Involvement International Journal of Business Vol. 2 No. 3, 2025
eISSN: 3032-485X
DOI: https://doi.org/10.62569/iijb.v2i3.137
Received: January 2025/ Revised: July 2025/ Accepted: July 2025

Involvement International Journal of Business
https://ejournal.agungmediapublisher.com/index.php/iijb

Advancing Automotive Business Strategy through Multimodal Aspect-Based Sentiment Analysis Using SSLU-GRU and YOLO

Prapullakumar Gowtham Pyate[1]*, Balaji Srinivasan[2]

[1]*Department of Management studies, Interdisciplinary, SCSVMV, Kanchipuram, 631561, India*
[2]*Professor, HOD-Management Studies, SCSVMV, Kanchipuram, 631561, India*

Abstract. Sentiment analysis (SA) has become a key tool in understanding consumer feedback in the automotive industry. However, most existing models are limited to unimodal data and fail to capture fine-grained, aspect-level sentiments from multimodal sources such as text, images, and video. Additionally, privacy concerns related to user-generated content remain under-addressed. This study proposes a novel Multimodal Aspect-Based Sentiment Analysis (MASA) framework that integrates textual, visual, and video data for business decision-making in the automotive sector. The framework includes a BERT-based aspect dictionary for extracting domain-specific features, SCV-YOLOv5 for object segmentation in images and videos, and a GRU model enhanced with the Sinu-Sigmoidal Linear Unit (SSLU) activation function for sentiment classification. A K-Anonymity method augmented by Kendall's Tau and Spearman's Rank Correlation is employed to protect user privacy in sentiment data. The framework was evaluated using the MuSe Car dataset, encompassing over 60 car brands and 10,000 data samples per brand. The proposed model achieved 98.94% classification accuracy, outperforming baseline models such as BiLSTM and CNN in terms of Mean Absolute Error (0.14), RMSE (1.01), and F1-score (98.15%). Privacy-preservation tests also showed superior performance, with a 98% privacy-preserving rate and lower information loss than traditional methods. The results demonstrate that integrating multimodal input with deep learning and privacy-aware techniques significantly enhances the accuracy and reliability of sentiment analysis in automotive business contexts. The framework enables better alignment of consumer feedback with strategic decisions such as product development and targeted marketing.

Keywords: Multimodal Sentiment Analysis; SSLU-GRU; YOLOv5; K-Anonymity; Automotive Business; Aspect-Based Sentiment

1. Introduction

The automotive industry is undergoing rapid transformation, driven by technological innovation, changing consumer expectations, and an increased emphasis on environmental sustainability (Ghadge et al., 2022; D. Lin et al., 2018; Llopis-Albert et al., 2021; Lukin et al., 2022; Xu et al., 2021). In parallel, the proliferation of user-generated content especially in the form of online reviews, social media posts, and video

*Corresponding author's email: gowtham.prapullakumar@gmail.com, Telp.: -

testimonials has created new opportunities for businesses to tap into public sentiment for strategic decision-making (Guo et al., 2021).

Sentiment Analysis (SA), also known as opinion mining, is a branch of Natural Language Processing (NLP) that focuses on identifying, extracting, and categorizing emotional tone within text (Nazir et al., 2022; Tan et al., 2020). SA provides insights into consumer opinions and market trends, supporting both operational and strategic decisions (Maitama et al., 2020). However, most conventional SA approaches are limited to textual data and struggle to capture nuanced emotions particularly in domains like the automobile industry where multimodal reviews (combining text, images, and video) are prevalent.

Multimodal Aspect-Based Sentiment Analysis (MASA) has emerged as a more advanced alternative, combining data modalities such as text and visuals to provide deeper, aspect-specific insights (Ligthart et al., 2021; Xu et al., 2021). MASA is particularly effective in scenarios where users express sentiments both linguistically and visually such as describing a car's interior while also sharing images or video clips (Bayoudh et al., 2022; Jena, 2020; Xiao et al., 2021). These models often rely on deep learning and machine learning frameworks (Alshuwaier et al., 2022; Barua et al., 2023), which are adept at learning complex patterns from diverse datasets.

Despite growing interest, several studies in the automotive sentiment space still exhibit limitations. Traditional sentiment analysis techniques frequently rely on big data and social media inputs but often fail to link reviews accurately with specific car models or features (Shoumy et al., 2020). Predefined dictionaries also limit semantic depth, while noise in textual and visual data affects the reliability of outputs (Motz et al., 2022). Moreover, privacy concerns surrounding user-generated content have not been sufficiently addressed in existing frameworks.

Several recent works highlight these gaps. Guo et al. (2021) underline the challenges in multimodal data alignment and the lack of robust datasets. Zhou et al. (2021) and Zhao et al. (2021) emphasize the limitations in sentiment polarity labeling and the sensitivity of models to image/text noise. Buscemi and Proverbio (2024) shows that even advanced language models like LLaMA2 or ChatGPT still face difficulties with fine-grained, aspect-level sentiment classification. Meanwhile, Zhang et al. (2020) stress the need for better fusion of external signals (e.g., macroeconomics or audio-visual data), while Lin et al. (2021) note the absence of explicit semantic representations.

To fill these gaps, this study proposes a novel and privacy-aware framework for Multimodal Aspect-Based Sentiment Analysis, specifically tailored for automotive reviews. The system integrates an aspect dictionary built on BERT, vehicle and component detection using YOLOv5, and sequential sentiment classification using GRU enhanced with the Sinu-Sigmoidal Linear Unit (SSLU) activation. It also embeds privacy protection mechanisms using K-Anonymity, refined by Kendall's Tau and Spearman's Rank Correlation.

This research aims to develop a Multimodal Aspect-Based Sentiment Analysis (MASA) framework capable of accurately processing and analyzing car reviews in various formats, including text, images, and video. To enhance sentiment classification, the model integrates the efficiency of the SSLU-GRU algorithm with a BERT-based aspect dictionary, enabling a more nuanced understanding of user opinions. Furthermore, the framework incorporates a tailored anonymization approach based on K-Anonymity to ensure user privacy throughout the data analysis process. By improving the alignment between reviews and specific car models, the proposed system is expected to provide actionable

insights for car manufacturers and marketers, supporting more informed and strategic business decisions.

2. Methods

This study proposes a robust and privacy-aware framework called MASA (Multimodal Aspect-Based Sentiment Analysis) designed to support data-driven business decision-making in the automotive sector. The methodology comprises multiple integrated stages: data collection, preprocessing, keyword extraction, dictionary creation using BERT, image and video segmentation, vehicle classification, sentiment analysis, and user privacy protection. These components work together to extract fine-grained sentiment insights from multimodal car reviews.

The process begins with collecting multimodal data from user-generated content, including textual reviews, car images, and video recordings. These data are processed separately at first to retain modality-specific features. In the case of text, the reviews typically express sentiments regarding aspects such as performance, comfort, fuel efficiency, and safety. Once gathered, the text undergoes preprocessing, which includes tokenization, stopword removal, and stemming. The raw review is first broken down into individual words (tokens) represented as:

$$T = \{t_1, t_2, \dots, t_{n\}}$$

After tokenization, non-informative common words (stopwords) are removed to enhance the signal-to-noise ratio in the analysis:

$$T_{\{clean\}} = T - Stopwords$$

Next, stemming is applied to reduce words to their base forms, yielding:

$$S = \{s_1, s_2, \dots, s_{n\}}$$

These processed terms are then passed to the keyword extraction stage, which employs multiple techniques. Aspect-Based Keyword extraction (e.g., using TF-IDF or RAKE), Named Entity Recognition (NER), Morphological Segmentation (MS), Hypernym/Hyponym detection using WordNet, and Word Sense Disambiguation (WSD). The goal is to extract both domain-specific terms and generalizable structures from the text for later matching with visual and sentiment contexts.

The extracted keywords are embedded into vector representations using a fine-tuned Domain-Specific BERT (DS-BERT) model. Each tokenized keyword is converted into an embedded vector:

$$E = BERT(T)$$

DS-BERT incorporates self-attention mechanisms to learn relationships between words. The self-attention score (SAS) is computed as:

$$SAS = softmax\left(\backslash frac\{QK^T\} \left\{ \sqrt{\{d_k\}} \right\} \right)$$

where Q, K, and V are the query, key, and value matrices, and dkd_kdk is the dimensionality of the key vectors. The outputs from BERT are then used to build a numerical dictionary of aspect-based sentiment scores.

In parallel, videos from reviews are preprocessed by converting them into individual frames, removing duplicates, and extracting keyframes that contain meaningful content. These frames are treated similarly to images and passed to the segmentation stage. Both still images and video frames are analyzed using an enhanced object detection model: SCV-YOLOv3. This version of YOLO incorporates Scaled Correlation Vectors (SCV) to improve detection accuracy, particularly for car parts with irregular aspect ratios. The object detection output includes bounding boxes BBB, where each box is defined by parameters such as location and dimensions:

$$x = \sigma(t_x) + c_x, \qquad y = \backslash sigma(t_y) + c_y$$

$$w = p_w \cdot e^{\{t_w\}}, \qquad h = p_h \backslash cdot\, e^{\{t_h\}}$$

where t_x, t_y, t_w, t_h, are the predicted box parameters, p_w, p_h, are the anchor box dimensions, and σ denotes the sigmoid function. Class probabilities are then computed for each box:

$$P_{\{class\}} = softmax(s)$$

The segmented car parts are then described using Gray Level Co-occurrence Matrix (GLCM) features, such as Contrast, Dissimilarity, and Correlation, to generate numerical descriptors for classification.

The classified car parts and features are passed to a GRU (Gated Recurrent Unit) neural network enhanced with SSLU (Sinu-Sigmoidal Linear Unit) activation. The SSLU activation provides smoother gradient flow and faster convergence compared to traditional functions like tanh or ReLU. The GRU processes temporal patterns through the reset and update gates as follows:

$$r_t = \sigma\left(W_r x_t + U_r h_{\{t-1\}}\right), \qquad z_t = \backslash sigma\left(W_z x_t + U_z h_{\{t-1\}}\right)$$

A candidate hidden state is computed using SSLU:

$$\tilde{}\{h\}_t = SSLU\left(W_h x_t + U_h\left(r_t \odot h_{\{t-1\}}\right)\right)$$

The final hidden state output is given by:

$$h_t = (1 - z_t) \odot h_{\{t-1\}} + z_t \odot \tilde{}\{h\}_t$$

The predicted car models—e.g., BMW, Hyundai, Kia—are then matched with textual reviews using Jaro-Winkler similarity to ensure alignment. This is calculated as:

$$JW(s_1, s_2) = Jaro(s_1, s_2) + \left(l \cdot p \cdot \left(1 - Jaro(s_1, s_2)\right)\right)$$

where l is the length of common prefix, and ppp is a scaling factor. Only reviews with high similarity to classified models proceed to sentiment analysis, again handled by SSLU-GRU using scores from the DS-BERT-generated dictionary. Sentiments are categorized by aspect, with positive sentiments indicating strong features (e.g., fuel economy), while negative ones highlight weaknesses (e.g., noise or maintenance cost).

To ensure user data privacy, the system integrates K-Anonymity enhanced by a KSQ (Kendall's Tau and Spearman's Rank Correlation)-based quasi-identifier selection. Sensitive personal information is suppressed:

$$Q' = Suppress(Q)$$

and generalized where necessary:

$$G = Generalize(A)$$

allowing users to share reviews without fear of identity exposure. This ensures that the analysis process remains ethical and trustworthy while preserving the value of the data for business intelligence.

3. Results and Discussion

The performance of the proposed MASA framework was evaluated using the MuSe Car dataset, which is specifically curated for multimodal sentiment analysis in the automotive domain. This dataset includes a diverse collection of car reviews and associated metadata such as manufacturer, model, and car type. It provides a reliable basis for testing multimodal systems because it integrates textual feedback, images, and even video inputs relevant to real consumer experiences. In this study, the system was implemented and tested using the Python programming language, taking advantage of its deep learning and computer vision libraries.

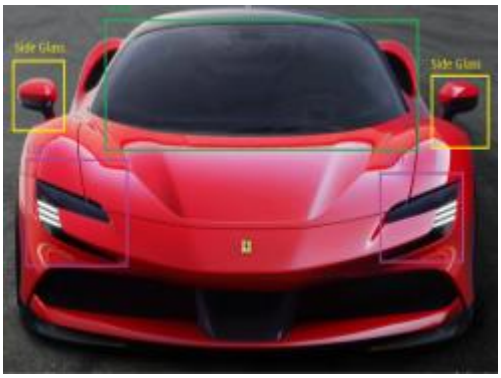
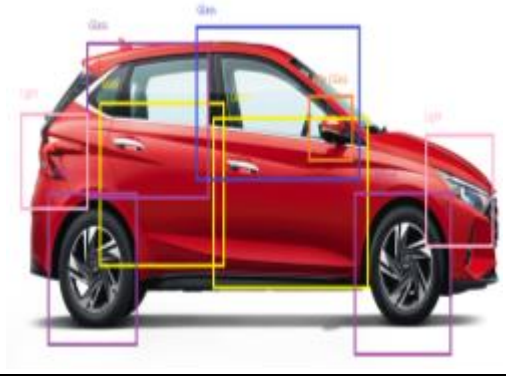
To ensure balanced evaluation, data from 62 well-known car brands were utilized. These include widely recognized names such as BMW, Ferrari, Hyundai, Kia, and Volkswagen. For each brand, 10,000 data samples were extracted, resulting in a robust training corpus. The dataset was split using an 80:20 ratio, with 80% used for training the model and 20% reserved for testing its performance. This split ensures that the model is trained with a diverse range of inputs while maintaining an unseen subset for fair evaluation.

One of the standout capabilities of the MASA (Multimodal Aspect-based Sentiment Analysis) system lies in its advanced processing of visual data, particularly car images. Rather than treating images as static visual content, the MASA system employs sophisticated image segmentation techniques to break down the visual representation of a vehicle into specific, identifiable components such as wheels, doors, windows, and headlights. This granular segmentation allows the system to isolate and extract meaningful visual cues from distinct car parts, which can be crucial for understanding customer sentiment that is targeted toward a specific feature or part of the vehicle. For example, if a reviewer comments negatively about the "headlights being too dim," the system is capable of focusing on the segmented headlight region in the image, thereby aligning the sentiment with the correct visual feature.

Beyond segmentation, MASA enhances interpretability by pairing the extracted visual data with corresponding textual reviews in a synchronized and structured manner. This

model-review pairing ensures that the opinions expressed in the text are grounded in visual context, reducing the ambiguity that often arises when interpreting sentiment from text alone. For instance, a review that states “the doors feel flimsy” can be mapped directly to the visual component of the car door in the image, reinforcing the sentiment with visual evidence. This alignment between modalities text and image not only improves the accuracy of aspect-level sentiment classification but also enables a richer and more context-aware analysis. Such multimodal integration is particularly valuable in domains like automotive reviews, where users frequently describe specific physical components whose evaluation can benefit from visual confirmation.

Table 1 Image results for proposed work

Car Brand	Input Image	Segmented Image
BMW		
Ferrari		
Hyundai		

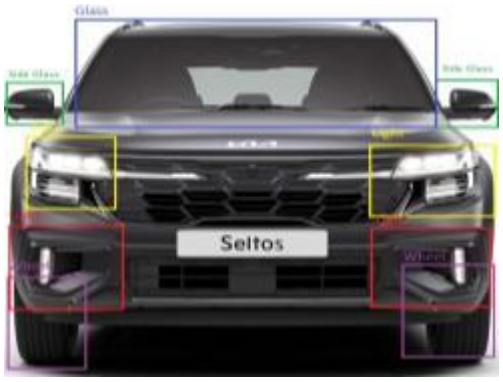
Kia		
Volkswagen		

Table 1 presents a sample of results for five car brands. For each brand, two visual representations are shown: the original input image and the segmented output image, processed through the SCV-YOLOv3 algorithm. This table demonstrates the system's capability to detect and extract vehicle components across different car types and visual conditions.

From the input-segmented image pairs, it is evident that the model is capable of handling a range of body styles and color contrasts, confirming the flexibility of the segmentation process. For instance, Ferrari and BMW brands with notably different car aesthetics were both successfully processed by the system, with key parts like grilles, doors, and bumpers clearly isolated. This segmentation accuracy is critical because it directly feeds into the feature extraction and vehicle classification pipeline using SSLU-GRU, which relies on high-quality visual cues to match reviews with car models.

The results validate the effectiveness of the proposed MASA framework in dealing with real-world, multimodal automotive data. The successful segmentation across various brands as displayed in Table 1 reinforces the robustness of the SCV-YOLOv3 algorithm, which improves upon traditional YOLO limitations. These image-level results serve as a foundational component for further steps in the MASA pipeline namely sentiment classification and privacy-aware analysis ensuring that both textual and visual content contribute meaningfully to business insights in the automotive industry.

4.2. Performance Analysis

This section presents a detailed evaluation of the performance of the proposed techniques, specifically the SSLU-GRU, supported by SCV-YOLO V3 and KSQ-KAnonymity. The focus is on how well the proposed SSLU-GRU model performs in sentiment analysis compared to several widely adopted deep learning architectures, namely GRU, BiLSTM,

LSTM, and RNN. This comparative analysis aims to demonstrate the superiority of the proposed approach in terms of prediction accuracy, error rate, and prediction time.

Table 2 Comparative Analysis for SSLU-GRU regarding sentiment analysis

Techniques	Prediction Accuracy (%)	Error Rate (%)	Prediction Time (ms)
Proposed SSLU-GRU	98.9428	0.1875	36458
GRU	96.7548	1.3256	42547
BiLSTM	93.8457	2.4578	46597
LSTM	91.5687	5.7821	51487
RNN	89.8745	8.3654	56871

Table 2 presents a comparative overview of the proposed SSLU-GRU model alongside four widely used deep learning models, GRU, BiLSTM, LSTM, and RNN based on three performance indicators: prediction accuracy, error rate, and prediction time. The SSLU-GRU model stands out with the highest accuracy at 98.94%, surpassing GRU (96.75%) and significantly outperforming RNN (89.87%), demonstrating its strong ability to interpret complex sentiment patterns. In terms of reliability, it achieves the lowest error rate of just 0.1875%, well below the average 4.48% seen in the baseline models, indicating a notable reduction in misclassification. Efficiency-wise, SSLU-GRU records a faster prediction time (36,458 ms) compared to LSTM (51,487 ms) and RNN (56,871 ms), while also slightly outperforming GRU and BiLSTM, maintaining a performance edge in speed.

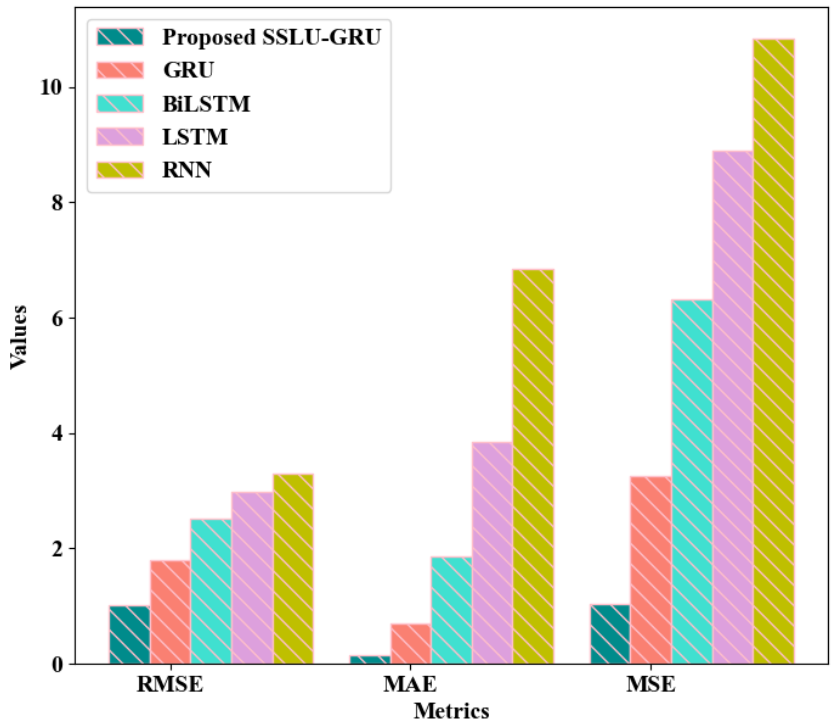


Figure 1 Graphical Representation regarding MAE, MSE, and RMSE

Figure 1 provides a graphical comparison of the proposed SSLU-GRU model against four benchmark models, GRU, BiLSTM, LSTM, and RNN based on three error metrics: RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MSE (Mean Squared

Error). The results clearly indicate that the SSLU-GRU model consistently achieves the lowest error values across all three metrics, highlighting its superior precision in sentiment prediction tasks. In contrast, the RNN and LSTM models exhibit the highest error rates, particularly in MSE where RNN reaches above 11, suggesting less stable and less accurate performance. GRU and BiLSTM perform moderately, but still fall short compared to SSLU-GRU. These findings reinforce the effectiveness of the proposed approach in minimizing predictive error, making it a more reliable model for sentiment analysis.

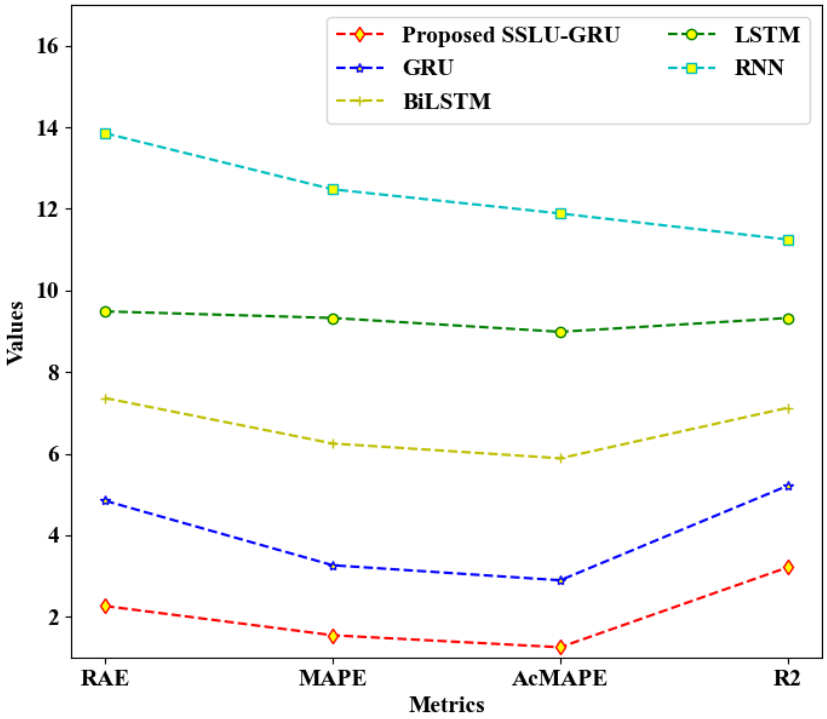


Figure 2 Comparative Analysis of the proposed classifier

Figure 1 and Figure 2 compare metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), Mean Absolute Percentage Error (MAPE), Accuracy Mean Absolute Percentage Error (AcMAPE), and R-Squared (R2) for the proposed model and existing works, namely GRU, BiLSTM, LSTM and RNN. The SA was predicted utilizing the proposed classifier, where the Sinu-Sigmoidal Linear Unit activation function was utilized. Hence, the proposed work yielded superior outcomes of 0.1475 MAE, 1.0354 MSE, 1.01754 RMSE, 2.2658 RAE, 1.5487 MAPE, 1.2547 AcMAPE, and 3.2145 R2 values. When analogized to the proposed work, the prevailing works obtained high values with an average of 3.3138 MAE, 7.329 MSE, 2.648 RMSE, 8.886 RAE, 7.829 MAPE, 7.414 AcMAPE, and 8.228 R2 values. Thus, the proposed classifier obtained low metric values. Therefore, the proposed classifier performed better when analogized to prevailing models. The proposed classifier's comparative analysis regarding vehicle prediction is explained below,

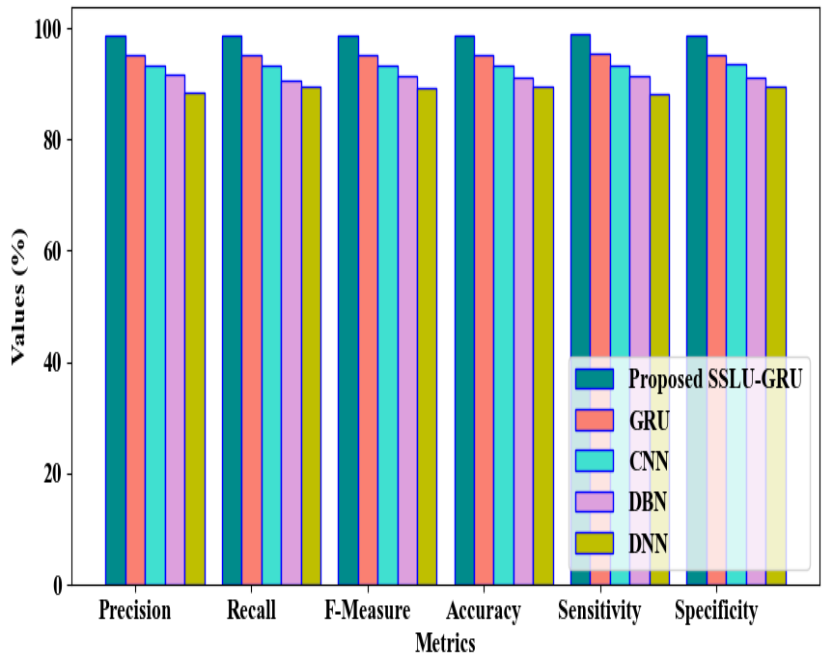


Figure 3 Comparative Analysis Regarding Vehicle Prediction

Figure 3 displays the proposed classifier's comparison with the prevailing models, such as GRU, BiLSTM, LSTM, and RNN regarding vehicle prediction. The proposed approach predicted the vehicles with 98.6659% accuracy, 98.7548% precision, 98.6478% recall, 98.7145%, F-Measure, 98.8754% sensitivity, and 98.7206% specificity, which is high compared to the existing works. The features extracted from the car parts aided the proposed system in predicting the vehicles with higher values. Hence, when analogized to the prevailing classifier, the proposed approach efficiently detected the vehicle. The comparative analysis for car parts segmentation is given as follows,

Table 3 Comparative Analysis of SCV-YOLO V3

Methods	Average Precision (%)						
	Logo	Light	Wheel	Side Glass	Glass	Steering	Seat
Proposed SCV-YOLO V3	98.98	98.84	98.85	98.75	98.88	98.72	98.87
YOLO V3	95.65	96.45	96.12	95.67	95.32	96.34	95.47
MobileNet	92.65	93.48	93.48	92.87	93.15	92.83	92.57
SSD	90.21	90.54	91.24	90.65	91.21	90.27	90.22
CNN	89.65	88.54	89.65	88.54	87.35	89.65	88.51

Table 3 presents a comparative analysis of object detection performance using various methods, specifically focusing on the detection accuracy of different car components such as the logo, light, wheel, side glass, glass, steering, and seat. The proposed SCV-YOLO V3 model consistently outperforms the other approaches across all categories, achieving an exceptionally high average precision above 98% for each component. In comparison, standard YOLO V3 performs slightly lower with precision

values ranging between 95%–96%, while MobileNet follows with scores in the 92%–93% range. The performance further declines with SSD, which averages around 90%–91%, and CNN, which lags behind with the lowest precision values between 87% and 89%. These results highlight the effectiveness of the SCV-YOLO V3 model in accurately identifying and segmenting fine-grained car parts, making it a robust visual detection tool within the MASA framework.

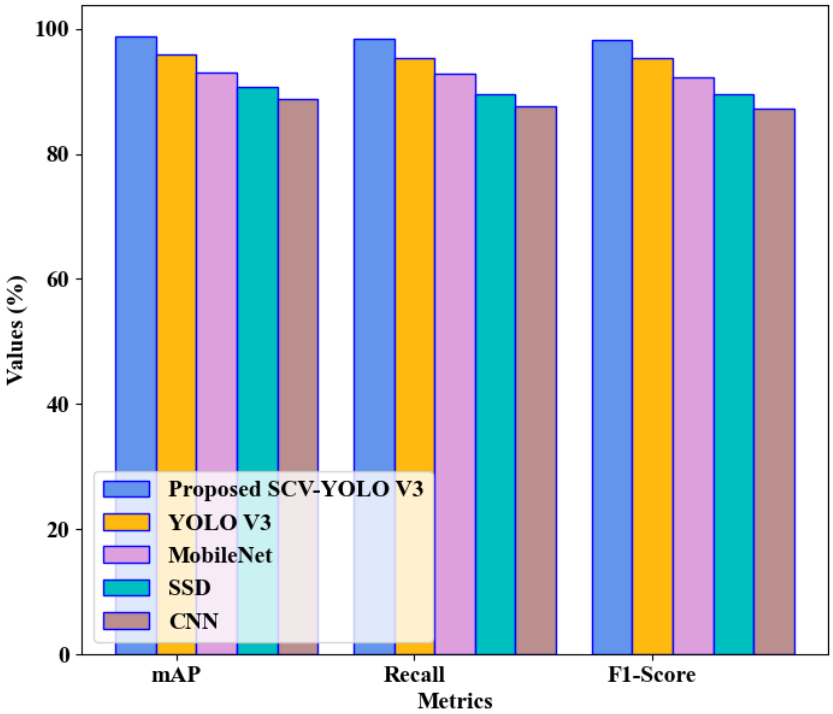


Figure 4 Comparative Analysis regarding mAP, Recall, and F1-Score

Figure 4 illustrates the comparative performance of five object detection models. Proposed SCV-YOLO V3, YOLO V3, MobileNet, SSD, and CNN based on three key evaluation metrics: mean Average Precision (mAP), Recall, and F1-Score. The proposed SCV-YOLO V3 consistently achieves the highest values across all metrics, indicating superior accuracy, completeness, and balance in detection. YOLO V3 performs second best, followed by MobileNet and SSD, while CNN registers the lowest scores in all categories. This performance gap underscores the robustness of SCV-YOLO V3 in object detection tasks, as it not only detects more relevant instances (high Recall) but also maintains precision and consistency (high F1-Score and mAP), making it the most reliable model among those compared.

Segmenting and identifying car parts using the Scaled Correlation Vector (SCV)-centric bounding box begins with pre-processing the input images before feeding them into the SCV-YOLO V3 model. As a result, the proposed method achieved high average precision scores for various car components: logo (98.98%), light (98.84%), wheel (98.85%), side glass (98.75%), glass (98.88%), steering (98.72%), and seat (98.87%), as shown in Table 3, outperforming the existing models. In addition, as illustrated in Figure 4, the proposed system recorded 98.84% mean Average Precision (mAP), 98.45% recall, and 98.15% F1-score, while baseline models averaged only 92.08% (mAP), 91.29% (recall), and 91.04% (F1-score). Compared to conventional systems such as YOLO V3, MobileNet, SSD, and CNN, the proposed approach delivers significantly better accuracy in

both detection and segmentation of car parts. The next section presents a comparative analysis focusing on privacy preservation from the user perspective.

Table 4 Comparative Analysis of KSQ-KAnonymity

Techniques	Information Loss (kb)
Proposed KSQ-KAnonymity	486
K-Anonymity	873
L-Diversity	1358
T-Closeness	1842
Randomization	2436

Table 4 presents a comparative analysis of different privacy-preserving techniques based on the metric of Information Loss (in kilobytes). The lower the information loss, the better a method is at preserving data utility while ensuring privacy. The proposed KSQ-KAnonymity method demonstrates the lowest information loss at 486 kb, indicating that it maintains more of the original data's usefulness compared to the others. In contrast, standard K-Anonymity results in a higher loss at 873 kb, followed by L-Diversity at 1358 kb, and T-Closeness at 1842 kb. The Randomization technique shows the highest information loss among all, reaching 2436 kb, which suggests a significant reduction in data quality and utility. Overall, this table clearly supports the effectiveness of KSQ-KAnonymity in balancing privacy protection with minimal loss of meaningful information, making it a more efficient and data-friendly approach for privacy preservation.

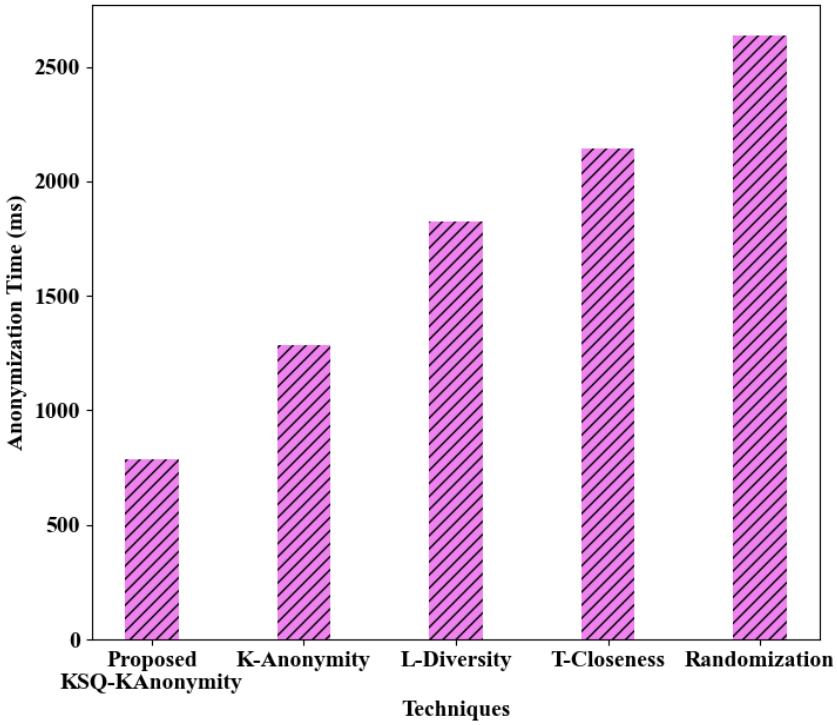


Figure 5 Comparative Analysis Regarding Anonymization Time

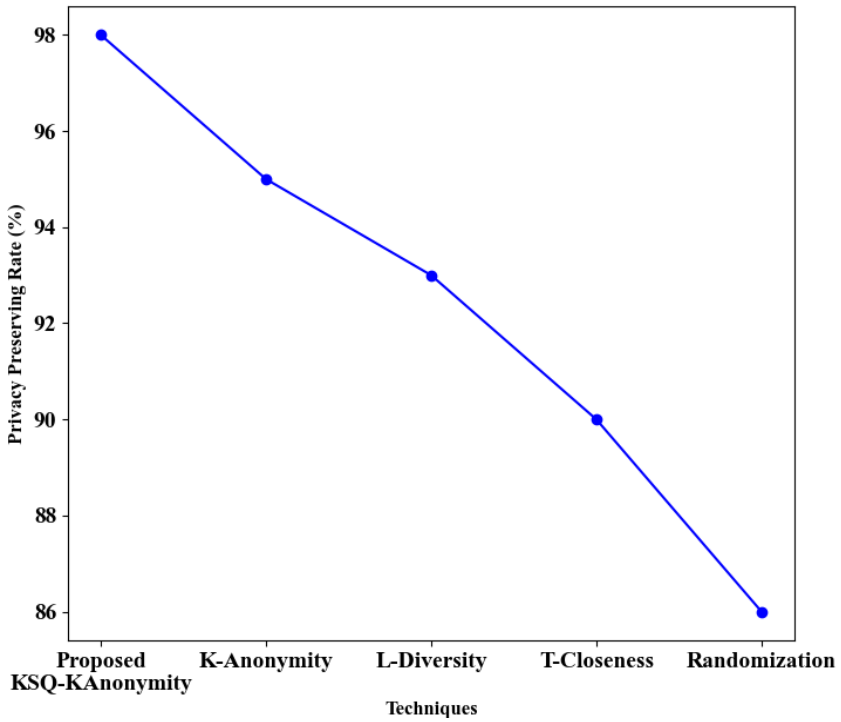


Figure 6 Graphical Analysis Regarding Privacy Preserving Rate

Figure 5 and Figure 6 illustrate the efficiency and effectiveness of the proposed KSQ-KAnonymity method in preserving user privacy during sentiment analysis. Specifically, Figure 5 shows that the system anonymizes user data significantly faster completing the process in just 784 ms, while other models such as K-Anonymity, L-Diversity, T-Closeness, and Randomization require an average of 1972.5 ms. Meanwhile, Figure 6 highlights the Privacy Preserving Rate, where the proposed model achieves a remarkable 98%, outperforming the 91% average achieved by existing methods. To ensure this high level of privacy, the KSQ-KAnonymity approach integrates Kendall's Tau, Spearman's Rank Correlation, and a Quasi Identifier strategy to safeguard user information effectively. As reflected in Table 4, this method results in minimal information loss only 486 kb out of 10,000 kb whereas the baseline methods lose an average of 1628 kb. These results confirm that the proposed model not only anonymizes data more quickly but also maintains a higher level of data integrity and privacy than traditional approaches.

Table 5 Comparative Analysis of Existing Works

Study	Method	Accuracy (%)	Error Rate (%)	RMSE
Proposed work	SSLU-GRU	98.9428	0.1875	1.0175
(Yang et al., 2022)	LSTM	91.03	2.45	3.142
(Li et al., 2023)	CNN	97.3	0.934	-
(Lin et al., 2020)	Corpus	93.84	-	6.67
(Park et al., 2021)	GSSL	95.7	0.956	-
(Li et al., 2021)	DCGAN	94.234	8.3	5.21

Table 5 presents the proposed study is analogized with the prevailing works. The prevailing studies utilized LSTM, CNN, Corpus classification, Graph-based Semi-Supervised Learning (GSSL), and Deep Convolutional Generative Adversarial Network

(DCGAN), and the proposed model utilized the SSLU-GRU technique. In the proposed work, the videos were converted to frames and then pre-processed. Then, the car segmentation was done for the pre-processed frames. The features extracted from the segmented car parts were given to the classifier that provided 98.9428% accuracy. However, approaches like LSTM, Corpus classifier, and GSSL concentrated only on aspect-centric classification, resulting in lower accuracy values of 91.03%, 93.84%, and 95.7% than the proposed classifier. The work, where CNN was utilized, took the data that has a very low negative review. Thus, the system acquired misclassification with a 0.934% error rate. In addition, the DCGAN technique utilized the pre-defined dictionary that couldn't detect the new car model types from the user's review and provided a 5.21 RMSE value. Thus, the proposed system investigated the sentiments of the car reviews for car classification and helped the business decisions.

4. Conclusions

This study introduced an integrated and efficient framework for multimodal sentiment analysis (SA) on car reviews by combining three key techniques: SSLU-GRU for textual sentiment classification, SCV-YOLO V3 for visual segmentation of car parts, and KSQ-KAnonymity for user privacy protection. The experimental results demonstrated strong performance across multiple evaluation metrics. For instance, car part segmentation achieved an F1-score of 98.15%, while vehicle classification reached an impressive accuracy of 98.94%. The sentiment classification component, based on aspect-centric keyword mapping from preprocessed textual input, produced robust results with a low Mean Squared Error (MSE) of 1.0354. Additionally, the proposed privacy module effectively safeguarded user information, achieving a 98% Privacy Preserving Rate and reducing information loss to just 486 kb, outperforming traditional methods in both speed and data protection. Collectively, these outcomes highlight the framework's potential not only in enhancing the precision of sentiment interpretation across modalities but also in supporting business decisions through accurate customer feedback while maintaining data privacy.

Despite the promising results, the current model does not incorporate additional valuable data sources such as historical car sales records and complaints from service station personnel, both of which could offer deeper insights into customer sentiment and product reliability. These omissions represent a limitation in capturing a holistic view of user experience and market response. For future research, the integration of such data will be prioritized to enrich the sentiment analysis process. By combining user-generated reviews with operational data from dealerships and service centers, future iterations of the framework aim to provide more comprehensive, actionable insights to automotive businesses and policymakers.

References

Alshuwaier, F., Areshey, A., & Poon, J. (2022). Applications and Enhancement of Document-Based Sentiment Analysis in Deep learning Methods: Systematic Literature Review. *Intelligent Systems with Applications*, *15*, 200090. https://doi.org/10.1016/j.iswa.2022.200090

Barua, A., Ahmed, M. U., & Begum, S. (2023). A Systematic Literature Review on Multimodal Machine Learning: Applications, Challenges, Gaps and Future Directions. *IEEE Access*, *11*, 14804–14831. https://doi.org/10.1109/ACCESS.2023.3243854

Bayoudh, K., Knani, R., Hamdaoui, F., & Mtibaa, A. (2022). A survey on deep multimodal

learning for computer vision: advances, trends, applications, and datasets. *The Visual Computer*, *38*(8), 2939–2970. https://doi.org/10.1007/s00371-021-02166-7

Buscemi, A., & Proverbio, D. (2024). *ChatGPT vs Gemini vs LLaMA on Multilingual Sentiment Analysis*.

Ghadge, A., Mogale, D. G., Bourlakis, M., M. Maiyar, L., & Moradlou, H. (2022). Link between Industry 4.0 and green supply chain management: Evidence from the automotive industry. *Computers and Industrial Engineering*, *169*. https://doi.org/10.1016/j.cie.2022.108303

Guo, W., Zhang, Y., Cai, X., Meng, L., Yang, J., & Yuan, X. (2021). LD-MAN: Layout-Driven Multimodal Attention Network for Online News Sentiment Recognition. *IEEE Transactions on Multimedia*, *23*, 1785–1798. https://doi.org/10.1109/TMM.2020.3003648

Jena, R. (2020). An empirical case study on Indian consumers' sentiment towards electric vehicles: A big data analytics approach. *Industrial Marketing Management*, *90*, 605–616. https://doi.org/10.1016/j.indmarman.2019.12.012

Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*, *54*(7), 4997–5053. https://doi.org/10.1007/s10462-021-09973-3

Lin, D., Lee, C. K. M., Lau, H., & Yang, Y. (2018). Strategic response to Industry 4.0: an empirical investigation on the Chinese automotive industry. *Industrial Management and Data Systems*, *118*(3). https://doi.org/10.1108/IMDS-09-2017-0403

Lin, Y., Fu, Y., Li, Y., Cai, G., & Zhou, A. (2021). Aspect-based sentiment analysis for online reviews with hybrid attention networks. *World Wide Web*, *24*(4), 1215–1233. https://doi.org/10.1007/s11280-021-00898-z

Llopis-Albert, C., Rubio, F., & Valero, F. (2021). Impact of digital transformation on the automotive industry. *Technological Forecasting and Social Change*, *162*. https://doi.org/10.1016/j.techfore.2020.120343

Lukin, E., Krajnović, A., & Bosna, J. (2022). Sustainability Strategies and Achieving SDGs: A Comparative Analysis of Leading Companies in the Automotive Industry. In *Sustainability (Switzerland)* (Vol. 14, Issue 7). https://doi.org/10.3390/su14074000

Maitama, J. Z., Idris, N., Abdi, A., Shuib, L., & Fauzi, R. (2020). A Systematic Review on Implicit and Explicit Aspect Extraction in Sentiment Analysis. *IEEE Access*, *8*, 194166–194191. https://doi.org/10.1109/ACCESS.2020.3031217

Motz, A., Ranta, E., Calderon, A. S., Adam, Q., Alzhouri, F., & Ebrahimi, D. (2022). Live Sentiment Analysis Using Multiple Machine Learning and Text Processing Algorithms. *Procedia Computer Science*, *203*, 165–172. https://doi.org/10.1016/j.procs.2022.07.023

Nazir, A., Rao, Y., Wu, L., & Sun, L. (2022). Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey. *IEEE Transactions on Affective Computing*, *13*(2), 845–863. https://doi.org/10.1109/TAFFC.2020.2970399

Shoumy, N. J., Ang, L.-M., Seng, K. P., Rahaman, D. M. M., & Zia, T. (2020). Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals. *Journal of Network and Computer Applications*, *149*, 102447. https://doi.org/10.1016/j.jnca.2019.102447

Tan, X., Cai, Y., Xu, J., Leung, H.-F., Chen, W., & Li, Q. (2020). Improving aspect-based sentiment analysis via aligning aspect embedding. *Neurocomputing*, *383*, 336–347. https://doi.org/10.1016/j.neucom.2019.12.035

Xiao, G., Tu, G., Zheng, L., Zhou, T., Li, X., Ahmed, S. H., & Jiang, D. (2021). Multimodality

Sentiment Analysis in Social Internet of Things Based on Hierarchical Attentions and CSAT-TCN With MBM Network. *IEEE Internet of Things Journal*, *8*(16), 12748–12757. https://doi.org/10.1109/JIOT.2020.3015381

Xu, J., Li, Z., Huang, F., Li, C., & Yu, P. S. (2021). Social Image Sentiment Analysis by Exploiting Multimodal Content and Heterogeneous Relations. *IEEE Transactions on Industrial Informatics*, *17*(4), 2974–2982. https://doi.org/10.1109/TII.2020.3005405

Zhang, C., Tian, Y.-X., Fan, Z.-P., Liu, Y., & Fan, L.-W. (2020). Product sales forecasting using macroeconomic indicators and online reviews: a method combining prospect theory and sentiment analysis. *Soft Computing*, *24*(9), 6213–6226. https://doi.org/10.1007/s00500-018-03742-1

Zhao, N., Gao, H., Wen, X., & Li, H. (2021). Combination of Convolutional Neural Network and Gated Recurrent Unit for Aspect-Based Sentiment Analysis. *IEEE Access*, *9*, 15561–15569. https://doi.org/10.1109/ACCESS.2021.3052937

Zhou, J., Zhao, J., Huang, J. X., Hu, Q. V., & He, L. (2021). MASAD: A large-scale dataset for multimodal aspect-based sentiment analysis. *Neurocomputing*, *455*, 47–58. https://doi.org/10.1016/j.neucom.2021.05.040