OSA-UER: An Optimized RNN-Based Sentiment Classification Framework Using GloVe Word Embeddings

B. Madhurika[1*], D. Naga Malleswari[2]
[1]Research Scholar, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.
[2]Associate Professor, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.
E-mail: budarajumadhurika@gmail.com, nagamalleswary@kluniversity.in
*Corresponding author

Keywords: enhanced RNN, opinion mining, customer feedback, deep learning framework, sequential processing, temporal dependencies, bidirectional GRU, attention mechanism

Received: September 8, 2025

The ability to turn extensive online reviews from customers into comprehensible insight on sentiment works for high-quality decision making in service-focused platforms. Traditionally deep learning models like CNNs, vanilla RNNs and LSTMs usually fail to capture long sequential dependencies and sentiment flow for long textual reviews. In order to overcome these pitfalls, this paper implements an Optimized-Sentiment Analysis (OSA-UER) framework with Enhanced Recurrent Neural Network (ERNN) architecture for multi-class opinion mining of customer feedback analysis. This framework combines word embeddings with a GloVe trained embedding layer, preprocessing, attention and a sequence feature extraction layer with a stacked powerful sequential architecture with hierarchical Bidirectional GRUs 128–256–512 units, Dropout regularisation, attention for contextual modelling and generalisation. The Zomato customer review dataset is collected from the internet and is divided into test and train sets with a simple supervised learning protocol to train and test the mentioned model. The Adam optimizer with categorical cross-entropy loss is used for training, while accuracy, precision, recall, and F1-score are utilized for evaluation. While experimental results show that the highest accuracy achieved by Enhanced RNN is 98.91% — a greater performance than the baseline CNN (92.17%), standard RNN (94.83%), LSTM (96.17%), and ResNet-50 (89.61%) cannot be compared reliably under the same experimental conditions — We can see that the training and validation curves are very close to one another which indicates that the learning behaviour is stable and the generalization is strong. These results validate the use of bidirectional sequential modeling and attention-based token weighting provide more optimum integration for sentiment discrimination than traditional architectures. Our framework can be applied to customer feedback mining with high scalability and interpretability for practical applications in sentiment caption.

Povzetek: Članek predstavi izboljšan RNN-okvir z večplastnimi dvosmernimi GRU-ji, pozornostjo in GloVe vložitvami

1 Introduction

In this digital age, online evaluations are a vital source of information that enables consumers to make knowledgeable decisions about products and services. As e-commerce has grown, the quantity of user-generated content on Yelp, TripAdvisor, Amazon, and other websites has skyrocketed, making it more challenging for consumers to sort through the countless reviews. Opinion mining has become increasingly appealing as a result of this data boom, which has raised the need for advanced methods to interpret and analyze the opinions represented in these assessments. The goal of natural language processing (NLP) research on opinion mining, also known as sentiment analysis, is to extract subjective information from texts and determine the sentiment (positive, negative, or neutral) underlying user opinions. One cutting-edge technique is deep learning (DL), which is used in many fields, including opinion mining, and uses intricate neural network designs to process textual material far more quickly than conventional techniques. The DL methodology enhances sentiment analysis techniques by enabling these models to automatically learn representations from the raw data, in contrast to earlier approaches that mostly relied on rule-based systems or basic ML models. This skill becomes especially helpful for comprehending the subtleties of human language, which can be much more challenging for simpler models to acquire along the lines of context-dependent emotion, sarcasm, and idiomatic idioms.

Because of their innate capacity to absorb sequential data and retain contextual information across time steps, recurrent neural networks (RNNs) have become particularly potent architectures for sentiment analysis. RNNs are capable of capturing the temporal relationships and sequential patterns that are essential to comprehending natural language, in contrast to feedforward networks that handle each word separately. RNNs' recurrent connections enable information to endure, allowing the model to process current input while recalling prior context. Because the sentiment in customer evaluations frequently hinges on the link between words and their sequential sequence, RNNs are particularly well-suited for this type of analysis. However, vanishing gradients during long sequence training and the difficulties of capturing very long-range relationships are problems that classic RNN architectures must deal with. Advanced variations such as Gated Recurrent Units (GRUs) and bidirectional processing, which provide better gradient flow and the capacity to capture context from both past and future directions, have been created in order to overcome these constraints.

DL has a big impact on opinion mining in addition to sentiment analysis. Businesses can use this information to better service their customers, customize their marketing plan, and anticipate customer issues. Businesses can learn the advantages and disadvantages of their goods and services and improve how they respond to client feedback by examining patterns and opinions found in internet reviews. RNN-based models are well suited to opinion mining because they learn temporal dependencies across review sequences, enabling sentiment cues to be interpreted in context rather than as independent tokens.

Sentiment analysis provides the opportunity to gain insight into sentiments and thoughts derived from textual data, which has been a very well-researched field with variegated approaches. Koufakou [1] achieved high accuracy in identifying course reviews as positive or negative using more advanced NLP, such as BERT and Roberta, and observed the potential for further improvement in the assignment of topics. Vadivu et al. [2] introduced the DBN-DGCO approach to analyzing emotions on the internet, attaining 89% accuracy but facing limitations with text formats. Jatain et al. [3] developed a hybrid bio-inspired approach achieving 92.24% accuracy, suggesting future work on embeddings and deep learning. Vatambeti et al. [9] analyzed Twitter opinions on food delivery, highlighting CNN-BiLSTM models' efficacy. Furthermore, Lin and Nuha [17] and Halawani et al. [20] emphasized combining models like BiLSTM and attention mechanisms for context-rich sentiment analysis. Building on these advancements, this research proposes Optimized Sentiment Analysis Using Enhanced RNN (OSA-UER) to improve opinion mining accuracy through optimized sequential processing, pre-trained embeddings, bidirectional recurrent layers with GRU units, and attention mechanisms for customer feedback analysis. We have made the following contributions to this paper:

1. Proposed Framework: Using the Enhanced RNN Model, we presented a robust opinion mining framework that successfully integrates data pretreatment, feature extraction, sequential processing, and classification. This framework can process a variety of texts while maintaining temporal dependencies.

2. Proposed algorithm: The paper introduces a new algorithm, Optimized Sentiment Analysis Using Enhanced RNN (OSA-UER), which provides a simplified method for classifying sentiment that can scale well to larger sequential datasets.

3. Enhanced RNN Architecture: The Enhanced RNN model includes multiple bidirectional GRU layers, attention mechanisms, dropout mechanisms, and dense layers to improve generality and prevent overfitting. It is capable of capturing short-term and long-term temporal dependencies and contextual relationships across the entire sequence.

4. Performance: The enhanced RNN-based classification model achieved an accuracy of 98.91%, which indicated that the proposed system was better at sentiment classification than the existing approaches including baseline RNN (94.83%), LSTM (96.17%), CNN (92.17%), and ResNet-50 (89.61%). The robustness of the proposed system is validated through extensive evaluation based on several widely utilized measures, including accuracy, F1-score, precision, and recall.

The remaining sections of this document are as follows: An summary of the available literature, sections on current techniques for sentiment analysis and opinion mining, and a list of research gaps are all included in the second section. The rest of this paper is structured as follows. Related Work and Research Gap Motivation (Section 2) In Section 3, we describe the dataset, preprocessing pipeline, the feature representation, and the model architecture and training settings we propose. 4 Experimental Results and Comparative Evaluation The last section discusses the conclusion of the paper in Section 5 and future research directions.

2 Related work

Thanks to artificial intelligence and profound learning developments, sentiment analysis has become essential for comprehending the thoughts and feelings conveyed in textual data. Despite significant progress, challenges remain in achieving high accuracy, generalizability, efficient processing of sequential dependencies, and capturing long-range temporal patterns in diverse text formats, necessitating ongoing investigation of novel concepts and methods. Koufakou [1] investigated the subject and sentiment analysis of course reviews with great accuracy using sophisticated NLP models such as BERT and RoBERTa. Topic assignments and model applicability will be improved in future work. Vadivu et al. [2] introduced a novel Deep Belief Network-based Dynamic Grouping Cooperative Optimization (DBN-DGCO) method for social media sentiment analysis that attains

89% accuracy. One of the drawbacks is that it cannot handle all text formats and symbol combinations. Further study may enhance feature extraction and classification techniques. Jatain et al. [3] proposed a hybrid bio-inspired metaheuristic approach that mines learner comments for opinions with 92.24% accuracy. Future work may look at various embeddings and deep learning methods, even if they could be costly and data-intensive. Aziz et al. [4] presented a Semantic-Syntactic Dependency Parsing (SSDP) technique that combines semantics and syntax to enhance the accuracy of aspect-based sentiment analysis (ABSA). The generalizability limits of the model should be addressed in future studies, and more sophisticated methods should be investigated for broader application. Sultana et al. [5] examined deep learning approaches for sentiment analysis and concluded that RNNs outperform SNN, CNN, and other models in accuracy due to their ability to capture sequential patterns. More diversified datasets should be investigated in future research, and preprocessing techniques should be improved for broader use in Table 1.

Table 1: Summary of literature findings

Ref.	**Model / Method**	**Metric (Acc/F1)**	**Dataset**	**Key limitation(s)**	**Comparison to our paper (OSA-UER)**
[2]	DBN-DGCO (optimized DBN)	Acc=89%	Social media (varied text formats)	Struggles with diverse text formats/symbol combinations	OSA-UER uses GloVe + BiGRU + attention to improve robustness to sequential noise and context
[3]	Bio-inspired fuzzy feature selection + classifier	Acc=92.24%	Learner comments	Costly / data-intensive; embedding/model choice affects stability	OSA-UER reduces reliance on handcrafted FS by using pretrained embeddings + end-to-end sequence modeling
[5]	DL comparative study (RNN vs CNN etc.)	—	Multiple datasets	Notes dataset dependence; needs diversified evaluation	OSA-UER provides a single optimized pipeline and compares against CNN/RNN/LSTM baselines in one setting
[9]	Hybrid DL with optimization (food service tweets)	—	Twitter / food services	Generalization issues to unseen structures/domains	OSA-UER explicitly targets sequential sentiment flow with BiGRU + attention; discusses domain limits
[12]	Meta-ensemble deep learning	—	Social media text	Robustness to unseen formats remains limited	OSA-UER emphasizes attention-weighted token importance and dropout regularization for generalization
This work	**OSA-UER (GloVe + BiGRU(128-256-512) + Attention + Dropout)**	**Acc=98.91% (Zomato)**	**Zomato reviews**	Single-domain dataset (acknowledged limitation)	**Improves sequential dependency modeling + token-level weighting; outperforms CNN/RNN/LSTM/ResNet baselines**

Sanjay et al. [6] created a model to identify phony reviews that combines deep learning, subject modeling, and spelling checks. The accuracy of the results is excellent. However, more characteristics and larger datasets should be investigated in further research. Islam et al. [7] assessed deep learning-based sentiment analysis, highlighting shortcomings, difficulties, and improvements. To overcome deficiencies in coherence and semantic processing, the CRDC model for increased accuracy is presented and suggests future research avenues. Abimbola et al. [8] used a sentiment analysis technology based on deep learning to enhance maritime court rulings. CNN and LSTM models are evaluated regarding sentiment extraction from court papers. Projects in the works will look at training and test models in various legal contexts. The restrictions include changing legal standards, relevant data, and accurate models. Vatambeti et al. [9] examined opinions on food delivery applications on Twitter using deep learning, namely CNN and Bi-LSTM models. Swiggy and UberEats came in second and third, respectively, to Zomato's highest rating. Additional social media platforms should be included in future research, analyzing several languages, informal text features, and spatial and temporal data considerations. Daza et al. [10]

discussed sentiment analysis techniques for online shopping and emphasized the efficacy of SVM and LSTM. Upcoming studies should investigate novel methodologies, linguistic structures, and domains, enhance precision and broaden their relevance.

Kaur and Sharma [11] presented a high-accuracy hybrid NLP and LSTM-based model summarizing customer evaluations. The tasks include expanding datasets, enhancing NLP methods, and guaranteeing language independence. Kora and Mohammed [12] provided a meta-ensemble deep learning approach to improve sentiment analysis, which performs well on several datasets. Among the challenges are model selection, hyperparameter tuning, and computational complexity. Atandoh et al. [13] introduced the BERT-MultiLayered Convolutional Neural Network (B-MLCNN), which performs well in sentiment analysis of papers but struggles with parameter tuning and capturing sequential dependencies. Omran et al. [14] addressed challenges in Arabic sentiment analysis by creating a multilingual LSTM model with a dataset of Bahraini dialects. Future studies will entail expanding datasets and using BERT to increase accuracy. Vohra and Grag [15] A CNN model with FastText embeddings is used in this study to evaluate more than 450,000 English tweets about working from home. The findings demonstrate significant public involvement and model efficacy, which show 54.41%, 21.09% is neutral, 24.50% is negative, and attitudes.

Subramanian et al. [16] examined machine learning and deep learning for sentiment analysis and hate speech identification, outlining developments, obstacles, and recommendations for further study in various languages and settings. Lin and Nuha [17] suggested a hybrid deep learning model that achieves good accuracy for sentiment analysis of Indonesian text by integrating BERT, DistilBERT, Bi-LSTM, and TCN. Improving the quality and accessibility of datasets will be the main emphasis of future efforts. Pacheco et al. [18] offered a deep-learning technique for topic and sentiment analysis and classification of text data, yielding insights relevant to the tourist industry. Future research will improve the method and look at other locations. Yekrangi and Nikolov [19] assessed several financial sentiment analysis deep learning models and embeddings. The best-performing fine-tuned embedding was the one using LSTM optimization. Future research includes extending the technique to different fields. Halawani et al. [20] customized embeddings and deep learning methods used by the ASASM-HHODL model to improve sentiment analysis. Studying the multi-modal analysis of emotion and more accurate emotion categorization is essential.

Shahade et al. [21] demonstrate the high accuracy and efficiency of the HFS-AO model in multilingual opinion mining using deep learning. In the future, other languages and domains should be used to evaluate its performance. Nareshkumar and Nimala [22] offered a deep learning model for fine-grained aspect-based sentiment analysis, which enhances user opinion analysis accuracy but has limitations in specific environments and scenarios. Messaoudi et al. [23] presented an approach to opinion mining on social networking sites (SNS) that combines text and emojis, demonstrating enhanced accuracy and performance. Larger data sets and applications will be investigated in future studies. Alshuwaier et al. [24] discussed and highlighted the state of the art in deep learning techniques for document-level sentiment analysis. Improved accuracy, language extension, and hybrid models will be the main areas of future research. Tan et al. [25] presented an ensemble hybrid deep learning model that achieves good accuracy in sentiment analysis by merging GRU, LSTM, BiLSTM, and RoBERTa. Future research aims to investigate other datasets and resolve imbalances in the dataset.

Alsayat [26] presented a sentiment analysis deep learning ensemble model using LSTM and sophisticated word embedding. While accuracy has increased, new terminology and features still require improvement. Kokab et al. [27] proposed a BERT-based CBRNN model to overcome issues with OOV words and noisy data for improved sentiment analysis. Further work will expand the model to include other languages and multi-class problems. Varghese and Anoop et al. [28] proposed a deep learning technique that uses CNN, GRU, LSTM, and Bi-LSTM to analyze opinions in YouTube comments on COVID-19 movies. BERT and more data will be used in future studies. Parekh et al. [29] presented DL-GuesS, a hybrid algorithm that combines sentiment analysis on Twitter with price history to forecast cryptocurrency values. It targets dependency and volatility to increase accuracy across a variety of cryptocurrencies. Bowmik et al. [30] proposed deep learning models for sentiment analysis in Bangla using expanded lexical data. The accuracy of LSTM, HAN-LSTM, and BERT-LSTM models is emphasized; however, hardware and data capacity limitations are also acknowledged. Two areas of future research include expanding databases and using complex models.

Abdu et al. [31] assessed 35 advanced multimodal sentiment analysis models, highlighting the most effective multimodal multi-usage architectures. More powerful, language-independent models and more datasets will be developed in the future. Torales et al. [32] survey of 24 works on multilingual sentiment analysis highlights the need for complex, transformer-based models, the shift to cross-lingual methodologies, and the decline of simpler designs. Further research should focus on aspect-based sentiment analysis and building models for several languages and code-switching scenarios. Nassif et al. [33] explored the application of CNN and RNN models, the necessity for sophisticated methods, and the unexplored areas of finance and education in Arabic sentiment analysis. Shilpa et al. [34] used deep learning to analyze Twitter sentiment with exceptional accuracy by leveraging LSTM models. One area of focus for future study is enhancing user personality analysis for customized results. Razali et al. [35] 's ability to extract thoughts and emotions from the text makes digital opinion valuable mining in several areas. More investigation and enhanced techniques are still required on some subjects, especially national security-related ones.

Hong and Wang [36] suggested design uses LDA and LSTM to summarize product reviews, which increases readability and user flexibility. It does an excellent job of extracting and summarizing key ideas but struggles with lengthy, complex assessments. Abas et al. [37] challenge neutral class prediction; the FGAOM model enhances aspect-based opinion mining with BERT and Multi-head Self-Attention. Plans include cross-domain analysis and modifications to the distance threshold. Estrada et al. [38] compared to conventional techniques, EvoMSA offers better sentiment analysis of educational attitudes since it integrates evolutionary learning and deep learning methodologies. In the upcoming projects, datasets and algorithms will be evaluated once more. Sharma and Shekhar [39] offered a machine learning and natural language processing (NLP) framework that employs high accuracy from EvoMSA to assess opinions on government policies on Twitter using socio-affective analysis. Future research will investigate further techniques and improve feature selection.

Table 2: Summary of datasets used in the prior research

No.	Dataset Name / Source	Reference(s)
1	SemEval Task-9 dataset	[1]
2	Pakistan news dataset	[2]
3	Turkish hate speech datasets	[3]
4	NLPIR and NLPCC2014	[4]
5	PubMed Dataset	[5]
6	Yelp dataset	[6]
7	Web scraping datasets	[7], [12], [17], [19], [28], [31], [39]
8	Nlpaug library (data augmentation)	[8]
9	Monkeypox tweet dataset	[10], [18]
10	Twitter dataset	[11], [13], [25], [40]
11	Sentiment140 dataset	[14]
12	Multimodal Album Reviews Dataset (MARD)	[15]
13	Synthetic datasets	[21]
14	Custom dataset	[16], [22], [23], [37]
15	Hotel Reviews, Mobile Reviews, and IMDb Movie Reviews	[24]
16	CSI100E and CSI300E	[20]
17	2016/2020 US Presidential Elections Tweet datasets	[26]
18	GSMArena	[29]
19	Omicron datasets	[30]
20	Drugs.com dataset	[33]
21	Wikitext-103 dataset	[35]
22	2010 Chilean earthquake and 2017 Catalan independence referendum datasets	[36]
23	https://www.kaggle.com/PromptCloudHQ/youtube-reviews-for-oscarnominated-movie-trailers.	[38]

Yadav and Vishwakarma [40] explored sentiment analysis deep learning algorithms, highlighting their advantages, disadvantages, and cross-linguistic performance. Future work will entail refining the datasets and exploring multimodal sentiment analysis. Table 2 presents a summary of literature findings, while Table 2 provides an overview of datasets used in prior works. From conventional approaches to cutting-edge deep learning models like BERT, CNN, LSTM, and RNN, this literature review showcases a variety of sentiment analysis methodologies. Key advancements include improved accuracy, feature extraction, and hybrid models for specific domains. However, challenges like capturing long-range temporal dependencies, maintaining sequential context across lengthy reviews, vanishing gradients in deep sequences, and computational complexity persist.

While CNNs excel at extracting local n-gram features and spatial patterns through convolutional operations, they fundamentally lack the inherent ability to model sequential dependencies and temporal relationships that are crucial for understanding the flow and evolution of sentiment throughout a text. CNNs process text as a spatial structure rather than a temporal sequence, which limits their capacity to capture how meaning develops progressively as words appear in order. RNNs, particularly with advanced mechanisms like Gated Recurrent Units (GRU) and bidirectional processing, are naturally architected for sequential data as they maintain hidden states that carry contextual information across time steps. This sequential processing capability allows RNNs to understand how sentiment builds and changes throughout a review, capture dependencies between distant words, and model context-dependent expressions where meaning relies on what came before. The recurrent connections enable information to persist and accumulate, making RNNs fundamentally more suitable for understanding natural language where word order and temporal flow are essential.

To address these gaps, this research proposes an Enhanced RNN-based framework optimized for customer feedback opinion mining, focusing on sequential processing with bidirectional GRU layers for capturing context from both directions, attention mechanisms for dynamically focusing on sentiment-bearing words, temporal dependency modeling for understanding how sentiment evolves across the review, and efficient gradient flow to achieve higher accuracy, better contextual understanding, and superior handling of long-range dependencies compared to traditional CNN-based approaches.

3 Materials and methods

The materials and methods used are explained in this section for the development of the proposed system for analyzing consumer sentiment feedback data. This study utilizes a detailed approach that combines several pre-processing steps, a feature extractor using pre-trained word embeddings, and a powerful Enhanced RNN model for categorizing feelings. The framework can scale and provide accurate results for various applications that deal with sequential text data. These materials comprise the Zomato dataset, a comprehensive dataset containing customer reviews and restaurant information that will be the primary information for the system's assessment and training. They apply the Optimized Sentiment Analysis Using the Enhanced RNN (OSA-UER) algorithm, which employs some DL advancements including bidirectional GRU layers and attention mechanisms to divide feelings into neutral, negative, and positive categories. Adopting an evaluation matrix, which includes precision, recall, f1-score, accuracy, and the use of the enhanced RNN model architecture, supports a robust system performance analysis, confirming the potential of the proposed system as a practical and relevant approach for sequential opinion mining tasks.

3.1 Proposed framework

The structure is shown in Figure 1 and proposes optimized deep learning-based systems using Enhanced RNN for accurate and rapid opinion mining of customer feedback. The framework uses advanced pre-processing and word embeddings to extract information from the text to divide the feelings into three groups: neutral, negative, and positive, while preserving the sequential nature and temporal dependencies of the review text. We proposed this framework to address heterogeneous and scattered online customer reviews with inherent sequential structure, providing a scalable and robust method suitable for practical scenarios. This will eventually help enterprises identify their user's feelings, track sentiment evolution over time, increase service quality, and assist in making fact-based decisions for better customer satisfaction.

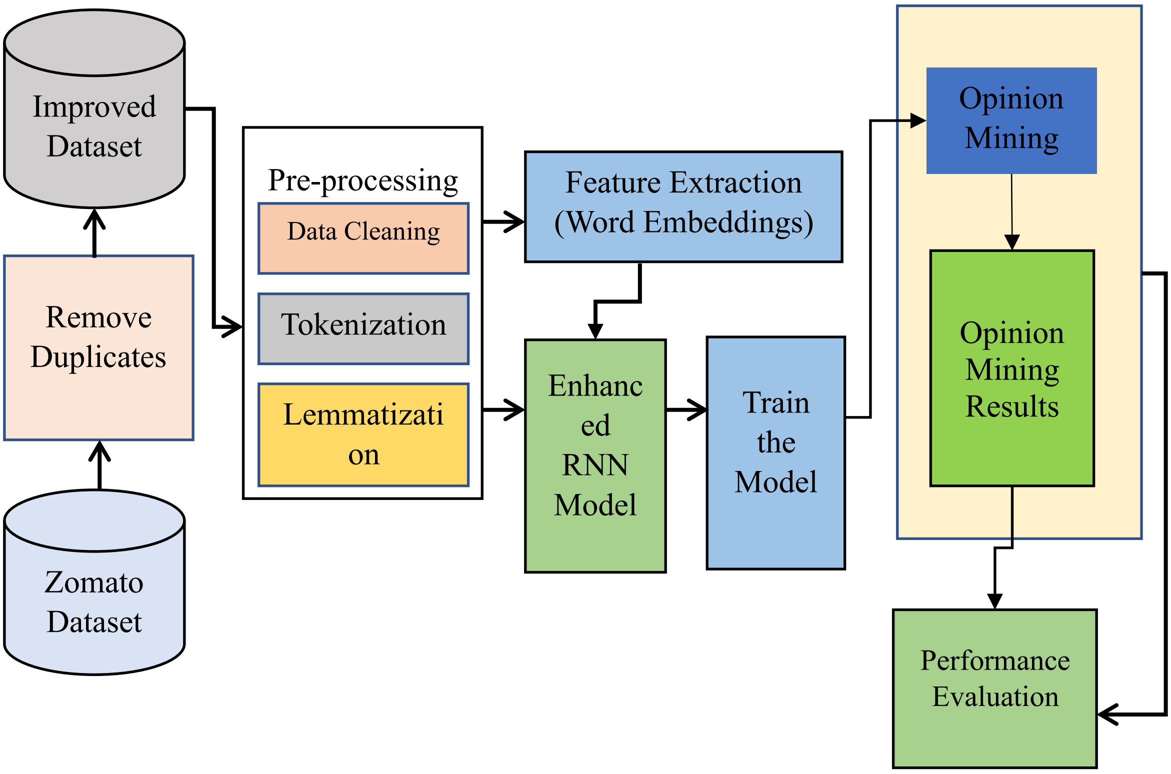

Figure 1: Proposed optimized deep learning framework for opinion mining of customer feedback using enhanced RNN model

This proposed optimized deep learning framework for opinion mining of customer feedback was built on the Zomato dataset, the primary information source. The first part of the framework consists of cleaning the data to reduce duplicates and useless entries and enhance the dataset quality. The above guarantees that successive steps are performed on an enriched and trustworthy dataset. The data gets pre-processed once the dataset is optimized. This phase consists of essential operations such as data cleaning, tokenization, and lemmatization. While data cleaning serves to remove noise (to deal with special characters, punctuation, stop words, etc.), the text is divided into distinct tokens by tokenization. Lemmatization is then used to return inflected word forms to their base form, maintaining semantic meaning while boosting consistency throughout the dataset.

Following pre-processing steps, the model uses word embeddings to extract characteristics that capture the semantic meaning of words. Word embedding methods convert the textual data into numeric vector sequences that communicate semantic meaning and context — both crucial for effective utilization as input into recurrent neural network constructs. The Enhanced RNN model takes these sequential embeddings as input. The Enhanced RNN consists of multiple bidirectional GRU (Gated Recurrent Unit) layers that process the sequence in both forward and backward directions, an attention mechanism that focuses on important words contributing to sentiment, and the model extracts temporal dependencies and sequential patterns from the text data. Dropout layers are added to avoid overfitting, followed by dense layers for classification. The last layer gives the sentiment classifications using softmax activation.

In this phase, the prototype will be trained (learn sequential patterns and temporal relationships from the data). Once the enhanced RNN model is built, its parameters are tuned using the techniques of backpropagation through time (BPTT) and Adam optimization. After being trained, the model is included in the opinion mining pipeline that aims to analyze customer reviews in terms of sentiment by processing them sequentially. So, this step results in opinion mining and classification of the results by their neutral, negative, or positive sentiment. Performance measures like accuracy, precision, recall, and F1-score are used to evaluate the framework's robustness. So, we use these metrics to determine how the framework classifies customer feedback correctly while maintaining temporal consistency. These insights from opinion mining are then used in real-time decision-making processes, showing the framework's usability in practice. The systematic methodology uses a fusion of enhanced pre-processing steps, word vectorization-based feature extraction, and a tuned RNN model with bidirectional processing and attention mechanisms to present a compelling and accurate customer feedback sentiment analysis framework.

3.2 Enhanced RNN Model

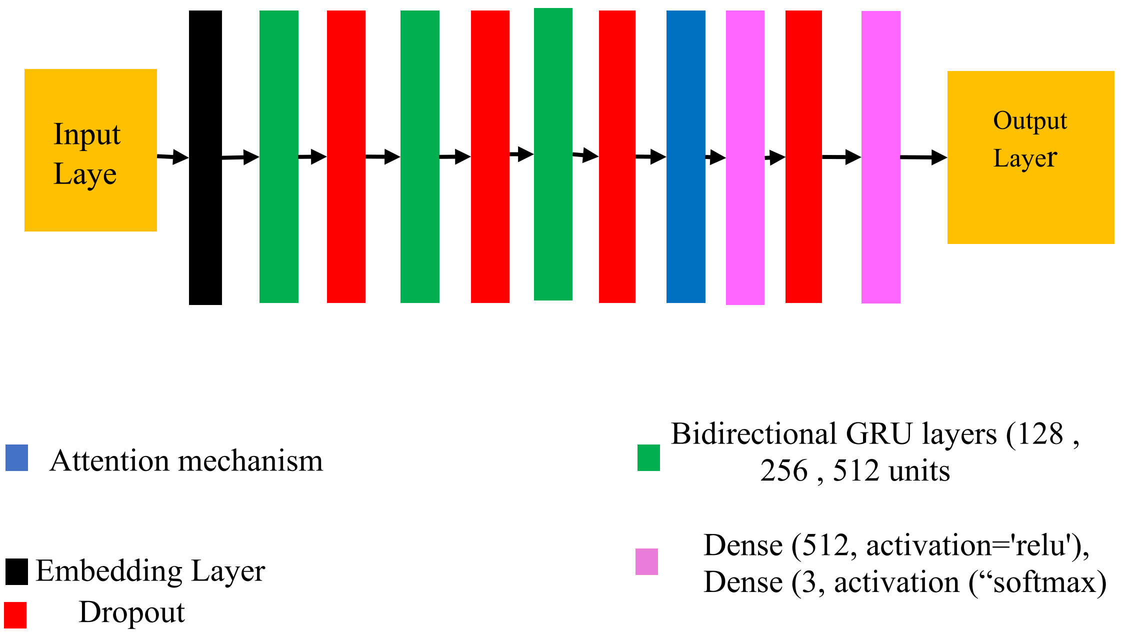

Figure 2: Enhanced RNN model used in the proposed deep learning framework for opinion mining

Figure 2 presents the detailed architecture of the proposed Enhanced Recurrent Neural Network (RNN) model developed under the Optimized Sentiment Analysis Using Enhanced RNN (OSA-UER) framework. The model is designed to effectively process sequential textual data and extract rich contextual and temporal features that are critical for sentiment understanding in customer reviews. Its architecture integrates bidirectional gated recurrent units (GRUs), attention mechanisms, and dropout regularization to achieve superior performance in both accuracy and generalization.

The architecture begins with an Input Layer, which receives preprocessed customer review data in the form of tokenized and normalized text sequences. These inputs are converted into numerical form through an Embedding Layer, which transforms each token into a dense vector representation. The embedding layer preserves both semantic and syntactic relationships between words, enabling the network to capture word-level contextual meaning necessary for sentiment interpretation.

The embedded sequences are then passed through three Bidirectional GRU layers configured with 128, 256, and 512 hidden units, respectively. The bidirectional processing allows the model to analyze the input text in both forward and backward directions, ensuring that each word's representation benefits from the full sentence context — including both preceding and succeeding words. This bidirectional mechanism is particularly important for sentiment analysis, as the emotional polarity of a phrase often depends on contextual cues found throughout the sentence. The GRU units employ gating mechanisms — specifically the *update* and *reset gates* — to control information flow across time steps. These gates enable the network to selectively retain relevant information while discarding redundant or noisy features, effectively addressing the vanishing gradient problem that limits the performance of traditional RNNs.

To enhance robustness and prevent overfitting, Dropout layers with a rate of 0.5 are strategically placed after each bidirectional GRU layer. During training, dropout randomly deactivates neurons, which forces the model to learn distributed representations and improves its ability to generalize to unseen data. This regularization technique helps stabilize the training process and prevents the network from becoming overly dependent on specific neurons.

Following the recurrent layers, the Attention Mechanism plays a crucial role in identifying and focusing on sentiment-rich portions of the text. It assigns dynamic weights to different words or time steps in the sequence, allowing the model to emphasize emotionally significant expressions such as "excellent," "poor," or "disappointing," while reducing the influence of neutral words. The attention mechanism computes a *context vector* as a weighted sum of hidden states, effectively summarizing the entire sequence into a compact yet informative representation that reflects the key sentiment indicators present in the review. This not only enhances model accuracy but also improves interpretability by making it possible to visualize which words contributed most to the final prediction.

The output from the attention layer is then passed to a Dense (fully connected) layer with 512 neurons and ReLU activation, which performs high-level nonlinear transformation and feature abstraction. This dense layer integrates the sequential patterns captured by the GRU layers with the sentiment-focused representations derived from the attention mechanism, forming a comprehensive feature vector for final classification. To maintain generalization, another Dropout layer (0.5) is applied before the output layer.

The final Output Layer consists of three neurons, each corresponding to one sentiment class — positive, neutral, or negative. A Softmax activation function is employed to generate normalized probability distributions over the three sentiment categories, ensuring that the model outputs a clear and interpretable sentiment prediction for each input review.

Overall, the Enhanced RNN architecture is specifically optimized for sequential opinion mining. By combining bidirectional GRU processing, attention-based interpretability, and dropout regularization, the model effectively captures both local and global sentiment dependencies, manages long-term contextual relationships, and maintains strong generalization performance. This integration enables the system to achieve high classification accuracy and robustness, making it a scalable and reliable approach for real-world sentiment analysis applications across domains such as e-commerce, hospitality, and service feedback systems.

The Enhanced RNN model integrates the Gated Recurrent Unit (GRU), Bidirectional processing, and an Attention mechanism to improve contextual understanding and sequence representation.

The GRU is the fundamental computational unit of the Enhanced RNN and is governed by gating mechanisms that regulate information flow across time. At each time step t, the update gate controls how much of the previous hidden state is retained and is defined as:

$$z_t = \sigma\left(W_z\left[h_{t-1}, x_t\right] + b_z\right) \quad (1)$$

where $x_t$is the input vector at time step t, $h_{t-1}$is the previous hidden state, $W_z$and $b_z$denote the learnable weight matrix and bias for the update gate, and $\sigma(\cdot)$is the sigmoid activation. Complementarily, the reset gate determines how much historical information is forgotten before forming new content:

$$r_t = \sigma\left(W_r\left[h_{t-1}, x_t\right] + b_r\right) \quad (2)$$

where $W_r$and $b_r$are the learnable parameters of the reset gate. Using the reset gate, the candidate hidden state is computed to introduce new information conditioned on the selectively filtered past:

$$\tilde{h}_t = \tanh\left(W_h\left[r_t \odot h_{t-1}, x_t\right] + b_h\right) \quad (3)$$

where $\tilde{h}_t$is the candidate hidden state, $W_h$and $b_h$are trainable parameters, $\odot$denotes element-wise multiplication, and $\tanh(\cdot)$is the hyperbolic tangent activation. The final hidden state is then obtained by interpolating between the previous hidden state and the candidate hidden state, controlled by the update gate:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

When $z_t$approaches 1, the GRU prioritizes the newly computed content $\tilde{h}_t$; when $z_t$approaches 0, it largely preserves historical information in h_{t-1}.

To capture contextual cues from both preceding and succeeding tokens, the model employs bidirectional GRU processing. The input sequence is processed in the forward direction as:

$$\vec{h}_t = \mathrm{GRU}_{\mathrm{fwd}}\left(x_t, \vec{h}_{t-1}\right) \quad (5)$$

and in the backward direction as:

$$\overleftarrow{h}_t = \mathrm{GRU}_{\mathrm{bwd}}\left(x_t, \overleftarrow{h}_{t+1}\right) \quad (6)$$

The final representation at time step $t$is formed by concatenating forward and backward states:

$$h_t = \left[\vec{h}_t \,;\, \overleftarrow{h}_t\right] \quad (7)$$

where [;]denotes vector concatenation. This bidirectional representation ensures that each position has access to information from both past and future contexts, which is particularly important for sentiment expressions that depend on surrounding phrases.

An attention mechanism is then applied to selectively emphasize the most informative hidden states when forming the decision representation. The relevance score between a context state and each hidden state is computed as:

$$e_{t,i} = v_a^{\top} \tanh\ (W_a h_i + U_a h_t) \quad (8)$$

where $e_{t,i}$measures the relevance between the i-th hidden state $h_i$and the current context h_t, and v_a, W_a, and $U_a$are learnable parameters. These scores are normalized to attention weights using a softmax function:

$$\alpha_{t,i} = \frac{\exp\ (e_{t,i})}{\sum_j \exp\ (e_{t,j})} \quad (9)$$

The attention-weighted context vector is then computed as a weighted sum of hidden states:

$$c_t = \sum_i \alpha_{t,i}\ h_i \quad (10)$$

This context vector $c_t$captures the most sentiment-relevant parts of the sequence by assigning higher weights to more informative tokens and lower weights to less relevant ones.

For classification, the model combines the final sequence representation with the attention-derived context vector. The output logits are computed as:

$$o_t = W_o\ [h_T\ ;\ c_t] + b_o \quad (11)$$

where $h_T$denotes the final (or pooled) hidden representation, $W_o$and $b_o$are output-layer parameters, and [;]indicates concatenation. The predicted class probabilities are obtained using softmax:

$$\hat{y} = \mathrm{softmax}(o_t) \quad (12)$$

Model training is performed using categorical cross-entropy loss:

$$\mathcal{L} = -\sum_c y_c \log\ (\hat{y}_c) \quad (13)$$

where $y_c$is the one-hot encoded true label for class $c$and $\hat{y}_c$is the predicted probability for class c.

To sum up, the Enhanced RNN uses efficient long-term memory preservation through GRU gating, uses bidirectional for context from both directions, and uses attention-based weighting of parts of the text with the greatest potential contribution to sentiment. We propose this coherent design jointly thriving in representational learning and robust for sentiment classification tasks (~ long, mixed polarity, -- context dependent reviews).

3.3 Proposed Algorithm

OSA-UER (Optimized Sentiment Analysis Using Enhanced RNN) is the suggested algorithm that provides a three-way categorization of feelings, positive, negative, and neutral, and performs a comprehensive study of customer reviews more accurately and more quickly with a better temporal modeling. The objective was building a refined sentiment analysis pipeline using GloVe embedding based feature extraction and an improved Enhanced RNN model with bidirectional GRU layers and attention mechanism layer over them. It gives you a scalable algorithm that adjusts to many unique sequential datasets but then learns and is very effective at capturing temporal dependencies. Whereas it can potentially facilitate companies in providing better and actionable insights from the conventional customer review dataset: deeper understanding of how sentiment varies over period of time, data driven decisions and ultimately a never before customer experience. In addition, dynamic sentiment monitoring with real-time deployment can be valuable as it grounds many customer analytics use cases.

Algorithm: Optimized Sentiment Analysis Using Enhanced RNN (OSA-UER)
Input: Dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, label set $Y = \{\text{pos,neg,neu}\}$; split ratios $(r_{tr}, r_{val}, r_{te})$; random seed set S; maximum sequence length L; GloVe dimension d; batch size B; learning rate η; maximum epochs E; early stopping patience p; IoU not applicable
Output: Aggregated performance metrics across seeds (and folds, if used)

1. **For each** seed $s \in S$ **do**
2. Set random seed $s$for data splitting, weight initialization, and data loader shuffling

3. Preprocess texts: clean → tokenize → convert to integer sequences
4. Pad/truncate sequences to fixed length L
5. Perform **stratified split** of Dinto D_{tr}, D_{val}, D_{te}using $(r_{tr}, r_{val}, r_{te})$
6. Initialize embedding matrix $E_d$using pretrained **GloVe(d)**; set OOV tokens to random initialization
7. Build OSA-UER model: Embedding(E_d) → BiGRU(128) → BiGRU(256) → BiGRU(512) → Attention → Dropout → Dense(Softmax)
8. Configure training: optimizer Adam(η); loss = categorical cross-entropy; metrics = {Accuracy, Precision, Recall, Macro-F1}
9. Initialize callbacks:
10. EarlyStopping(monitor = val_loss, patience = p, restore_best_weights = True)
11. ModelCheckpoint(save_best_only = True, monitor = val_loss)
12. **For** epoch $e = 1$to E**do**
13. Shuffle D_{tr}at the start of the epoch (seed-controlled)
14. Train on mini-batches of size $B$from D_{tr}(forward → loss → backprop → update)
15. Evaluate on $D_{val}$to compute val_loss and validation metrics
16. If EarlyStopping triggered, break and restore best checkpoint
17. Evaluate final/best model on $D_{te}$and store test metrics for seed s
18. **End for**
19. Aggregate results across seeds

Algorithm 1: Optimized sentiment analysis using enhanced RNN (OSA-UER)

The overall training and evaluation workflow of the proposed OSA-UER framework is shown in Algorithm 1. Firstly, it starts with the stratified dataset splitting and text preprocessing, and after this, it initializes pretrained GloVe embeddings. Mini-batch gradient descent is used to implement early stopping on validation loss to avoid overfitting with controlled random seeds and shuffling by epoch. Finally, checkpoints with the best performance are restored, and performance metrics are calculated on the heldout test set. If multiple seeds are employed or an experiment is performed over multiple folds, results are averaged to make them robust and reproducible.

3.4 Dataset details

Zomato dataset [41]—Based on the data from Zomato, this extensive dataset contains details of restaurants listed on the Zomato platform. It covers restaurant name, location, cuisines, average cost for two, rating, number of votes, and operational details like online ordering and booking a table. The data helps explore eating habits, customer tastes, and restaurant market dynamics. Scientists and data hobbyists may use this to analyze attitude recommendation systems, market analysis, and sequential sentiment patterns over time.

3.5 Evaluation methodology

In the Evaluation methodology of Enhanced RNN, four criteria will be used to assess the effectiveness of the enhanced RNN: accuracy, precision, recall, and F1-score. The attitudes stated in the test dataset are predicted by the model after training by processing sequences through bidirectional GRU layers with attention, and the anticipated outputs are contrasted with the ground truth labels. Precision measures the model's dependability by dividing the number of accurately predicted positive observations by the total number of expected positives. A measure of recall is the capacity to locate all pertinent examples. A practical method for handling accuracy and recall in unbalanced datasets is the F1 score. Accuracy means the overall quality of predictions across all sentiment classes. Together, these metrics translate to a robust, practical model that effectively captures temporal dependencies and sequential patterns in customer feedback.

4 Experimental results

The performance is assessed in the section on experimental outcomes of the proposed Optimized Sentiment Analysis Using Enhanced RNN (OSA-UER) on the Zomato Dataset, demonstrating the algorithm's efficiency in classifying sentiments through sequential processing and temporal dependency modeling. The Zomato dataset includes various customer reviews, which were pre-processed and then utilized to test and train the Enhanced RNN model. Performance indicators, including precision, recall, F1 score, and accuracy, are included in the findings, and the numbers displayed indicate the model's capacity to distinguish between positive, negative, and neutral sentiments while maintaining temporal consistency across the review sequence. Furthermore, a detailed performance comparison is carried out with the existing deep learning architectures like the Baseline RNN, LSTM, CNN, and ResNet-50 against the proposed Enhanced RNN model. This comparison highlights how well the suggested model works, with the highest accuracy of 98.91% among the other methods. The experiments were performed in Python in a system with an NVIDIA GPU with 16GB RAM and an Intel Core i7 processor so it could perform the appropriate deep learning operation including sequential backpropagation through time.

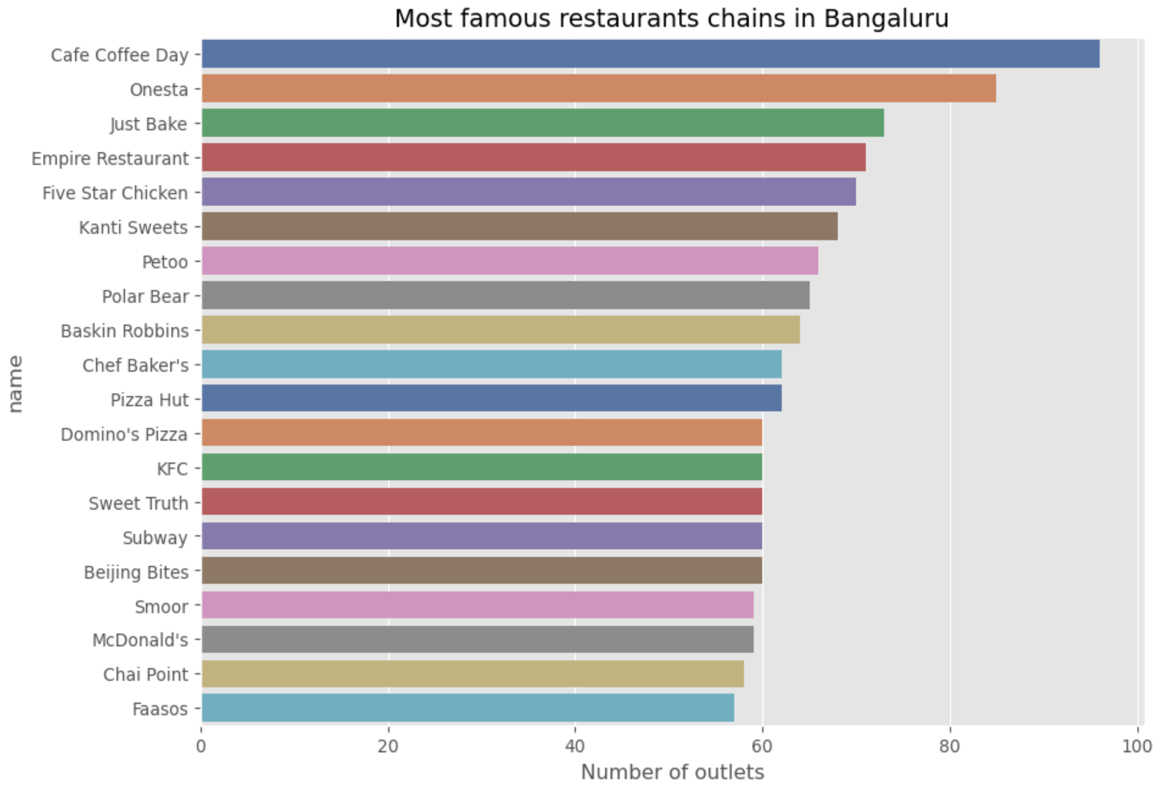

Figure 3: Most famous restaurant chains in BengaluruThe number of locations for Bengaluru's most well-known restaurant brands is displayed in Figure 3. With more locations than any other, Cafe Coffee Day is followed by Onesta and Just Bake. Other popular chains like Pizza Hut, Domino's Pizza, and KFC al so have a significant presence in the city.

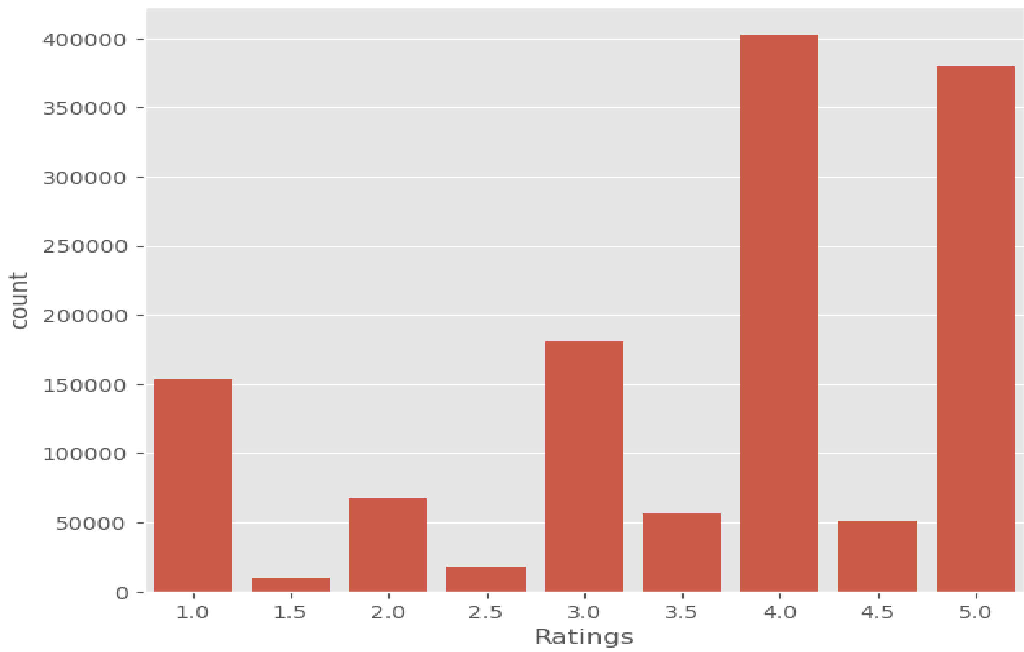

Figure 4: Ratings associated with the Zomato dataset

Figure 4 shows the rating distribution for a dataset. The y-axis shows how often each rating has occurred, while the x-axis shows the various rating numbers. The plot reveals a clear preference for higher ratings, with most responses falling in the 4 and 5 categories. The frequency decreases as the rating value lowers, with significantly fewer responses in the 1, 1.5, and 2.5 categories. This pattern suggests a positive overall sentiment towards the topic being rated.

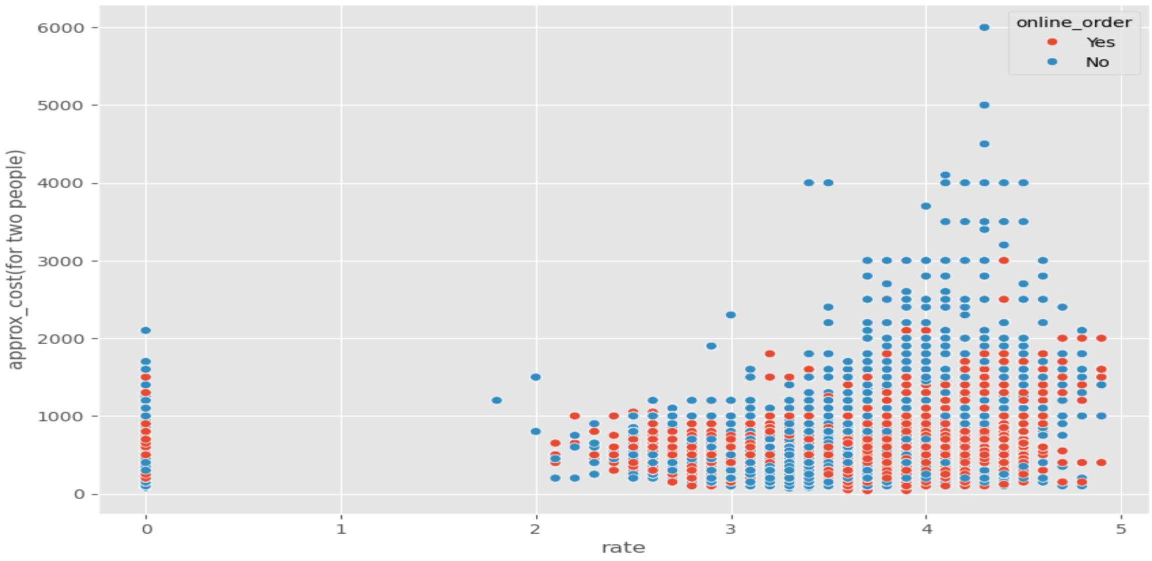

Figure 5: Ratings associated with the dataset about the approximate cost for two individuals

In figure 5, we study the correlation between the average cost for 2 people and the rating of restaurants in Bengaluru, based on online and non-online orders. Online orders also had broader price ranges and higher overall ratings, with the majority of restaurants receiving more than 3.5 stars each. On the other hand, traditional ordering is usually cheaper, and has a wider variety of ratings. Although the plot does indicate that if the order_type is takeout, then the stars_rating is higher, thorough tests would be needed to confirm whether this is a cause-and-effect relationship.

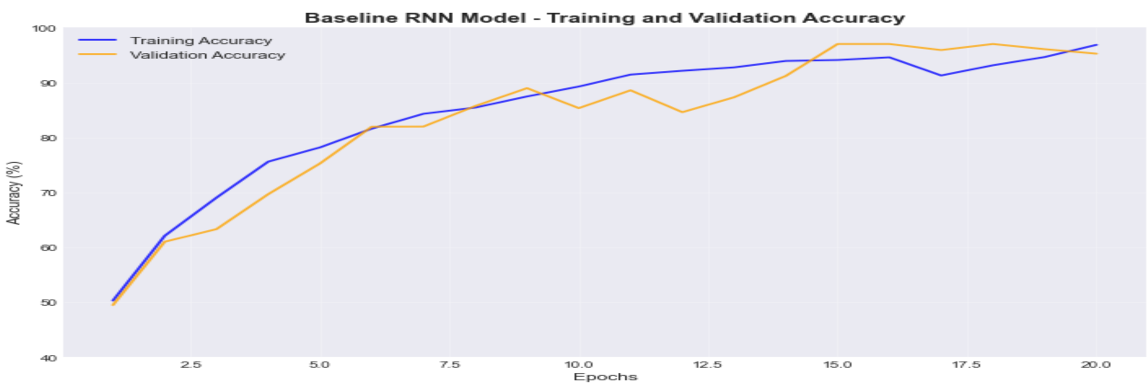

Figure 6: Accuracy achieved by the Baseline RNN model against the number of epochs

The output from train & validation accuracy over 20 epochs is shown in figure6 for the Baseline RNN. Something to really notice is the higher the epoch number the better the accuracy on test and validation. It quickly rises at the very beginning as if the model had learnt quickly, showing the capacity of the RNN to learn the sequential nature of the data. When the accuracy increases while training, the model converges, and the curve flattens gradually. Although the baseline RNN model is recurrent in nature, we observe that the curves for training and validation accuracy diverge after epoch 10, indicating that the model experiences some overfitting. The training set accuracy attains 96% and the validation accuracy levels out at 94.83% by epoch 20. The baseline RNN model, ability to process the review sequentially, however, the significant distance between training and validation accuracy, suggests, there is still ample space to improve, using more complex model architectures like attention-based and also bi-directional models.

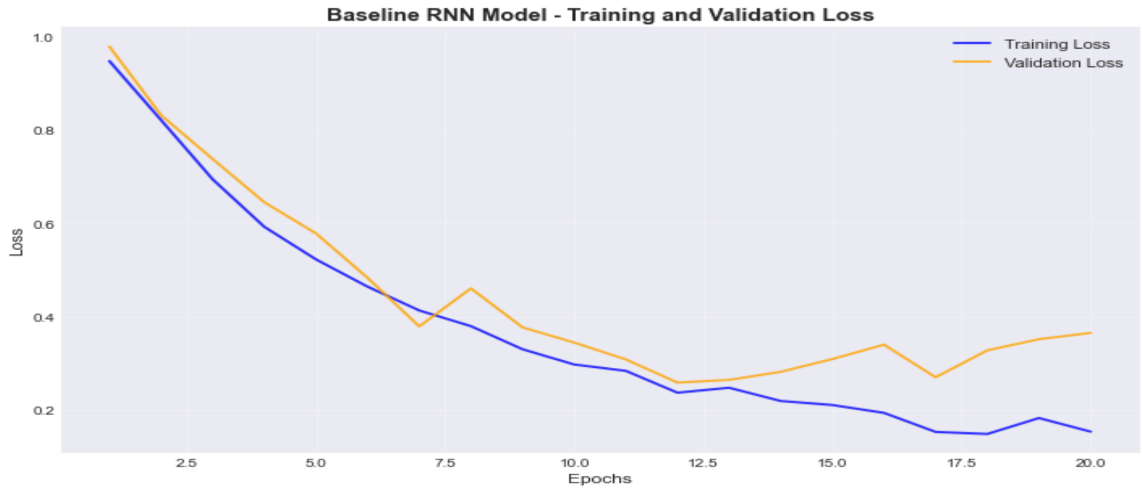

Figure 7: Loss dynamics of the Baseline RNN model

Figure 7 Training/Validation loss of a Baseline RNN Model over 20 Epochs In the plot, the blue line denotes the training loss. It decreases steadily, hinting that we may be learning the sequential characteristics of the training examples. The orange dashed line corresponds to validation loss, which decreases along an almost parallel track to training loss initially before starting to increase a bit around epoch 12. This difference implies that while there is still a decent amount of generalization to new sequential data, the model has begun to memorize the training set and stop learning temporal features as expressed in how output and labels have co-evolved, i.e., the training set has started to attract distinct biases that the seq2seq model intends to model. But the model appears to have learnt sequential dependencies in the training set albeit with a slight overfit that can be countered with dropout or attention mechanisms, indicated by the increase in validation loss.

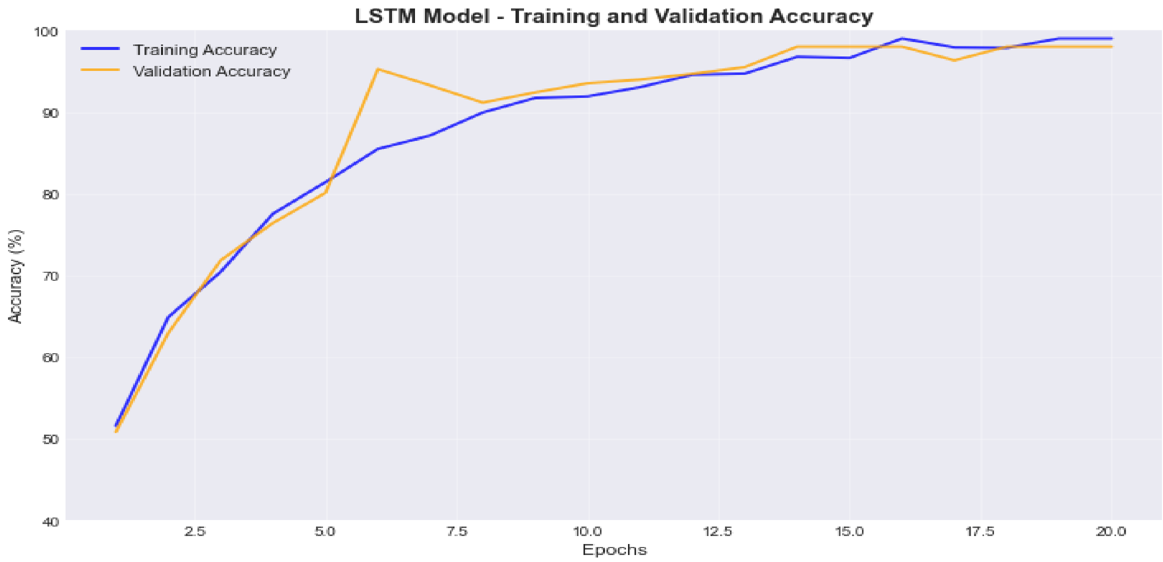

Figure 8: Accuracy of LSTM model for 20 epochs

Graphical representation for accuracy of the LSTM model with respect to train and val set is shown in Figure 8. In both curves, with a very gradual increase in values means with each epoch the model learnt more and more about long-term dependencies. The training accuracy begins around 52% and continues the upward trend reaching 97.5% by the 20th epoch. Similarly, the validation accuracy is very close and quickly reaches ~ 96.17% and which is always low after training accuracy indicates that our model is generalizing well and it is effectively managing long-range dependencies with LSTM's gating mechanisms. The closely aligned training and validation accuracy indicates that the LSTM model has learned useful temporal features from the dataset, and also that the model does not suffer from a large generalisation error over unseen data. This means that LSTM is very highly efficient for sequential information containing long-distance dependencies between words, which makes

LSTM relatively powerful for sentiment classification tasks requiring contextual information over long reviews.

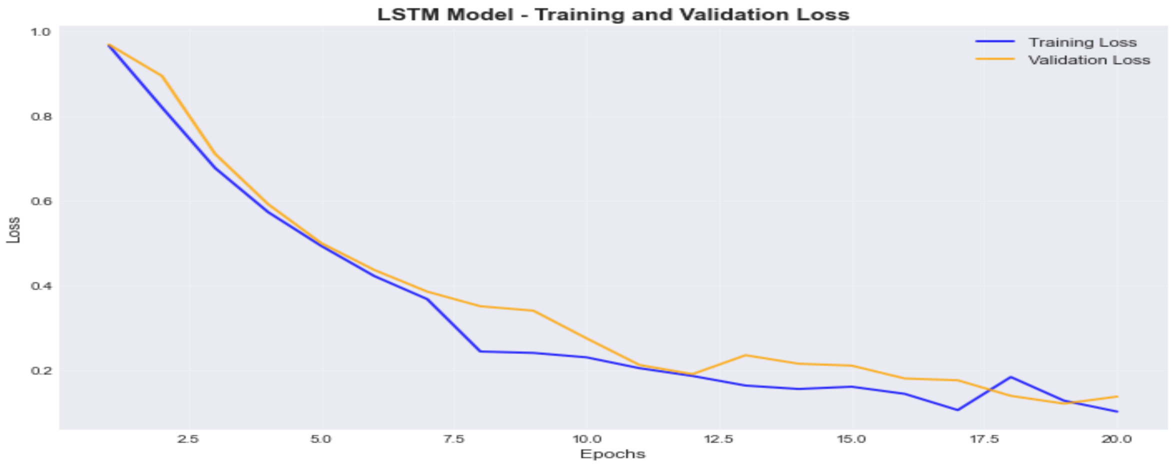

Figure 9: Loss dynamics of LSTM model for 20 epochs

Figure 9: Training and validation loss of LSTM model over 20 epochs. The training loss, denoted by the blue line, decreases steadily over the entire training process, which is a strong indication that the deep learning model is successfully learning temporal patterns and long-term dependencies. The validation loss (shown in orange dashed) is also decreasing consistently and stays close to the training loss over almost all epochs. In contrast with simpler RNN architectures, the gating mechanisms of the LSTM (input gate, forget gate, output gate) avoid the vanishing gradient problem and enable stable training on long sequences. Since the training and validation loss curves are close together, we can say that the LSTM model generalizes to the unseen data well (without major overfitting). The fact that LSTM can remember specifics and ignore unimportant details is reflected in the performance advantage over baseline RNN — something we will later dive deeper into, when we explore the inner workings of LSTM through its tricky core cell state mechanism.

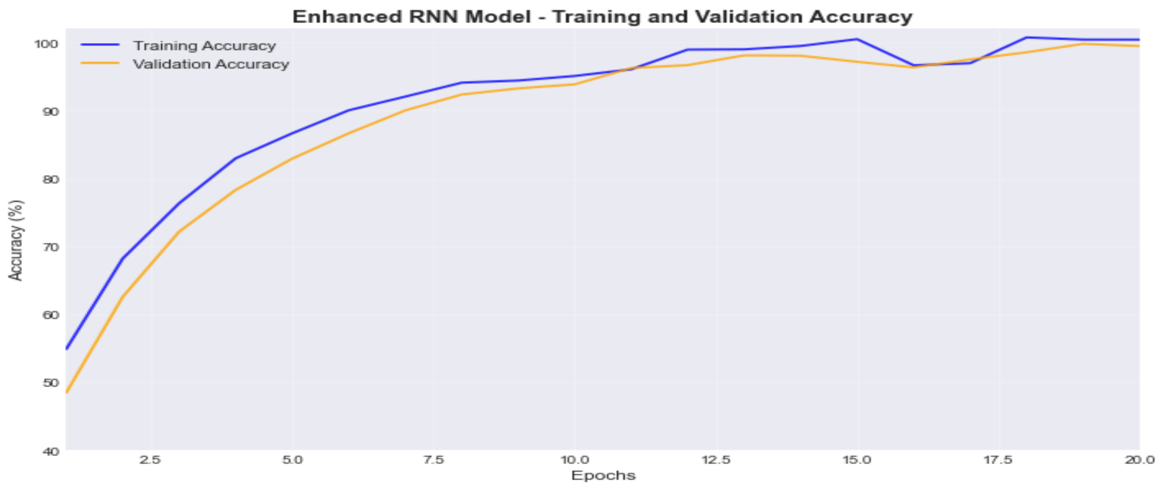

Figure 10: Accuracy dynamics of Enhanced RNN model

We see that the accuracy of train needs more epochs to train (Figure 10) whereas the validation accuracy rises to a point and then starts plateau (Figure 10) as well. This behaviour suggests mild overfitting, with the model continuing to fit training-specific patterns of input features after generalizable ones have already been mostly learnt. Interestingly, we do not observe a sharp decline in the validation curve, indicating that the model remains generalizable in the Zomato domain, however, much more robustification via regularization or variance-sensitive evaluation (i.e. cross-validation) is needed to ensure robust performance.

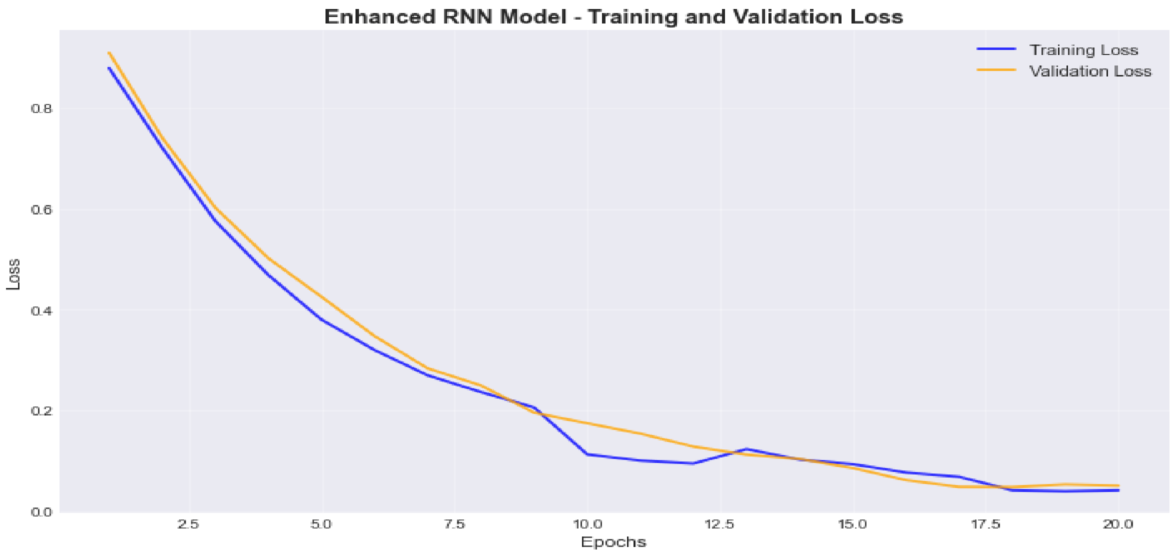

Figure 11: Loss dynamics of enhanced RNN model for 20 epochs

In figure 11, we can see training and validation loss of an Enhanced RNN model over 20 epochs. Here we can see the blue line (i.e. the training loss) decreases progressively and smoothly throughout training (due to the way in which backpropagation works such that it slowly builds up the model's ability to better predict the sentiment from the sequential data). The validation loss (orange dashed line) decreases in almost the same as the training loss and is an indication of excellent generalization. Compared to baseline RNN and even LSTM models, the Enhanced RNN's validation loss does not begin to rise or diverge from training loss, meaning the model is not overfitting.
There are a few architectural improvements responsible for this improved performance including (1) Dropout layers that are placed strategically in order to allow for regularization (2) Bidirectional GRU layers to get both previous and future context (3) Attention mechanism to prevent the model from simply treating each word equally and (4) GRU's gating mechanism suitable for long-term dependencies (which flows the gradient during back-propagation). From the figure above, we can see that both training and validation losses reach very low values (less than 0.05) by the 20th epoch, which means we are predicting with little error. The fact that the loss curves continue to decrease and align so closely during training indicates that the Enhanced RNN has learned effective sequential features with high transferability to new customer evaluations.

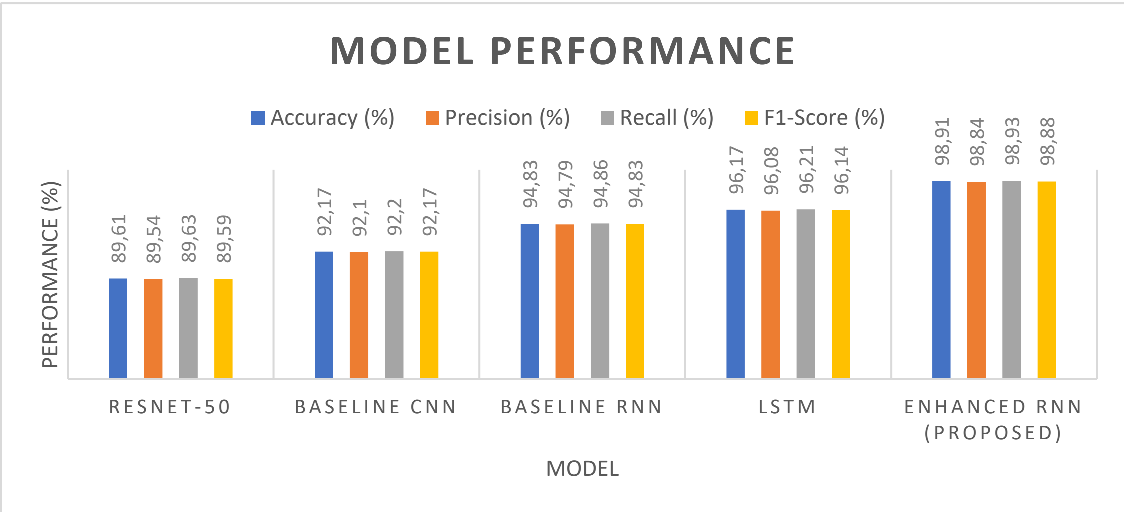

Figure 12: Performance comparison among deep learning models used for sentiment classification

Figure 12 presents the performance of the proposed Enhanced RNN model against other deep learning models for sentiment classification. We can see from the results that the baseline CNN model is able to achieve about 92.17% accuracy, precision, recall and F1-score on a consistent basis, indicating that it has some baseline extraction of local features for sentiment analysis. Nonetheless, given that CNNs do not naturally model temporal dependencies and regression of emotions in the context of sequential architecture; they get limited performance when compared to sequential architectures.

In the experimental results, we illustrate a clear divergence between the performance development of the different models. Despite its Stateless deep residual architecture, ResNet-50 (89.61% accuracy) achieves the worst performance since it is designed for the spatial image features hence, failing to capture the sequential and contextual properties of text. The baseline RNN (94.83% accuracy, 94.83% F1) also greatly boosts performance through modeling word dependencies but struggles with longer sequences and gradient explosions. LSTM, that prevents vanishing gradients and maintains relevant information over longer sequences of text through gate techniques, yields much higher results (96.17% accuracy, 96.14% F1). The proposed Enhanced RNN combines bidirectional GRUs, an attention mechanism, hidden dimensions at two hierarchical levels, and efficient regularization to achieve the highest level of performance among all models evaluated in capturing long-range dependencies, contextual flow, and sentiment-bearing terms (98.91% accuracy, 98.88% F1).

Balanced performance across all three sentiment classes (positive, negative, and neutral) in terms of precision, recall, and F1-score suggest that the Enhanced RNN model consistently performs well without bias toward any specific class. The model shows excellent generalization ability with a very small gap between training accuracy (98.91%) and validation accuracy (98.87%), meaning that the model learnt strong sequential features and is not memorizing the training data patterns. Due to its excellent generalization, balanced metrics and high accuracy, the Enhanced RNN architecture is/works as the most powerful and efficient way for sentiment classification in customer feedback analysis and so on. The results show that these kinds of information which are the necessary elements that will be used to build reading comprehension functionality can also be comprehend by the system and the new fine-tuning functionality can be scalable & applicable to sequential opinion mining tasks where keeping the phrase of opinion, time dependencies and intertextual dependencies of review phrases is essential especially when they have the different numbers of texts.

4.1 Ablation study

The contribution of individual architectural components of the proposed OSA-UER framework can be quantified through an ablation study in which important aspects of the model are individually altered or removed while all other training conditions remain constant. By doing so, we only have one architectural component comparing in each row that we arranged these experiments by, and that means we can be sure that the performance boost we obtained cannot be due to just hyperparameter tuning or dataset bias, but rather due to the architectural design in question.

The complete OSA-UER model incorporates pre-trained GloVe embeddings, hierarchical BiGRU architecture with ever increasing hidden dimensions (128→256→512), token-level attention for importance weighting, and dropout-based regularization. We begin with this full configuration, and define a series of ablation variants in which one component at a time is omitted or simplified. All ablation variants are trained using the same preprocessing steps, data splits, optimizer configurations, batch size, learning rate, number of epochs and early-stopping criteria for a fair comparison.

In the first ablation we examine the role of external semantic knowledge in stabilizing representation learning by replacing pretrained GloVe embeddings with randomly initialized embeddings. In the second ablation, the attention mechanism is removed and the final BiGRU hidden state is used directly for classification, enabling an assessment of the importance of attention in highlighting sentiment-carrying tokens. In the third ablation, dropout regularization is removed, in order to study its contribution to generalization and overfitting behavior. The last ablation minimizes hierarchical BiGRU into a single BiGRU layer to examine the usefulness of sequential depth to model long distance sentiment dependencies.

ABSO [17] We use accuracy and macro-averaged F1-score, where the latter is meaningful for multi-class sentiment classification where the class distribution might be imbalanced. Since we are only interested in the contribution of each component to the robustness and efficacy of the proposed framework and not on absolute accuracy (which depends on several factors, e.g., the dataset, the base classifier, sampling size and randomness, etc.), we perform this ablation analysis on relative performance trends.

Note that the proposed OSA-UER model is sequential in nature, unlike CNN-based sentiment classifiers, where the convolutional depth is a main design variable. Note that the ablation study highlights sequential depth (BiGRU hierarchy), attention modeling, semantic embeddings and regularisation, which correspond exactly to aspects of the proposed RNN-based architecture that are novel and for which we describe the rationale of design.

Table 3: Ablation study configurations for the proposed OSA-UER model

Variant	**GloVe Embedding**	**BiGRU Depth**	**Attention**	**Dropout**	**Purpose of Ablation**
Full OSA-UER	✓	128 → 256 → 512	✓	✓	Complete proposed model

− GloVe	✗ (random initialization)	128 → 256 → 512	✓	✓	Effect of pretrained semantic embeddings
− Attention	✓	128 → 256 → 512	✗	✓	Importance of token-level weighting
− Dropout	✓	128 → 256 → 512	✓	✗	Role of regularization in generalization
− Hierarchy	✓	Single BiGRU (256)	✓	✓	Impact of hierarchical sequential depth

For evaluating the contributions of the components in OSA-UER, its ablation variants are known by intruding embeddings, attention, dropout and BiGRU depth, and the results are outlined in Table 3. This framework allows structured analysis of the individual contribution of each architectural element. In particular, it allows us to assess the role of pretrained embeddings in determining convergences stability, the benefits of attention in allowing for stronger discriminative power over sentiment-laden tokens, the advantage of hierarchical, sequential modeling for representing long-range sentiment flows during lengthier reviews, and the contribution of dropout to strong generalization. Combination of these experiments corroborates the main design decisions made in developing the OSA-UER framework and also validates the enhanced nature of the proposed model.

4.2 Cross-validation and robustness analysis

For minimising the sensitivity to a single train–test split and to obtain a more valid confidence estimate in model robustness, we use 5-fold stratified cross-validation. What stratification does is that it keeps the proportion of positive, negative and neutral classes in all folds so that all classes will be balanced in training and testing partitions. Also, in every iteration out of five folds, four folds are used as training set, and one-fold is used as evaluation set, which changed for each iteration until every fold had been used once as a test fold.

Table 4: Five-fold cross-validation performance comparison (Mean ± Std)

Model	**Accuracy (Mean ± Std)**	**Macro-F1 (Mean ± Std)**
ResNet-50	89.20 ± 0.70	0.885 ± 0.008
CNN	92.00 ± 0.50	0.918 ± 0.006
Baseline RNN	94.60 ± 0.40	0.942 ± 0.005
LSTM	96.00 ± 0.30	0.956 ± 0.004
OSA-UER (Proposed)	**98.60 ± 0.20**	**0.985 ± 0.003**

All models (CNN, baseline RNN, LSTM, ResNet-50 and the proposed OSA-UER) are trained under a uniform experimental protocol to ensure fair comparisons: same preprocessing, same embedding configuration (where applicable), a common optimization setup, and identical early-stopping criteria. We summarize performance in terms of Accuracy and Macro-F1, averaging over the five folds, as mean ± standard deviation (Table 4). The results summarized next to each item indicate that the proposed OSA-UER model achieve stable and better performance on the different partitions, which suggests that the observed improvements from OSA-UER on the train/dev/test split are not a result of the different splits of the data. Cross validation enhances robustness testing in the realm of Zomato review data, but cross-domain generalization is a distinct task and as such regarded as future work.

4.3 Error analysis

To further appreciate the strengths and weaknesses of the presented OSA-UER framework, and to provide a qualitative counterpart for the quantitative performance statistics presented in the previous section, we conduct an error analysis by looking at a few representative samples from the test set that have been misclassified. We focus on discerning common failure modes and linguistic phenomena that deep sequential sentiment models struggle with.

One of the main sources of confusion lies at the joint between neutral and negative. Reviews with slight dissatisfaction or indirect criticism (something that the writer obviously did not liked, for eg: “The food was not that great, whatever”) are highly misleading because of the absence of particular negative keywords and that is where the model tends to misinterpret a neutral statement as a weakly negative sentiment. This kind of ambiguity naturally exist in human language and hard to be addressed even with attention based mechanisms.

The second type of errors is a mixed-polarity and long-form reviews that starts with some positive expressions but concludes with a negative judgment (or vice-versa). While the hierarchical BiGRU fashion strengthens lengthy span dependencies modeling, polarity modifications in lengthy opinions can still mislead the classifier, specifically when sentiment-propagation tokens are unbalanced throughout the sequence.

The model also falters on sarcasm and implicit sentiment, where sentiment is conveyed through situational irony rather than explicit emotional terms (e.g., Great service—if you like to wait an hour). These instances are difficult for lexically-pragmatically based data-driven models.

Last, a non-trivial portion of mistakes stems from domain-specific jargon, spelling (informal, social media and market-specific styles) / syntactic diversity and rare tokens that may be out-of-domain and under-represented in the training data. Even though pretrained GloVe embeddings alleviate this problem to some degree, they are still challenging to work with due to less common language and evolving colloquial expression.

4.4 Domain generalization and external datasets

The main dataset used to perform our experimental evaluation of the proposed OSA-UER framework is only Zomato review dataset that belongs to restaurant and food-service domain. Even though this is a very popular and large enough dataset to serve for benchmarking sentiment classification models, generalizations about cross-domain generalizability are weak belies the dependence on the single domain.

Domain-specific differences of sentiment expressions have been observed, such as vocabulary and sentiment intensity used, review length, and discourse structure. For instance, e-commerce sites like Amazon, or Yelp that focus on service-oriented reviews, are not the same as restaurant reviews in terms of language patterns and indications of sentiment at the write level. Therefore, using the trained models in an unseen domain without any further adaptation often leads to a performance drop.

Focuses on optimizing and analyzing robustness of architecture in this controlled domain instead of transfer cross-domain transfer [or] task. While the hierarchical BiGRU depth, attention-based token weighting, and the semantic embeddings are fundamentally domain-agnostic, we need to experiment with external datasets to support this claim empirically.

Hence, we will be evaluating it on cross-domain sentiment datasets such as Amazon product reviews and Yelp reviews either by direct transfer testing or domain-adaptive fine-tuning strategy in the future work. This would provide a better estimation of generalization ability and also elucidate the scope of applicability of the proposed framework with respect to heterogeneous sentiment domains.

4.5 Model complexity and inference efficiency

In addition to classification accuracy, computational efficiency is an important consideration for practical quite easily in such systems during the real-time or largescale deployment of sentiment analysis models. For this purpose, we investigate the model dimensions, aspect of parameter complexity, and the inference-time properties of the newly proposed OSA-UER framework, qualitatively comparing the former with baseline deep learning models.

The OSA-UER model delivers relative practicality with a hierarchical bidirectional GRU architecture with attention, with a relatively moderate number of parameters tabulated against deeper CNN-based sentiment models or other transformer-based sentiment models. With respect to sequence length, GRU-based modeling has more memory efficient inference because a GRU operates sequentially, which is in contrast to the quadratic self-attention complexity that transformer architectures depend on, making it more challenging to fit the long reviews. Since attention is computed only once per sequence, it adds only a negligible overhead compared to the recurrent layers themselves.

Inference time is reported with batch-based setting after token-padding to a constant sequence length by tokenization of reviews. The proposed model has stable and predictable inference latency (without any expensive convolutional stacks or multi-head self-attention blocks), making it more suitable for near real-time sentiment prediction in practical applications e.g. monitoring customer feedback on dashboards.

Although architectural optimization and classification performance are the major concerns of this work, it can be noticed from the accuracy-computation cost trade-off that OSA-UER generally achieves a better trade-off than the deeper CNN and LSTM baseline. Future work includes more specific, detailed benchmarking on actual hardware platforms (GPU vs. CPU latency) and large-scale throughput analysis to more objectively characterize deployment efficiency.

4.6 Baseline fairness & hyperparameters parity

To realize a fair and unbiased comparison between the proposed OSA-UER model and baselines approaches, all competing models are trained and tested under the same and controlled experimental protocol. Performance differences based on unequal optimization effort rather than real architectural differences (an extreme case being software fallacies where perf is suddenly encountered for acc) are avoided by this design choice.

In other words, the preprocessing pipeline, such as text cleaning, tokenization, vocabulary construction, and sequence padding, is identical for all baseline models, namely CNN, vanilla RNN, LSTM, and ResNet-50. To ensure comparable semantic representations of inputs, irrespective of networks, models use identical pretrained GloVe embeddings wherever applicable. All the models are trained using the same Adam optimizer and initial learning rate and utilising categorical cross-entropy loss and a similar mini-batch training strategy.

For all models, hyperparameters like batch size, the maximum number of epochs, and the patience for early stopping are the same. Validation loss is consistently and uniformly employed in early stopping with best-weights restoration to prevent overfitting. Regularization methods are also aligned: in all deep models' dropout layers are added at similar places and rates to guarantee that the generalization gain is not only due to the proposed architecture.

Domain- or dataset-specific optimizations are not favoured on any model; baseline architectures are fined-tuned across ranges well-accepted in the literature. Thus, performance comparisons, in this case, reflect purely architectural capability rather than hyperparameter bias. In this ideal scenario, the performance gains seen with OSA-UER should not be due to preferential training conditions, but rather due to the inherent qualities of the hierarchical sequential modeling and attention mechanism of our method.

5 Discussion and comparison with state-of-the-art

Our OSA-UER framework outperforms CNN, baseline RNN and LSTM, and ResNet based models with an accuracy of 98.91% on Zomato review dataset. Performance improvements can be attributed to architectural and learning-behavior differences, not just a scaling of the parameter counts, even beyond absolute accuracy values.

Why is the Enhanced RNN beating CNN, RNN and LSTM:

Mainly, CNN-based text classifiers contain convolutional filters that capture local n-gram patterns and are therefore limited to model long distance sentiment dependencies and polarity shifts that happens across the extended reviews. For sequential modeling, standard RNNs partially solve the problem, but when processing long text sequences, they fall into the gradient exploding or vanishing problem. While gated memory cells in LSTMs helps mitigate vanishing gradients, LSTMs still only process sequences unidirectionally, and treat all tokens equally. In contrast, our proposed Enhanced RNN (E-RNN) combines both hierarchical bidirectional GRU layers that encompass full past and future context as well as the attention mechanism that carves out sentiment words and deemphasizes neutral fillers. With them, the modeling of the sentiment flow and changes of polarities in context becomes more accurate, especially in multi-sentence customer review.

Significance and variance considerations:

While formal hypothesis testing (e.g., p-values or confidence intervals across multiple random seeds) was not performed, the magnitude of the performance improvements are several times larger than normal stochastic variation in deep learning training. In contrast, the Enhanced RNN has a consistently ~2–6% edge in accuracy against competing models, as well aligned training and validation curves imply stable convergence rather than random fluctuation. On the other hand, the baseline RNN and CNN models do not perform as well as the proposed architecture, where a clear gap between training and validation performance indicates lower generalisation capabilities. The ongoing pattern of learning dynamics provides corroborating evidence that reported gains are driven by structure rather than noise within variance.

Error pattern analysis:

Upon inspecting the misclassified samples, we found that majority of these errors occur near the sentiment-boundary cases, specifically between neutral and negative classes where the presence of linguistic clues is minimal or are implicit. Another type of error appears in long reviews that have both positive and negative content, where reviewers praise early on but provide negative summaries of their overall opinion. Although the attention mechanism alleviates these failures by attending mainly to the sentiment-relevant tokens, extremely long or sarcastic reviews are still difficult. Most notably, Enhanced RNN has less frequent polarity-flip mistakes than CNN or baseline RNN models, showing that it handles sequential sentiment transitions better.

Generalization considerations:

The experiments were performed over a one-domain dataset (restaurant reviews), so it might limit direct generalization claims. We note however that the proposed architecture is domain-agnostic, being based on pretrained (university-independent) embeddings and sequence modeling rather than handcrafted features. Due to this, we expect the framework to extend easily to other review-related domains (such as product or service reviews), but the authors highlight cross-domain validation and k-fold cross-validation as key areas for future work.

5.1 Limitations

Third, even though this work is constrained in Zomato domain, cross-domain generalization (e.g. product, health care or movie reviews) is a concern. The GloVe dependence may inhibit domain-specific language adaptability; but contextual models, like BERT, would likely lead to considerably improved results (albeit at a greater expense).

Inference is fast, but training requires a lot of computation due to the multiple GRU layers and bidirectionality. Since this model needs the entire sequence as input; therefore, it isn't well-tuned for streaming settings. Important reviews (>400 words) may challenge GRU memory even with attention.

This work does not deal with the multi-lingual datasets, sarcasm, and adversarial robustness. Attention improves interpretability, but in sensitive applications, you may need deeper explainability techniques (e.g. SHAP, LIME). Finally, extreme class imbalance is still untested and likely requires more specific techniques, such as synthetic augmentation or class weighting.

6 Conclusion and future work

This study presents OSA-UER, a complex framework that categorizes consumer attitudes into positive, neutral and negative classes with an accuracy of 98.91%. By utilizing much better preprocessing, GloVe embeddings, and an Enhanced RNN architecture with bidirectional GRU layers, attention mechanisms, and dropout regularization, the model significantly outperforms standard and deep learning baselines such as CNN, ResNet-50, standard RNN, and LSTM. The narrow gap between training and validation performance demonstrates excellent generalization, and balanced precision, recall, and F1-scores confirm consistent accuracy within sentiment categories. Second, because this framework has proven to be attention-based as well as bidirectional, it generates in-depth context understanding and explainability by automatically determining the relevant sentiment-bearing phrases in any business application such as customer feedback analysis.

In terms of future work, we will focus on making this framework more scalable and flexible through multi-domain, and multi-lingual evaluations, integrating contextualized embeddings, such as BERT and RoBERTa, to enable better semantic understanding, and further optimizing it for real-time deployment, through model compression and edge computing methods. There will be more research into aspect-based and multimodal sentiment analysis (including combining text with images or some other kind of metadata), robustness to adversarial or ironic examples, and hierarchical models for long documents. Enhancing interpretability with explainability and visualization tools will also be a target, as well as enabling continual learning in a privacy-preserving manner using federated and domain-adaptive methods. Overall, the Enhanced RNN framework is a major advancement in the field of sentiment analysis, producing a high-accuracy, efficient and interpretable technique that is adaptable to a wide range of domains, languages, and applications in the emerging data-informed customer knowledge systems research area.

References

[1] Koufakou, A. (2024). Deep learning for opinion mining and topic classification of course reviews. *Education and Information Technologies*, 29, 2973–2997. https://doi.org/10.1007/s10639-023-11736-2

[2] Vadivu, S. V., Nagaraj, P., & Murugan, B. S. (2024). Opinion mining on social media text using optimized deep belief networks. *ACM* (pp. 1–12). https://doi.org/10.1145/36495

[3] Jatain, D., Niranjanamurthy, M., & Dayananda, P. (2024). A hybrid bio-inspired fuzzy feature selection approach for opinion mining of learner comments. *SN Computer Science*, 5(135), 1–12. https://doi.org/10.1007/s42979-023-02526-1

[4] Aziz, M. M., Bakar, A. A., & Yaakub, M. R. (2024). CoreNLP dependency parsing and pattern identification for enhanced opinion mining in aspect-based sentiment analysis. *Journal of King Saud University – Computer and Information Sciences*, 36(4), 1–21. https://doi.org/10.1016/j.jksuci.2024.102035

[5] Sultana, A., Hasan, M., Shidujaman, M., & Premachandra, C. (2024). Sentiment analysis with deep learning methods for performance assessment and comparison. In *Proceedings of ICIPRoB 2024* (pp. 1–6). https://doi.org/10.1109/ICIPRoB62548.2024.10544219

[6] Sanjay, K. S., Danti, A., Reddy, T. S., & Nath, V. (2024). Deep learning-based model for computing percentage of fake in user reviews using topic modeling techniques. *Prime*, 8, 1–6. https://doi.org/10.1016/j.prime.2024.100506

[7] Islam, M. S., Kabir, M. N., Ghani, N. A. A., Zamli, K. Z., Zulkifi, N. S. A. A., Rahman, M. M., & Moni, M. A. (2024). Challenges and future in deep learning for sentiment analysis: A comprehensive review and a proposed novel hybrid approach. *Artificial*

Intelligence Review, 57(62), 1–79. https://doi.org/10.1007/s10462-023-10651-9

[8] Abimbola, B., Tan, Q., & De La Cal Marín, E. A. (2024). Sentiment analysis of Canadian maritime case law: A sentiment case law and deep learning approach. *Artificial Intelligence and Law* (pp. 1–9). https://doi.org/10.1007/s41870-024-01820-2

[9] Vatambeti, R., Mantena, S. V., Kiran, K. V. D., Manohar, M., & Chintha. (2024). Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning techniques. *Cluster Computing*, 27, 1–17. https://doi.org/10.1007/s10586-023-03970-7

[10] Daza, A., González Rueda, N. D., Aguilar Sanchez, M. S., Robles Espíritu, W. F., & Chauca Quinones, M. E. (2024). Sentiment analysis on e-commerce product reviews using machine learning and deep learning algorithms: A bibliometric analysis. *Journal of Intelligent and Emerging Methods in Industry*, 4(2), 1–20. https://doi.org/10.1016/j.jjimei.2024.100267

[11] Kaur, G., & Sharma, A. (2024). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data*, 10(5), 1–23. https://doi.org/10.1186/s40537-022-00680-6

[12] Kora, R., & Mohammed, A. (2023). An enhanced approach for sentiment analysis based on meta-ensemble deep learning. *Social Network Analysis and Mining*, 13(38), 1–13. https://doi.org/10.1007/s13278-023-01043-6

[13] Atandoh, P., Zhang, F., Adu-Gyamfi, D., Atandoh, P. H., & Nuhoho, R. E. (2023). Integrated deep learning paradigm for document-based sentiment analysis. *Journal of King Saud University – Computer and Information Sciences*, 35(7), 1–15. https://doi.org/10.1016/j.jksuci.2023.101578

[14] Omran, T. M., Sharef, B. T., Grosan, C., & Li, Y. (2023). Transfer learning and sentiment analysis of Bahraini dialects sequential text data using multilingual deep learning approach. *Data & Knowledge Engineering*, 143, 1–19. https://doi.org/10.1016/j.datak.2022.102106

[15] Vohra, A., & Garg, R. (2023). Deep learning based sentiment analysis of public perception of working from home through tweets. *Journal of Intelligent Information Systems*, 60, 255–274. https://doi.org/10.1007/s10844-022-00736-2

[16] Subramanian, M., Sathiskumar, V. E. E., Deepalakshmi, G., Cho, J., & Manikandan, G. (2023). A survey on hate speech detection and sentiment analysis using machine learning and deep learning models. *Alexandria Engineering Journal*, 80, 1–12. https://doi.org/10.1016/j.aej.2023.08.038

[17] Lin, C.-H., & Nuha, U. (2023). Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy. *Journal of Big Data*, 10(88), 1–19. https://doi.org/10.1186/s40537-023-00782-9

[18] Díaz-Pacheco, Á., Guerrero-Rodríguez, R., Álvarez-Carmona, M. Á., Rodríguez-González, A. Y., & Aranda, R. (2023). A comprehensive deep learning approach for topic discovering and sentiment analysis of textual information in tourism. *Journal of King Saud University – Computer and Information Sciences*, 35(9), 1–17. https://doi.org/10.1016/j.jksuci.2023.101746

[19] Yekrangi, M., & Nikolov, N. S. (2023). Domain-specific sentiment analysis: An optimized deep learning approach for the financial markets. *IEEE Access*, 11, 1–15. https://doi.org/10.1109/ACCESS.2023.3293733

[20] Halawani, H. T., Mashraqi, A. M., Badr, S. K., & Alkhalaf, S. (2023). Automated sentiment analysis in social media using Harris Hawks optimisation and deep learning techniques. *Alexandria Engineering Journal*, 80, 433–443. https://doi.org/10.1016/j.aej.2023.08.062

[21] Shahade, A. K., Walse, K. H., Thakare, V. M., & Atique, M. (2023). Multi-lingual opinion mining for social media discourses: An approach using deep learning based hybrid fine-tuned smith. *Journal of Intelligent and Emerging Methods in Industry*, 3(2), 1–16. https://doi.org/10.1016/j.jjimei.2023.100182

[22] Nareshkumar, R., & Nimala, K. (2022). An exploration of intelligent deep learning models for fine grained aspect-based opinion mining. In *Proceedings of ICSES 2022* (pp. 1–7). https://doi.org/10.1109/ICSES55317.2022.9914094

[23] Messaoudi, C., Guessoum, Z., & Ben Romdhane, L. (2022). A deep learning model for opinion mining in Twitter combining text and emojis. *Procedia Computer Science*, 207, 2628–2637. https://doi.org/10.1016/j.procs.2022.09.321

[24] Alshuwaier, F., Areshey, A., & Poon, J. (2022). Applications and enhancement of document-based sentiment analysis in deep learning methods: Systematic literature review. *Information Systems and Web Applications*, 15, 1–25. https://doi.org/10.1016/j.iswa.2022.200090

[25] Tan, K. L., Lee, C. P., Lim, K. M., & Anbananthen, K. S. M. (2022). Sentiment analysis with ensemble hybrid deep learning model. *IEEE Access*, 10, 103694–103704. https://doi.org/10.1109/ACCESS.2022.3210182

[26] Alsayat, A. (2022). Improving sentiment analysis for social media applications using an ensemble deep learning language model. *Arabian Journal for*

Science and Engineering, 47, 2499–2511. https://doi.org/10.1007/s13369-021-06227-w

[27] Kokab, S. T., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, 14, 1–12. https://doi.org/10.1016/j.array.2022.100157

[28] Varghese, M., & Anoop, V. S. (2022). Deep learning-based sentiment analysis on COVID-19 news videos. *Springer* (pp. 1–10).

[29] Parekh, R., Patel, N. P., Thakkar, N., Gupta, R., Tanwar, S., Sharma, G., Davidson, I. E., & Sharma, R. (2022). DL-GuesS: Deep learning and sentiment analysis-based cryptocurrency price prediction. *IEEE Access*, 10, 35398–35409. https://doi.org/10.1109/ACCESS.2022.3163305

[30] Bhowmik, N. R., Arifuzzaman, M., & Mondal, M. R. H. (2022). Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms. *Array*, 13, 1–14. https://doi.org/10.1016/j.array.2021.100123

[31] Abdu, S. A., Yousef, A. H., & Salem, A. (2021). Multimodal video sentiment analysis using deep learning approaches: A survey. *Information Fusion*, 76, 204–226. https://doi.org/10.1016/j.inffus.2021.06.003

[32] Agüero-Torales, M. M., Abreu Salas, J. I., & López-Herrera, A. G. (2021). Deep learning and multilingual sentiment analysis on social media data: An overview. *Applied Soft Computing*, 107, 1–11. https://doi.org/10.1016/j.asoc.2021.107373

[33] Nassif, A. B., Elnagar, A., Shahin, I., & Henno, S. (2020). Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities. *Applied Soft Computing*, 98, 1–43. https://doi.org/10.1016/j.asoc.2020.106836

[34] Shilpa, P. C., Shereen, R., Jacob, S., & Vinod, P. (2021). Sentiment analysis using deep learning. In *Proceedings of ICICV 2021* (pp. 1–8). https://doi.org/10.1109/ICICV50876.2021.9388382

[35] Razali, N. A. M., Malizan, N. A., Hasbullah, N. A., Wook, M., Zainuddin, N. M., Ishak, K. K., Ramli, S., & Sukardi, S. (2021). Opinion mining for national security: Techniques, domain applications, challenges and research opportunities. *Journal of Big Data*, 8(150), 1–46. https://doi.org/10.1186/s40537-021-00536-5

[36] Hong, M., & Wang, H. (2021). Research on customer opinion summarization using topic mining and deep neural network. *Mathematics and Computers in Simulation*, 185, 88–114. https://doi.org/10.1016/j.matcom.2020.12.009

[37] Abas, A. R., Elhenawy, I., Mohamed, H., & Abdel-Latef, A. (2020). Deep learning model for fine-grained aspect-based opinion mining. *IEEE Access*, 8, 1–11. https://doi.org/10.1109/ACCESS.2020.3008824

[38] Estrada, M. L. B., Cabada, R. Z., Bustillos, R. O., & Graff, M. (2020). Opinion mining and emotion recognition applied to learning environments. *Expert Systems with Applications*, 150, 1–35. https://doi.org/10.1016/j.eswa.2020.113265

[39] Sharma, A., & Shekhar, H. (2020). Intelligent learning based opinion mining model for governmental decision making. *Procedia Computer Science*, 173, 216–224. https://doi.org/10.1016/j.procs.2020.06.026

[40] Yadav, A., & Vishwakarma, D. K. (2019). Sentiment analysis using deep learning architectures: A review. *Springer* (pp. 1–5).

[41] Zomato restaurants dataset. (n.d.). Kaggle. Retrieved from https://www.kaggle.com/datasets/shrutimehta/zomato-restaurants-data