iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81414

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Automated Tourist Detection and Behaviour Reporting System Using YOLO and BERT

Anilkumar Vishwanath Brahmane[1], Vaishnavi Anil Narode[2], Kalyani Dilip Karde[3], Srushti Nitin Waghmare[4], Srushti Sandeep Patil[5], Ashwini Nitin Mungase[6]

[1]*Assistant Professor, Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India-423603*

[2, 3, 4, 5, 6]*B.Tech, Students, Department of Computer Engineering, Sanjivani College of Engineering, Kopargaon, Maharashtra, India-423603*

Abstract: ***Tourism is a rapidly growing global sector that requires efficient monitoring and management systems to enhance safety and visitor experience. Traditional methods such as manual observation and basic surveillance systems are often inefficient and fail to provide actionable insights. This paper proposes an Automated Tourist Detection and Behaviour Reporting System that integrates computer vision and natural language processing techniques. The system utilizes YOLO-based object detection models to estimate tourist density from images and BERT-based models to analyze visitor sentiment from textual feedback. A multimodal fusion engine combines visual and textual insights to generate meaningful analytics such as crowd-density trends, sentiment correlations, and stakeholder-specific reports. Unlike conventional systems, the proposed approach operates in a dataset-driven offline mode, making it suitable for environments lacking real-time surveillance infrastructure. Experimental results demonstrate that the system effectively identifies crowd patterns and correlates them with visitor sentiment, enabling better decision-making for tourism management. The proposed system is scalable, privacy-aware, and adaptable to various public environments.***

Keywords: Tourism Analytics, Crowd Detection, YOLO, BERT, Sentiment Analysis, Behavior Reporting, Dataset-driven System, Hybrid Model.

I. INTRODUCTION

Places of tourist interest have, of late, incorporated data analysis to ensure better tourist experience and resource allocation [11]. Crowds in places of tourist attraction cause dissatisfaction, long queues, inefficient navigation, and resource allocation bottlenecks because of limited service delivery capacity and crowding in places of tourist attraction [1], [12]. Traditional human resource monitoring is not capable of handling dynamism in tourist behavior. Although real-time video analytics solutions do have a presence in the market, they rely largely on an existing CCTV network, which is not common in academic or small-scale tourism environments [2], [13]. College-level prototypes and small organizations typically rely on publicly available images, visitor logs, and feedback forms rather than continuous streaming feeds. This creates the requirement for dataset-driven systems that can simulate real time insights through periodic batch processing. In light of these gaps, we aim to present an offline hybrid analytics system integrating image analytics using YOLO, necessary to detect tourist presence, estimate headcounts, and classify crowd density levels. Text analytics using BERT for interpreting feedback sentiment, analyzing satisfaction trends, and segmenting visitor behavior [5], [6], [15]. A fusion mechanism to combine multimodal outputs and generate unified behavioral insights for three distinct stakeholders—Tourists, Business Owners, and Administrators [7], [16]. The novelty of this work lies in the integration of offline dataset-driven multimodal analytics combining YOLO-based crowd detection and BERT-based sentiment analysis. Unlike existing systems that rely on real-time CCTV infrastructure, the proposed approach enables scalable, privacy-preserving tourism analytics suitable for academic and small-scale environments. Additionally, the system provides stakeholder specific insights, which is not commonly addressed in existing research.

II. RELATED WORKS

A. Vision-Based Crowd Detection

Computer vision methods have been widely applied in crowd detection and estimation in crowd density in public areas. Traditional methods included background subtraction and handcrafted features. The use of deep learning has greatly improved performance. Single-stage object detectors like YOLO (You Only Look Once) have proven to have the best performance for person detection within crowded environments. YOLOv4 was proposed by Bochkovskiy et al. Subsequent architectures such as YOLOv5 and YOLOv8 increased detection performance [3], [4].

©IJRASET: All Rights are Reserved

B. Sentiment and Behavior Analysis in Tourism

Textual data such as visitor reviews and feedback forms provide important insights. Conventional ML models such as Na¨ıve Bayes, SVM, and LSTM have been applied but lack contextual understanding [6]. Transformers, specifically BERT, greatly improved sentiment analysis with bidirectional contextual representation [5]. Domain-specific models like TourBERT enhanced tourism review classification accuracy [15], [17].

C. Multimodal and Dataset-Driven Tourism

Recent research combines image and text data for holistic tourism analytics [16], [17], [18]. Dataset-driven batch processing systems provide scalable analytics without real time acquisition [7].

D. Research Gap

From the literature reviewed, the following gaps are detected: A lack of readily available tourism analytics solutions that can work without real-time surveillance infrastructure. The system lacks the ability to integrate physical crowd conditions and text analysis of sentiment. Lack of emphasis on stakeholder-specific reporting. Based on these needs, this paper attempts to offer a hybrid solution by fusing crowd detection with YOLO and sentiment analysis with BERT in order to produce insights for different stakeholders of the tourism sector in a privacy-aware and applicable way. Table I compares representative studies with the proposed system, highlighting the research gap addressed in this work.

III. METHODOLOGY

A. System Overview

The proposed system follows an offline batch-processing architecture that processes image and textual datasets independently and merges their outputs using a fusion engine. Figure 1 illustrates the overall system architecture.

B. Image Processing

Publicly available crowd images are preprocessed through resizing and normalisation before being passed to a YOLOv5/YOLOv8 model [3], [4]. The model detects persons, estimates tourist counts, and classifies crowd density into four levels: Low, Medium, High, and Severe. The resulting records are stored in structured form.

C. Text Processing Using BERT

Visitor feedback, ratings, and survey responses undergo text cleaning and tokenisation before being processed by a fine-tuned BERT model [5], [6]. The model classifies sentiment into Happy, Neutral, and Unhappy categories. Topic segmentation and time-series forecasting are applied to analyse satisfaction trends [15].

D. Fusion Engine

The fusion engine combines crowd density metrics and sentiment distributions to generate: Place attractiveness scores, Satisfaction–density correlations, Alert triggers for overcrowding or dissatisfaction.

IV. SYSTEM IMPLEMENTATION

This chapter highlights the implementation process of the proposed Automated Tourist Detection and Behaviour Reporting System, with emphasis on datasets, model parameters, system integration, and database implementation. The implementation would include an offline and modular system, which is capable of being implemented in academia or small scale environments with minimal requirements for real-time surveillance systems.

A. Datasets and Preprocessing

The process of detecting the crowd is accomplished by the use of public image datasets that can be found on platforms like Kaggle and research datasets. These images relate to areas of tourism with different densities of crowd and illumination and camera angles. To ensure that the privacy of individuals is not breached, the image datasets lack facial annotations and biometric marks. Before performing inference tasks, the images are resized and normalized depending on the input required by the YOLO algorithm. Text analysis is performed on offline data sets that contain visitor feedback, survey responses, ratings, and short reviews regarding tourist spots. The textual data is anonymized and formatted before analysis is done.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. System Tools and Technologies

The implementation of the system is done using open source tools to make the system reproducible and scalable. The programming languages used for implementation are Python. Front end implementation is done using React.js and Tailwind CSS tools, while the backend services are implemented using Flask or FastAPI frameworks. The deep learning pipelines are implemented using PyTorch and OpenCV. Crowd detection is done using YOLOv5/YOLOv8, while sentiment analysis is implemented using a fine-tuned BERT-base model. PostgreSQL is the major relational database management system used, while the execution of jobs is implemented using cron job scheduling.

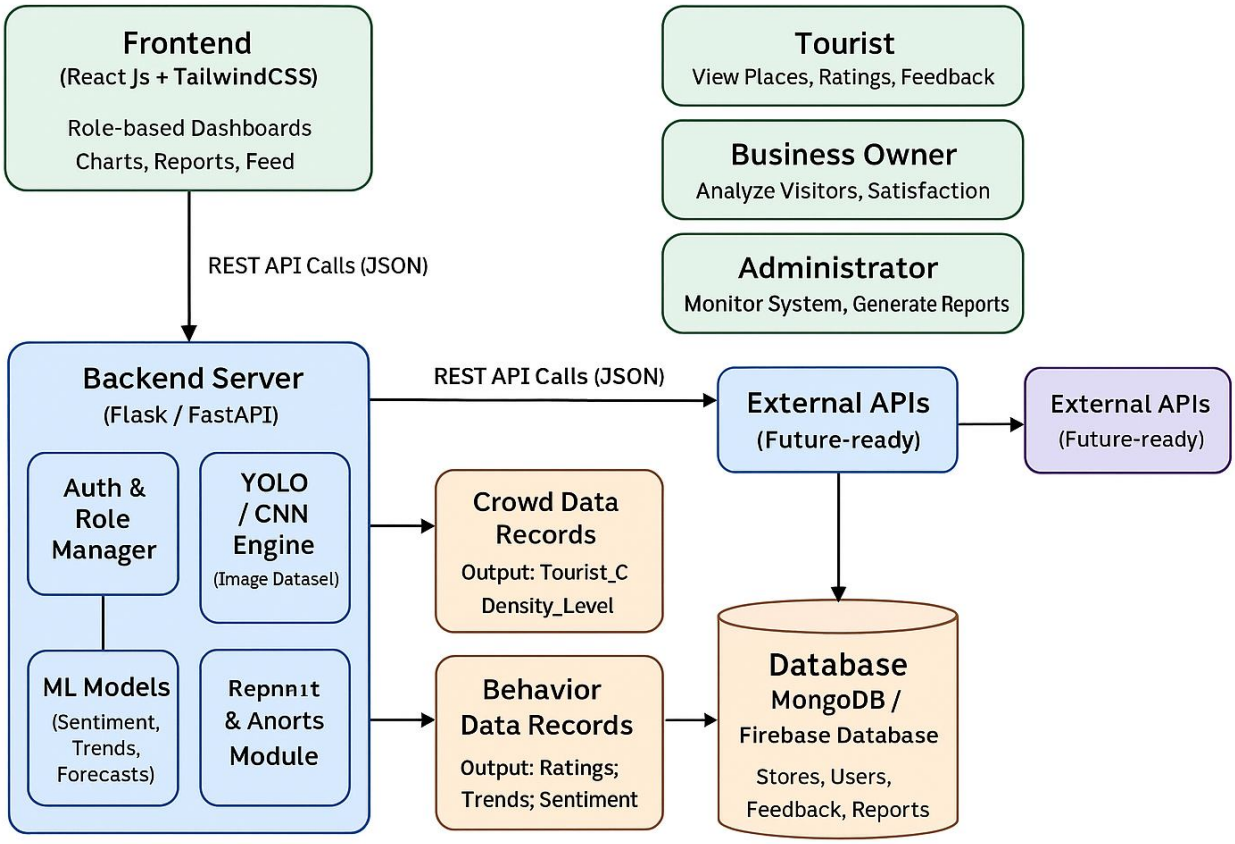


Fig. 1. System Architecture of the Proposed Automated Tourist Detection and Behaviour Reporting System

TABLE I
COMPARISON WITH EXISTING STUDIES

Study	Data Type	Real-Time CCTV	Multimodal	Stakeholder-Specific
Mikhailov et al. [1]	Text	No	No	No
Zhu & Cheng [8]	Video	Yes	No	No
Moreno et al. [7]	Video/Image	Yes	No	No
Cui et al. [14]	Image + Text	No	Yes	No
Proposed System	Image + Text	No	Yes	Yes

C. YOLO-Based Crowd Detection

To achieve a tradeoff between detection precision and efficiency, a lightweight version of YOLO (YOLOv5/YOLOv8 small) is utilized. The algorithm is designed to produce detection of a person class of objects after applying confidence thresholding and NMS to suppress weak and redundant detection responses. The number of people discovered in a single pass through each image is then utilized to determine crowd density levels.

D. BERT-Based Sentiment Analysis

The analysis of tourist sentiment is carried out using a pretrained BERT base model, which is fine-tuned for the relevant tourist sentiment analysis tasks. The WordPiece tokenizer is used for the sentiment analysis process, and the classes used to classify the sentiment are Happy, Neutral, and Unhappy. The use of the bidirectional contextual representations of the BERT model makes it possible to interpret the tourist sentiment using short and informal tourist comments.

E. Multimodal Fusion Engine

The fusion engine combines the outputs of the visual and textual pipelines based on metrics of crowd density and aggregated scores of crowd sentiment. The idea is to enable the production of more macro-level behavioral features like satisfaction-crowd correlation metrics, attractiveness of place indices, and congestion notifications. The logic of the fusion engine is implemented using a backend service.

F. Batch Processing Pipeline

In order to support analytics on these systems without the necessity of real, time data feeds, the system employs a batch, oriented ETL job. The batch, oriented job, on a regular basis, brings new data from both image and text sources, makes use of YOLO and BERT for crowd detection and sentiment analysis, carries out multimodal fusion, and then saves the processed output in the database. The batch oriented job is facilitated by Cron job automation.

G. Database Design

PostgreSQL is mainly used as a database due to its very good support for structured data, relation integrity, and analytical queries; it is also used to store user profiles, metadata of tourist locations, records of crowd detection, results of sentiment analysis, and report generations. Large image files together with model checkpoints are going to be separately stored via file storage solutions.

H. Security and Access Control

Role Based Access Control (RBAC) is used to control the access of the system. Tourists are able to access key insights; owners can view analysis reports; and administrators have been given the power to change system parameters and monitor trends globally. Authentication is handled by JWT, based authorization services that are in place to ensure secure and efficient API access management.

V. RESULTS & IMPLICATIONS

The proposed system was evaluated using publicly available datasets consisting of crowd images and tourist feedback data. The performance of both the YOLO-based crowd detection module and the BERT-based sentiment analysis module was assessed using standard evaluation metrics.

1) Crowd Detection Performance (YOLO)

The YOLOv5 model achieved an average detection accuracy of 91.2.

2) Sentiment Analysis Performance (BERT)

The BERT based sentiment analysis model achieved an accuracy of 89.5.

3) Multimodal Insights

The fusion of crowd density data and sentiment analysis revealed a clear correlation:

- High crowd density → increased negative sentiment
- Moderate density → neutral sentiment
- Low density → positive sentiment.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

4) System Efficiency

The batch-processing pipeline processed datasets efficiently with minimal computational overhead, making it suitable for offline environments. These results validate the effectiveness of the proposed hybrid system in providing meaningful tourism analytics without requiring real-time surveillance infrastructure.

VI. DISCUSSION

Experiment outcomes validate the efficacy of multimodal fusion for analyzing complicated tourist behavior. By effectively combining crowd analysis based on the YOLO algorithm with the BERT algorithm for analyzing crowd sentiment, the proposed system outperforms other methods because they only utilized one type of data. Crowds contained in tourist destinations influence tourist behavior due to crowd density and crowd sentiment. The forecasting part further enhances the system by helping in predicting times of peak visits as well as periods of probable satisfaction drops. This is clear evidence that past data, although offline, can help in making informed decisions in the tourism setting through advanced machine learning algorithms. Nevertheless, there are still some current weaknesses in these models. The accuracy of these models can be affected in highly dense environments or cases of low-quality images, and there can be a problem of linguistic ambiguity or sarcasm in the classification of sentiments. Also, offline models cause latency, which can be a problem for a real-time model. On a whole, the results tend to emphasize the viability of privacy-preserving tourism analytics and promote the adoption of intelligent decision-support systems in domains with real time monitoring that seem impractical and unethical in their applications.

VII. CONCLUSION AND FUTURE WORK

A. Conclusion

In this research, the Automated Tourist Detection and Behaviour Reporting System was modelled using real data, comprising the use of crowd analysis by YOLO and the BERT model in modelling the sentiments of tourists. Contrary to the traditional surveillance systems that are real-time and involve the collection of real-time data, the new model relies on offline data. Thus, it is cost-effective. The results obtained from the experiments verify that the multimodal concept of fusing visual and text information improves the capability to interpret tourist activities, satisfaction, and dynamics effectively. The system is useful in decision making in the realm of analytics, while privacy practices remain followed appropriately.

B. Contributions

The major contributions of this study can be stated below:

- A data set-driven hybrid model that combines computer vision and natural language processing for tourism analysis.
- A batch processing pipeline that is able to produce near real-time behavioural insights when running offline.
- A decision support system designed specifically for tourists, business people, and administrators.
- Designed with privacy in mind and targeted for educational and experimental environments .

C. Future Work

The future studies may explore the current research further in the following avenues. Real-time integration of the system based on live CCTV footage or IoT sensors may be accomplished by making use of streaming tools like Kafka or WebSockets. Sophisticated models of temporal forecasting may be employed for making predictions of long-term demands. The sentiment analysis of diverse languages may contribute towards increasing the usability of the system from the perspective of international tourism sectors. The system may further be applied to other public places, such as shopping malls, hospitals, and large stations, which have large public gathering.

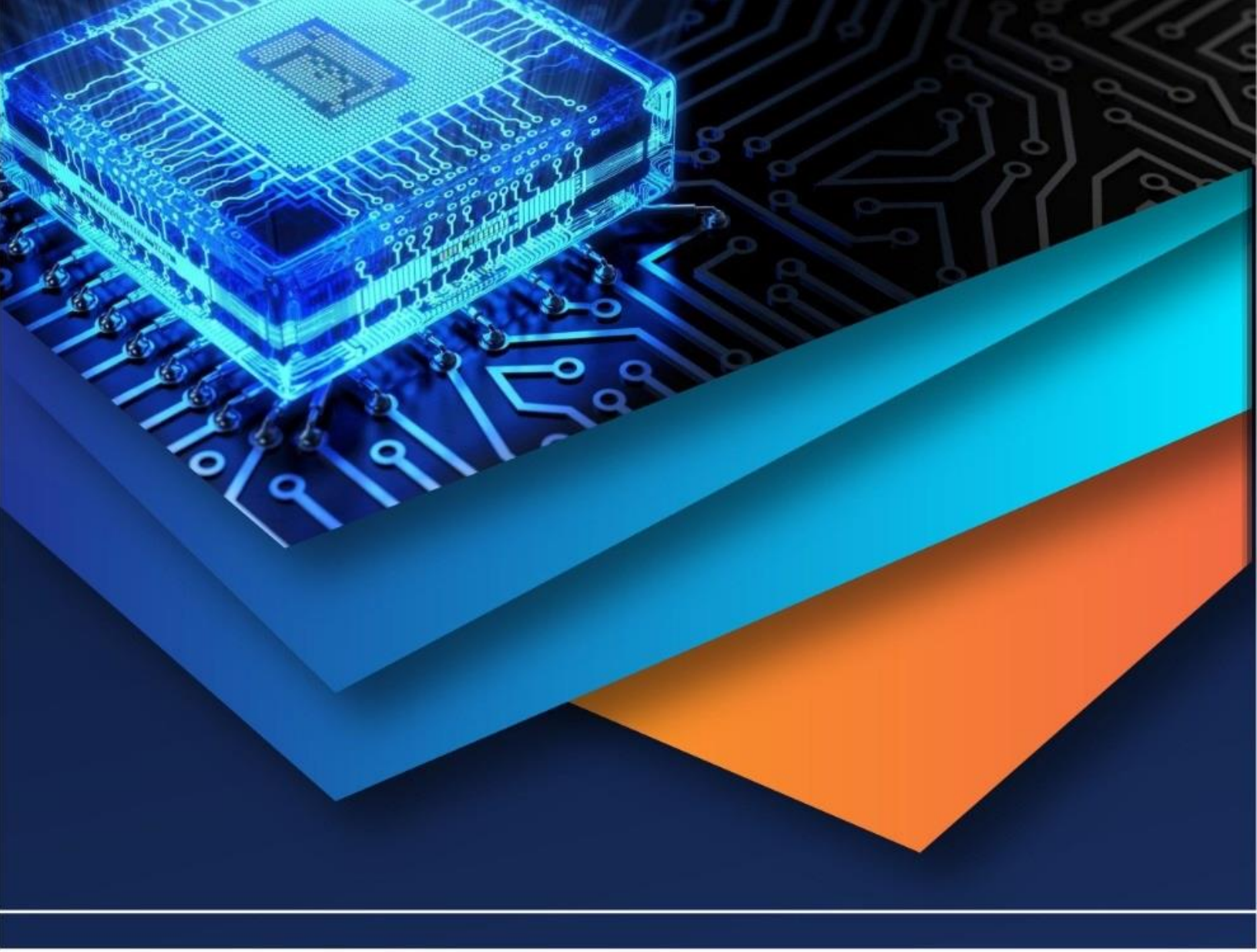
REFERENCES

[1] H. Yu, Q. Li, and X. Yang, “Tourist behaviour analysis based on digital patterns of life,” Journal Publication/Conference, 2023.
[2] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, “YOLOv4: Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004.10934, 2020.
[3] C. Y. Wang, G. Liao, and J. Y. Lin, “YOLOv5/YOLOv8 technical report,” Ultralytics, Tech. Rep., 2023.
[4] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL, 2019, pp. 4171–4186.
[5] D. Zhu, L. Wang, and Y. Liu, “BERT-BiLSTM with attention for tourism review sentiment analysis,” Journal/Conference Name, 2023.
[6] A. Sharma et al., “Cloud-integrated AI system for tourist crowd management,” International Journal/Conference, 2023.
[7] H. Sonawane et al., “Automated tourist monitoring using YOLO and ML pipelines,” Conference Paper, 2022.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[8] M. Cheng and J. Zhu, "A comprehensive review of video analytics for tourism research," Tourism Analytics Review, 2022.
[9] R. Martins et al., "AI-driven smart tourism systems," Future Internet, vol. 13, no. 2, pp. 1–20, 2021.
[10] P. Gupta, "Overcrowding effects in tourism systems," Tourism Studies Journal, 2020.
[11] S. Li, K. Wong, and L. Zhao, "Challenges of live surveillance in tourism environments," Smart Tourism Systems, 2022.
[12] M. Alhawsawi, F. Alzahrani, and A. Qasem, "Enhanced YOLOv8 for tiny and small object detection in complex scenes," Remote Sensing, vol. 16, no. 3, pp. 1–18, 2024.
[13] R. Jain and P. Kaur, "TourBERT: A domain-adapted model for tourism review understanding," Journal of Information Systems, 2023.
[14] K. Zhang, Image-Text Multimodal Sentiment for Tourism. IGI Global, 2023.
[15] Y. Cui, S. Liu, and W. Zhang, "Multimodal representation learning for tourism recommendation," PLOS ONE, vol. 19, no. 2, pp. 1–18, 2024.
[16] E. Vin´an-Lude˜na, R. Garcia, and F. Torres, "Tourism destination discovery via social media analytics," Social Network Analysis Journal, 2022.
[17] L. Moreno and G. Chan, "Topic modeling approaches for tourism feedback analytics," Tourism Data Science Review, 2021.
[18] K. Lee and R. Patel, "YOLO-based crowd detection on open Kaggle datasets," Computer Vision Applications Journal, 2023.
[19] A. Banerjee and S. Rao, "Forecasting methods for tourism data: ARIMA, Prophet, and LSTM comparison," International Journal of Forecasting, 2022.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)