datasets.
- Visualization: It offers a comprehensive visualization of sentiment data trends, enabling users to monitor shifts in customer opinion over time. [14]
- Predictions: The platform provides insightful predictions for future sentiment trends, allowing for proactive business strategy and marketing adjustments.

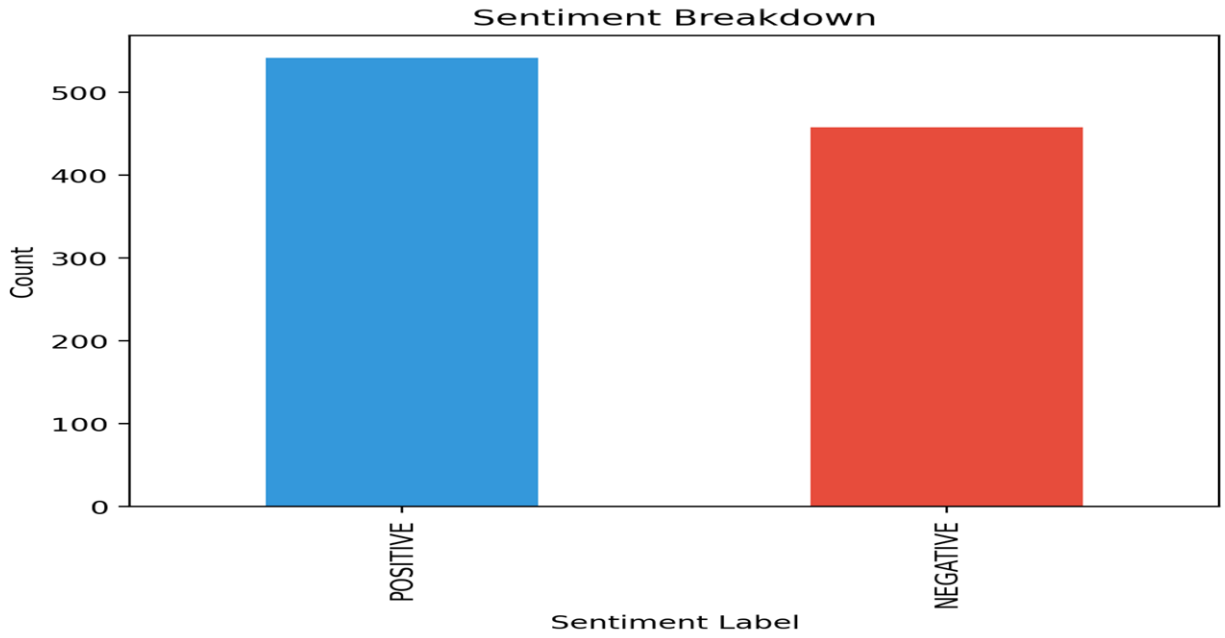

Fig.3 System sentiment label

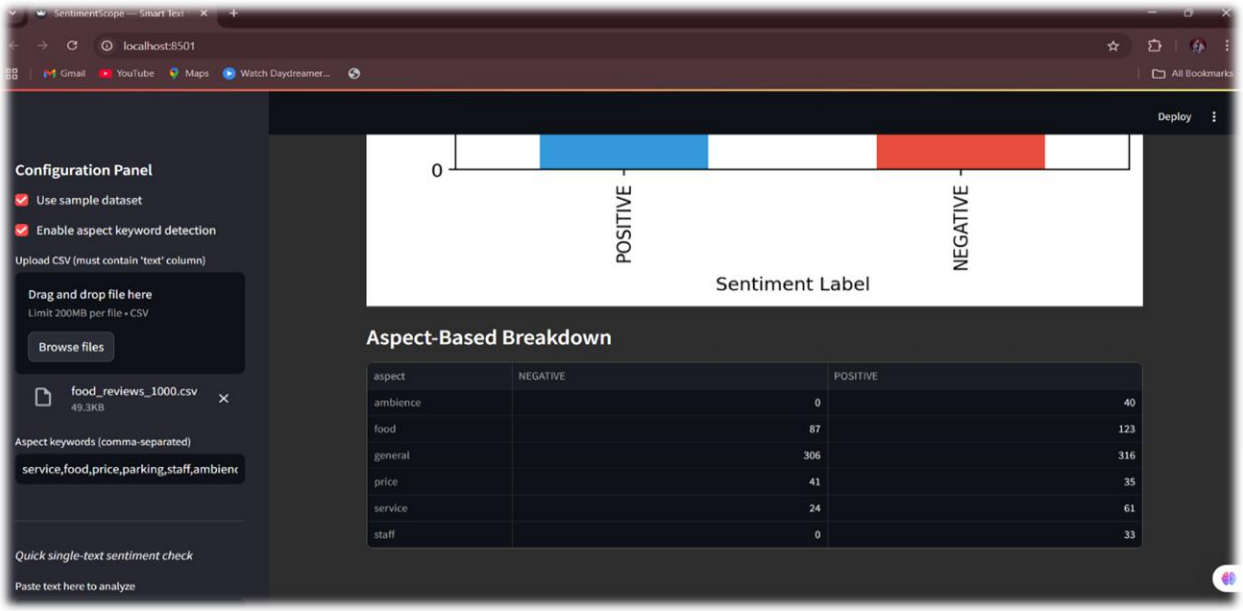

Fig.4 Aspect Based Breakdown

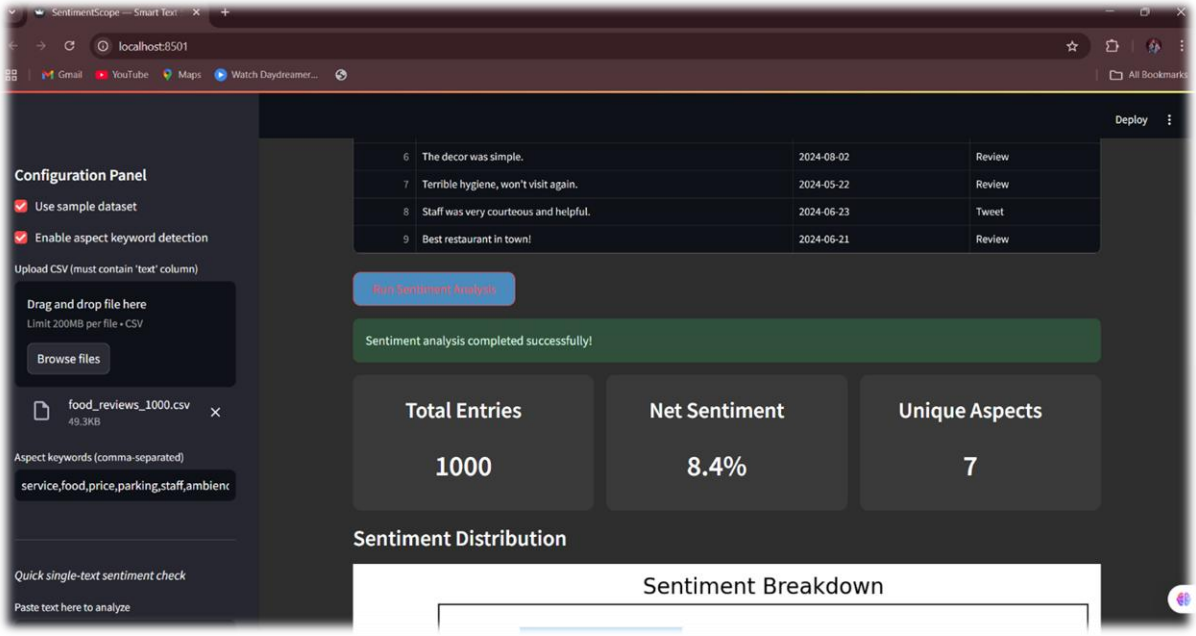

Fig.5 Accuracy Score

Overall, the system provides both insights and speed, enabling the rapid analysis of social media sentiment to facilitate data-driven decisions for businesses across multiple platforms. [15]

V CONCLUSION AND FUTURE WORK

5.1 Conclusion

This project successfully demonstrates the design and implementation of an advanced Predictive Analytics Platform that integrates Artificial Intelligence (AI) and data analytics to deliver actionable insights. The system effectively bridges the gap between unstructured social media data and meaningful business intelligence, transforming large volumes of raw text into structured, interpretable, and predictive information.

Through the application of natural language processing (NLP) and machine learning models, the platform enables organizations to analyse public sentiment, identify emerging trends, and forecast potential market movements. This predictive capability not only supports data-driven decision-making but also enhances strategic planning, customer engagement, and brand reputation management.

Overall, the project proves that **predictive analytics, when powered by AI**, can serve as a transformative tool for businesses across industries—helping them unlock hidden insights, anticipate consumer behaviour, and make more informed, timely, and intelligent decisions.

5.2 Future Enhancements for Predictive Analytics

While the current platform effectively demonstrates the potential of AI-driven predictive analytics, several enhancements can further elevate its performance, scalability, and real-world applicability.

- Voice and Speech Data Integration: Future versions of the system will incorporate advanced speech recognition and natural language understanding (NLU) modules. This will enable the platform to process voice recordings, customer service calls, video reviews, and other audio-based content. By analysing tone, emotion, and spoken sentiment, the system will provide deeper insights into customer feedback and engagement patterns.
- Real-Time Monitoring and Alerts: The next phase of development will focus on real-time sentiment tracking and trend analysis. This enhancement will allow continuous monitoring of social media feeds, news sources, and online discussions, offering instantaneous sentiment

alerts and visual dashboards. Such real-time insights will help organizations react swiftly to public opinion shifts, product feedback, and emerging trends.

- Multilingual and Cross-Cultural Support: To expand global reach, the platform will be enhanced to support multiple languages and cultural contexts. Incorporating multilingual NLP models will make it possible to analyse sentiments and trends across diverse regions, helping international brands understand local perspectives and tailor their communication strategies accordingly.

These enhancements aim to transform the platform into a more versatile, intelligent**,** and globally adaptive predictive analytics ecosystem, capable of handling multiple data formats, languages, and real-time scenarios. Ultimately, they will position the system as a comprehensive solution for businesses seeking to leverage AI for strategic foresight and competitive advantage.

REFERENCE

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 20151.

[2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv:1409.1556, 20142.

[3] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, “Object detection with deep learning: A review,” IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212–3232, 20193.

[4] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, “A review of semantic segmentation using deep neural networks,” International Journal of Multimedia Information Retrieval, vol. 7, no. 2, pp. 87–93, 20184.

[5] X. X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer, “Deep learning in remote sensing: A comprehensive review nd list of resources,” IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 4, pp. 8–36, 20175.

[6] H. Parikh, S. Patel, and V. Patel, “Classification of SAR and PolSAR images using deep learning: a review,” International Journal of Image and Data Fusion, vol. 11, no. 1, pp. 1–32, 20206.

[7] S. Chen and H. Wang, “SAR target recognition based on deep learning,” in International Conference on Data Science and Advanced Analytics (DSAA), 20147.

[8] L. Wang, A. Scott, L. Xu, and D. Clausi, “Ice concentration estimation from dual-polarized SAR images using deep convolutional neural networks,” IEEE Transactions on Geoscience and Remote Sensing, 20148.

[9] P. Wang, H. Zhang, and V. Patel, “SAR image despeckling using a convolutional neural network,” IEEE Signal Processing Letters, vol. 24, no. 12, pp. 1763–1767, 20179.

[10] N. Anantrasirichai, J. Biggs, F. Albino, P. Hill, and D. Bull, “Application of machine learning to classification of volcanic deformation in routinely generated InSAR data,” Journal of Geophysical Research: Solid Earth, 201810.

[11] L. Hughes, M. Schmitt, L. Mou, Y. Wang, and X. X. Zhu, “Identifying corresponding patches in SAR and optical images with a pseudosiamese CNN,” IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 784–788, 201811.

[12] K. Ikeuchi, T. Shakunaga, M. Wheeler, and T. Yamazaki, “Invariant histograms and deformable template matching for SAR target recognition,” in Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 1996, pp. 100–10512.

[13] V. U. Rathod, N. P. Sable, N. N. Thorat, and S. N. Ajani, "Deep Learning Techniques Using Lightweight Cryptography for IoT Based E-Healthcare System, " in 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1–5, doi:10.1109/CONIT59222.2023.10205808.

[14] A. P. Pimpalkar, N. N. Thorat, M. B. Gulame, D. A. Lokare, N. Kulal, and P. Khune, Partitions of Liver Cancer by Enhanced Feature Extraction and Mapping with Improved Transfer Developing Methods, " in 2024 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2024, pp. 1–5, doi: 10.1109/ESCI59607.2024.10497258.

[15] N. N. Thorat, P. Dhotre, N. Kulal, A. P. Pimpalkar, V. S. Patil, and S. M. Kolekar, "Tamper-forgery Image Detection in Security Digital

Image: A Review," in 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Pune, India, 2024, pp. 1–5, doi: 10.1109/ICBDS61829.2024.10837567.

[16] N. N. Thorat, N. M. Kulal, K. V. Baburao, S. A. Hirve, S. Kolekar, and R. Sangore, "AI based Security in computer vision and machine learning algorithms for hand sign recognition, " in 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Pune, India, 2024, pp. 1–5, doi: 10.1109/ICBDS61829.2024.10837275.

[17] S. A. Hirve, N. N. Thorat, R. L. Arya, V. R. Neharkar, and A. G. Hake, "GeoDrishti: Automatic Road Extraction and Alert Generation for New Roads from Satellite Imagery, " in 2024 International Conference on Progressive Innovations in Intelligent Systems and Data Science (ICPIDS), Pattaya, Thailand, 2024, pp. 469–475, doi: 10.1109/ICPIDS65698.2024.00079.

[18] S. M. Kolekar, A. P. Pimpalkar, R. P. More, S. A. Hirve, M. B. Gulame, and N. N. Thorat, "Digital Image Tamper-Forgery Detection in Security, " in 2024 2nd International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC), Hyderabad, India, 2024, pp. 41–46, doi: 10.1109/ICMACC62921.2024.10894413.

[19] N. N. Thorat, A. Singla, and T. Dhaigude, “Sharing Secret Colour Images with Embedded Visual Cryptography Using the Stamping Algorithm and OTP Procedure”, IJRITCC, vol. 11, no. 6, pp. 63–70, Jul. 2023. N. N. Thorat, et al., "Cotton Plants Diseases Detection Using CNN, " International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 10, pp. 294–299, 2023, doi: 10.17762/ijritcc.v11i10.8492.