Taylor & Francis
Taylor & Francis Group



Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: 0005-1144 (Print) 1848-3380 (Online) Journal homepage: www.tandfonline.com/journals/taut20

Fusion of deep recurrent neural network models and fuzzy decision support system for tweet sentiment analysis and classification

A. Sherin, S. Lokesh, S. N. Deepa & I. Jasmine Selvakumari Jeya

To cite this article: A. Sherin, S. Lokesh, S. N. Deepa & I. Jasmine Selvakumari Jeya (2025) Fusion of deep recurrent neural network models and fuzzy decision support system for tweet sentiment analysis and classification, Automatika, 66:4, 28-50, DOI: 10.1080/00051144.2025.2561424

To link to this article: https://doi.org/10.1080/00051144.2025.2561424

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 07 Oct 2025.

Submit your article to this journal

Article views: 191

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=taut20

AUTOMATIKA
2025, VOL. 66, NO. 4, 28–50
https://doi.org/10.1080/00051144.2025.2561424

OPEN ACCESS Check for updates

Fusion of deep recurrent neural network models and fuzzy decision support system for tweet sentiment analysis and classification

A. Sherin[a], S. Lokesh[b], S. N. Deepa[c] and I. Jasmine Selvakumari Jeya[d]

[a]Department of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore, India; [b]Department of Artificial Intelligence and Data Science, PSG Institute of Technology and Applied Research, Coimbatore, India; [c]Department of Electrical Engineering, National Institute of Technology Calicut, Kerala, India; [d]School of Computing Science and Engineering, VIT Bhopal University, Sehore, India

ABSTRACT

With the advent of social media and networking, Tweet Sentiment Analysis (TSA) has become a significant methodology to extract useful information from the "X" platform users. The extraction of useful information from the tweets enables us to develop a decision support process from which necessary opinion mining can be carried out. These opinions help improve business models, product reviews, customer satisfaction, and thereby improve services and quality of any system or product. In this research paper, an innovative fusion of a deep recurrent neural network (DRNN) and a fuzzy decision support system (FDSS) is done to evaluate customer satisfaction based on tweets. The proposed ensemble Bi-directional Long Short-Term Memory (Bi-LSTM) and Gated Recurrent Units (GRU), which are deep recurrent neural network models, were employed to attain the polarity of the tweet sentiments (positive, negative or neutral). The innovative fuzzy decision support system determines the contentment level of customers based on the tweets. FDSS with ensemble Bi-LSTM-GRU (eBi-LSTM-GRU) handles uncertainties and imprecision in tweet sentiments, enhancing performance. The developed fusion model attained an accuracy of 84.33%, 96.92% and 93.81% for the sentiment 140 dataset, the T4SA dataset and the airline Twitter dataset, respectively. The proposed fusion eBi-LSTM-GRU-FDSS model outperforms previous baseline approaches.

ARTICLE HISTORY
Received 30 January 2025
Accepted 11 September 2025

KEYWORDS
Ensemble bi-LSTM model; gated recurrent unit; fuzzy decision support system; sentiment analysis; sentiment classification; performance metrics

1. Introduction

The advancements of technology and social networking have enabled millions of interested people to exchange and share their opinions despite their different geographic locations and multilingual cultures on various products, systems, business models, customer services and so on. To share their opinions, people are in need of quick, simple and fast tools to meet their requirements. Social networks like "X" (Formerly known as Twitter), Facebook, and review platforms fulfil user requirements to meet the increasing demand. The number of user accounts on social media platforms increases each second, and each individual has their own concern to post their views on the domain they are focused on.

Twitter, "X" is one of the most widely used social media platforms as of September 2024, nearly 586 million people are using it worldwide [1,2]. The tweet data has become a vital source of information for several manufacturers, professionals, practitioners in various fields like the health sector, entertainment and media, government officials, politicians, academicians, judicial experts, developers and so on. In 2022 and 2023, Twitter's daily active users were 237.8 and 245 million, respectively and as of 2024, its daily active users are 245 million [3–5]. Twitter gained 1.6 million daily active users over the past week of November 2024 (Source: https://x.com). Further, as of 2024, 39.1% of Twitter users are women, while 60.9% are men worldwide. 30% of consumers use Twitter to look at brands, and thus, the user opinions are given topmost priority [6].

With immense ways for people to express and share their feelings online, there are powerful tools that are in demand by organizations for monitoring the opinions shared by the public on their products and services in a real-time dynamic scenario. Tweet Sentiment Analysis (TSA) has evolved into a broad and continuously expanding research field, in which the sentiments and opinions of users on different brands, individuals, particular or general societal issues, organizations and so on are analysed based on the tweet data content [7]. TSA has become more important with the existence of the "X" platform to track the positive, negative and neutral sentiments and thereby do the branding of the specific event or product. TSA provides clarity with respect to the opinions of the end-use customers and is more supportive of both manufacturers and the public by

CONTACT A Sherin sherinameer@gmail.com Department of Computer Science and Engineering, Hindusthan College of Engineering and Technology, Coimbatore 641032, India

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00051144.2025.2561424.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

providing useful information based on the online tweet reviews. TSA becomes significant as it delivers objective results using customer reviews, enables real-time brand reputation monitoring and enhances the scalability of various intelligence models and their features.

As companies integrate Sentiment Analysis (SA) into their processes to examine tweet interactions and conversations, they gain better insights into customer friction points and can monitor each stage of the customer journey more effectively. The recent increase in online shopping has made the need to analyse whether a product is good or not, and this has made tweet sentiment analysis even more significant.

Deep learning models, predominantly Recurrent Neural Networks (RNNs), have been extensively used for sentiment analysis. Among these, Bidirectional Long Short-Term Memory (Bi-LSTM) and Gated Recurrent Units (GRU) have proven effective in capturing sequential dependencies in modelling text. While Bi-LSTM processes input in both forward and backward directions, GRU simplifies the learning process by reducing computational complexity. However, existing RNN-based approaches struggle to handle complex linguistic structures and require enhancements for improved sentiment classification. To address this, we propose an ensemble model integrating Bi-LSTM and GRU to better capture sentiment dependencies in tweets. In the context of Twitter sentiment classification, handling complex linguistic patterns, contextual dependencies and sequential relationships in short and often informal text is crucial. Both BiLSTM and GRU have been widely used in Natural Language Processing (NLP) tasks due to their capability to model long-range dependencies in sequential data. However, each model has distinct strengths:

- *BiLSTM (Bidirectional Long Short-Term Memory)* effectively learns equally past and future contextual information, making it well-suited for understanding sentiment in tweets where word order significantly impacts meaning.
- *GRU (Gated Recurrent Unit)* is a computationally efficient substitute to LSTM, requiring fewer parameters while maintaining comparable performance, making it well-suited for real-time sentiment classification tasks.

By combining BiLSTM and GRU, our approach leverages the strengths of both architectures – BiLSTM's ability to capture bidirectional dependencies and GRU's efficiency in learning sequential representations. This hybrid structure enhances the model's capacity to understand sentiment nuances in short, noisy and context-sensitive Twitter text, leading to improved classification accuracy while maintaining computational efficiency.

This research integrates a Fuzzy Decision Support System (FDSS) to enhance decision-making accuracy in tweet sentiment classification. Given the imprecise and subjective nature of tweet vocabulary, Fuzzy Logic (FL) effectively handles uncertainties by capturing user opinions from tweet data. The ensemble Bi-LSTM-GRU model first evaluates sentiment scores (positive, neutral or negative) and provides them as inputs to the FDSS, which then assesses customer contentment levels. The FDSS defines membership functions for input and output variables, formulates a fuzzy rule base using "If–Then" statements and processes sentiment scores through a fuzzy inference mechanism to determine customer satisfaction. The final step, defuzzification using the centroid method, converts fuzzy outputs into crisp values, yielding an accurate measure of customer contentment. The design and implementation of the FDSS, including membership function definitions and rule formulation, are detailed in this study. The efficacy of our FDSS model is established through comparative analysis with other fuzzy approaches, showing significant improvements in sentiment classification accuracy and decision-making reliability.

The design of the new eBi-LSTM-GRU-FDSS model is done to obtain the text features, attain the polarity of the texts and then decide on the user's contentment level. The research objective in this study is to model an improved tweet sentiment analysis such that its performance is enhanced. To attain this objective, this work focuses on devising a new ensemble Bi-LSTM with the GRUs in fusion with a fuzzy decision support system wherein the opinion of the tweets is computed with the eBi-LSTM-GRU model and the customer content satisfaction is obtained from the designed fuzzy decision support model. The developed model is intended to employ the sentiment parsers obtained and retain the most relevant data from the tweet sentiments. Employing the proposed eBi-LSTM-GRU-FDSS technique for tweet sentiment analysis, the long-term dependencies of the parsers are derived, which enables effective tweet sentiment opinion classification. Figure 1 illustrates the workflow of the proposed model.

The devised new ensemble deep learning recurrent model, eBi-LSTM-GRU-FDSS approach, is tested for its applicability over the Twitter datasets, Twitter for Sentiment Analysis (T4SA) dataset, Sentiment 140 and Airline Twitter datasets. The devised new ensemble Bi-LSTM-GRU with FDSS shows superior performance metrics compared to the baseline TSA approaches available prior [8–15]. The hybridization of the eBi-LSTM-GRU with the FDSS captures the polarity and the contentment level of the consumers, thereby facilitating the manufacturers or administrators in having lucid, critical remarks and opinions on their products. Employing the new e-Bi-LSTM-GRU-based FDSS model, the sentiment classification accuracy has been

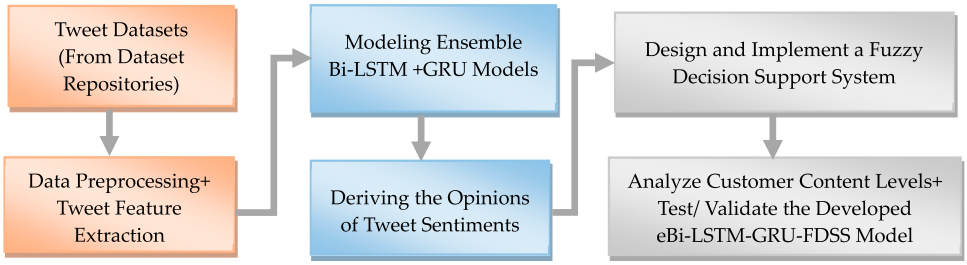


Figure 1. Workflow of the proposed system.

significantly increased by 27% compared to the baseline TSA approaches. The evaluation of the proposed model includes standard and advanced performance metrics such as Recall, Precision, F1 Score, Accuracy (%) and Cohen's Kappa (CK) Score. Additionally, Area Under the Curve (AUC) comparison plots and convergence plots are analysed to assess model stability. The statistical significance of the results is validated using *P*-values computed via the Wilcoxon Rank Sum Test, while computational efficiency is examined through a detailed Computational Complexity Analysis and ablation study.

To summarize, the contributions made in this paper are as follows:

(1) An Effective pre-processing method is performed on the tweet data, including tokenization and removal of special characters and other irrelevant stop-words.
(2) A Deep Learning (DL) based technique is devised with an Ensemble Bi-LSTM model along with Gated Recurrent Units (GRU), wherein all the sentiment feature vectors are extracted and opinion mining is carried out to attain a better tweet classifier.
(3) A new Fuzzy Decision Support System is proposed to capture the satisfaction level of the customers based on the derived opinions of the DL model. The Customer Contentment Level (CCL) is derived using this FDSS model combined with the ensemble DL model.
(4) The FDSS model is formulated with suitable membership function definitions and formation of a rule base using the sentiment feature vectors.

The remaining section of the paper is systematized as follows: Section 2 presents the related work conducted in tweet sentiment analysis. Section 3 presents the proposed deep learning approach combined with the FDSS model developed for tweet sentiment analysis. Section 4 details the datasets employed in this paper and the experimental computation of tweet sentiment classification and customer content level evaluated, using the benchmark datasets and the contributions made in the paper. Section 5 provides a detailed discussion and comparison of the proposed TSA classifier with the existing baseline sentiment classifier models. Section 6 gives the closing remarks of the research made on the tweet sentiment classification and the attainment of the customer content level.

2. Related work

Text sentiment analysis has taken its own perspective and importance in natural language processing. And with the growing technology and fast communication, the natural language becomes highly significant and communicative, in which the text sentiment analysis plays a vital role in determining the emotions, various opinions, attitudes and their understanding over numerous customer services. This section presents an overview of recent research conducted on text sentiment analysis over the past few years.

Priyadarshini & Cotton [16] propose a novel Long Short-Term Memory (LSTM) – Convolutional Neural Networks (CNN) – grid search-based deep neural network model for sentiment analysis. Hyperparameter optimization has also been carried out, which resulted in better output than the baseline models [16]. Sharaf Al-Deen et al. [17] developed a hybrid model that combines a deep neural network with a multi-head attention mechanism (DNN – MHAT) [17]. Here, an improved deep neural network was designed to capture the text's actual context and extract the local features of position invariants by combining recurrent bidirectional long short-term memory units (Bi-LSTM) with a convolutional neural network (CNN). Kamyab et al. [8] modelled an attention-based model that utilizes CNNs with LSTM (named ACL-SA) for sentiment analysis [8]. Additionally, the model used an integrated bidirectional LSTM to capture long-term dependencies and was tested on Sentiment140, US-Airline, Sentiment140-MV and SA4A datasets.

Kumar [18] proposed a hybrid Deep Learning (DL) model which leverages the strengths of the BiLSTM and CNN model classes [18]. The sentiments were predicted through a proposed hybrid Deep Learning (DL) model, which leverages the strengths of the BiLSTM and CNN model classes, resulting in an accuracy of 90%. Khan et al. [19] used rule-based, machine learning (SVM, NB, Adaboost, MLP, LR and RF) and deep learning (CNN-1D, LSTM, Bi-LSTM, GRU and Bi-GRU)

Table 1. Review of related works in sentiment analysis.

S.No	Authors	Methodologies employed	Remarks
1.	Priyadarshini & Cotton [16]	Long short-term memory hybrid convolutional neural networks model	96% overall accuracy attained. Requires an estimate of other parameters as well.
2.	Sharaf Al-Deen et al. [17]	Deep neural network with a multi-head attention mechanism (DNN – MHAT). Used a sigmoid classifier for performing sentiment polarity classification.	Should be extended for long reviews.
3.	Kumar [18]	BiLSTM + CNN approach	90% accuracy and analysed geotagged tweets as well.
4.	Khan et al. [19]	Rule-based, machine learning (SVM, NB, Adaboost, MLP, LR and RF) and deep learning approaches	Attained an F1 score of 81.49% and tested on newly created Urdu Tweet datasets
5.	Yan et al. [21]	Sentiment knowledge-based BERT model	Different dimensions of representations were generated for the classification.
6.	Hammi et al. [23]	CBRS (CNN-Bi-RNN-SVM) model for aspect-based sentiment analysis	The F1 value was 85.7% for the considered 2016 restaurant datasets
7.	Jia et al. [24]	BERT-CBLBGA model	Tested on four different datasets and displayed 92.3% accuracy.
8.	Alzaid and Fkih [26]	Fuzzy logic with bidirectional long short-term memory (BiLSTM)	Resulted in an accuracy of 86% and F1 score of 85%. Fuzzy logic was able to handle the uncertainties.
11.	Galal et al. [27]	Design of BERT final layer and hidden embedding layer	F1-score of 64.41 for sarcasm detection and FPN of 75.78 for sentiment classification
12.	Fatima et al. [28]	CNN, LSTM, BiLSTM, GRU, BiGRU, RNN and BiRNN SA classifiers	90%, 91%, 92%, 91%, 91%, 75%, and 89% for each of the mentioned DL algorithms.
18.	Chandrasekaran et al. [30]	CNN + Bi-LSTM with attention networks	Identifying sarcasm remarks with an accuracy rate of 91.76%
19.	Ambreen et al. [31]	BiLSTM combined with fuzzy logic	Accuracy of 95% was attained, outperforming earlier baseline techniques.
20.	Motevalli et al. [32]	Proximate support vector machine + Transductive long short-term memory (TLSTM)	F-measures for English restaurant and Hindi datasets was 90.12 and 84.10 respectively.
21.	Fadhel et al. [33]	HHO + SVM and HHO + X-Boost	For airline tweet reviews, 94.35% with SVM and 96.07% with X-Boost was attained.
23.	Darraz et al. [34]	Hybrid recommendation system using BERT + model-based deep matrix factorization.	Root Mean Squared Error (RMSE) reduced to 0.1 for the restaurant datasets.

techniques for their proposed Urdu dataset with 9312 reviews [19]. Soni and Mathur [20] made a fusion of Long short-term memory (LSTM) and an Encoder with attention for sentiment analysis and experimented on the Twitter Sentiment140 dataset [20].

Yan et al. [21] introduced sentiment knowledge, SK – Bi-directional Encoder Representation from Transformers (BERT), that integrates sentiment knowledge words into independent sequences and then encodes the sequence and context into static and dynamic vectors with the BERT pretrained models [21]. Sun et al. [22] proposed a stock prediction hybrid model named ISI-CNN-LSTM, considering investor sentiment based on the combination of long short-term memory (LSTM) and a convolutional neural network (CNN) [22]. Hammi et al. [23] combined two deep learning models [a Convolutional Neural Network (CNN) and a Bidirectional Recurrent Neural Network (Bi-RNN)] with the classical machine learning model Support Vector Machine (SVM) and tested them on 8000 smartphone reviews [23]. Jia et al. [24] proposed a fused model of text sentiment classification called BERT-CBLBGA, which combines BERT with multiple feature-extraction deep learning models [24]. They used self-attention to fuse the output of TextCNN, BiLSTM and BiGRU to enhance the ability to extract text sentiment features.

Ranjan and Daniel [25] proposed an attention-based model using Dual-LSTM and ConvNet for sentiment classification [25]. Alzaid and Fkih [26] proposed a model that combines fuzzy logic with Bidirectional Long Short-Term Memory (BiLSTM) for the sentiment analysis of students' textual feedback on e-learning [26].

Galal et al. [27] used BERT final layer and hidden layers' embeddings; by proposing different aggregation architectures for text classification tasks such as sentiment analysis and sarcasm detection, BERT also acts as a feature extractor [27]. Fatima et al. [28] applied the deep learning algorithms to 55,000 crawled end-user comments and attained 90%, 91%, 92%, 91%, 91%, 75% and 89% with CNN, LSTM, BiLSTM, GRU, BiGRU, RNN and BiRNN classifiers, respectively [28].

Dervenis et al. [29] employed Machine Learning (ML) and Lexicon-based models for sentiment analysis using an open-source data mining platform [29]. Chandrasekaran et al. [30] employed the deep Bi-directional Long-Short Memory (Bi-LSTM) and a hybrid architecture of the Convolution Neural Network + Bi-LSTM (CNN + Bi-LSTM) with attention networks for identifying sarcastic remarks in a corpus [30].

Ambreen et al. [31] proposed a hybrid deep learning technique enriched with fuzzy logic for an Arabic sentiment-based system [31]. The level of consumer contentment was determined using fuzzy logic, and ablation studies demonstrate the need for an attention mechanism. Motevalli et al. [32] employed multiple kernels for word embedding, with a Transductive Long Short-Term Memory (TLSTM) for sentence representation that considered samples near the test point for better model refinement. The developed model was tested on Hindi datasets [32].

Fadhel et al. [33] selected a small subset of features out of the thousands of features for accurate classification of spam review detection using Harris Hawks Optimization (HHO) in sentiment analysis. This was tested on airline tweet reviews [33].

Darraz et al. [34] integrates sentiment analysis into recommendation systems using a dataset from Yelp, focusing on two domains: restaurants and hotels and leveraged the power of BERT [34]. The authors also integrated sentiment analysis into content-based recommendation using clustering, resulting in improved recommendations with a higher silhouette score.

Table 1 presents a brief review of the earlier baseline models as applied for the tweet sentiment analysis and classification. From the detailed review made, it has been understood that DL models have paved their way into the sentiment analysis and classification of tweet datasets. DL models also lack in attaining the complete semantic information or the customer content level that is highly significant for sentiment classifier performance. A few of the lacunae that the existing sentiment analysis approaches come across include:

(1) Suffers in handling the higher-dimensional feature vectors.
(2) Deliberates on different features as belonging to the same.
(3) Unable to attain the CCL from the tweets.
(4) For different and higher-dimensional features, the accuracy rate decreases significantly.
(5) Delayed convergence because of dimensionality issues.
(6) Poor sentiment classification in case of highly complex tweet sentences with more stop-words and other symbols.

Under these conditions, there is continually a need to develop an improved and enhanced version of the tweet sentiment analysis and classifications by better feature extraction and capturing the required relevant features from tweet sentences. In this manner, it paves the way to attain a more accurate TSA model using the DL ensemble models for opinion mining and the FDSS model for analysing the customer content level. Due to this, this paper focused on the development of the most effective and efficient TSA model using the proposed ensemble Bi-LSTM – GRU combined FDSS system for the considered tweet datasets.

3. Proposed methodology

An innovative fusion of an ensemble bidirectional long short-term memory network with gated recurrent units, combined with a fuzzy decision support system model, is designed and implemented in this research study to perform as an effective tweet sentiment analyser and classifier. The devised new eBi-LSTM-GRU-FDSS model possesses the adaptability and ability to learn the most prominent tweet features from the tweet sentences and thereby perform the classification of the tweet opinions. The new FDSS intends to determine the customer contentment level based on the classified opinions and derive the satisfied or unsatisfied level of customers. The new ensemble Bi-LSTM-GRU with its six significant modules is capable of classifying the tweets based on their prominent features and attains superior performance metrics. The generalization ability and learning ability are significantly improved by employing the new eBi-LSTM-GRU-FDSS model for the tweets collected from the X platform.

3.1. Pre-processing and embedding

In the preprocessing stage of Twitter Sentiment Analysis (TSA), multiple steps are undertaken to clean and normalize tweets before they are used for training sentiment classification models. Since tweets often contain noise in the form of emojis, slang, repeated characters, and special symbols, preprocessing is essential to improve model accuracy. The steps adopted for preprocessing are as follows:

(1) *Tokenization*: Tweet sentences are broken down into individual words, known as tokens. This step is performed using NLTK's word_tokenize function, ensuring proper separation of words, punctuation, and contractions (e.g. *"I can't"* is tokenized as *["I", "ca", "n't"]*).
(2) *Noise handling*: Tweets often contain unwanted characters, special symbols, and HTML tags. To address this, URLs, mentions (@username), hashtags (#topic), and special characters are removed using regular expressions. Additionally, duplicate spaces and unnecessary punctuation are eliminated to ensure clean text.
(3) *Normalization*: To ensure consistency, multiple transformations are applied:
    - *Lowercasing*: All text is converted to lowercase to maintain uniformity (e.g. *"GREAT"* → *"great"*).
    - *Diacritic removal*: Accented characters are converted to their standard ASCII equivalents using Unicode normalization (e.g. *"café"* → *"cafe"*).
    - *Handling elongated words*: Words with repeated characters are shortened to two consecutive occurrences using regex (e.g. *"goooood"* → *"good"*).
    - *Slang replacement*: Common internet slang and abbreviations are replaced with their standard forms using a predefined dictionary (e.g. *"u r gr8"* → *"you are great"*).
(4) *Stemming*: Words are reduced to their root forms by means of Porter Stemmer, which removes

common suffixes (e.g. *"running"* → *"run"*, *"happily"* → *"happi"*). This step helps minimize vocabulary size and improve text generalization.

(5) *Stop-word removal*: Words that carry little semantic meaning, such as "is," "the," "in," and "at," are removed using Natural Language Toolkit's predefined list of English stop words. Eliminating these words reduces noise and improves the relevance of textual features in sentiment classification.

(6) *Handling emojis*: Emojis often carry significant sentiment information. Instead of completely removing them, they are either
- *Replaced with their textual meaning* using emoji dictionaries (e.g. "😍" → *"love"*) OR
- *Removed entirely* if they do not contribute to sentiment analysis.

(7) *Handling slang and abbreviations*: Since Twitter users often use informal language, a predefined dictionary-based approach is used to replace slang words with their full forms (e.g. *"lol"* → *"laughing out loud"*).

After the execution of the pre-processing steps as above, including the parsing of tweets, linguistic word analysis, tokenization, normalization, stemming and stop word removal, the meaningful tweet words with the relevant features get extracted after removal of special characters and the execution of pre-processing steps. Extracted features shall be passed on to the formation of word tokens. The word features shall be identified by each of the tokens, and then these word features are passed on to the Bi-directional Encoder Representation from Transformers (BERT) embedding layer. The preprocessing and embedding process in the proposed TSA classifier model is shown in Figure 2.

Using BERT in the proposed TSA classifier model, a fixed representation of Word2Vec is used, and a text context model is not used. BERT is capable of operating in both directions, and this study uses BERT's masked language model (MLM) to represent the input data sequence by utilizing the extracted word feature vector. In this study, the BERT language model is utilized to generate sentiment-enhanced word embeddings. Sentiment word tokens captured are used in BERT, and these are converted into a word vector using the data features. From the BERT embedding layer, the word vectors are then processed using the Ensemble Bi-LSTM-GRU model for obtaining the positive, negative or neutral opinions, and then the FDSS model acts to detect the customer contentment level.

3.2. Ensemble Bi-LSTM-GRU model development

A bi-directional long short-term memory network and gated recurrent units are combined to form an ensemble Bi-LSTM-GRU model, merging the merits of both Bi-LSTM and GRUs. Bi-LSTM operates in a bi-directional mode and it captures the previous, past and future context information corresponding to the input sequence elements. Further to this, the GRU model, comprising fewer parameters and a simple architecture, holds the sequential tweet data in memory selectively and/or operates in forget mode for a particular time duration. GRU individually is likely to have overfitting issues and does not operate for long-term dependencies, and Bi-LSTM requires large amounts of training data for representing the input tweet sequence in a meaningful manner. On sufficient training data not presented to the model, the required generalization ability shall not be attained when new tweet data are presented. Under these conditions, this research study formulated a novel ensemble version of Bi-LSTM and GRUs combined to overcome the overfitting issues of GRUs and improve the generalization ability of Bi-LSTM models by presenting them with high volumes of tweet data. This section details the formulation of the ensemble Bi-LSTM-GRU model.

3.2.1. Bi-LSTM architecture and learning process

A deep recurrent learning neural model is the bi-directional long short-term memory network that is capable of processing the sequential information of the sentences. Bi-LSTM comprises two LSTM layers – one LSTM layer processing the input sequence in the forward direction and the other layer processing the sequence in the backward direction, wherein the future context of the sequence gets captured. A combination of these two forward and backward LSTM layers formulated the bi-directional LSTM, which accesses the significant representation of the input sequence data, thereby improving the overall performance of the model [35]. Figure 3 shows the architecture of the bi-directional LSTM learning model.

Bi-LSTM models effectively manage long-term dependencies in sequential data, utilizing their internal memory state to retain information over extended periods.

The past information gets captured in the forward pass of the Bi-LSTM model, and the future data sequence gets captured in the backward pass i.e. in the upcoming time steps. The Bi-LSTM model operates on evaluating the following parameters based on weight updates and the activation functions employed.

$$in_k = (v^L_{xi} x_k + v^L_{hi} o_{k-1} + bs_i) \tag{1}$$

$$fg_k = (v^L_{xf} x_k + v^L_{hf} o_{k-1} + bs_f) \tag{2}$$

$$yn_k = (v^L_{xy} x_k + v^L_{hy} o_{k-1} + bs_y) \tag{3}$$

$$g_{tk} = tanh(v^L_{xg} x_k + v^L_{hg} h_{k-1} + bs_g) \tag{4}$$

$$c_{tk} = (fg_k . c_{tk-1} + in_k . g_{tk}) \tag{5}$$

$$on_{tk} = (yn_k . tanh(c_{tk})) \tag{6}$$

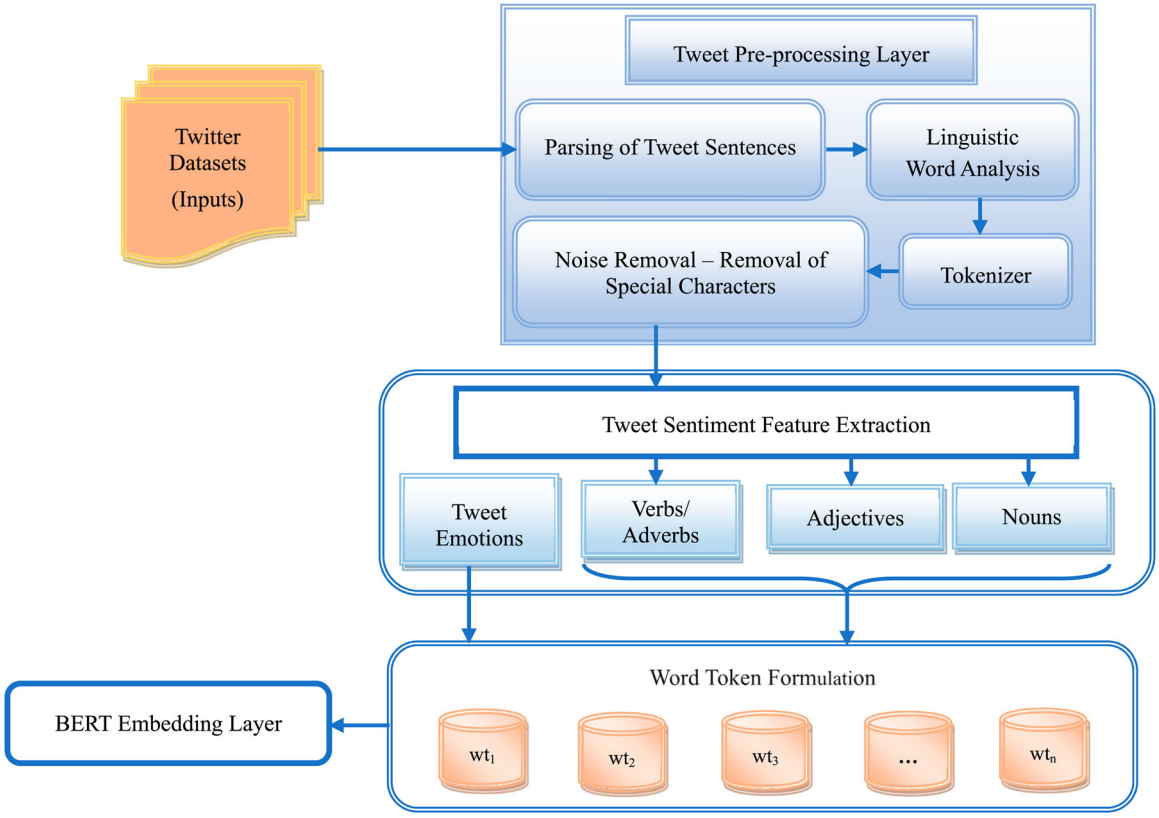


Figure 2. Preprocessing and embedding process in the proposed TSA classifier model.

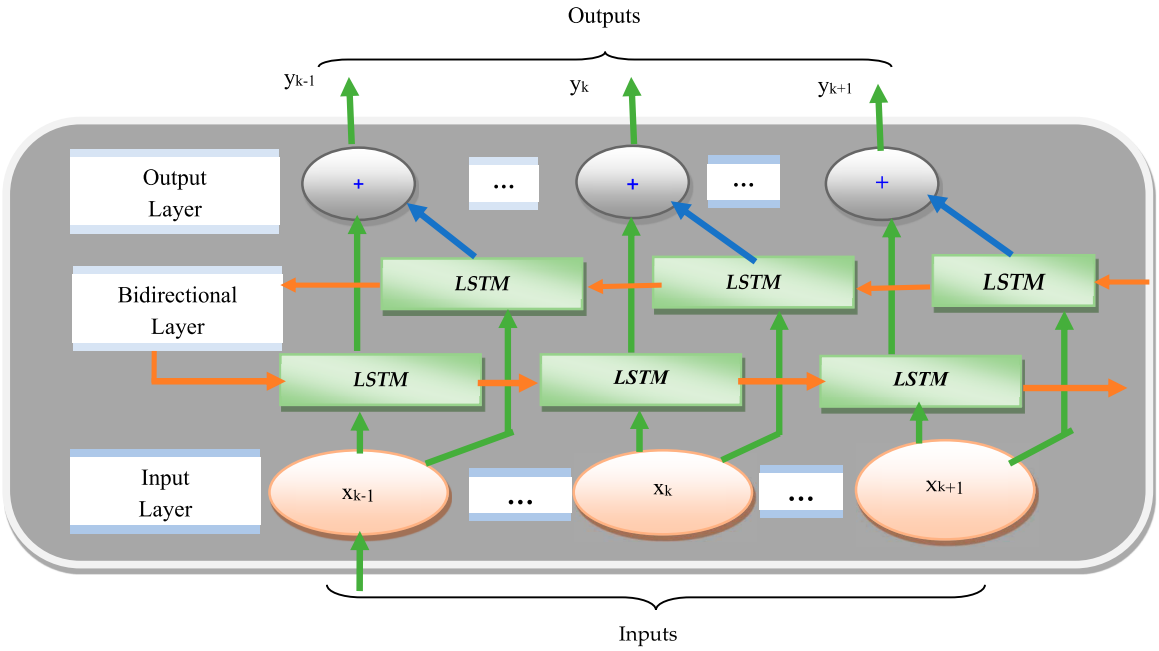


Figure 3. Bi-LSTM recurrent learning architecture.

Referring to Equations (1)–(6), "σ" specifies sigmoidal function to evaluate the output of the model and "*tanh*" denotes the tangential function that determines the output pertaining to the layers, "*o*" is the cell memory output, "*c*" denotes the intermediate gate and "*in*", "*fg*" and "*on*" specifies input gate, forget gate and output gate respectively, "*k*" represents the time steps, "*bs*" is the bias levels, "*v*" is the weights between all the layers and "*L*" indicates the window length for the given data time sequence. The Bi-LSTM model processes information in both forward and backward directions, retaining cell memory output while also applying the activation function to compute the final result. The forget gate of the model identifies the data that has to be retained or forgotten (removed) and proceeds towards determining the final output from the model.

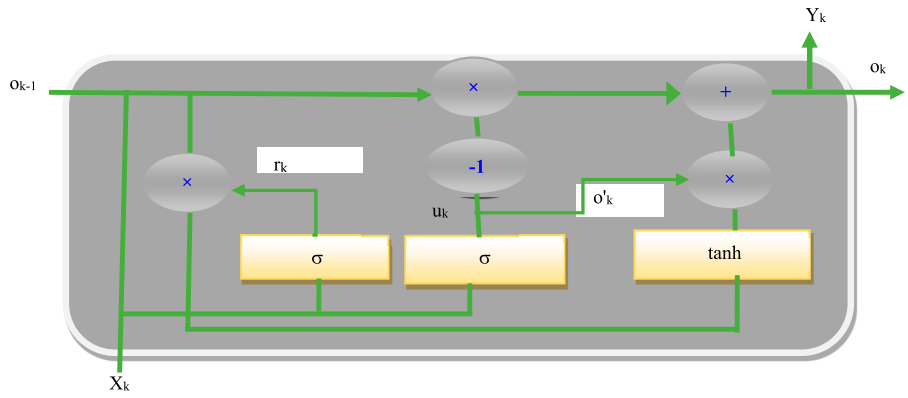


Figure 4. GRU learning architecture.

The Bi-LSTM model in this study is configured as given below:

(1) Number of layers: 2 layers, one in the forward pass and the other in the backward pass, for capturing the long-term complex dependencies. The input sequence is fed to both layers.
(2) Hidden Neuron Units: In each LSTM layer, there are 156 hidden neuronal units that resulted in better performance metrics. The fixation of these neuron units in the hidden layer was done initially on a trial-and-error basis, and subsequently, based on the training process, it was fixed to be 156 hidden units in the two LSTM layers.
(3) Output evaluation: For the gates, "sigmoidal" activation function is utilized for the cell state, "tanh" activation function is employed.

3.2.2. GRU architecture and model

The architecture of a deep recurrent learning model with a minimal number of gates is the gated recurrent unit model. The gates in GRU are minimized compared to those of the Bi-LSTM recurrent learning model. Figure 4 shows the architecture of the GRU deep learning model.

In the layer design of the GRU model, the input gate and forget gate are controlled by a single gate, thereby the simplification of the GRU architecture is attained. In Figure 4, when ok = 1, the input gate closes, and a new input sequence is permitted to pass through the gate and the forget gate opens. When ok = 0, the input gate opens to permit a new sequence and the forget gate closes. The new input data combines with that of the memory cells to attain the output corresponding to the new state with the help of the reset gate. The following equations are used to obtain the output from the GRU deep learning model.

$$r_{tk} = (v^L_{xr} x_k + v^L_{hr} o_{k-1} + bs_r) \quad (7)$$

$$u_{tk} = (v^L_{xu} x_k + v^L_{ou} o_{k-1} + bs_u) \quad (8)$$

$$y_{tk} = (u_{tk}.y_{tk-1} + (1 - u_{tk}).y_{tk-2}) \quad (9)$$

where "r_{tk}" is the reset gate, "u_{tk}" indicates update gate, "y_{tk}" corresponds to the output gate, "k" denotes the time step, "L" is the word length, "v" is the weights between the layers and "bs" represents the bias level employed in the model.

3.2.3. Ensemble Bi-LSTM-GRU model

Ensemble learning attempts to be a combination of multiple deep learning models to produce a more accurate and robust classifier in comparison with that of an individual learning model. The fundamental idea behind ensemble learning is to leverage the merits by combining and aggregating the classifications of multiple models, each with its own merits and demerits. This led to improving the generalization ability and enhanced performance. Ensemble deep learning is a compensation for the deep learning algorithms that are computationally expensive, and numerous individual models are designed to learn from the same input sequences and result in an aggregation of outputs from which the final classification output is computed. Building on the advantages of ensemble learning, this study proposes an ensemble model combining bidirectional long short-term memory networks and gated recurrent units to analyse sentiment in benchmark tweet datasets. Figure 5 illustrates the structure of the ensemble Bi-LSTM-GRU model.

The Bi-LSTM model processes information in both forward and backward directions, so as to learn the discriminative data feature from the input sequence for each time step. 30% neurons are dropped out in the dropout layer, such that the generalization ability is maintained. 156 hidden neuron cells exist in the Bi-LSTM and GRU model, thereby the contextual features are extracted. In GRU processing, the dropout layer is executed with 20% neurons being dropped out. Due to these operations, the over-fitting occurrences are completely avoided, and a flattened layer is operated to compute a single column word feature vector. The final layer of the eBi-LSTM-GRU model is a fully connected dense layer with 60 neurons, followed by an output layer with 3 neurons, corresponding to the three opinion classes.

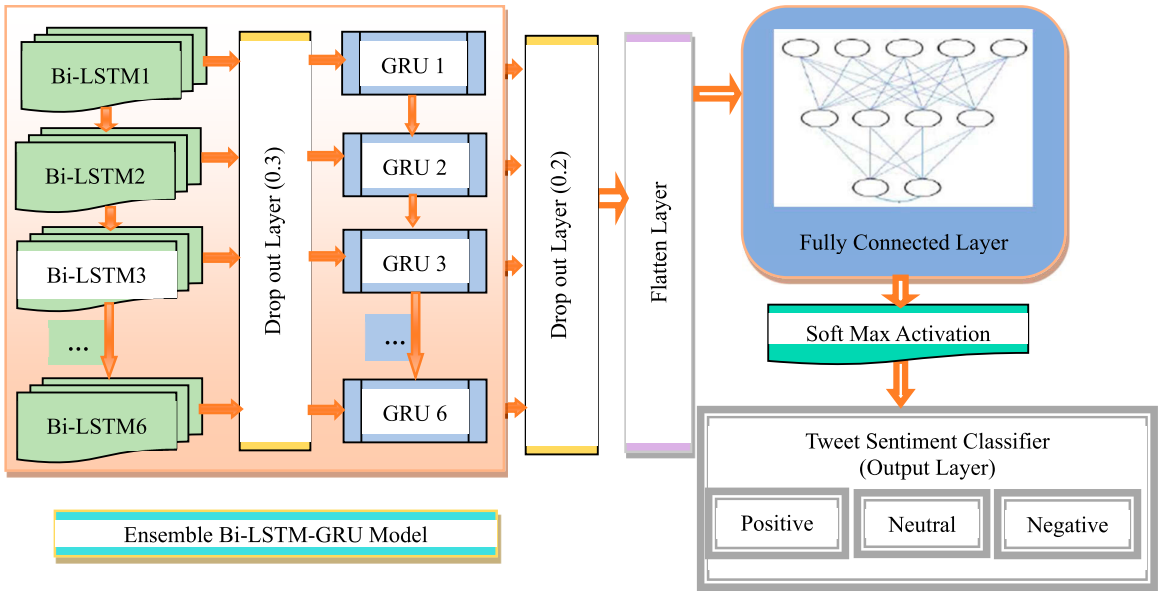


Figure 5. Configuration of ensemble Bi-LSTM-GRU model.

The class labels in the output layer correspond to three output neurons – "*+1 – positive*", "*0 – neutral*" and "*−1 – negative*" for executing the sentiment classification and the "*softmax*" function is utilized to generate the probability of the word sequences. The defined activation functions in the modelled new eBi-LSTM-GRU model are given by

$$= \frac{1}{1 + e^{-yp}} \tag{10}$$

$$tanh = \frac{e^{yp} - e^{-yp}}{e^{yp} + e^{-yp}} \tag{11}$$

$$softmax(x_k) = \frac{e^{x_n}}{\sum_{c=1}^{N} e^{x_c}} \tag{12}$$

where "x" denotes the input vector, "c" is the number of classes, "N" denotes the total number of samples and "n" corresponds to the respective sample number.

3.3. Modelling the fuzzy decision support system model

Building on our earlier framework [36], this study incorporates a Fuzzy Decision Support System to enhance decision-making accuracy in sentiment classification. Tweet sentiment analysis generally contains imprecise and subjective vocabulary of the language used in the tweets. An approach to handle the uncertainties and imprecision is the Fuzzy Logic (FL), wherein it is capable of capturing the user opinions that are revealed from the considered tweet data. Human reasoning ability is employed in the fuzzy logic approach to handle the existing ambiguity and uncertainty. For tweet sentiment analysis, fuzzy systems are more suitable to provide robust solutions in sentiment classification applications. This research study focuses on modelling a Fuzzy Decision Support System (FDSS) model to evaluate and analyse the customer contentment level based on the tweet opinions (Figure 6). The ensemble Bi-LSTM-GRU model evaluates the tweet sentiment scores (positive/neutral/negative) and presents them to the modelled FDSS. FDSS model utilizes these sentiment scores as input, customer contentment level as output and defines membership functions for all the variables used for input and output. A fuzzy rule base is formulated based on the membership functions defined, and then the output delivers the satisfaction level of customers from the tweets. Design and development of the FDSS is done by membership function definitions and framing fuzzy rules, and are presented in detail in this section [37].

3.3.1. Implementing the steps of the FDSS model

Figure 7 shows the flowchart depicting the procedural steps adopted for the design and development of a fuzzy decision support system model. The steps adopted for implementing the proposed FDSS model are as follows:

(1) *Input variable definitions* – The sentiment scores from the ensemble Bi-LSTM-GRU are used as inputs for the FDSS model. These scores indicate the degree of opinions of the users as indicated in the tweets. In this step, linguistic variables are well-defined for both the input parameters and the output parameters.
(2) *Fuzzification* – This step transforms the crisp input variables into the fuzzy variables and defines the membership functions for each of the input and output variables identified. The membership function definition is carried out to bring out the subtlety of the sentiment scores and the customer contentment levels. In this study, triangular membership functions are defined for both the

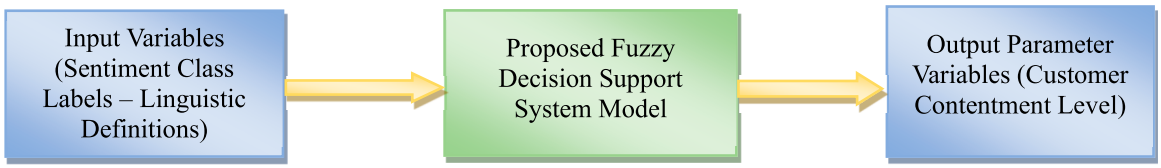


Figure 6. FDSS configuration with input/output variables.

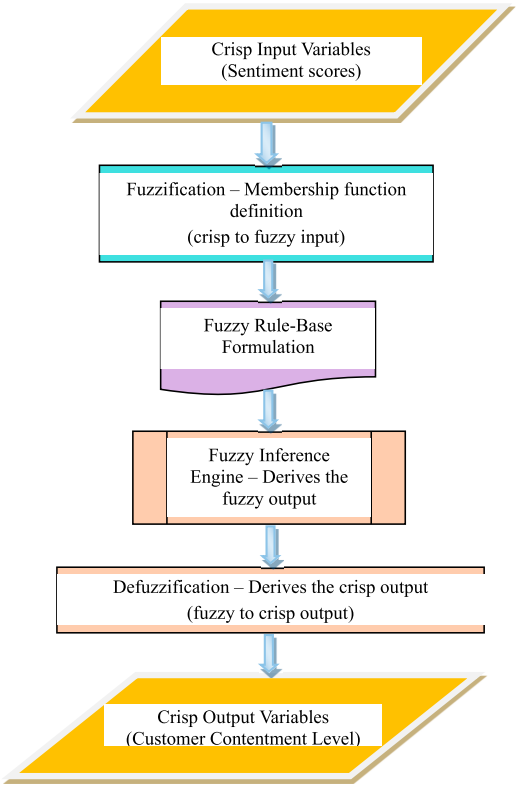


Figure 7. Process flow of the FDSS model.

sentiment scores and CCL output. The proposed FDSS is numerically expressed using,

(3) $P = \{x, \mu_P(x) | x \in X\}$ (13)

(4) In Equation (13), $\mu_P(x)$ represents the membership function of x in P, and the universal set is X.

(5) *Fuzzy rule-base formulation* – The fuzzy decision support system model constructs its rule base using "*If–Then*" rule statements. The rules are formulated based on the domain knowledge of the problem and the level of expert understanding. The rules pertain to the specific combinations of the defined fuzzy input and output variables with respect to the customer contentment level.

(6) *Fuzzy Inference mechanism* – In this mechanism, the fuzzy engine operates to define input and output variables and use the formulated fuzzy rules to derive the fuzzy output. On completion of the fuzzy inference step, the fuzzy output will be obtained, depicting the CCL opinion.

(7) *Defuzzification* – Lastly, the centroid defuzzification method is employed to transform the fuzzy output into a crisp value, enabling the determination of the precise customer contentment level based on these actual crisp values.

Table 2. Linguistic variable definition.

Input/output variable	Input/output parameters	Linguistics
Input variable	Sentiment Scores	Negative Neutral Positive
Output variable	Customer contentment level	High – Satisfaction Low – Dissatisfaction

3.3.2. Membership function definitions and fuzzification

For the input variables, which are the output from the eBi-LSTM-GRU model, the sentiment scores (positive/neutral/negative) are to be defined with the linguistics and their membership function values. The output variable is the customer contentment level (satisfaction/dissatisfaction) for which the linguistic and membership assignment has to be made. Table 2 provides the linguistic-defined for the input and output variables.

The fuzzification process is executed, and the classical input is transformed into the corresponding fuzzy inputs. The linguistics are defined for the input variable (sentiment scores) and output variable (customer contentment level) as in Table 2. The input variable is assigned three linguistic variables, and the output variable CCL is assigned two linguistic parameters. In the proposed study, the FDSS model utilizes a triangular membership function to define its membership function. The membership degree of each linguistic term for an input variable is specified within the range of 0–1, so that it gets transformed to a fuzzy set. The Triangular Membership Function (TMF) has been chosen for fuzzification in the FDSS model due to its computational efficiency, simplicity and suitability for sentiment classification tasks. Unlike the Gaussian function, which requires complex parameter tuning, TMF is easy to interpret and implement, as it only requires three parameters – lower bound, peak and upper bound – making it ideal for handling linguistic variables like Low, Medium and High customer contentment. Additionally, TMF provides well-defined boundaries, ensuring a smooth yet distinct transition between sentiment categories, which is crucial in sentiment analysis. Its linear nature enables faster computations, making it suitable for large-scale datasets like T4SA and Sentiment140. Figures 8 and 9 show the membership

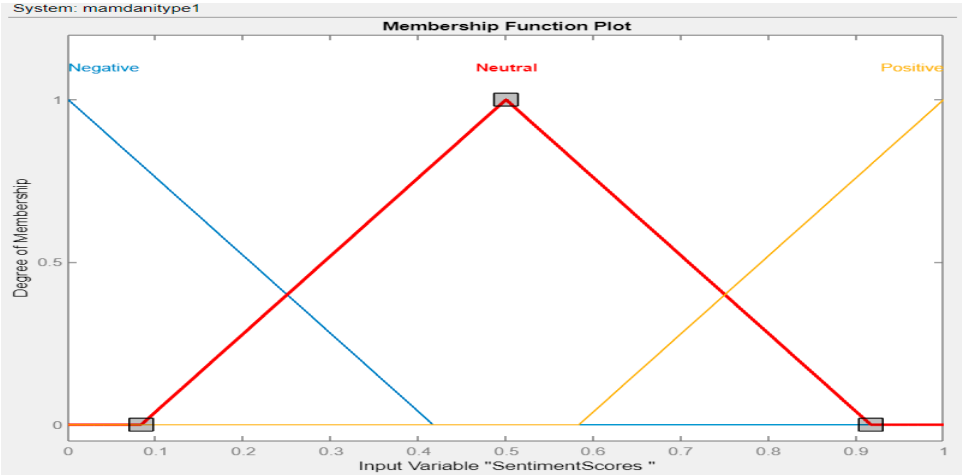


Figure 8. Input membership function definition.

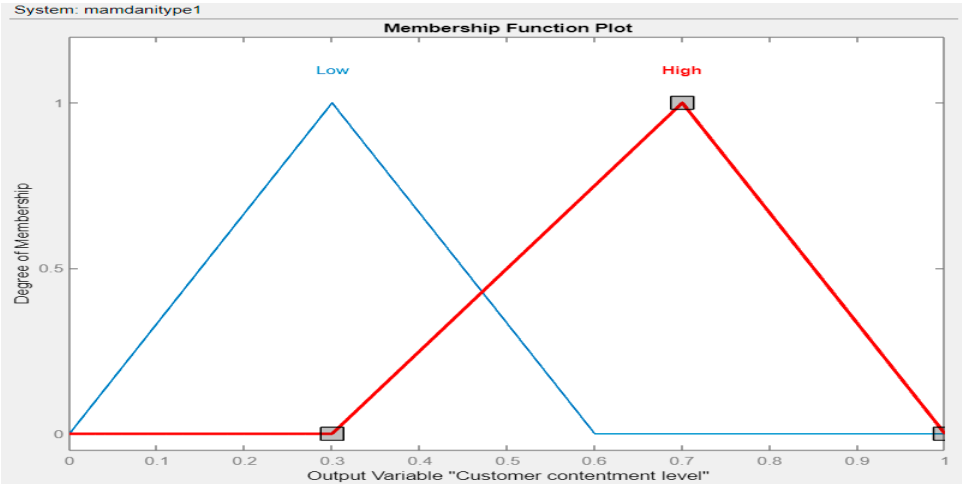


Figure 9. Output membership function definition.

functions defined for the input and output variables based on their linguistic assignments.

3.3.3. *Fuzzy rule-base formulation*

A fuzzy inference engine is designed by formulating the fuzzy rule-base and deriving the fuzzy output. Fuzzy rules are framed using the "IF–THEN" statements, wherein "IF $<$ antecedent $>$ " specifies the input and "THEN $<$ consequent $>$ " indicates the attainment of the outputs. The general expression for fuzzy rule base formation is given by

$$IF < antecedent - input\ condition > THEN$$
$$< Consequent - output\ action >$$

Table 3 shows the formation of fuzzy rules for the analysis of customer satisfaction level using the FDSS model.

The Fuzzy Decision Support System (FDSS) applies sentiment-based rules to classify customer contentment based on sentiment scores derived from the T4SA (Twitter for Sentiment Analysis) dataset. For instance, a tweet like "*Worst restaurant experience ever. The food was cold, and the service was awful!*" with a sentiment score of −0.82 (negative) is classified under Rule 1, leading to Low Customer Contentment. Similarly, a neutral tweet such as "*The hotel stay was okay, nothing special but not bad either*," with a sentiment score of 0.08, is processed under Rule 2, also resulting in Low Customer Contentment, since it lacks strong positive sentiment. On the other hand, a highly positive tweet like "*Absolutely loved the new gadget! Works perfectly and looks amazing*", With a sentiment score of 0.89, it falls under Rule 3, indicating High Customer Contentment. These examples demonstrate how FDSS systematically determines customer satisfaction based on sentiment analysis from real-world datasets.

Table 3. Fuzzy rules for TSA.

S.No	Rule	Weight	Name
1	If the Sentiment Scores are Negative, then Customer Contentment is Low	1	rule1
2	If the Sentiment Scores are Neutral, then Customer Contentment is Low	1	rule2
3	If the Sentiment Scores are Positive, then Customer Contentment is High	1	rule3

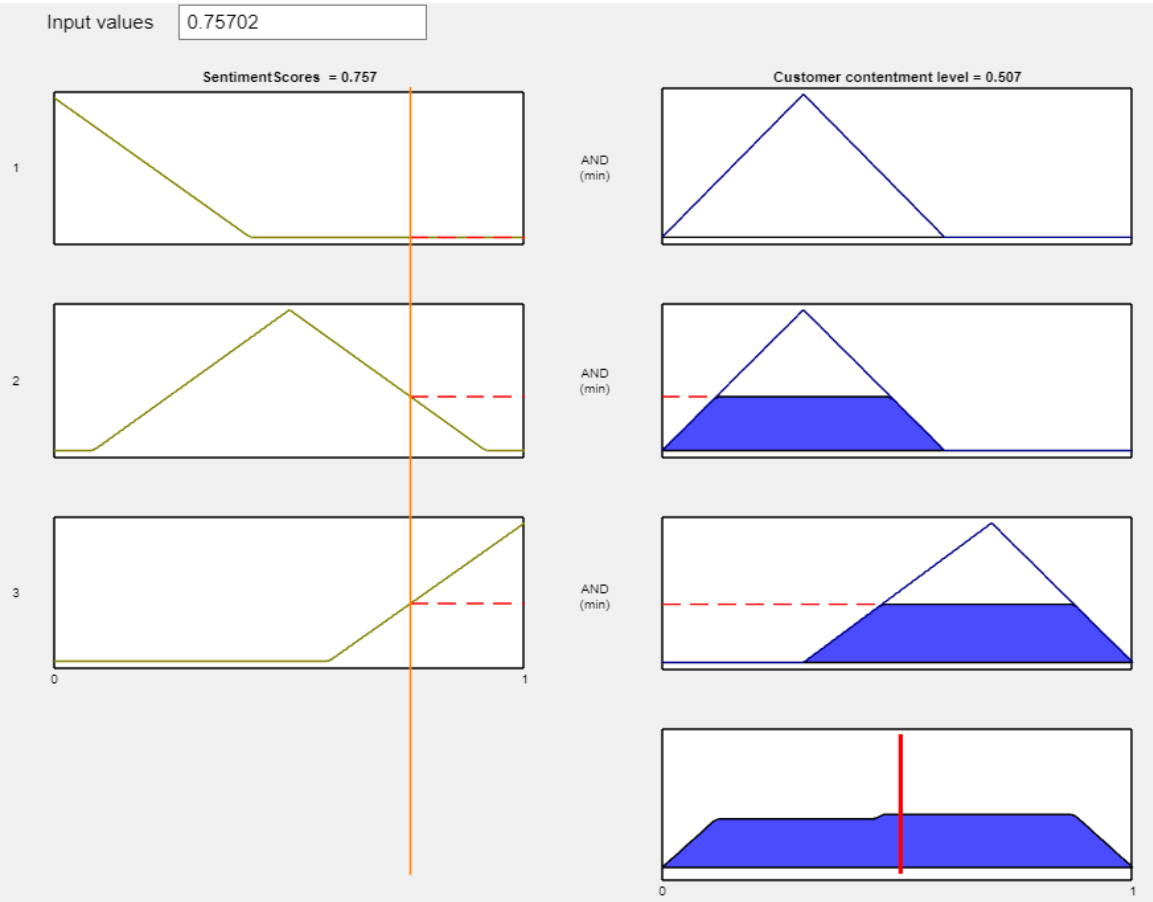


Figure 10. Rule inference for customer contentment level.

3.3.4. Output linguistic determination and defuzzification

In this study, the Mamdani fuzzy inference system is utilized, and for determining the final crisp output, the centroid approach is used. The fuzzy outputs are transformed into numerical classical outputs using

$$x^* = \frac{\int x_P(x)dx}{\int_P(x)dx} \tag{14}$$

The fuzzy set membership obtained from the fuzzy rule-base will be μP(x) and this is transformed to numerical outputs employing Equation (14). Figure 10 shows the rule inference process that provides evaluated numerical output for the customer contentment level using the proposed FDSS model. Figure 10 presents a clear inference between the input sentiment scores and the output customer contentment level. For example, in Figure 10, a sentiment score of 0.757 results in a customer contentment level of 0.507 which is considered satisfactory.

3.4. Fusion of eBi-LSTM-GRU with the FDSS model

To start with, the tweet sentences corresponding to the datasets pass through the pre-processing stage and feature extraction is carried out. The pre-processing steps – parsing, linguistic word analysing, tokenization and noise removal are done and the prominent tweet words after removing all stop words are passed on to formulating the word tokens. The tokens generated corresponding to the word features are then presented to the Bi-directional Encoder Representation from Transformers (BERT) embedding layer. BERT operating in both directions uses a masked language model to represent the input tweet sequence along with a feature vector. In this work, the BERT embedding layer is utilized to generate word embeddings, where the word vector is formed based on the extracted word features. The newly modelled ensemble Bi-LSTM-GRU model takes the word vectors generated from the BERT layer as input.

The eBi-LSTM-GRU model is designed with 6 Bi-LSTM and GRU models, with the respective dropout layers to handle the computations. The flattened layer obtains the single-column feature vector, which is then forwarded to the fully connected dense layer. The output from the eBi-LSTM-GRU model is the sentiment scores denoted by the class labels as "*+1 – positive*", "*0 – neutral*" and "*−1 – negative*" for performing the sentiment classification and "*softmax*" function to attain the probability of the word sequence. The output sentiment scores from the ensemble Bi-LSTM – GRU are presented to the new fuzzy decision support system for obtaining the customer satisfaction level.

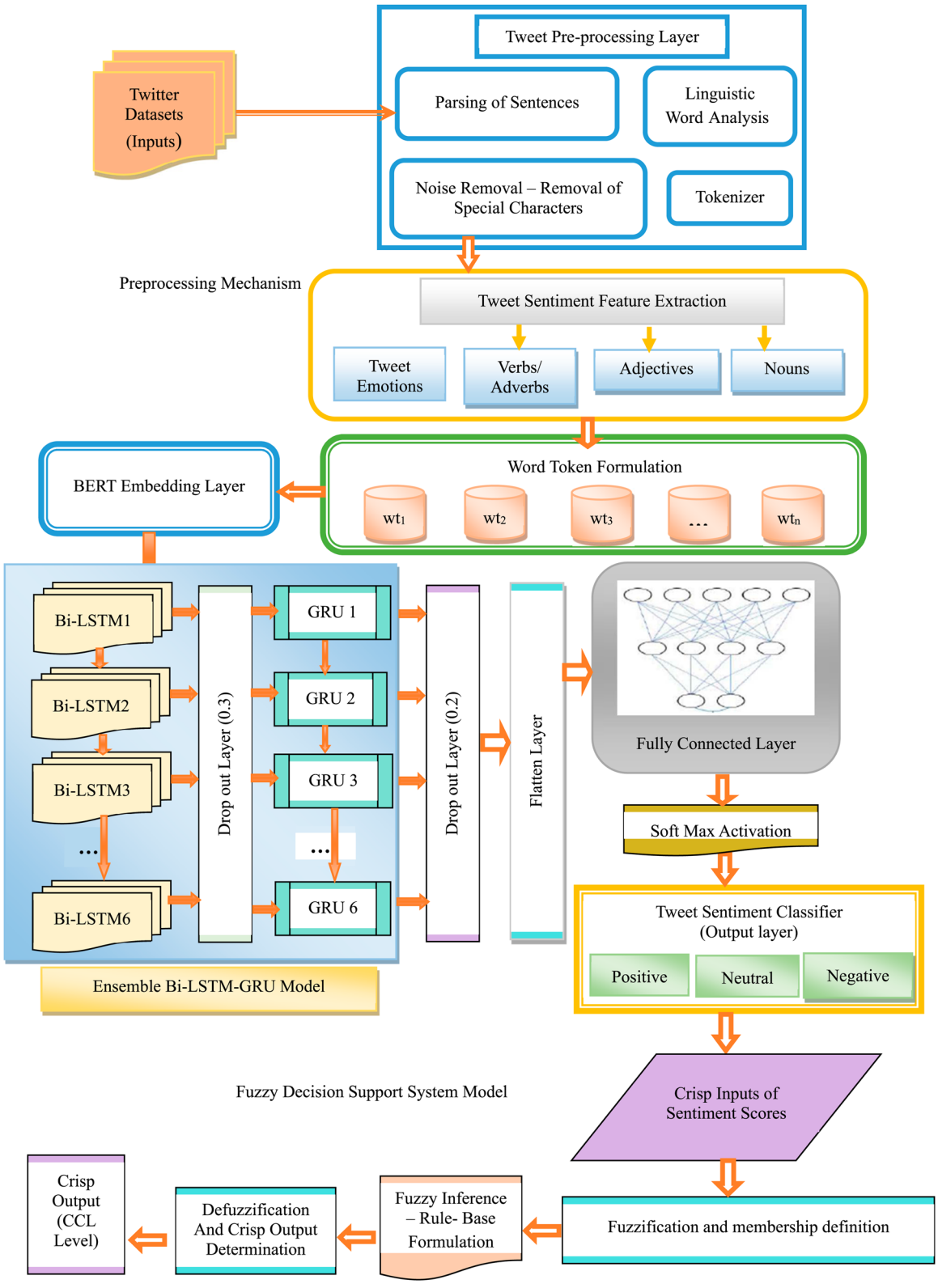


Figure 11. Modelled new eBi-LSTM-GRU-FDSS as a sentiment classifier.

The input sentiment score and the output customer contentment level are assigned with suitable membership functions. Subsequently, the fuzzy rules are formulated to derive the satisfaction of the customer based on their tweet sentiment scores. The fusion of the eBi-LSTM-GRU with the FDSS model is designed to attain an accurate sentiment score based on the tweet sentences and also to evaluate the satisfaction level of customers. The modelled new eBi-LSTM-GRU-FDSS, combining the features of the Bi-LSTM-GRU and the fuzzy system model, results in deriving the prominent features and sentiment scores from the input tweet sentences and thereby determining the contentment level of the customers. The combination of all these models is intended to derive the accurate sentiment class labels of the tweet opinion and better predictions on

Table 4. Parameter setting of ensemble Bi-LSTM-GRU model.

Parameters	Parametric values
Parameter Optimization	Random
Dropout (Bi-LSTM and GRU)	0.4
Dropout (GRU and Fully Connected Layer)	0.3
Learning rate	0.01
Word embedding mechanism	BERT
Dimension of word embedding	100
Mini-batch size	124
Convergence	10^{-6}

Algorithm: Proposed Ensemble Bi-LSTM-GRU – FDSS model

1. Input: Tweet data sequence
2. Output: Classifier labels, positive (+1), neutral (0), negative (−1)
   Customer contentment levels (satisfied +1, dissatisfied −1)
3. Set Learning rate, mini-batch size, word embedding dimension, dropout percentage, convergence criterion, fuzzy linguistic definitions
4. Pre-processing (Noise removal, parsing, linguistic word analysis, tokenizer)
5. Word token formulation and synthesis (wt1, wt2, wt3, … ,wtn)
6. BERT Word Embedding layer (emotion sequence vector, word sequence vector)
7. Bi-LSTM (hidden layer size, batch size)
8. Bi-LSTM Weight and Bias Optimization (EAQ optimizer)
9. Dropout Layer (0.3)
10. GRU (hidden layer size, batch size)
11. GRU Weight and Bias Optimization (EAQ optimizer)
12. Dropout Layer (0.2)
13. Flatten Layer (Single column word feature vector)
14. Fully Connected (FC) Layer (60 neurons)
15. FC (three classes – positive, neutral, negative)
16. FC(Softmax)
17. Output (Sentiment Classifier Output – sentiment scores)
18. Invoke: Fuzzy decision support system model
19. Define – Membership functions (sentiment scores)
20. Formulate the fuzzy rule base
21. Derive fuzzy output from the fuzzy inference engine
22. Apply defuzzification(centroid) – derive the crisp output (customer contentment level)
23. Output (CCL – Satisfied/Dissatisfied)

the customer contentment levels (satisfied/dissatisfied). Figure 11 illustrates the complete workflow of the proposed FDSS model, which is based on an ensemble of Bi-LSTM and GRU. Table 4 gives the parameter setting corresponding to the ensemble Bi-LSTM-GRU model.

The proposed hybrid deep recurrent learning algorithm with fuzzy system design is implemented with the following steps.

4. Results and discussion

In this study, the modelled new deep ensemble Bi-LSTM-GRU-based fuzzy decision support system model is trained and tested for the tweet datasets. The noise present in the tweet datasets is removed in the pre-processing stage, and after the tokens are generated in the tokenization, they are passed through the BERT embedding layer to attain the word feature vector. The generated word feature vectors are then presented as input to the new eBi-LSTM-GRU-FDSS model to attain the tweet opinions and the customer content level. Simulations of the proposed ensemble deep learning recurrent bi-directional LSTM and GRU model with the fuzzy model are run in MATLAB 2023a environment.

4.1. Tweet datasets used in the model

The following Twitter datasets are employed in this study to carry out the tweet sentiment analysis and classification.

(1) *Sentiment 140 dataset* [38]: Graduate students of Stanford University developed this dataset. This dataset consists of 1,600,000 tweets, where each tweet is classified into one of three polarity categories: negative, neutral or positive.
(2) *T4SA dataset* [39]: Twitter for Sentiment Analysis dataset comprises 1,179,957 tweets and is categorized into negative and positive classes. The T4SA dataset is the accumulation of the tweets carried out from July to December 2016.
(3) *Airline Twitter dataset* [40]: 14,641 tweets are present in the airline Twitter dataset pertaining to the services of the US airlines, that is, the tweets analyse the travellers' tweets during February 2015 expressing their feedback with 2363 positive, 3099 neutral and 9179 negative reviews.

4.2. Model parameter – initialization

The proposed new Ensemble Bi-LSTM-GRU with the fuzzy decision support system model, along with the earlier classic methods for the considered datasets, are coded in the MATLAB environment, and the steps of pre-processing, tokenizing and BERT embedding are executed for the datasets. In the beginning, for the tweet datasets, the first 30 words of tweet comments with a padding size of 30 are considered, and the dimension of embedding is 100 in the embedding layer. The FDSS model is designed with the triangular membership functions (Figures 8 and 9) for the tweet opinions (input) and customer contentment level (output), along with the defined fuzzy rule formulation, as shown in Table 3.

For the eBi-LSTM-GRU model, mini-batch size is set as 124, and the dropout in Bi-LSTM and GRU is 0.4, and that between GRU and the fully connected dense layer is 0.3. The training and testing are carried out until the best mean accuracy is attained, and this is computed for convergence. To handle overfitting issues, 4-fold cross-validation is done, and for the training process, the learning rate is set to 0.01.

4.3. Evaluation metrics used in the study

The modelled ensemble deep learning recurrent neural network with the FDSS model is tested for its effectiveness and validity with the Evaluation indices – Recall, Precision, F1 Score (F1), Classification Accuracy (CA)

Table 5. Evaluation indices.

Indices	Criterion for evaluation
Recall	$Recall = \frac{TP}{TP + FN}$
Precision	$Precision = \frac{TP}{TP + FP}$
F1 Score (F1)	$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision}$
Classification Accuracy (CA)	$CA = \frac{TP + TN}{TP + FP + TN + FN}$
Cohen-Kappa Score (CK Score)	$CK = \frac{2 \times (TP \times TN - FP \times FN)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (TN + FN)}$

∗ TP – True Positive, TN – True Negative, FP – False Positive, FN – False Negative.

and Cohen-Kappa Score (CK). These evaluation metrics are widely used for text and tweet sentiment analysis [7–9,41–46], and the formula used for evaluation is given in Table 5.

4.4. Results and comparative analysis

To establish the effectiveness and the superiority of the developed new ensemble Bi-LSTM and GRU model hybridized with the fuzzy model, simulation experimentation was done on the benchmark Twitter datasets, as given in Section 4.1. Results are presented and detailed in this section for the proposed sentiment classification approach and its suitability and superiority is proven on comparison with the other baseline approaches from the prevailing literature; which includes, CNN [10], LSTM [11], Bi-LSTM [12], GRU [13], Attention-CNN + Bi-LSTM [8], Quantum Neural Network (QNN) [9], Type 1 Fuzzy Model with triangular membership function[14] and Type 1 Fuzzy Model with trapezoidal membership function[15]. Preprocessing steps are carried out initially, and tokens are generated and presented to the ensemble Bi-LSTM-GRU model. The BERT embedding layer works on pre-processed tweets and also sends the input words to the deep learning model. The ensemble of the bidirectional LSTM and gated recurrent unit activates in a way so that no word features are lost and are captured and processed for opinion mining.

The word tokens and the word feature vector generated by the BERT embedding layer carry the most suitable tweet word features based on the tokens and pass it to the ensemble Bi-LSTM-GRU model. The existence of both Bi-LSTM and GRU sequencing results in capturing the past information and operating on the present information, with the unnecessary data being forgotten and passing the relevant word feature vector only. Tables 6–8 presents the evaluation results of the proposed model and existing baseline models on short Twitter datasets.

Implementation of the developed eBi-LSTM-GRU-FDSS model was done on the tweet datasets: Sentiment 140, T4SA and Airline Twitter with short tweets and the results, as tabulated in Tables 6–8. The modelled sentiment classifier has attained a classification accuracy improvement of 5.3%, 8.7% and 7.9% as in Tables 6–8 for Sentiment 140, T4SA and Airline Twitter datasets, respectively. Pertaining to the F1 score, the eBi-LSTM-GRU-FDSS model has achieved an improvement of 6.03%, 3.87% and 5.06% for positive classes and 3.97%, 2.81% and 3.03% for negative classes of the three datasets, respectively. Considering the neutral classes, 2.63% and 3.27% improvement were attained for T4SA and Airline datasets, respectively. The Kappa Coefficient confirms to be 4.91%, 6.59% and 5.72% better than all other models of the short tweet three datasets, respectively. The results substantiate the efficacy of the developed eBiLSTM-GRU-FDSS over the compared models, CNN [10], LSTM [11], Bi-LSTM [12], GRU [13], Attention-CNN + Bi-LSTM [8], Quantum Neural Network (QNN) [9], Type 1 Fuzzy Model with triangular membership function [14] and Type 1 Fuzzy Model with trapezoidal membership function [15] with respect to classification accuracy, F1 score and Kappa coefficient. As tweet datasets consist of a small number of words and henceforth less amount of improvement gets noticed, but a significant classification structure can be noticed for this sentiment analysis model.

Positive reviews have evolved better classification accuracy than the negative and neutral reviews, as given in Tables 6–8. This is because each and every tweet possesses 12 words on average, which is more than one tweet sentence. In respect of this, there are very short sentences with other sentiment polarity getting changed and between negations, there are no longer gaps. The kappa coefficient measures the inter-rater reliability for the class instances, and the kappa coefficient value varies from 0 to 1. For all three datasets, the Kappa value was 0.7411, 0.7301 and 0.7496 for the Twitter datasets Sentiment 140, T4SA and Airline Twitter datasets, respectively.

The Kappa coefficient value K is to be greater than 0.6, and for all the datasets pertaining to the proposed model, it proves its effectiveness for the corresponding class instances and concurrence with that of the predicted and actual classes. In the developed sentiment and opinion classifier, the Kappa coefficient value is

Table 6. Evaluated results – sentiment 140 dataset.

Deep learning models	Class	Recall	Precision	F1 score	Accuracy (%)	CK score
CNN [10]	Positive	0.8006	0.7546	0.7987	72.59	0.6741
	Negative	0.7564	0.8249	0.7247		
LSTM [11]	Positive	0.7995	0.7142	0.7564	74.66	0.6859
	Negative	0.7619	0.8176	0.7361		
Bi-LSTM [12]	Positive	0.8016	0.7019	0.7844	75.03	0.7014
	Negative	0.7846	0.8343	0.7433		
GRU [13]	Positive	0.7948	0.7348	0.7936	70.54	0.6988
	Negative	0.7457	0.8429	0.7394		
Attention-CNN + Bi-LSTM [8]	Positive	0.8049	0.7416	0.7831	78.49	0.7199
	Negative	0.7702	0.8466	0.7522		
QNN [9]	Positive	0.8154	0.7633	0.8147	78.61	0.7263
	Negative	0.7567	0.8346	0.7948		
Type 1 Fuzzy Model with triangular membership function [14]	Positive	0.8061	0.7509	0.7922	77.41	0.7199
	Negative	0.7499	0.8147	0.7741		
Type 1 Fuzzy Model with trapezoidal membership function [15]	Positive	0.7946	0.7461	0.7864	76.03	0.7051
	Negative	0.7166	0.8006	0.7549		
Modelled neweBi-LSTM-GRU-FDSS	**Positive**	**0.9065**	**0.7805**	**0.8491**	**84.33**	**0.7411**
	Negative	**0.8761**	**0.8864**	**0.8179**		

Table 7. Evaluated results – T4SA dataset.

Deep learning models	Class	Recall	Precision	F1 score	Accuracy (%)	CK score
CNN [10]	Positive	0.9306	0.9425	0.9396	92.78	0.7146
	Neutral	0.9189	0.9361	0.9316		
	Negative	0.9506	0.9374	0.9421		
LSTM [11]	Positive	0.9347	0.9436	0.9407	92.55	0.7265
	Neutral	0.9203	0.9375	0.9327		
	Negative	0.9517	0.9387	0.9405		
Bi-LSTM [12]	Positive	0.9315	0.9465	0.9419	93.65	0.7122
	Neutral	0.9269	0.9387	0.9355		
	Negative	0.9598	0.9403	0.9430		
GRU [13]	Positive	0.9265	0.9411	0.9365	92.61	0.6922
	Neutral	0.9144	0.9362	0.9300		
	Negative	0.9548	0.9369	0.9324		
Attention-CNN + Bi-LSTM [8]	Positive	0.9351	0.9584	0.9475	95.46	0.7094
	Neutral	0.9302	0.9422	0.9389		
	Negative	0.9611	0.9405	0.9461		
QNN [9]	Positive	0.9386	0.9658	0.9518	94.22	0.7154
	Neutral	0.9321	0.9541	0.9477		
	Negative	0.9647	0.9467	0.9529		
Type 1 Fuzzy Model with triangular membership function [14]	Positive	0.9752	0.9425	0.9606	95.57	0.7228
	Neutral	0.9694	0.9375	0.9532		
	Negative	0.9612	0.9582	0.9597		
Type 1 Fuzzy Model with trapezoidal membership function [15]	Positive	0.9800	0.9448	0.9624	96.25	0.7265
	Neutral	0.9720	0.9398	0.9556		
	Negative	0.9589	0.9615	0.9602		
Modelled neweBi-LSTM-GRU-FDSS	**Positive**	**0.9465**	**0.9847**	**0.9694**	**96.92**	**0.7301**
	Neutral	**0.9557**	**0.9684**	**0.9522**		
	Negative	**0.9847**	**0.9503**	**0.9681**		

inferred to be less than the classification accuracy rate, and this is due to the unbalanced nature of the dataset.

The area under the curve (AUC) evaluated for the developed sentiment classifier model and state-of-the-art models for the three tweet datasets is shown in Figure 12. AUC computed using the new ensemble Bi-LSTM-GRU with fuzzy decision support system model has resulted in 0.9006, 0.9195 and 0.9459 for the Sentiment 140, T4SA and Airline Twitter datasets, respectively. This confirms the validity of the modelled new sentiment classifier based on the ensemble design of Bi-LSTM and GRU models and the fuzzy decision support system model. Figure 12 confirms that the sentiment classification performed with the new ensemble + FDSS model has resulted in better AUC value proving its superiority over other compared CNN [10], LSTM [11], Bi-LSTM [12], GRU [13], Attention-CNN + Bi-LSTM [8], Quantum Neural Network (QNN) [9], Type 1 Fuzzy Model with triangular membership function [14] and Type 1 Fuzzy Model with trapezoidal membership function [15]. The convergence plot using the proposed eBi-LSTM-GRU + FDSS model as applied for the three Twitter datasets is shown in Figure 13. The cost function is the mean square error evaluated for attaining the convergence of the developed sentiment classifier model, and the minimized MSE confirms the attainment of better classification accuracy for the evolved number of iterations.

In comparison with the baseline models, the developed ensemble Bi-LSTM-GRU with the FDSS model for the metrics classification accuracy and Cohen-Kappa Score is shown in Figure 14 for all three tweet datasets. From the tweets, the word features are extracted, and the tokens are generated to perform the training process of the ensemble deep learning

Table 8. Evaluated results – airline Twitter dataset.

Deep learning models	Class	Recall	Precision	F1 score	Accuracy (%)	CK score
CNN [10]	Positive	0.9296	0.9151	0.9218	91.21	0.7028
	Neutral	0.8371	0.8315	0.8275		
	Negative	0.7591	0.7945	0.7844		
LSTM [11]	Positive	0.9318	0.9287	0.9365	90.52	0.6854
	Neutral	0.8354	0.8467	0.8312		
	Negative	0.7542	0.8039	0.7908		
Bi-LSTM [12]	Positive	0.9361	0.9306	0.9411	90.87	0.7349
	Neutral	0.8409	0.8482	0.8365		
	Negative	0.7709	0.8118	0.7964		
GRU [13]	Positive	0.9268	0.9244	0.9322	89.64	0.7255
	Neutral	0.8317	0.8361	0.8497		
	Negative	0.7645	0.8049	0.7861		
Attention-CNN + Bi-LSTM [8]	Positive	0.9405	0.9361	0.9546	92.83	0.7164
	Neutral	0.8506	0.8520	0.8563		
	Negative	0.7961	0.8329	0.8149		
QNN [9]	Positive	0.9416	0.9498	0.9511	91.67	0.7384
	Neutral	0.8523	0.8493	0.8517		
	Negative	0.8107	0.8217	0.8047		
Type 1 Fuzzy Model with triangular membership function [14]	Positive	0.9573	0.9470	0.9566	92.74	0.7440
	Neutral	0.9070	0.9024	0.9052		
	Negative	0.8715	0.8606	0.8556		
Type 1 Fuzzy Model with trapezoidal membership function [15]	Positive	0.9610	0.9497	0.9591	93.28	0.7468
	Neutral	0.9332	0.9273	0.9314		
	Negative	0.8954	0.8856	0.8792		
Modelled neweBi-LSTM-GRU-FDSS	**Positive**	**0.9524**	**0.9647**	**0.9622**	**93.81**	**0.7496**
	Neutral	**0.8641**	**0.8653**	**0.8609**		
	Negative	**0.8195**	**0.8416**	**0.8246**		

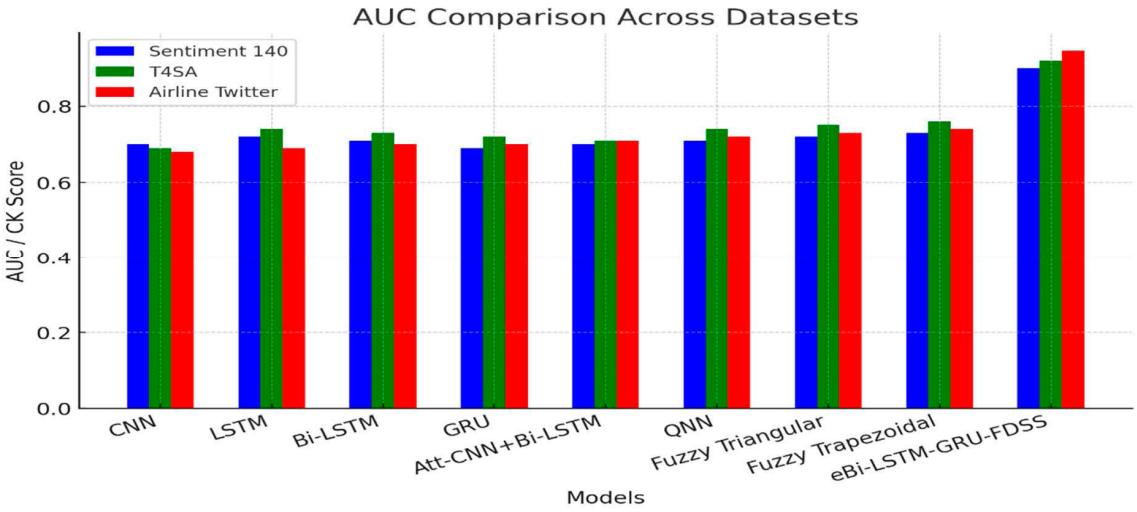


Figure 12. AUC comparison plots for the three Tweet datasets.

recurrent model. Reflecting on the opinions, then the customer contentment level is classified as positive or negative class, and hence the classification accuracy and CK score are attained.

The developed ensemble model got an accuracy of 93.81%, 96.92% and 84.33% for airline Twitter, T4SA and sentiment 140 datasets, respectively. For the sentiment 140 dataset, with the QNN model having an accuracy of 78.61% the accuracy has been increased to 5.72%. The Cohen-Kappa Score using the modelled sentiment classifier is 0.7496, 0.7301 and 0.7411 for the airline Twitter, T4SA and sentiment 140 datasets, respectively.

The fuzzy decision support system model, with its formulated triangular memberships and based on the formulated rule base, evolved the customer contentment level. Ensemble Bi-LSTM-GRU model mines the positive, negative and neutral opinions and then performs the customer contentment level identification for the tweet sentiment datasets. Figure 15 presents the comparison plot of the Recall, Precision and F1 score for the three classes, positive, negative and neutral, using the baseline models with that of the proposed model for the three datasets.

With the considered sentiment 140, T4SA and airline tweet datasets, the modelled new ensemble Bi-LSTM – GRU with FDSS for determining customer contentment level has reported better evaluation indices compared with the baseline models.

Ensemble modelling resulted in learning from all the word features generated, and resulted in better opinion mining and the customer contentment level was

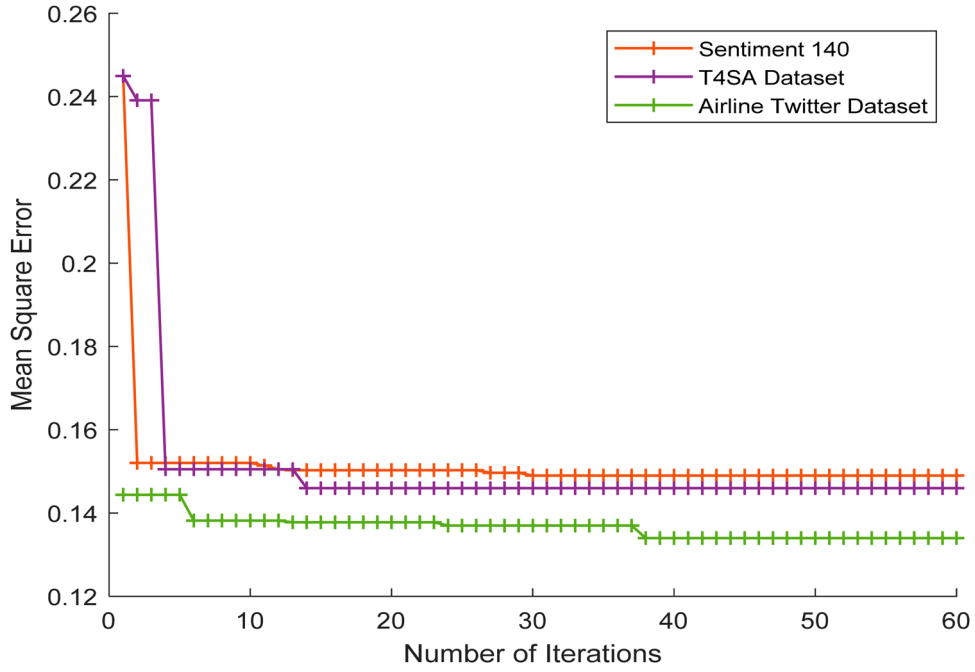


Figure 13. Convergence plot of proposed model for the three Twitter datasets.

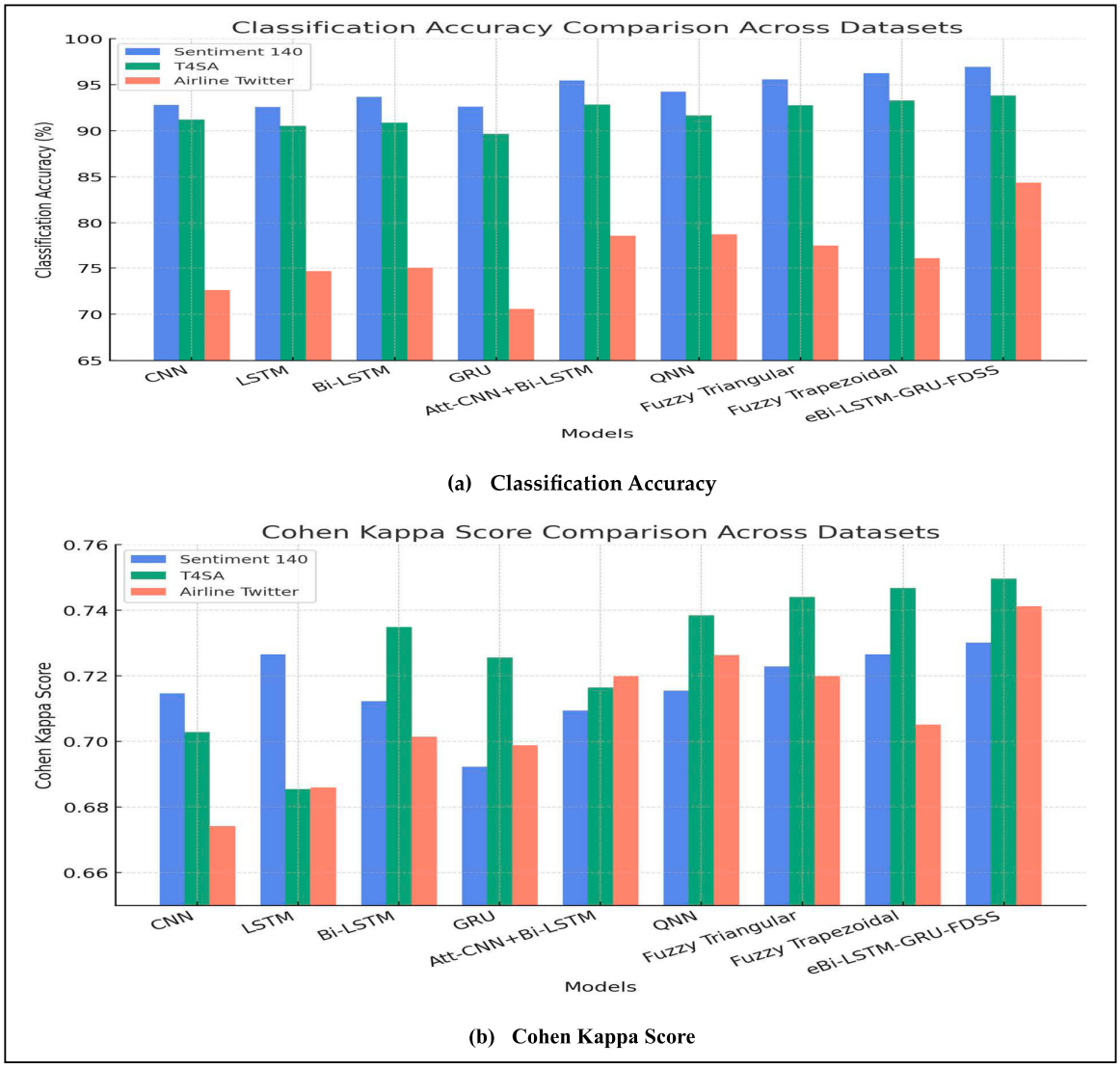
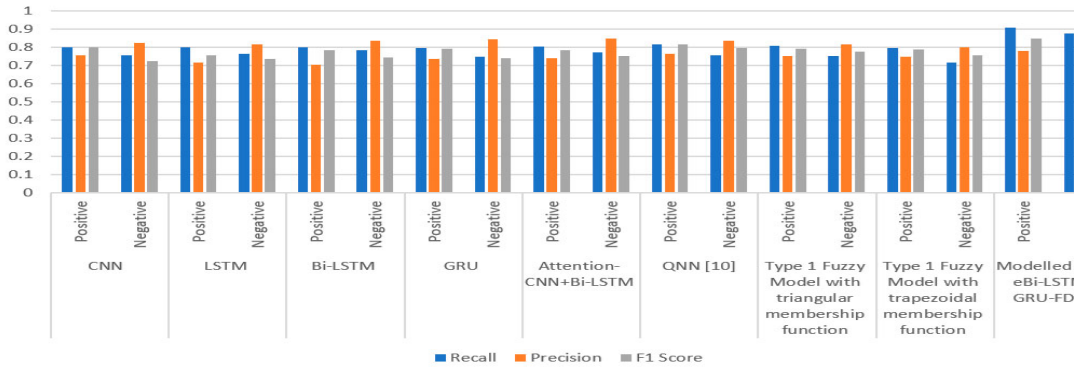
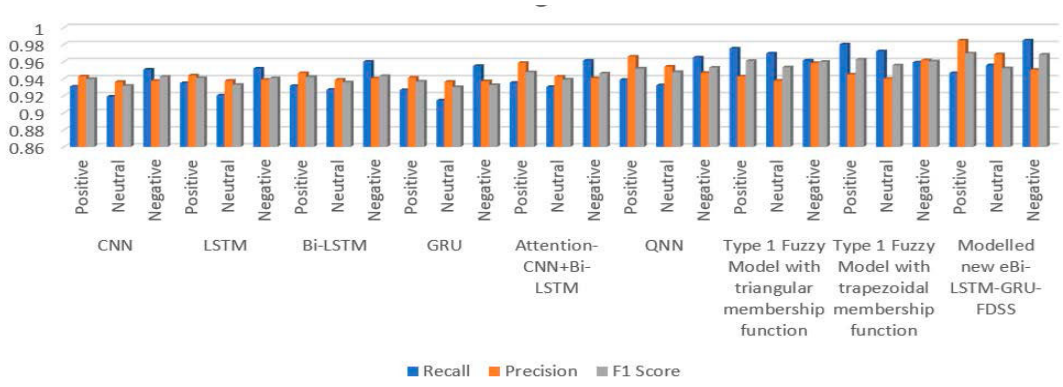


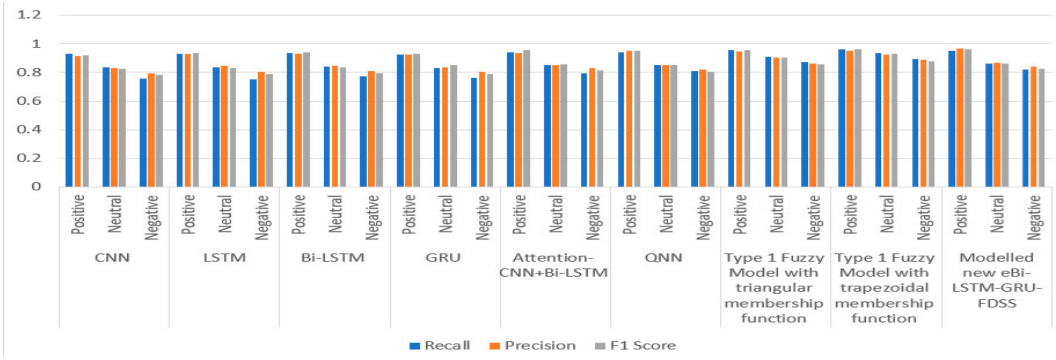
Figure 14. Comparison of baseline models for three datasets (a) classification accuracy (b) Cohen-Kappa score.



(a) Sentiment 140 dataset



(b) T4SA dataset



(c) Airline Twitter dataset

Figure 15. Comparison of baseline models for evaluation indices – recall, precision and F1 score (a) sentiment 140 dataset (b) T4SA dataset (c) airline Twitter dataset.

evaluated with the fuzzy decision model. Statistical validation of the modelled DL ensemble model is highly required for validation of the attained results of opinion mining and contentment level. In this research work, the Wilcoxon Rank Sum test [47,48] is applied for validating the proposed models. The Wilcoxon test calculates the *p*-value to statistically assess the difference between two samples. *P*-value is to be less than 0.05 for validation of the developed model, and in this proposed research, the model was executed for 36 trial runs, and the classification results were employed for running the Wilcoxon test. Table 9 gives the evaluated *P*-values. The values are less than 0.05, and this confirms that the statisticalvalidation of the developed eBi-LSTM-GRU-FDSS model.

Table 10 provides the evolved computational complexity parameters – depth of the network, number of parameters, batch size, number of epochs and computational time elapsed (seconds) of the proposed model and that of the other baseline models as applied for the three tweet datasets. With only 0.72 M parameters and a depth of 54, the number of epochs elapsed is 64, which makes the modelled eBi-LSTM-GRU-FDSS computationally simple and easy in comparison with the other

Table 9. *P*-values evaluated with wilcoxon rank sum test of baseline models.

Compared models	*P*-value
CNN [10]	0.00416
LSTM [11]	0.00367
Bi-LSTM [12]	0.00314
GRU [13]	0.00259
Attention-CNN + Bi-LSTM [8]	0.00205
QNN [9]	0.00237
Type 1 Fuzzy Model with triangular membership function [14]	0.00219
Type 1 Fuzzy Model with trapezoidal membership function [15]	0.00196
Modelled new eBi-LSTM-GRU-FDSS	0.00112

baseline models. A minimum batch size of 141 and a 12.09 s elapsed execution time of the new ensemble Bi-LSTM-GRU-FDSS model was noted. This is because the effective and optimal sentiment features were identified and fed into the developed ensemble deep learning and fuzzy decision support system model.

To evaluate the contribution of the Fuzzy Decision Support System (FDSS) to the overall performance of our proposed eBi-LSTM-GRU-FDSS model, we conducted an ablation study by removing or modifying key components and analysing their impact on accuracy, precision, recall and F1-score.

We trained and estimated different versions of our model on the same dataset using identical hyperparameters. The variations tested include the model without FDSS, which uses only the eBi-LSTM-GRU architecture; the model without Bi-LSTM, which consists of only GRU and FDSS and the model without GRU, which retains Bi-LSTM and FDSS. Finally, the full model (eBi-LSTM-GRU-FDSS) integrates both Bi-LSTM and GRU along with FDSS to leverage their combined strengths. The findings are demonstrated in Tables 11–13 for the three datasets.

The ablation study evaluates the contribution of different components in the eBi-LSTM-GRU-FDSS model by systematically removing key elements (FDSS, Bi-LSTM and GRU) and analysing the performance impact across multiple datasets. The full model consistently achieves the maximum accuracy and F1 score, indicating the effectiveness of integrating Bi-LSTM, GRU and FDSS.

Removing FDSS leads to a notable drop in performance, particularly in CK Score and F1 Score, highlighting its role in refining sentiment classification. The absence of GRU results in a further decline, showing its contribution to capturing long-range dependencies. The Bi-LSTM exclusion has the most significant negative impact, demonstrating its essential role in bidirectional context learning.

Across datasets (Sentiment140, T4SA and Airline Twitter), the trend remains consistent – each component contributes to improving the classification accuracy, recall and precision, reinforcing the necessity of all three components in achieving optimal sentiment analysis performance.

The proposed fuzzy rule-based approach demonstrates notable effectiveness in capturing sarcasm, irony, and ambiguity by leveraging linguistic and contextual cues. These rules enhance sentiment classification by allowing for nuanced interpretation of complex textual expressions.

- *Sarcasm detection*: The model effectively identifies sarcasm by analysing sentiment polarity shifts and exaggerated expressions. It performs well in explicit cases where positive sentiment masks negative intent. While implicit sarcasm dependent on external knowledge or multimodal cues remains challenging, the fuzzy rules offer a structured and interpretable approach to handling sarcastic content.
- *Irony detection*: The fuzzy rules successfully recognize irony by detecting contradictions between

Table 10. Computational complexity analysis of the developed sentiment classifier and existing models.

Techniques	Depth	# No. of parameters (CC)	Computational time (s)	Batch size	Epochs
CNN [10]	161	14.8 M	44.91	198	127
LSTM [11]	154	14.2 M	46.81	209	136
Bi-LSTM [12]	165	15.4 M	54.33	232	148
GRU [13]	124	10.2 M	41.25	218	159
Attention-CNN + Bi-LSTM [8]	106	8.3 M	37.18	221	164
QNN [9]	118	5.5 M	26.49	192	123
Type 1 Fuzzy Model with triangular membership function [14]	102	6.1 M	29.34	198	128
Type 1 Fuzzy Model with trapezoidal membership function [15]	108	6.5 M	31.12	202	132
Modelled new eBi-LSTM-GRU-FDSS	54	0.72 M	12.09	141	64

Table 11. Ablation study results for eBi-LSTM-GRU-FDSS model – sentiment 140 dataset.

Model variant	Class	Recall	Precision	F1 score	Accuracy (%)	CK score
Full eBi-LSTM-GRU-FDSS	Positive	**0.9065**	**0.7805**	**0.8491**	**84.33**	**0.7411**
	Negative	**0.8761**	**0.8864**	**0.8179**		
Without FDSS	Positive	0.8921	0.7450	0.8324	82.45	0.7195
	Negative	0.8620	0.8613	0.8021		
Without GRU	Positive	0.8713	0.7658	0.8287	81.92	0.7114
	Negative	0.8417	0.8725	0.7932		
Without Bi-LSTM	Positive	0.8452	0.7301	0.8105	80.31	0.6958
	Negative	0.8154	0.8432	0.7689		

Table 12. Ablation study results for eBi-LSTM-GRU-FDSS model – T4SA dataset.

Model variant	Class	Recall	Precision	F1 score	Accuracy (%)	CK score
Full eBi-LSTM-GRU-FDSS	**Positive**	**0.9465**	**0.9847**	**0.9694**	**96.92**	**0.7301**
	Neutral	**0.9557**	**0.9684**	**0.9522**		
	Negative	**0.9847**	**0.9503**	**0.9681**		
Without FDSS	Positive	0.9283	0.9551	0.9447	94.87	0.7108
	Neutral	0.9385	0.9432	0.9408		
	Negative	0.9702	0.9321	0.9510		
Without GRU	Positive	0.9201	0.9497	0.9345	94.02	0.7012
	Neutral	0.9287	0.9342	0.9314		
	Negative	0.9628	0.9225	0.9422		
Without Bi-LSTM	Positive	0.9024	0.9320	0.9163	92.65	0.6854
	Neutral	0.9105	0.9214	0.9158		
	Negative	0.9483	0.9102	0.9286		

Table 13. Ablation study results for eBi-LSTM-GRU-FDSS model – airline Twitter dataset.

Model variant	Class	Recall	Precision	F1 score	Accuracy (%)	CK Score
Full eBi-LSTM-GRU-FDSS	**Positive**	**0.9524**	**0.9647**	**0.9622**	**93.81**	**0.7496**
	Neutral	**0.8641**	**0.8653**	**0.8609**		
	Negative	**0.8195**	**0.8416**	**0.8246**		
Without FDSS	Positive	0.9342	0.9510	0.9425	91.45	0.7293
	Neutral	0.8505	0.8531	0.8457		
	Negative	0.8032	0.8221	0.8123		
Without GRU	Positive	0.9221	0.9420	0.9318	90.62	0.7198
	Neutral	0.8327	0.8413	0.8369		
	Negative	0.7890	0.8107	0.7995		
Without Bi-LSTM	Positive	0.9040	0.9278	0.9157	89.27	0.7041
	Neutral	0.8101	0.8250	0.8175		
	Negative	0.7704	0.7982	0.7840		

expected and actual meanings, particularly in cases where sentiment polarity is intentionally reversed. The approach works well for clear instances of irony but can be further refined to handle subtle, context-dependent irony in longer, complex statements.

- *Ambiguity handling*: The incorporation of fuzzy logic enhances robustness in sentiment classification by allowing degrees of uncertainty, reducing misclassification in ambiguous statements. While effective in many general cases, domain-specific ambiguities and context-heavy interpretations could benefit from additional refinements.

Overall, the fuzzy rule-based approach provides an interpretable and adaptable solution for sentiment analysis, particularly in handling nuanced expressions. Future enhancements, such as integrating contextual embeddings, leveraging sarcasm-specific datasets, or incorporating multimodal sentiment cues, could further strengthen its ability to process complex linguistic phenomena.

5. Conclusion

A modelled new ensemble bidirectional long-short term memory – gated recurrent unit hybridized with the fuzzy decision support system approach has been developed in this research study to act as a sentiment classifier and customer contentment level identifier. The proposed ensemble Bi-LSTM-GRU-FDSS model has been trained and tested for the three Twitter datasets – Sentiment 140, T4SA and Airline Twitter datasets. The applicability of the developed ensemble deep learning model with the fuzzy system has been carried out for sentiment analysis and classification, and also the identification of the customer contentment level. The word features get extracted, and the embedded tokens from the BERT embedding layer after pre-processing are the inputs to the ensemble Bi-LSTM and GRU model. The combination of both Bi-LSTM and GRU acts so that no word features are lost from the tweet sentences and stores all the past features in order to perform an effective sentiment classification. Once the positive, neutral and negative sentiment classification is done with the eBi-LSTM-GRU model, then the fuzzy decision support system model activates to attain the customer contentment level (positive or negative). For the classification and contentment level performed, the evaluation indices are evaluated to prove the efficacy of the proposed model. The modelled new ensemble Bi-LSTM-GRU-FDSS model has been trained, tested and validated on the three Twitter datasets. The proposed model has proved its superiority and effective performance over the other baseline models compared. Better results have been attained for the T4SA and Airline Twitter dataset in comparison with the Sentiment dataset. Results substantiate that the deep learning ensemble models, which are used in numerous image processing applications, are also suited for sentiment analysis and classification. In future research, transformer-based models such as BERT and GPT can be explored to enhance the proposed hybrid framework by incorporating self-attention mechanisms for improved contextual understanding. Integrating these models with fuzzy logic could further refine sentiment classification by combining

deep contextual representations with techniques for managing uncertainty.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] O'reilly T, Milstein S. The Twitter book. Sebastopol, CA: O'Reilly Media, Inc.; 2011.

[2] Wu S, Hofman JM, Mason WA, et al. Who says what to whom on twitter. In: Proceedings of the 20th International Conference on World Wide Web; 2011. p. 705–714. doi:10.1145/1963405.1963504

[3] Weller K, Bruns A, Burgess J, et al. Twitter and society: an introduction. Digit Form. 2014;89:xxix–xxxviii. https://katrinweller.net/wp-content/uploads/2012/08/twitter-and-society-introduction-2014.pdf.

[4] Burgess J, Baym NK. Twitter: a biography. New York: NYU Press; 2022.

[5] Nakagawa K, Yang NT, Wilson M, et al. Twitter usage among physicians from 2016 to 2020: algorithm development and longitudinal analysis study. J Med Internet Res. 2022;24(9):e37752. doi:10.2196/37752

[6] X/Twitter: number of users worldwide 2024 | Statista.

[7] Zimbra D, Abbasi A, Zeng D, et al. The state-of-the-art in twitter sentiment analysis: a review and benchmark evaluation. ACM Trans Manage Inf Syst. 2018;9(2):1–29. doi:10.1145/3185045

[8] Kamyab M, Liu G, Adjeisah M. Attention-based CNN and Bi-LSTM model based on TF-IDF and GloVe word embedding for sentiment analysis. Appl Sci. 2021;11(23):11255. doi:10.3390/app112311255

[9] Jia K, Meng F, Liang J. Hierarchical graph contrastive learning framework based on quantum neural networks for sentiment analysis. Inf Sci (Ny). 2024;690(2025):121543. doi:10.1016/j.ins.2024.121543

[10] Khare S. Accuracy enhancement during sentiment analysis in twitter using CNN. In: 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO); Noida, India; 2022. p. 1–7. doi:10.1109/ICRITO56286.2022.9964733

[11] Nath Saha B, Senapati A. Long short-term memory (LSTM) based deep learning for sentiment analysis of English and Spanish data. In: International Conference on Computational Performance Evaluation (ComPE); Shillong, India; 2020. p. 442–446. doi:10.1109/ComPE49325.2020.9200054

[12] Xie J, Chen B, Gu X, et al. Self-attention-based BiLSTM model for short text fine-grained sentiment classification. IEEE Access. 2019;7:180558–180570. doi:10.1109/ACCESS.2019.2957510

[13] Atassi A, El Azami I, Sadiq A. The new deep learning architecture based on GRU and word2vec. In: International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS); Kenitra, Morocco; 2018. p. 1–3. doi:10.1109/ICECOCS.2018.8610611

[14] Azam MH, Hasan MH, Hassan S, et al. A novel approach to generate type-1 fuzzy triangular and trapezoidal membership functions to improve the classification accuracy. Symmetry (Basel). 2021;13(10):1932. doi:10.3390/sym13101932

[15] Jefferson C, Liu H, Haig E. Fuzzy approach for sentiment analysis. In: Proceedings of the IEEE International Conference on Fuzzy Systems; Naples, Italy; July 2017. doi:10.1109/FUZZ-IEEE.2017.8015577

[16] Priyadarshini I, Cotton C. A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis. J Supercomput. 2021;77(12):13911–13932. doi:10.1007/s11227-021-03838-w

[17] Sharaf Al-deen HS, Zeng Z, Al-sabri R, et al. An improved model for analyzing textual sentiment based on a deep neural network using multi-head attention mechanism. Appl Syst Innov. 2021;4(4):85. doi:10.3390/asi4040085

[18] Kumar V. Spatiotemporal sentiment variation analysis of geotagged COVID-19 tweets from India using a hybrid deep learning model. Sci Rep. 2022;12(1):1849. doi:10.1038/s41598-022-05974-6

[19] Khan L, Amjad A, Ashraf N, et al. Multi-class sentiment analysis of urdu text using multilingual BERT. Sci Rep. 2022;12(1):5436. doi:10.1038/s41598-022-09381-9

[20] Soni J, Mathur K. Sentiment analysis based on aspect and context fusion using attention encoder with LSTM. Int J Inf Technol . 2022;14(7):3611–3618. doi:10.1007/s41870-022-00966-1

[21] Yan H, Yi B, Li H, et al. Sentiment knowledge-induced neural network for aspect-level sentiment analysis. Neural Comput Appl. 2022;34(24):22275–22286. doi:10.1007/s00521-022-07698-0

[22] Sun Y, Sun Q, Zhu S. Prediction of Shanghai stock index based on investor sentiment and CNN-LSTM model. J Syst Sci Inf. 2022;10:620–632. doi:10.21078/JSSI-2022-620-13

[23] Hammi S, Hammami SM, Belguith LH. Advancing aspect-based sentiment analysis with a novel architecture combining deep learning models CNN and bi-RNN with the machine learning model SVM. Soc Netw Anal Min. 2023;13(1):117. doi:10.1007/s13278-023-01126-4

[24] Jia K, Meng F, Liang J, et al. Text sentiment analysis based on BERT-cblbga. Comput Electr Eng. 2023;112:109019. doi:10.1016/j.compeleceng.2023.109019

[25] Ranjan R, Daniel AK. Cobiat: A sentiment Classificati-Cobiat: a sentiment classification model using hybrid convnet- dual-lstm with attention mechanismon model using hybrid ConvNet- dual-LSTM with attention mechanism. Informatica. 2023;47(4):523–536. doi:10.31449/inf.v47i4.3911

[26] Alzaid M, Fkih F. Sentiment analysis of students' feedback on E-learning using a hybrid fuzzy model. Appl Sci. 2023;13(23):12956. doi:10.3390/app132312956

[27] Galal O, Abdel-Gawad AH, Farouk M. Rethinking of BERT sentence embedding for text classification. Neural Comput Appl. 2024;36(32):20245–20258. doi:10.1007/s00521-024-10212-3

[28] Fatima E, Kanwal H, Khan JA, et al. An exploratory and automated study of sarcasm detection and classification in app stores using fine-tuned deep learning classifiers. Autom Softw Eng. 2024;31(2):69. doi:10.1007/s10515-024-00468-3

[29] Dervenis C, Kanakis G, Fitsilis P. Sentiment analysis of student feedback: a comparative study employing lexicon and machine learning techniques. Stud Educ Eval. 2024;83:101406. doi:10.1016/j.stueduc.2024.101406

[30] Chandrasekaran G, Chowdary MK, Babu JC, et al. Deep learning-based attention models for sarcasm detection in text. Int J Electr Comput Eng. 2024;14(6):2088–8708. doi:10.11591/ijece.v14i6.pp6786-6796

[31] Ambreen S, Iqbal M, Asghar MZ, et al. Predicting customer sentiment: the fusion of deep learning and a fuzzy system for sentiment analysis of Arabic text. Soc Netw Anal Min. 2024;14(1):206. doi:10.1007/s13278-024-01356-0
[32] Motevalli MM, Sohrabi MK, Yaghmaee F. Aspect-based sentiment analysis: a dual-task learning architecture using imbalanced maximized-area under the curve proximate support vector machine and reinforcement learning. Inf Sci (Ny). 2025;689:121449. doi:10.1016/j.ins.2024.121449
[33] Fadhel Z, Attia H, Ali YH. Optimized and comprehensive fake review detection based on Harris Hawk's optimization integrated with machine learning techniques. J Cybersec Inf Manage. 2025;15(1):11–21. doi:10.54216/JCIM.150102
[34] Darraz N, Karabila I, El-Ansari A, et al. Integrated sentiment analysis with BERT for enhanced hybrid recommendation systems. Expert Syst Appl. 2024;261(4):125533. doi:10.1016/j.eswa.2024.125533
[35] Ghosh S, Ekbal A, Bhattacharyya P. Natural language processing and sentiment analysis: perspectives from computational intelligence. Computational intelligence applications for text and sentiment data analysis. Cambridge: Academic Press; 2023. p. 17–47. doi:10.1016/B978-0-32-390535-0.00007-0
[36] Sherin A, Jasmine Selvakumari Jeya I, Deepa SN. Enhanced aquila optimizer combined ensemble Bi-LSTM-GRU with fuzzy emotion extractor for tweet sentiment analysis and classification. IEEE Access. 2024;12:141932–141951. doi:10.1109/ACCESS.2024.3464091
[37] Ross TJ. Fuzzy logic with engineering applications. Chichester: John Wiley & Sons; 2005.
[38] Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. CS224N project report, Stanford. 2009;1(12):2009.
[39] Vadicamo L, Carrara F, Cimino A, et al. Cross-media learning for image sentiment analysis in the wild. In Proceedings of the IEEE International Conference on Computer Vision Workshops; 2017. p. 308–317.
[40] Airline Twitter sentiment dataset; 2019. https://data.world/crowdflower/airline-twitter-sentiment.
[41] Gao W, Sebastiani F. From classification to quantification in tweet sentiment analysis. Soc Netw Anal Min. 2016;6:1–22. doi:10.1007/s13278-016-0327-z
[42] Salur MU, Aydin I. A novel hybrid deep learning model for sentiment classification. IEEE Access. 2020;8:58080–58093. doi:10.1109/ACCESS.2020.2982538
[43] Fu X, Yang J, Li J, et al. Lexicon-Enhanced LSTM with attention for general sentiment analysis. IEEE Access. 2018;6:71884–71891. doi:10.1109/ACCESS.2018.2878425
[44] Khan J, Ahmad N, Khalid S, et al. Sentiment and context-aware hybrid DNN with attention for text sentiment classification. IEEE Access. 2023;11:28162–28179. doi:10.1109/ACCESS.2023.3259107
[45] Aydin CR, Güngör T. Combination of recursive and recurrent neural networks for aspect-based sentiment analysis using inter-aspect relations. IEEE Access. 2020;8:77820–77832. doi:10.1109/ACCESS.2020.2990306
[46] Basiri ME, Nemati S, Abdar M, et al. ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis. Future Gener Comput Syst. 2021;115:279–294. doi:10.1016/j.future.2020.08.005
[47] Wilcoxon F, Katti S, Wilcox RA. Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. In: Harter, Owen, editors. Selected tables in mathematical statistics. Chicago: Makham Publishing Co; 1970; 1. p. 171–259.
[48] Rosner B, Glynn RJ, Ting Lee ML. Incorporation of clustering effects for the Wilcoxon rank sum test: a large-sample approach. Biometrics. 2003;59(4):1089–1098. doi:10.1111/j.0006-341X.2003.00125.x