A Cohesive Transformer-Capsule Network Model for Sentiment and Aspect Analysis Using Contextual Embeddings

Laxmi Pamulaparthy
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana, India
laxmi.p16@gmail.com

C. H. Sumalakshmi
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad, Telangana, India
sumascarlet@gmail.com (corresponding author)

Received: 2 December 2025 | Revised: 28 January 2026 and 19 February 2026 | Accepted: 20 February 2026

 DOI: https://doi.org/10.48084/etasr.16650

ABSTRACT

In aspect-based sentiment analysis, identifying and associating sentiments with specific aspects is a significant challenge due to the complex interplay between context and language. Traditional methods often struggle to accurately identify sentiments associated with particular aspects, especially in texts characterized by dynamically evolving linguistic contexts. The present study introduces a novel Transformer-Capsule Network for Sentiment and Aspect Analysis (TCSA) model that integrates transformer architectures and Capsule Networks (CapsNets) designed to improve the precision of aspect analysis. This model combines the deep contextual understanding capabilities of RoBERTa with the dynamic routing efficiency of CapsNets. The key strategy employed involves dynamic contextual word embeddings generated by BERT, which are crucial for capturing semantic contexts contributing to accurate aspect analysis. Furthermore, the model uses sophisticated data augmentation algorithms, including synonym substitution, back-translation, and contextual augmentation to expand the training data. The TCSA model utilizes a multi-head self-attention mechanism in the transformer-capsule framework, enabling focused attention on the various data fragments, and thus interpreting complex interactions between textual aspects and their contexts. The model demonstrates excellent performance, achieving an accuracy of 97.15%, a precision of 97.20%, a recall of 98.12%, and an F1-score of 97.30%. The results not only demonstrate the effectiveness of the model for aspect analysis but also set a new standard in the use of dynamic contextual embeddings and complex network architectures for sentiment analysis.

Keywords-sentiment analysis; transformer networks; capsule networks; dynamic contextual embeddings; multi-head self-attention; RoBERTa; textual semantics

I. INTRODUCTION

TCSA encompasses fast applications such as market intelligence and social media analysis. Although the traditional methods of sentiment analysis treat text as documents or sentences, they fail to identify the subtle sentiments related to particular aspects [1]. The complexity of human language, with its nuanced semantic layers and contextual variations, presents a challenge in accurately capturing these sentiments.

Traditional sentiment analysis methods attempt to determine a document or sentence's overall sentiment without distinguishing between content-specific sentiments [2]. Static rule-based algorithms or shallow machine learning models cannot contextually adapt to evolving language structures and multi-aspect texts [3]. Advanced deep learning methods, such as BERT-based models, treat aspect extraction and sentiment classification as separate processes, resulting in inconsistent results and high computational costs [4]. The current aspect analysis methods require contextual awareness to interpret subtle linguistic cues and structural sensitivity to preserve part-whole relationships between aspects and their contexts; however, these needs are not addressed by existing solutions. CapsNets capture hierarchical features via dynamic routing but lack deep contextual understanding [4], while transformer models such as RoBERTa excel in contextual embedding but overlook spatial hierarchies due to attention alone [5]. Existing methods also require laborious domain-specific annotated

datasets and lose information due to CNN pooling layers or RNN sequential bottlenecks [6].

This study proposes the TCSA, which integrates RoBERTa's deep contextual modeling with CapsNets' dynamic routing to capture global context and local hierarchical semantics. The framework implicitly learns aspect representations within an end-to-end transformer–capsule architecture, enabling simultaneous aspect localization and sentiment polarity prediction. Advanced data augmentation strategies further enhance linguistic diversity and model robustness, making TCSA effective for complex aspect-based sentiment analysis tasks.

After preprocessing, a feature matrix is constructed using dynamic contextual embeddings generated by BERT [7], while RoBERTa integrated with CapsNets forms the TCSA model. This model combines RoBERTa's contextual understanding with CapsNets dynamic routing and multi-head self-attention to capture complex text relationships and aspect–sentiment interactions. Trained with optimized hyperparameters and cross-validation, the model achieves an accuracy of 97.15%, a precision of 97.20%, a recall of 98.12%, and an F1-score of 97.30%, demonstrating its effectiveness for aspect-based sentiment analysis and establishing a robust hybrid deep learning benchmark.

II. LITERATURE REVIEW

From rule-based approaches to deep learning models, sentiment analysis is widely applied in NLP applications, including business intelligence and social media monitoring [8]. Text emotion categorization uses traditional feature engineering as well as psychological and sociological deep learning. Recent methods identify delicate sentiments using complex algorithms, while earlier methods have used emotion dictionaries, tagged data, topic words, and polarity. However, a high-dimensional feature space requires feature selection [9]. Traditional sentiment categorization employs rules and machine learning [10]. Sentiment classification research has also utilized deep learning due to strong human feature extraction [11]. For example, authors in [12] improved LSTM feature extraction with multi-task training and parameter tuning; however, LSTMs are computationally inefficient. Authors in [13] proposed a bidirectional deep convolution network-based text classification model to classify text sequence information using causal convolution.

Transformers are commonly employed for text feature extraction, self-attention word dependence modeling, and efficient sentence structure capture. Authors in [14] used a CNN architecture with one hidden layer for Twitter sentiment analysis. Authors in [15] utilized Global Vectors for Word Representation (GloVe) with a Deep Convolutional Neural Network (DCNN) for Twitter sentiment analysis. The GloVe-DCNN model outperforms Bag of Words (BoW) combined with Logistic Regression (LR) and support vector machines. Authors in [16] proposed high-dimensional text categorization in sentiment analysis. Authors in [17] compared VD-CNNs with the Google's pre-trained BERT architecture [18]. Their results indicated that the model with the BERT architecture outperformed the VD-CNN model. However, VD-CNNs are simpler and more computationally efficient than the BERT, RoBERTa, DistilBERT, and Lite BERT models.

CapsNets preserve data hierarchical links and spatial hierarchies, making them a suitable alternative to neural network topologies for sentiment analysis. CapsNets with vector representations and dynamic routing can capture complicated text features that CNNs may overlook [19]. Authors in [20] used CapsNets for sentiment analysis and cross-domain categorization and proved that they can handle complex language patterns. Space and hierarchical relationships are maintained by vector capsules and dynamic routing in CapsNets. Authors in [1] utilized capsule-based variants T-Caps, which utilize CapsNets and transformers to reduce sentiment classification data loss, and achieved an excellent F1-score of 90.69% and an accuracy of 91.04%. SA-CapsNet reduced parameters and improved text classification accuracy on the IMDB and MR datasets by 1.43% by employing self-attention [21].

By modelling local spatial hierarchies with CapsNets for aspect-based sentiment analysis, XLNetCN outperformed BERT and conventional models on SemEval datasets [22]. These advances demonstrate CapsNets' ability to encode complicated language without max-pooling. Authors in [22] demonstrated how transformer technology prioritizes consecutive data via self-attention, while aspect-based sentiment analysis on benchmark datasets favored BERT-XLNetCN. In [23], transformers captured global dependencies without sequential processing utilizing self-attention, changing NLP, while BERT and XLNet excelled in ABCA. This ST-GCN simulated diverse textual input and handled feature-space dimensionality with graph structures and transformer-based embeddings [25].

The BERT model often fails due to the pretrain-finetune gap and interdependence of masked positions. In [24], XLNet with denoise auto-encoding pre-training outperformed BERT-based models on 17 datasets, including SQuAD, GLUE, and RACE, in lengthy document and text generation, language understanding, and question-answering tasks. Authors in [18] utilized deep contextualized word embeddings such as BERT and improved NLU tasks. BERT's intensive language model pre-training yields excellent results in question answering, sequence classification, and sequence tailoring classification.

III. PROPOSED SYSTEM

This study proposes the TCSA framework for accurate aspect-level sentiment analysis. After rigorous preprocessing, dynamic contextual embeddings generated by BERT prepare the data for analysis. Subsequently, RoBERTa and CapsNets jointly extract global contextual information via self-attention and local hierarchical semantics through dynamic routing, overcoming CNN pooling limitations. Trained with multi-head attention, optimized hyperparameters, and cross-validation, TCSA demonstrates accuracy, scalability, and linguistic adaptability, offering a robust and reliable solution for advanced aspect-based sentiment analysis. The proposed methodology is illustrated in Figure 1.

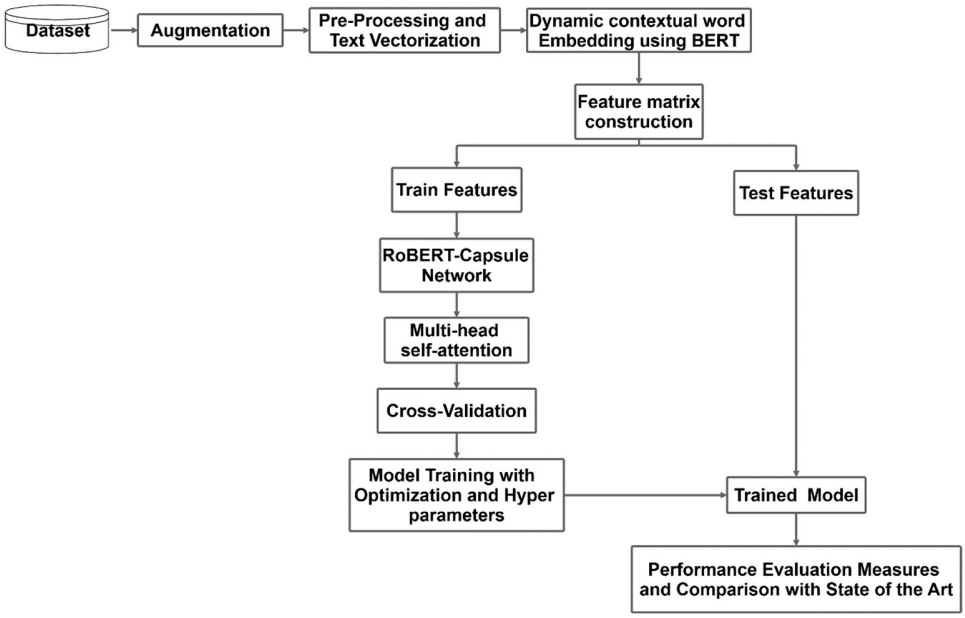

Fig. 1. Proposed methodology.

A. Dataset Acquisition and Augmentation

The 515K hotel reviews dataset [27] from Booking.com contains 515,000 reviews of 1,493 European hotels collected between August 2015 and August 2017, covering 227 nationalities. Each entry includes textual feedback and numerical ratings, with over 95% positive reviews averaging 8.6 and negative reviews averaging 3.86. The 17-field CSV file also provides contextual tags, such as "Leisure trip" and "Business trip", for structured analysis.

Several dataset augmentation methods are employed to improve the robustness and generalizability of the proposed TCSA model. Augmentation increases training data volume and introduces linguistic variations to strengthen the model [26]. The augmentation methods include:

- Synonym replacement: It involves replacing words in sentences with synonyms while retaining context. It helps the model avoid keywords and generalize across sentiment expressions.
- Back-translation: This technique involves translating a sentence from English to a foreign language (e.g., German) and then back to English. Syntactic diversity and sentence structure variations increase the linguistic diversity of the training data.
- Contextual augmentation: A more advanced technique that adjusts sentences based on context. Depending on the desired effect, contextual augmentation alters sentences' sentiments.

1) Contextual Augmentation

Contextual augmentation modifies sentences by considering the broader semantic and syntactic context in which specific words or phrases occur. Augmentation may involve more complex transformations than simple synonym replacement, including paraphrasing or changing certain parts of speech while retaining the original sentiment and meaning.

The steps involved in contextual augmentation are:

- Context analysis: In this step, the context and sentiment of each sentence in the dataset are identified, and key phrases or words that strongly influence the sentiment are determined. The context analysis process is represented as:

$$C = AnalyzeContext(S) \quad (1)$$

where S is the original sentence.

- Sentence transformation: In this step, transformations are applied to the sentence structure or vocabulary without changing its sentiment. Sentence transformation uses NLP tools to generate paraphrases that preserve the original sentiment but modify the expression or complexity of the sentence. Sentence transformation is defined in:

$$S' = Transform(S, C) \quad (2)$$

B. Pre-Processing and Text Vectorization

Pre-processing text data in the TCSA model standardizes its format and removes irrelevant noise that could affect model performance. For format and vocabulary consistency, the entire text is converted to lowercase:

$$S' = lowercase(S)$$

where S is the original text, and S is the converted text. Additionally, punctuations are removed:

$$S'' = \text{remove\_punctuation}(S')$$

Punctuations are removed for the model to be focused on word content rather than on syntax or formatting. $S''' = \text{strip\_whitespace}(S'')$ denotes the pre-processed text, in which leading and trailing spaces, as well as redundant spaces between words are removed.

C. Dynamic Contextual Word Embedding

In the proposed TCSA framework, contextual word embedding is performed using BERT, which is utilized to generate dynamic, token-level contextual embeddings that capture semantic dependencies within each sentence. These embeddings are precomputed during the preprocessing stage and organized into a structured feature matrix, which serves as the input to the downstream deep learning architecture. The BERT parameters are kept fixed during subsequent training to ensure stable contextual representations and to reduce computational complexity.

The embeddings generated by BERT are contextually aware, meaning that the representation of each word is influenced by the other words in the sentence, unlike static embeddings. This dynamic nature of BERT embeddings ensures that nuances and subtle meanings in the text are captured. Furthermore, each embedding is specifically tailored to reflect the word's usage in its particular context, enhancing the model's ability to capture and understand diverse expressions and sentiments in the text. The feature matrix M is defined as:

$$M[i,:] = E_i[i,:]$$

where $M[i,:]$ is the i^{th} row in the feature matrix M, and E_i is the embedding vector for the i^{th} word.

D. Feature Matrix Construction

After obtaining embeddings for each word, these embeddings are compiled into a feature matrix M. This matrix is structured such that each row corresponds to the embedding of a word or sentence, aligning all data into a format suitable for analysis by neural networks. This structured and meticulous approach to preprocessing, vectorizing, and constructing contextual embeddings ensures that the TCSA model is built on a foundation of clean, well-formatted, and diverse data, enabling effective and nuanced aspect-based sentiment analysis.

E. RoBERTa-CapsNet Application

In this stage, the TCSA model leverages the combined capabilities of RoBERTa for deep contextual understanding and CapsNets for recognizing and processing spatial hierarchies. RoBERTa processes the feature matrix M_{train} derived from the training set, applying its pre-trained contextual insights to enhance each vector's representation based on the surrounding textual context:

$$R = RoBERTa(M_{train}) \tag{3}$$

While BERT is used for initial contextual embedding generation, RoBERTa functions as the transformer backbone within the proposed TCSA architecture. RoBERTa is fine-tuned end-to-end together with the CapsNets, enabling task-specific adaptation of contextual representations. The self-attention mechanism in RoBERTa captures global contextual relationships, whereas the CapsNets preserve spatial and hierarchical dependencies between aspects and sentiment expressions through dynamic routing. This division of responsibilities allows the TCSA model to leverage stable contextual encoding and adaptive deep feature learning.

1) Application of CapsNet

The transformed matrix R is input into the CapsNets, which structure data hierarchically and capture relationships that are crucial for understanding the text structure:

$$C = CapsNet(R) \tag{4}$$

Aspect extraction in the TCSA framework is implicitly achieved through the dynamic routing process of the CapsNet. Tokens and phrases that exhibit semantic agreement are routed toward the same capsules, effectively grouping them into aspect-specific representations. The vector orientation of each capsule encodes the sentiment direction, while the capsule length represents the confidence of the detected aspect–sentiment pair. This representation enables simultaneous learning of aspect localization and sentiment classification without a separate extraction stage.

2) Squash Function

The squash function is used within the CapsNet, which squashes the vector outputs to a length between 0 and 1, maintaining the direction:

$$v_j = \frac{\|S_j\|^2}{1+\|S_j\|^2}\frac{S_j}{||S_j||} \tag{5}$$

Equations (3)-(5) emphasize the end-to-end differentiability of the TCSA architecture rather than stepwise procedures. The RoBERTa transformation maps contextual feature matrices to higher-level semantic representations via stacked self-attention and feed-forward layers, while dynamic routing and squash operations in the CapsNets are compactly represented as learned functions. This formulation aligns with deep learning practice, highlighting architectural integration and joint optimization through backpropagation while avoiding unnecessary low-level derivations.

F. Multi-Head Self-Attention Mechanism

After passing through RoBERTa and CapsNets, the model applies a multi-head self-attention mechanism to the output C. Through this mechanism, the model can attend to different parts of the data, improving its ability to identify important aspects and correlations to the text.

$$A = MultHeadSelfAttention(C) = Concat(head_1, head_2, \ldots . head_k)W^O \tag{6}$$

where each $head_i$ is computed as:

$$head_i = Attention\left(CW_i^Q, CW_i^k, CW_i^v\right) = softmax\left(\frac{(CW_i^Q)(CW_i^k)^T}{\sqrt{d_k}}\right)CW_i^v \tag{7}$$

G. Model Architecture

The proposed TCSA architecture, as displayed in Figure 2, illustrates the step-by-step process from raw text input to final sentiment classification. This method represents an advanced and efficient approach to aspect-based sentiment analysis, combining multiple deep learning techniques to achieve high accuracy and robustness.

The proposed TCSA model performs joint aspect-based sentiment analysis by implicitly learning aspect information

within the CapsNet, without explicit aspect extraction. Lower-level capsules capture local semantic patterns, while higher-level capsules aggregate them into aspect–sentiment representations through dynamic routing. This end-to-end design enables simultaneous aspect identification and sentiment polarity prediction.

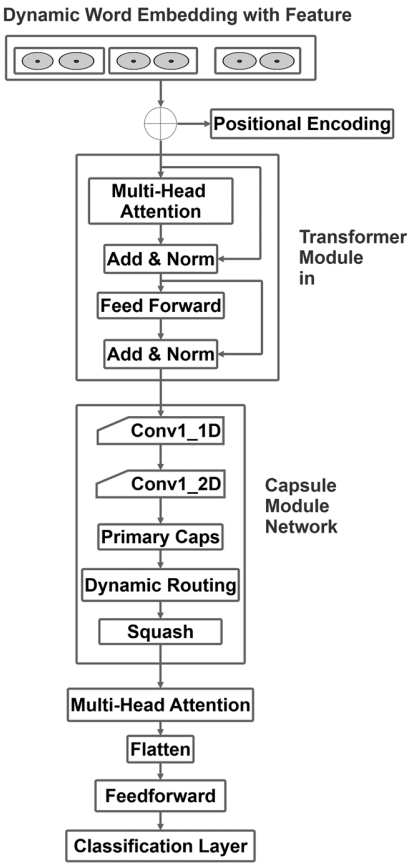

Fig. 2. Architecture of the proposed TCSA model.

As portrayed in Figure 2, the input text is first converted into a feature matrix using dynamic contextual word embeddings, followed by feature extraction through the transformer module with multi-head self-attention to capture complex contextual relationships. The extracted representations are then processed by the CapsNets with dynamic routing to preserve hierarchical semantic structures. The resulting features are subsequently passed through flattened and dense layers to prepare them for final sentiment classification using a softmax layer, enabling accurate and interpretable aspect–sentiment prediction.

The TCSA framework learns aspects implicitly within the CapsNets rather than explicitly extracting aspect terms. Lower-level capsules capture aspect-relevant semantics from BERT and RoBERTa representations, while higher-level capsules form aspect–sentiment entities through dynamic routing. Since aspects are embedded as latent capsule vectors, evaluation relies on sentiment metrics, such as Receiver-Operating Characteristic Curve (ROC), Area Under the Curve (AUC), confusion matrices, and calibration analysis, instead of explicit aspect-level precision and recall.

H. Proposed Model Algorithm

The proposed aspect-based sentiment analysis model employs Algorithm 1, RoBERTa, and CapsNets, with deep contextual analysis to manage global and local textual features. Using BERT, hotel reviews are meticulously preprocessed, augmented, and vectorized for complex neural network processing. The model also creates a sophisticated neural architecture that combines multi-head self-attention with dynamic routing to refine text-aspect relationships. This is followed by the cross-validation and performance evaluation using standard metrics.

Algorithm 1: TCSA Model for aspect-based sentiment analysis
Step 1: Initialize TCSA Model
Step 2: Data Acquisition and Augmentation
Load the dataset consisting of 515K hotel reviews.
Augment data using synonym replacement, back-translation, and contextual augmentation.
Step 3: Pre-processing and Text Vectorization
Normalize and preprocess dataset, and vectorize the preprocessed text using BERT to obtain contextual word embeddings.
Step 4: Feature matrix construction
Aggregate BERT embeddings into a feature matrix M_M, which captures the nuanced semantic context of each word, and organize data into a structured form, M_M, that is suitable for neural network processing.
Step 5: RoBERTa-Capsule Network Integration
Apply RoBERTa for deep contextual analysis on M_{train}. Utilize CapsNets to capture and process spatial hierarchies from RoBERTa's output.
Output CC: The feature matrix post integration.
Step 6: Apply multi-head self-attention
Implement Multi-Head Self-attention on C to refine features and focus on intricate text-aspect relationships.
Step 7: Model training and performance evaluation
Evaluate the TCSA model using the test dataset Mtest for accuracy (α), precision (p), and recall (r),
and F1-score (F1).

IV. RESULTS AND ANALYSIS

The proposed TCSA model is built using Tensorflow, one of the deep learning frameworks developed by Google. The framework is further tested to evaluate its performance. The

TensorFlow framework simplifies development by integrating the GRU model [28]. The proposed model was implemented using Python with TensorFlow and PyTorch on a system with an Intel i5 processor, P-100 GPU, and 16GB RAM, ensuring efficient deep learning computations. This setup enabled seamless training and optimization of the RoBERTa-capsule network for aspect-based sentiment analysis.

A. Performance Assessment

The model performance was evaluated using accuracy, recall, precision, and F1-score metrics. Model accuracy represents the compatibility with the data, while recall measures the model's ability to identify the true positives. The main objective of precision is to reduce the probability of false positives, while the F1-score helps in balancing the recall and accuracy. Collectively, these metrics provide insights into the strengths and weaknesses of the model and can be calculated using [29]:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TN} + \text{FP})} \quad (10)$$

$$\text{F1} - \text{score} = \frac{2\times(\text{Precision}\times\text{Recall})}{(\text{Precision} + \text{Recall})} \quad (11)$$

where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

To address severe class imbalance in the hotel review dataset, the proposed TCSA framework employs data augmentation, stratified cross-validation, and CapsNets dynamic routing to enhance minority sentiment learning. Model robustness is evaluated using precision, recall, F1-score, ROC–AUC, confusion matrices, and calibration analysis rather than relying on accuracy alone.

B. Performance Evaluation of the TCSA Model

Model training on the dataset evaluates accuracy and loss across epochs to optimize parameters and refine feature learning. Figure 3 illustrates performance improvements over 50 epochs using RoBERTa and CapsNets for feature extraction, aspect identification, and sentiment classification. The TCSA model is trained with a batch size of 32, Adam optimizer (learning rate 0.001), dropout 0.50, early stopping, learning-rate reduction, Sparse Categorical Cross-Entropy loss, and ReLU/Softmax activations.

C. Aspect-Based Sentiment Analysis Using TCSA

Training and validation of the TCSA model over 50 epochs demonstrate its effectiveness in aspect-based sentiment analysis, as shown in Figure 3. Validation accuracy matches the training curve, demonstrating model generalization without overfitting, while the close tracking of the curves suggests model robustness. Similar to accuracy trends, training and validation loss drop quickly and plateau. The loss reduction and plateau indicate early model stabilization in training. The small gap between training and validation loss during training supports model generalization without a significant performance difference between the training and validation sets. These trends reflect a balance between learning ability and data generalization. For a large dataset such as 515K hotel reviews, the proposed TCSA model effectively classified sentiment and aspects with high accuracy and low loss across training and validation phases.

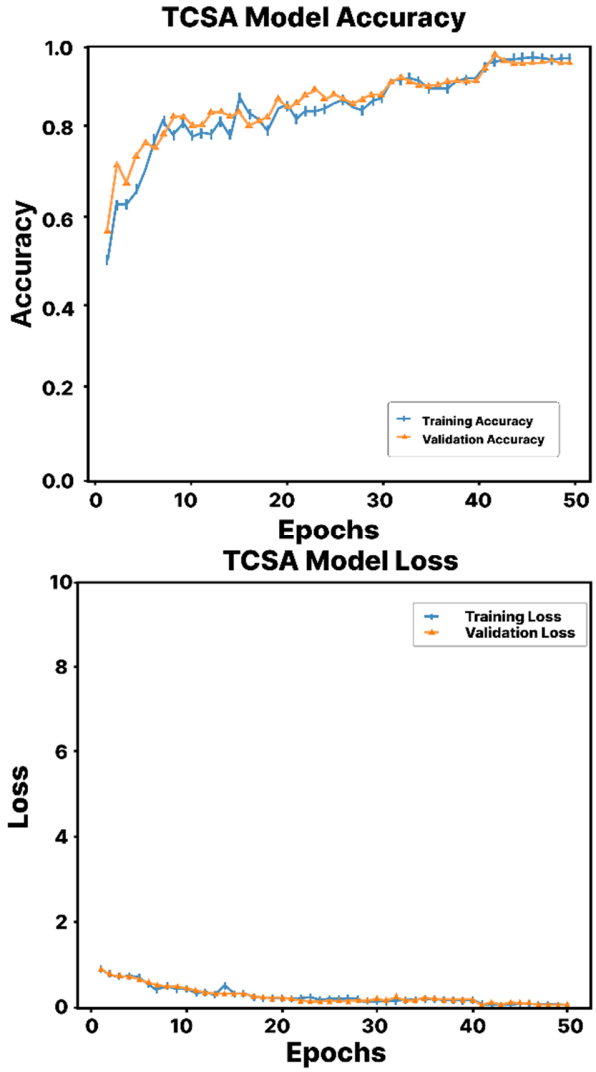

Fig. 3. Results from 50 iterations of the TCSA model on the 515K hotel review dataset.

D. ROC Analysis of TCSA Model

As depicted in Figure 4, the proposed TCSA model achieved a high AUC of 0.98 on the hotel review dataset, indicating almost perfect predictive output. For the proposed TCSA model, the ROC curve approaches the upper left corner, indicating high TP rates and low FP rates on different thresholds, indicating a good classification ability. The TCSA model's superiority over TextCNN (AUC = 0.85) and BiLSTM (AUC = 0.78) demonstrates its advanced class discrimination ability. The TCSA model outperforms BERT (AUC = 0.90) and DS-Caps (AUC = 0.91) in sentiment analysis tasks.

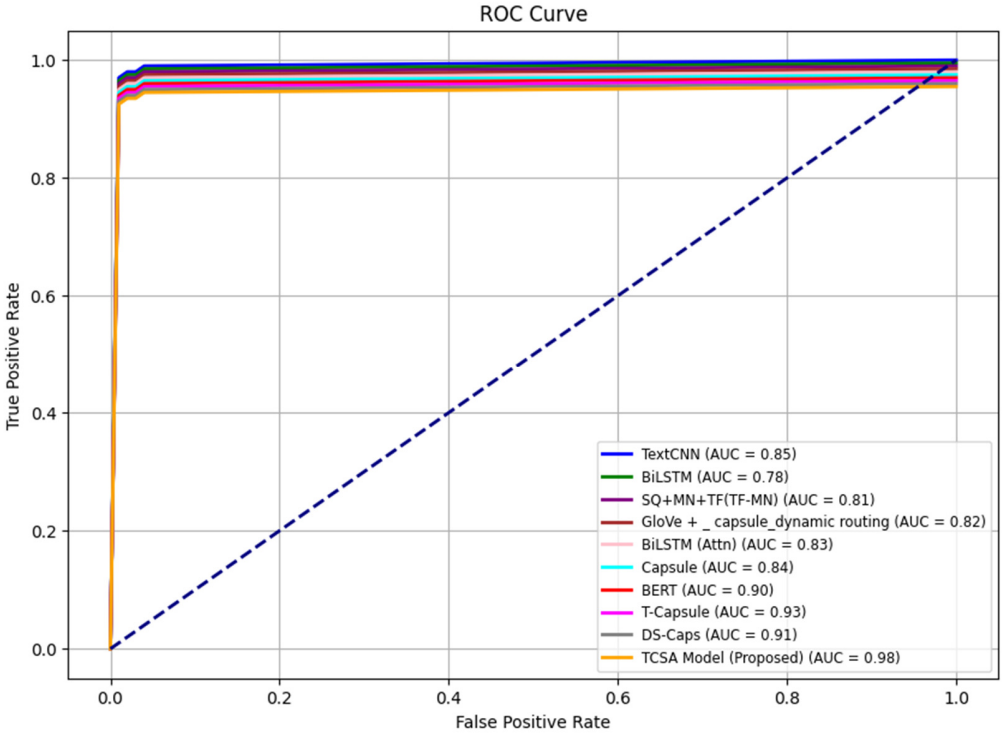

Fig. 4. ROC curve for TCSA on 515K hotel review dataset.

E. Calibration Analysis of TCSA Model

Figure 5 shows the TCSA model's calibration curve for the dataset, illustrating its probabilistic prediction ability. As observed in the TCSA calibration curve, the solid red line mostly follows the dashed line of perfect calibration, especially at the lower and higher probability spectra. This alignment suggests that the model's sentiment probabilities match the observed outcomes. In the mid-probability range (0.4-0.7), the TCSA model sometimes overpredicts or underpredicts positive sentiments. Despite these minor discrepancies, TCSA calibrates well, confirming its ability to predict sentiment and aspect outcomes across large datasets.

F. TCSA Model Predictions on Test Data

The confusion matrix of the proposed TCSA, as exhibited in Figure 6, demonstrates its strong sentiment and aspect classification performance on test data. From the 4,090 samples, the model correctly identified 3,111 as positive reviews and 819 as negative reviews. It also found 76 FP and 84 FN, where the positive samples were mislabeled. The model accurately distinguishes sentiments, but it often misclassifies TN as positives. Finally, TCSA handles diverse sentiment classifications with robust generalizability and precision.

G. Comparison of the TCSA Performance with Existing Models

Table I presents a comparative analysis of the proposed TCSA model with existing state-of-the-art approaches on the 515K hotel review dataset. Traditional models, such as TextCNN and BiLSTM, achieved F1-scores of 85.27% and 80.31%, respectively, while transformer-based baseline models, such as BERT and DeBERTa-V3-base-ABSA, showed improved performance with an F1-score of 89.51% and 89.55%, respectively. Capsule-based variants, such as T-Caps and DS-Caps, further enhanced performance, achieving F1-scores of 92.69% and 92.60%, respectively. In contrast, the proposed TCSA model significantly outperformed all compared methods, as shown in Figure 7, attaining a precision of 97.20%, a recall of 98.12%, an F1-score of 97.30%, and an accuracy of 97.15%, and thus demonstrating the effectiveness of integrating RoBERTa, CapsNets, multi-head attention, and data augmentation within a unified framework.

TABLE I. PERFORMANCE COMPARISON ON 515K HOTEL REVIEW DATASET

Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
TextCNN [1]	84.37	86.02	85.27	85.77
BiLSTM [1]	78.37	82.36	80.31	80.70
SQ+MN+TF (TF-MN) [10]	77.73	78.06	80.70	81.87
GloVe + Shallow Capsule [4]	80.26	79.56	80.26	79.79
BiLSTM (Attn) [1]	82.48	76.85	79.57	81.13
Capsule [1]	81.52	85.21	83.32	83.71
BERT [1]	88.59	90.45	89.51	89.86
DeBERTa-V3-base-ABSA [30]	90.94	90.32	89.55	84.91
RGAT-BERT [31]	80.92	78.40	66.18	71.13
T-Capsule [1]	91.69	93.70	92.69	94.04
DS-Caps [19]	92.50	92.10	92.60	92.80
TCSA model (proposed)	97.20	98.12	97.30	97.15

H. Ablation Study and Component-wise Analysis

Table II presents an ablation study on the hotel review dataset, showing that the TCSA model achieves the best

performance across all metrics. Removing the CapsNet or multi-head attention leads to a significant drop in precision, recall, F1-score, and accuracy, underscoring their roles in hierarchical aspect modeling and global context learning. Performance further degrades without data augmentation or when RoBERTa is replaced by static embeddings, confirming that all components are crucial to TCSA's effectiveness.

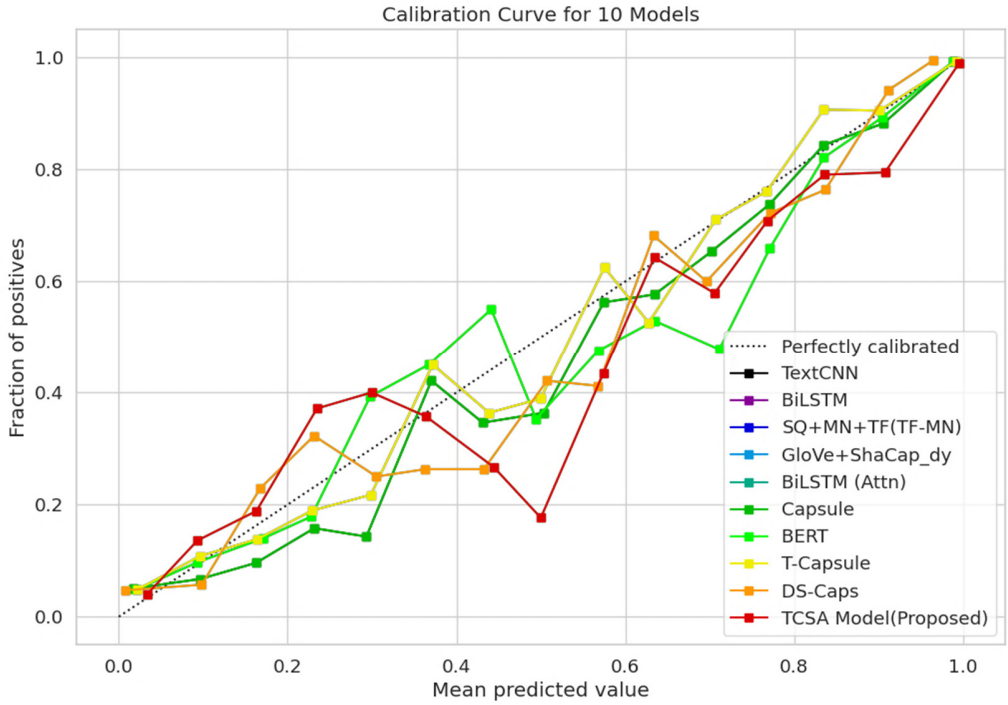

Fig. 5. Calibration curve for TCSA on the hotel review dataset.

TABLE II. COMPONENT-WISE IMPACT ON PERFORMANCE AND ABLATION STUDY

Model variant	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Full TCSA (RoBERTa + CapsNet + MH-Attn + Augmentation)	97.2	98.12	97.3	97.15
Without CapsNets (RoBERTa + MH-Attn)	91.4	92.1	91.7	92.0
Without multi-head attention (RoBERTa + CapsNet)	92.0	93.0	92.5	92.8
Without data augmentation	94.1	94.6	94.3	94.5
Without RoBERTa (Static contextual embeddings + CapsNet)	88.9	90.2	89.5	89.8

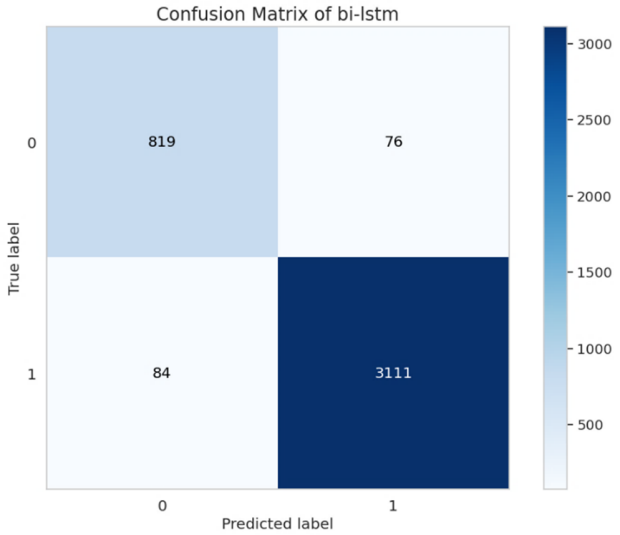

Fig. 6. Confusion matrix for TCSA on test data.

I. Cross-Domain Generalization Analysis

Table III demonstrates that the proposed TCSA model achieves 97.20% precision, 98.12% recall, and 97.30% F1-score on the 515K hotel review dataset, while maintaining competitive performance on the SemEval-2014 restaurant dataset with 92.45% precision, 93.10% recall, and 92.75% F1-score. The observed reduction of approximately 4–5% across metrics reflects expected domain shift effects. Overall, the results demonstrate the model's ability to generalize effectively across domains without domain-specific adaptation.

TABLE III. CROSS-DOMAIN PERFORMANCE COMPARISON

Dataset	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
515K hotel reviews dataset [27]	97.20	98.12	97.30	97.15
SemEval-2014 [32]	92.45	93.10	92.75	92.90

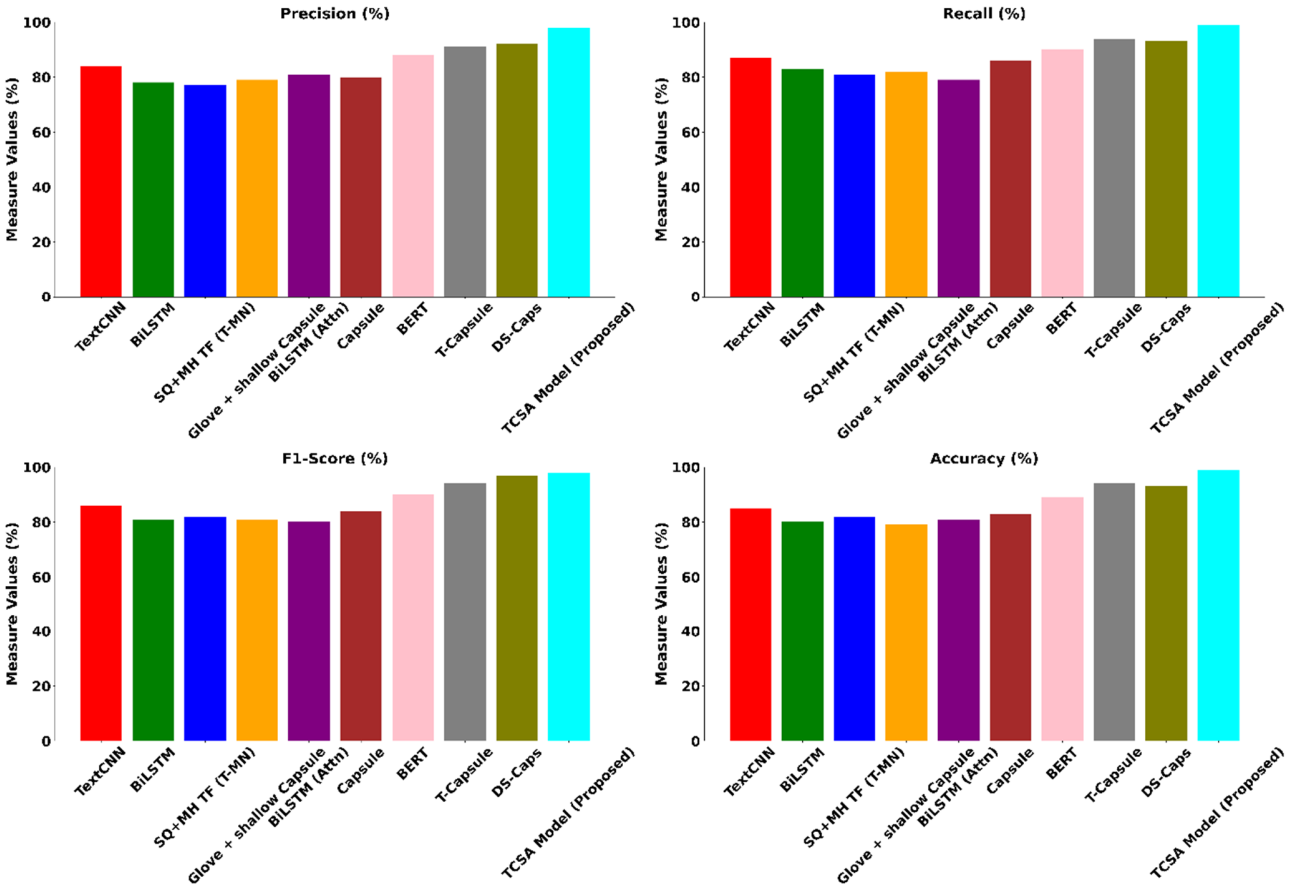

Fig. 7. Comparison of TCSA with other state-of-the-art algorithms.

V. CONCLUSIONS

This study presented the Transformer-Capsule Network for Sentiment and Aspect Analysis (TCSA) model, a novel framework that addresses key limitations of existing pipeline-based, CNN-RNN, and transformer-only approaches. The novelty of the proposed TCSA model lies in its end-to-end architecture that implicitly learns aspect–sentiment representations through capsule network dynamic routing, thereby eliminating explicit aspect extraction and reducing error propagation while preserving hierarchical part–whole relationships within text. The framework further introduces a dual-level contextual modeling strategy, where frozen BERT embeddings provide stable semantic grounding and fine-tuned RoBERTa captures task-adaptive global contextual dependencies. The integration of post-capsule multi-head self-attention enables refined focus on salient aspect–sentiment entities, effectively combining hierarchical reasoning with long-range contextual awareness.

In addition, the linguistically informed data augmentation strategies improve robustness under severe class imbalance and linguistic variability, enhancing generalization across domains. Extensive experiments on the large-scale 515K hotel review dataset demonstrate that the TCSA model achieves strong performance with an accuracy 97.15%, a precision of 97.20%, a recall of 98.12%, and an F1-score of 97.30%, outperforming strong transformer and capsule-based baseline models while maintaining strong cross-domain transferability. Beyond its current formulation, the TCSA architecture is inherently extensible. Thus, future research will focus on integrating Graph Neural Networks (GNNs) to explicitly model syntactic, dependency, and discourse-level relationships between aspect terms, enabling richer relational reasoning. Additionally, lightweight adaptation strategies, such as meta-learning and parameter-efficient fine-tuning, can be explored to improve scalability and effectiveness in low-resource, multilingual, and real-time sentiment analysis applications.

REFERENCES

[1] B. Chen, Z. Xu, X. Wang, L. Xu, and W. Zhang, "Capsule Network-Based Text Sentiment Classification," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 698–703, 2021, https://doi.org/10.1016/j.ifacol.2021.04.160.

[2] B. Liu, *Sentiment Analysis and Opinion Mining*. Cham, Switzerland: Springer International Publishing, 2012.

[3] Z. M. Zohreh Madhoushi, A. R. Hamdan, and S. Zainudin, "Aspect-Based Sentiment Analysis Methods in Recent Years," *Asia-Pacific Journal of Information Technology & Multimedia*, vol. 08, no. 01, pp. 79–96, Jun. 2019, https://doi.org/10.17576/apjitm-2019-0801-07.

[4] P. Demotte, K. Wijegunarathna, D. Meedeniya, and I. Perera, "Enhanced Sentiment Extraction Architecture for Social Media Content Analysis Using Capsule Networks," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 8665–8690, Mar. 2023, https://doi.org/10.1007/s11042-021-11471-1.

[5] P. Pookduang, R. Klangbunrueang, W. Chansanam, and T. Lunrasri, "Advancing Sentiment Analysis: Evaluating RoBERTa against Traditional and Deep Learning Models," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 20167–20174, Feb. 2025, https://doi.org/10.48084/etasr.9703.

[6] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for Target-Dependent Sentiment Classification," in *Proceedings of the 26th International Conference on Computational Linguistics*, Osaka, Japan, Dec. 2016, pp. 3298–3307.

[7] Z. Wang, H. Wu, H. Liu, and Q.-H. Cai, "Bert-Pair-Networks for Sentiment Classification," in *2020 International Conference on Machine Learning and Cybernetics*, Adelaide, Australia, Dec. 2020, pp. 273–278, https://doi.org/10.1109/ICMLC51923.2020.9469534.

[8] A. Yadav and D. K. Vishwakarma, "Sentiment Analysis Using Deep Learning Architectures: A Review," *Artificial Intelligence Review*, vol. 53, no. 6, pp. 4335–4385, Aug. 2020, https://doi.org/10.1007/s10462-019-09794-5.

[9] K. M. Karaoğlan and O. Fındık, "Extended Rule-Based Opinion Target Extraction with a Novel Text Pre-Processing Method and Ensemble Learning," *Applied Soft Computing*, vol. 118, Mar. 2022, Art. no. 108524, https://doi.org/10.1016/j.asoc.2022.108524.

[10] M. Jiang, J. Wu, X. Shi, and M. Zhang, "Transformer Based Memory Network for Sentiment Analysis of Web Comments," *IEEE Access*, vol. 7, pp. 179942–179953, 2019, https://doi.org/10.1109/ACCESS.2019.2957192.

[11] Q.-H. Vo, H.-T. Nguyen, B. Le, and M.-L. Nguyen, "Multi-Channel LSTM-CNN Model for Vietnamese Sentiment Analysis," in *2017 9th International Conference on Knowledge and Systems Engineering*, Hue, Vietnam, Oct. 2017, pp. 24–29, https://doi.org/10.1109/KSE.2017.8119429.

[12] P. Liu, X. Qiu, and X. Huang, "Recurrent Neural Network for Text Classification with Multi-Task Learning." arXiv, 2016, https://doi.org/10.48550/ARXIV.1605.05101.

[13] J. Han, J. Chen, P. Chen, J. Liu, and D. Peng, "Chinese Text Sentiment Classification Based on Bidirectional Temporal Deep Convolutional Network," *Computer Applications and Software*, vol. 36, no. 12, pp. 225–231, 2019.

[14] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "CNN for Situations Understanding Based on Sentiment Analysis of Twitter Data," *Procedia Computer Science*, vol. 111, pp. 376–381, 2017, https://doi.org/10.1016/j.procs.2017.06.037.

[15] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," *IEEE Access*, vol. 6, pp. 23253–23260, 2018, https://doi.org/10.1109/ACCESS.2017.2776930.

[16] R. Johnson and T. Zhang, "Effective Use of Word Order for Text Categorization with Convolutional Neural Networks," in Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, CO, USA, 2015, pp. 103–112, https://doi.org/10.3115/v1/N15-1011.

[17] M. Cai, "Sentiment Analysis of Tweets Using Deep Neural Architectures," in *32nd Conference on Neural Information Processing Systems*, Montréal, Canada, 2018, pp. 1–8.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Pre-Training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, MN, USA, 2019, pp. 4171–4186, https://doi.org/10.18653/v1/N19-1423.

[19] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," in *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 2017.

[20] P. Demotte and S. Ranathunga, "Dual-State Capsule Networks for Text Classification." arXiv, 2021, https://doi.org/10.48550/ARXIV.2109.04762.

[21] X. Yu, S.-N. Luo, Y. Wu, Z. Cai, T.-W. Kuan, and S.-P. Tseng, "Research on a Capsule Network Text Classification Method with a Self-Attention Mechanism," *Symmetry*, vol. 16, no. 5, Apr. 2024, Art. no. 517, https://doi.org/10.3390/sym16050517.

[22] J. Su, S. Yu, and D. Luo, "Enhancing Aspect-Based Sentiment Analysis with Capsule Network," *IEEE Access*, vol. 8, pp. 100551–100561, 2020, https://doi.org/10.1109/ACCESS.2020.2997675.

[23] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding." arXiv, Jan. 02, 2020, https://doi.org/10.48550/arXiv.1906.08237.

[24] J. Kim and J.-H. Lee, "Modeling Inter-Speaker Relationship in XLNet for Contextual Spoken Language Understanding." arXiv, 2019, https://doi.org/10.48550/ARXIV.1910.12531.

[25] B. AlBadani, R. Shi, J. Dong, R. Al-Sabri, and O. B. Moctard, "Transformer-Based Graph Convolutional Network for Sentiment Analysis," *Applied Sciences*, vol. 12, no. 3, Jan. 2022, Art. no. 1316, https://doi.org/10.3390/app12031316.

[26] S. Longpre, Y. Lu, Z. Tu, and C. DuBois, "An Exploration of Data Augmentation and Sampling Techniques for Domain-Agnostic Question Answering," in *Proceedings of the 2nd Workshop on Machine Reading for Question Answering*, Hong Kong, China, 2019, pp. 220–227, https://doi.org/10.18653/v1/D19-5829.

[27] J. Liu, "515K Hotel Reviews Data in Europe." Kaggle, 2014, [Online]. Available: https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe.

[28] A. Gulli and S. Pal, Deep Learning with Keras: Implement Neural networks with Keras on Theano and TensorFlow. Birmingham, UK: Packt Publishing, 2017.

[29] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[30] H. Yang, C. Zhang, and K. Li, "PyABSA: A Modularized Framework for Reproducible Aspect-Based Sentiment Analysis," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, Birmingham, UK, Oct. 2023, pp. 5117–5122, https://doi.org/10.1145/3583780.3614752.

[31] X. Bai, P. Liu, and Y. Zhang, "Investigating Typed Syntactic Dependencies for Targeted Sentiment Classification Using Graph Attention Neural Network," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 503–514, 2021, https://doi.org/10.1109/TASLP.2020.3042009.

[32] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation*, Dublin, Ireland, 2014, pp. 27–35, https://doi.org/10.3115/v1/S14-2004.

AUTHORS PROFILE

Dr. Sumalakshmi C H, assistant professor at KL University, is a Ph.D. holder, specializing in AI-based facial expression recognition and DL techniques, with 13+ years of experience, multiple awards, publications, and RPA certification.

Ms. Laxmi Pamulaparthy is an assistant professor in AI and ML at VBIT, Hyderabad, and a Ph.D. scholar at KLEF, with 11 years of academic experience and a research focus on NLP and DL.