Indonesian Journal of Electrical Engineering and Computer Science
Vol. 44, No. 1, October 2026, pp. 250~262
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v44.i1.pp250-262

Sentiment analysis in online course reviews using fusion-based sentence-BERT and BiLSTM

Priyanka Shukla, Jitendra Nath Singh, Shachi Mall
School of Computer Science and Engineering, Galgotias University, Greater Noida, India

Article Info

Article history:

Received Aug 27, 2025
Revised Jul 15, 2026
Accepted Sep 18, 2026

Keywords:

Aspect-based sentiment analysis
Bilstm
Classification
Online learning
SentenceBERT

ABSTRACT

The proliferation of online education has fundamentally transformed the accessibility of learning opportunities; however, the efficacy of digital educational platforms is frequently assessed based on the subjective evaluations provided by learners. These reviews typically contain diverse sentiments tied to specific aspects such as instructor quality, content delivery, and assessment methods. Conventional sentiment analysis models struggle to capture this aspect-level nuance, leading to generalized insights that lack actionable value. This research proposes an enhanced deep learning framework for aspect-based sentiment analysis (ABSA), designed to uncover fine-grained opinions embedded in student reviews. The core innovation lies in the fusion of contextual sentence-BERT embeddings with linguistically informed features, namely syntactic structure, part-of-speech patterns, and sentiment polarity, thereby enriching semantic representation. The combined feature space is then processed by a bidirectional long short-term memory (BiLSTM) classifier tailored for five-class sentiment classification: very positive, positive, neutral, negative, and very negative. Aspect identification is performed through a keyword-based mapping approach targeting key educational facets: content, instructor, platform, delivery, and assessment. Experimental results establish that the proposed fusion-based BiLSTM classifier significantly outperforms baseline algorithms and achieves 0.99 accuracy for five classes.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Priyanka Shukla
School of Computer Science and Engineering, Galgotias University
203201 Greater Noida, UP, India
Email: engpriyanka.8@gmail.com

1. INTRODUCTION

The educational sector is undergoing a transformative shift driven by the increasing incorporation of digital knowledges, particularly online learning platforms. Massive open online courses (MOOCs), virtual classrooms, and blended learning systems are revolutionizing how knowledge is disseminated and consumed [1]. Accessing and attending online classes has become much easier as technology advances, making online education a vital part of modern teaching methods, especially with global interruptions like the COVID-19 pandemic [2]-[9]. As more students adapt to online schooling, there has been a massive increase in the amount of digital course feedback, posts, ratings, and forum interactions [3]. These reviews are rich in data that can provide valuable insights into learners' experiences, preferences, and perceptions. Unfortunately, those methods that focus on a single, aggregate conclusion overlook the details embedded in textbooks and the numerous comments provided by students on different parts of their experiences. Take, for example, a student who finds a course entertaining and engaging, but critiques the way the teacher communicates or the

technical usability of the platform [4], [5]. In such situations, the ability to analyze specific dimensions of a complex situation becomes critical for generating helpful information and enhancing different parts of online education. The task falls under the domain of natural language processing (NLP) and is known as Sentiment Analysis, which aims to interpret and categorize opinions in textual formats. Its classical approach focuses on the general sentiment of a given text.

However, this coarse-grained approach is insufficient when applied to domains like education, where users frequently express diverse sentiments toward different elements within the same review. This has led to the emergence of aspect-based sentiment analysis (ABSA), which purposes to recognize the target aspects within a review and determine the sentiment associated with each feature [6].

Natural language processing (NLP) has recently seen high-level advancements providing new applications for sentiment analysis across many sectors. Xu *et al.* [7] presented a BiLSTM-based model that fully leveraged contextual relationships displayed in comment texts to classify sentiment. Also, using transformers and deep learning-based models for the social media context was a study by [8]. They showed that by using the attention mechanism, even the most subtle distinctions between their opinions could be classified. Arbane *et al.* [9] considered health-related commentary and applied a BiLSTM-based model for sentiment classification for comments on data arising from COVID-19 commentary, and was a viable architecture for a real-world, affectively laden dataset [10]. Take a hybrid solution that is based on deep learning and combines many models into a comprehensive model that returns more accurate supervised sentiment prediction. Salur *et al.* [11] and Rasappan *et al.* [12] took hybridization a step further to integrate feature selection and term weighting into model building for sentiment analysis using e-commerce data. Together, these studies collectively provided compelling evidence suggesting that advanced NLP techniques (like BiLSTM and transformer-based models) can be useful for student feedback, along with many other examples of user-generated feedback, from the nature of their existing interaction with a learning environment, or areas where learner sentiment holds potential for informative coaching to improve educational exploration.

The motivation for this study arises from the need to bridge this gap by deploying more context-aware models that can distinguish between sentiments related to specific educational features such as course content, instructor quality, assessment design, peer interaction, and platform experience. Online learning has become a cornerstone of modern education systems, offering diverse learners the opportunity to access quality education regardless of geographical or socioeconomic constraints. It supports personalized learning paths, accommodates varying paces of knowledge acquisition, and encourages the use of multimedia and interactive tools to enhance engagement. Institutions worldwide have integrated learning management systems (LMS) and platforms like Coursera, edX, Udemy, and SWAYAM to deliver structured and scalable education. These systems generate vast amounts of data, including log files, quiz results, engagement metrics, and especially learner-generated textual reviews [6]. This unstructured feedback is a goldmine for understanding user satisfaction and course effectiveness if mined and interpreted correctly. The significance of sentiment analysis in this context cannot be overstated. As education becomes increasingly learner-centric, institutions must prioritize student voice and experience in decision-making.

ABSA offers a mechanism to incorporate this feedback systematically and analytically, thereby closing the feedback loop between learners and educators. The contributions of this research work include:

- Perform a background study on ABSA using online reviews and their related methodologies, technologies, and frameworks, and compare them one by one.
- Design a new deep learning method for sentiment polarity classification using a fusion-based sentence-BERT(SBERT) feature extractor and a bidirectional long short-term memory (BiLSTM) Classifier.
- Multi-class sentiment classification performance evaluation of this study.

The organization of the paper is as follows: Section 2 shows the related work. Section 3 covers the proposed Methodology, which includes the algorithm framework, training, testing, and classification. Section 4 contains the experimental analysis that includes the dataset used, experimental setup, performance matrix, and experimental analysis. Section 5 covers the conclusion and future work.

2. RELATED WORK

In this section, a comprehensive examination of contemporary research concerning aspect-based sentiment analysis methodologies utilized across diverse fields, encompassing healthcare, e-commerce, digital services, and educational evaluations. Recent advances in aspect-based sentiment analysis (ABSA) have seen a paradigm shift toward deep learning and transformer-based architectures. a RoBERTa-based model that significantly enhanced sentiment prediction by leveraging contextual embeddings, which capture nuanced linguistic cues by [13]. This is consistent with the Author's emphasis on the importance of quality filtering and BERT-driven context generation in improving ABSA outcomes.

These approaches highlight the critical role of pretrained transformers in modelling semantic relationships between aspects and sentiments in complex textual data, particularly in user-generated content [14]. Employed gradient boosting trees and sentiment analysis to predict the satisfaction of online learners. Other deep learning strategies have demonstrated promising results by fusing multiple neural network components. A hybrid model, Bi-GRU-LSTM to extract aspect-based features, illustrating how combining recurrent architectures can enrich sequential understanding [15]. Similarly, an ensemble BiLSTM framework for extracting aspects from online reviews is introduced, offering robustness against linguistic variability [16]. Complementing this, Phan *et al.* [17] explored CNN over BERT-GCN for aspect-based sentiment analysis, showcasing the potential of integrating GCN with transformers to model inter-aspect dependencies effectively. Traditional machine learning methods have also been evaluated, particularly in educational contexts.

Setiawan *et al.* [18] compared various ensemble-based classifiers using random oversampling to predict student dropout, underscoring the need for class-balancing techniques in educational datasets. Khan *et al.* [19] and Atlas *et al.* [20] further demonstrated the utility of CNN-LSTM architectures for multilingual sentiment analysis across English and Roman Urdu, bridging the gap between classical and deep learning paradigms. Their work indicates that hybrid models can generalize well even across linguistically diverse corpora. Beyond model architecture, several studies focused on broader applications and feature extraction techniques. Dragoni *et al.* [21] introduced an unsupervised method for real-time aspect extraction. Aguero *et al.* [22] provided an extensive review of deep learning methods for multilingual sentiment analysis, emphasizing scalability and generalizability. These contributions collectively affirm the growing emphasis on context-aware, interpretable, and language-agnostic sentiment models for diverse applications, including online learning platforms. Table 1 represents the method, accuracy, dataset, and approach used in the previous research. Chauhan *et al.* [23] presented an upgraded RoBERTa-based ABSA model that reached advancement in terms of accuracy while compared to baseline models, especially when used with data set that were educational [24] presented a BERT-based ABSA framework with context generation and quality filtering, in both cases reaching improvement with precision and F1 scores, though the model only accepted English text-based input.

In the context of language diversity, [25] used supervised algorithms, to engage in sentiment analysis in Afaan Oromo, achieving moderate accuracy given the lack of linguistic resources and lack of sub-language particularities. Wan *et al.* [26] employed an association-constrained Latent Dirichlet Allocation (LDA) model to model the extraction of product aspects by sentiments, then outputting what would provide a proper topic-opinion pairing, missing a deeper representation of context as a limitation. In multimodal sentiment analysis, Yadav *et al.* [27] implemented a CNN-based emotion recognition model that outputs facial expression data that could reach high accuracy, while not traditional to the text model, it remained similar. Izzaddien *et al.* [28] provided extension to any transformer-based model to develop a multilabel ABSA approach with many embedding strategies to create transformer models such as RoBERTa for label detection, with mention to the computational costs of transformer models layers.

Table 1. Existing related work

Method	Precision	Recall	f1-score	Accuracy	ROC
BERT-BiLSTM [29]	93.78	93.8	93.79	-	-
BB-OAM [30]	0.906	0,920	0.914	0.865	-
CNN over BERT GCN [17]	-	-	78.76	85.25	
Bi-GRU+LSTM [15]	0.921769	-	0.924909	0.932787	-
RoBERT+AC5 [28]	0.862	-	0.674	0.89	-
Bi-GRU+LSTM [20]	0.966	0.987	0.9743	0.9879	yes
	85.43	79.51	82.36	91.28	-
Ensemble BiLSTM [16]	87.18	91.89	89.47	94.59	-
	47.25	62.35	53.76	95.85	-
	0.75	0.714	0.731	0.74	-
CNN+LSTM [19]	0.761	0.729	0.745	0.748	-
	0.85	0.84	0.844	0.841	-

3. PROPOSED STUDY

In this section, the proposed SBERT–BiLSTM model for Aspect-based sentiment analysis of online learning reviews is described in detail. The proposed methodology architecture, combining SBERT and BiLSTM for fine-grained sentiment classification, is illustrated in Figure 1. This study advocates a sentiment analysis methodology on digital course feedback, using a hybrid SBERT-BiLSTM model to classify sentiments into five categories: very negative (VN), negative (N), neutral (Neu), positive (P), or very positive (VP). The proposed methodology steps are as follows:

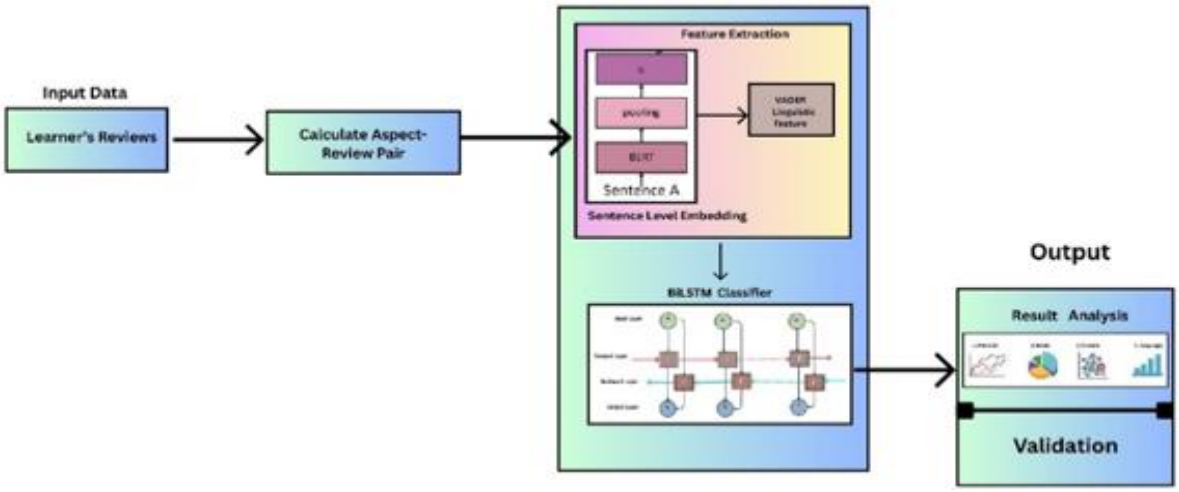

Figure 1. Proposed framework for sentiment classification

- Input: Initially, we collected a large number of online learners' reviews.
- Aspect mapping: With the help of keywords, predetermined aspects are matched with reviews.
- Feature extraction: To extract semantically rich embeddings, we use a pre-trained SBERT model that was specifically fine-tuned for classification tasks. Each review–aspect pair is concatenated into a single input string (e.g., "aspect: instructor [SEP] review: The instructor explained clearly") and passed through SBERT.

3.1. Sentence-BERT pretraining language model

Sentence-BERT (SBERT) is an adaptation of the base BERT model to produce semantically rich sentence-level embeddings. Although typical BERT has contextualized token embeddings, it does not have an effective way to compare or represent a whole sentence and takes more computational time. Sentence-BERT addresses this with a twin network architecture, where two BERT models have shared weights and handle a pair of sentences in parallel. These are then fed through a pooling layer, usually mean or max pooling, over the token embeddings to get a fixed-size vector representation for every sentence. Sentence-pair objectives are used to fine-tune the model, including natural language inference (NLI) and semantic textual similarity (STS), allowing it to recognize subtle semantic relations between sentences. After training, SBERT is capable of encoding sentences into dense low-dimensional vectors without losing their semantic information, thus being well-suited for downstream tasks for classification and ABSA. In the current work, SBERT is utilized to map every aspect–review pair to a 384-dimensional vector that is fed as input to the BiLSTM layer for further sentiment modeling (Figure 2).

The core objective of sentence-BERT is to transform input text into fixed-length, semantically rich vector representations suitable for tasks like semantic similarity and classification. Given a sentence $S = [w_1, w_2, \ldots, w_n]$, where w_i are tokens SBERT applies a transformer-based encoder BERT to generate contextual embeddings for each token calculated in (1).

$$H = [h_1, h_2, \ldots, h_n], \qquad h_i \in R^d \tag{1}$$

where h_i is the embedding of a token w_i, and d denotes the dimensionality of the transformer output (384). To distill this token-level information into a single sentence vector, SBERT applies a pooling strategy across the sequence. The most common approach is mean pooling, which computes the average over all token embeddings. To aggregate these into a single sentence embedding, SBERT applies mean pooling using (2).

$$s = \frac{1}{n}\sum_{i=1}^{n} h_i \tag{2}$$

This resulting vector $s \in R^d$ serves as a compressed semantic representation of the entire sentence, capturing both syntactic and contextual nuances. It is then used as input for downstream tasks, such as sentiment classification, thus preserving the strength of deep contextual modeling while enabling efficient comparison between sentence-level representations.

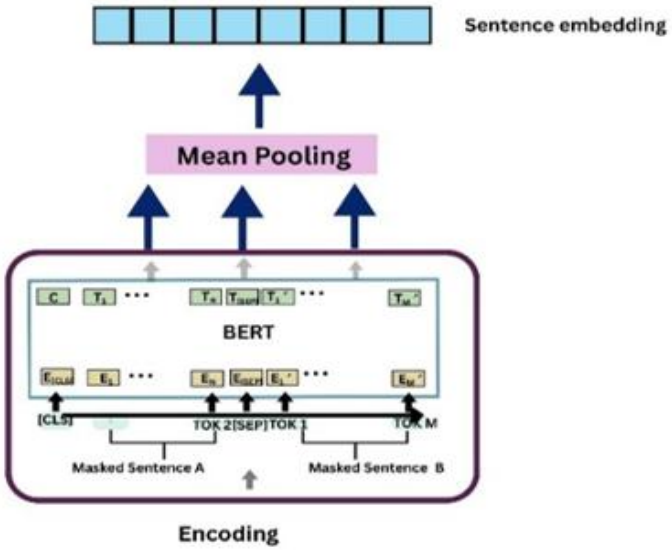

Figure 2. Sentence embedding (SBERT)

3.2. Calculate polarity and extract linguistic features

We apply Aspect and feature extraction to identify specific elements within online course reviews, such as content quality, instructor effectiveness, and course structure. This enables the model to link sentiments directly to distinct aspects, allowing for more accurate and fine-grained classification into very negative (VN), negative (N), neutral (Neu), positive (P), or very positive (VP) categories. Polarity is calculated using the valence-aware dictionary and sentiment reasoner (Algorithm 1). Figures 3 and 4 show the distribution of sentiments labelled in five classes, i.e., VN, N, Neu, P, and VP. Figure 5 represents the sentiment distribution for each aspect for the classes.

Algorithm 1. Polarity classification

```
Input: polarity ∈ [-1, 1]
Begin
    If polarity ≥ 0.6 then
        Sentimentlabel ← 4(VP)
    Else if polarity ≥ 0.2 then
        sentimentlabel ← 3 (P)
    Else if polarity > -0.2 then
        sentimentlabel ← 2 (N)
    Else if polarity > -0.6 then
        sentimentlabel (VN)
    Else
        sentiment_label ← 0 (Neu)
End
```

- Concatenation: sentence embedding and linguistic features are concatenated into one vector and will be input to the classifier.
- Training and classification: To further capture sequential dependencies and emotional tone, the output of fused-SBERT is passed through a BiLSTM layer, which models long-range contextual features from both directions. A softmax classifier then assigns one of the five sentiment labels. This combined model leverages SBERT's deep semantic embeddings and BiLSTM's sequence learning capabilities, enhancing the accuracy of fine-grained sentiment classification in course reviews. The reviews are classified into five classes.

3.3. Sequential modeling using bidirectional LSTM

The sequence of SBERT-generated embeddings is passed into a BiLSTM layer. The Bi-LSTM architecture epitomizes a subclass of recurrent neural networks (RNNs) that enhances the examination of sequential data by concurrently engaging in processing along both forward and backward trajectories. This framework integrates the functionalities of long short-term memory (LSTM) networks with bidirectional processing capabilities, thereby facilitating the extraction of contextual insights from both preceding and succeeding components within the input sequence. LSTM represents a sophisticated architecture of RNNs

specifically engineered to model sequential data and long-term dependencies effectively. In contrast to conventional RNNs, LSTMs possess the capability to acquire long-range temporal dependencies, attributed to their distinctive architecture, which incorporates a memory cell alongside three gating mechanisms: the forget gate, the input gate, and the output gate, shown in Figure 5. The equation for LSTM is as follows [31], [32]. Figure 6 shows the framework of Bidirectional LSTM and how it works.

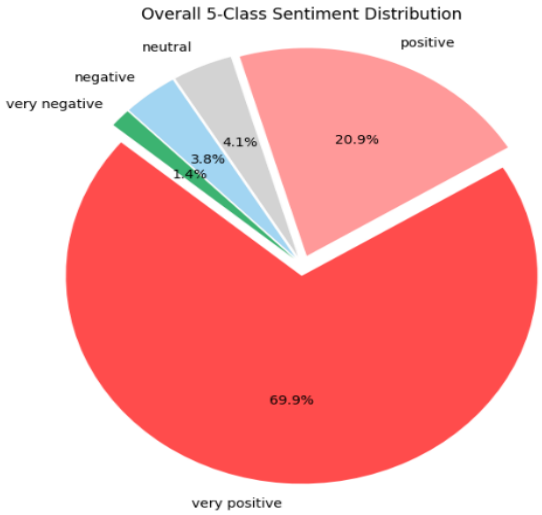

Figure 3. Overall Sentiment Distribution of Sentiments in 5 classes

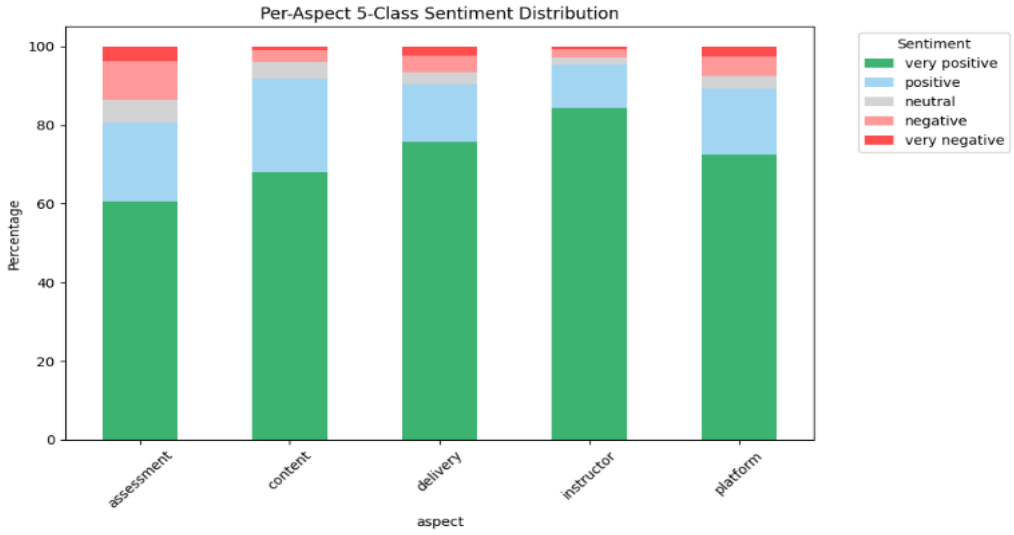

Figure 4. Sentiment distribution of aspects

$$f_t = \delta(W_f.[h_{t-1}, w_t] + b_f) \quad (3)$$

$$i_t = \delta(W_f.[h_{t-1}, w_t] + b_i) \quad (4)$$

$$o_t = \delta(W_o.[h_{t-1}, w_t] + b_o) \quad (5)$$

$$q_t = tanh(W_c.[h_{t-1}, w_t] + b_c) \quad (6)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot q_t \quad (7)$$

$$h_t = o_t \odot tanhc_f \quad (8)$$

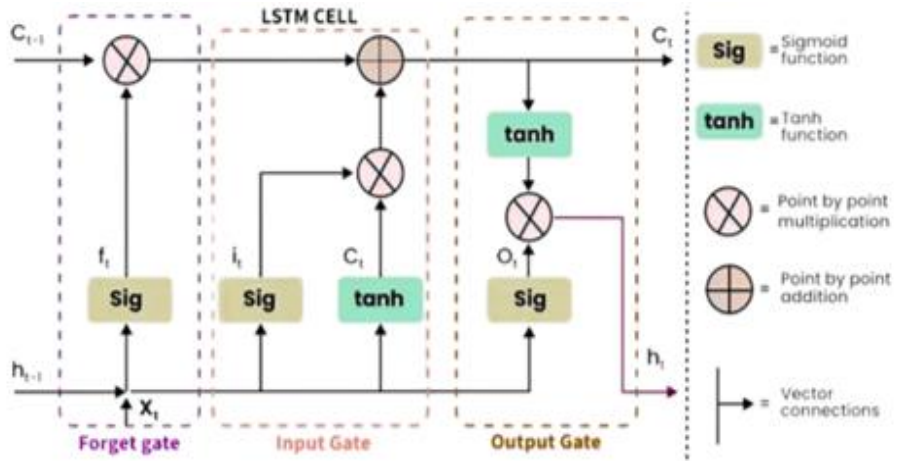

Figure 5. LSTM cell unit

Where f_t, i_t, o_t are the forget gate, input gate, and output gate, respectively. c_t is an intermediate state. h_t Is the output at time t. W_f, W_i and W_o show the weight of the model and b_f, b_i and b_o belong to model biases. δ is a common sigmoid. function, and ⊙ is multiplication. Unlike SBERT, which models attention-based relationships, BiLSTM captures sequential dependencies and long-range contextual patterns by reading the input in forward and reverse directions. This helps the model recognize nuanced sentiment features, such as negations or intensifiers, spread across the sentence. The BiLSTM produces a comprehensive output vector, either through pooling or by concatenating the final forward and backward hidden states. This vector is then passed into a fully connected layer that outputs five sentiment logits. A softmax activation transforms these logits into class probabilities corresponding to: VN, N, neutral, P, and VP. During training, the model utilizes cross-entropy loss and fine-tunes the BiLSTM and classifier layers (optionally, SBERT) to enhance classification accuracy. This SBERT+BiLSTM hybrid model effectively combines contextual and sequential insights, enhancing fine-grained sentiment classification performance.

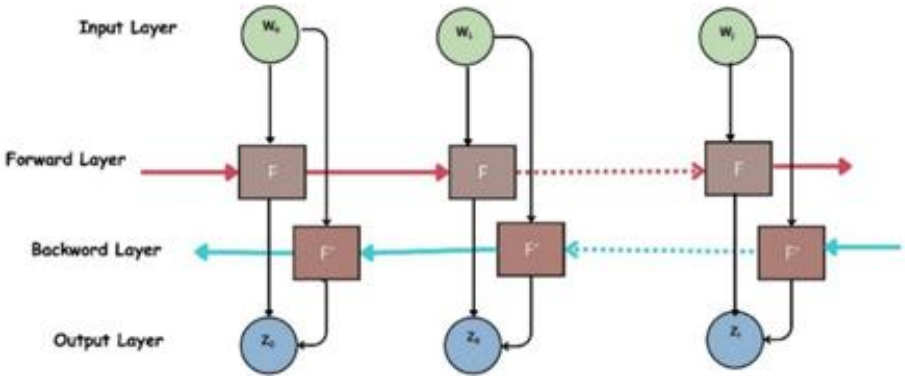

Figure 6. Architecture of bidirectional LSTM

- Input Layer: w_0, w_1, ..., w_j are the nodes that represent input vectors at each time step in the sequence.
- Forward layer: The sequence blocks in LSTM labeled F process from start to end, capturing information about the past context at each time step. The forward hidden layer vector is V. In (9) is given to calculate V.

$$V_i = LSTM(w_j, V_{i-1}) \tag{9}$$

- Reverse layer: This LSTM block labeled F′ processes the input sequence from the end to the start, capturing information about the future context at each time step. The backward hidden layer vector is calculated using in (10).

$$V_i' = LSTM(w_j, V_{i-1}') \tag{10}$$

- Output layer: These nodes' outputs (z_0, z_1, ..., z_j) from both the forward and reverse LSTMs are usually combined in some way; for instance, through operations such as concatenation or averaging, one can formulate the ultimate representation pertinent to each discrete time interval (Algorithm 2).

Algoritma 2. Pseudocode Bi-LSTM

```
Begin
 Initialize the sequential model
Add Input Layer
Add Bidirectional Layer
           return Sequences =False
Add DropoutLayer
Add a DenseLayer = 64 units and ReLU activation
 Add a DenseLayer = 5 units and softmax activation
          Set optimizer = 'adam'
         Set loss = 'crossentropy'
          Set metrics = ‘Precision’,
'accuracy',’Recall’, ‘f1-score’
End
```

- Evaluation: Model performance is evaluated using different metrics. Also evaluated this for each class as well. Additionally, the validation process was conducted to ensure generalizability.

4. EXPERIMENTAL SETUP AND RESULTS

4.1. Dataset

The course review dataset utilized in this study was sourced from Kaggle, consisting of course reviews with a total of 1.5 million reviews. Each review includes fields that contain the reviewer's Name, Date of review, and rating. There are 622 courses, including technical and non-technical courses. Reviews, course name, and institute name. Table 2 represents the description of the dataset. Only course reviews given by reviewers to the course are used in this research link to the dataset is: https://www.kaggle.com/datasets/imuhammad/course-reviews-on-coursera

Table 2. Deatails of the dataset field

Variable	Description
name	Name of course
institution	The name of reviewer who wrote the review
course_url	Link (URL) of course
course_id	Unique id of course
reviews	Reviews of course
reviewers	The name of the reviewer who wrote the review
date_reviews	Review posted date

4.2. Environmental setup

In the case of deep learning, an epoch is one full cycle through the entirety of the training data. For this specific model, it was trained over the course of five epochs. So that it would iteratively fine-tune its internal parameters. The batch size of 16 indicates that the model updated its weights after processing 16 samples. Although smaller batch sizes are associated with faster convergence, they also slow down learning due to noisier gradient estimates. In order to assess the model's results on new data, the validation split was set to 20% as shown in Table 3. This helps track how well the model is likely to perform on unseen data and helps in hyperparameter tuning or in applying early stopping techniques. To avoid overfitting, a dropout rate of 0.3 was added, which turns off 30% of the units in a given layer during training. This ensures that a model does not depend on particular features. To inject non-linearity into the model, the ReLU (Rectified Linear Unit) activation function is used as shown in (11). This is advantageous since it is computationally efficient and minimizes the likelihood of vanishing gradients. Also, the Adam optimizer was selected because of its adaptability and how well it works for training deep neural networks.

4.3. Matrices

In this paper, the different assessment parameters used to check the performance of the proposed algorithm are Precision, Recall, and accuracy, as shown in Figure 7 [12]. Accuracy: This can be measured by “The ratio of correctly predicted instances to total instances “.

Table 3. Experimental parameter setting

Parameters	value
Epochs	5
Batch-Size	16
Split	0.2
Drop-out	0.3
Activation function	Relu
optimizer	Adam
Transformer	all-MiniLM-L6-v2

$$f(x) = \max(0, x) \tag{11}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{12}$$

Precision: This is measured by "The ratio of correctly predicted positive instances to the total predicted positive instances".

$$Precision = \frac{TP}{TP+FP} \tag{13}$$

Recall: This is measured by "The ratio of correctly predicted positive instances to the total predicted instances that are actually labelled as positive".

$$Recall = \frac{TP}{TP+FN} \tag{14}$$

F1-score: is the harmonic mean of precision and recall, considering both the accuracy and comprehensiveness of the model.

$$F1\ score = \frac{2 \times Precision \times Recall}{Prescision + Recall} \tag{15}$$

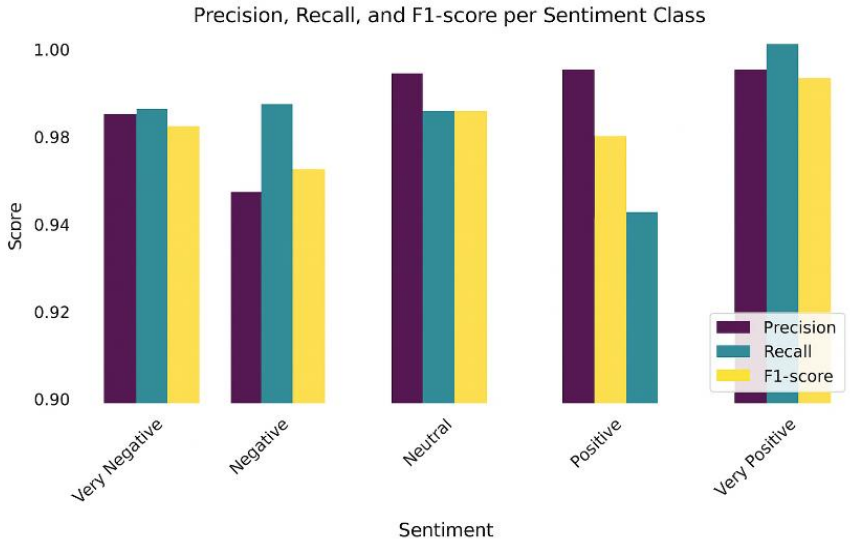

Figure 7. Precision, Recall, F1- Score per sentiment using the proposed model

4.4. Experiment analysis

The originality of the proposed work lies in the amalgamation of a hybrid architecture that combines fusion-based SBERT with (BiLSTM) for fine-grained sentiment classification, specifically applied to online course reviews. Unlike previous models that typically use sentence-BERT with a simple classification layer, our approach leverages BiLSTM to process SBERT's contextual embeddings further, enabling the model to get sequential patterns and long-range dependencies in textual data. Another key innovation is the

implementation of sentiment classification in five classes VN, N, Neu, P, VP, offering a more detailed sentiment spectrum than the binary or ternary classifications commonly used in prior studies.

Additionally, we incorporate aspect and feature extraction to associate sentiment with specific components of a course, such as content or instructor quality, thus enhancing interpretability and insight. Finally, we introduce a heuristic aggregation algorithm capable of determining the overall sentiment polarity from the SBERT+BiLSTM output vectors. This algorithm is flexible and can be adapted to binary, four-class, or five-class classification tasks, making the system robust and versatile across different sentiment analysis scenarios. Collectively, these contributions present a novel and comprehensive approach to analyzing user feedback in online educational platforms. Table 4 represents the performance matrix of the proposed model, and Table 5 shows the performance comparison of different models with our model.

Table 4. Performance matrix of this study

	Presission	Recall	F1- score	Accuracy
Very negative	0.97	0.98	0.97	
Negative	0.95	0.98	0.96	
Neutral	0.99	0.98	0.98	0.99
Positive	0.99	0.95	0.97	
Very positive	0.99	1	0.99	

Table 5. Performance comparison of models

Method	Precision	Recall	f1-score	Accuracy	ROC
BB-OAM [30]	0.906	0,920	0.914	0.865	-
BERT-BiLSTM [29]	93.78	93.8	93.79	-	-
CNN over BERT GCN [17]	-	-	78.76	85.25	-
Bi-GRU+LSTM [15]	0.92	-	0.92	0.93	-
Proposed model	0.978	0.978	0.974	0.99	Yes

Figures 8 and 9 show that the model's training paths improve, with both training and validation accuracies steadily increasing to approximately 98% at their peak. Concurrently, the loss functions for both datasets demonstrate a significant reduction, indicative of effective learning and minimal overfitting. The negligible inequality between the training and validation metrics across the epochs recommends that the model simplifies proficiently to previously hidden data, thereby corroborating its consistency and stability in the context of sentiment classification endeavors.

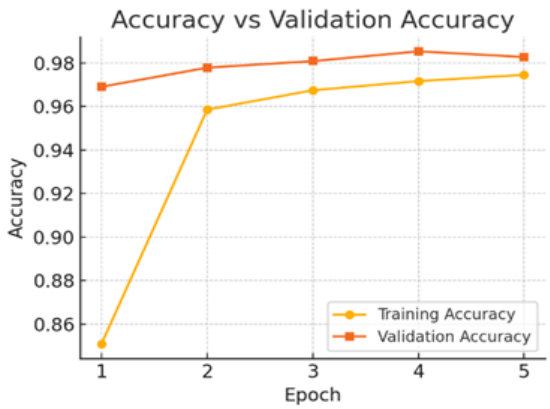

Figure 8. Accuracy Vs validation accuracy



Figure 9. Loss Vs validation loss

The ROC curve plot shown in Figure 10 depicts the performance of a multi-class sentiment analysis model over five sentiment classes: VN, N, Neutral, P, and VP. Each of these curves shows the TPR versus the FPR for a class. The area under the curve (AUC) for all classes remains consistently high (AUC = 0.99), indicating that the model has excellent discriminative ability and can effectively differentiate between various sentiment classes. The proximity of the curves to the top-left corner of the plot also depicts a good sensitivity vs specificity balance. All these results suggest that the model has reliable performance in classification across all sentiment classes and can be applied to precise sentiment prediction tasks.

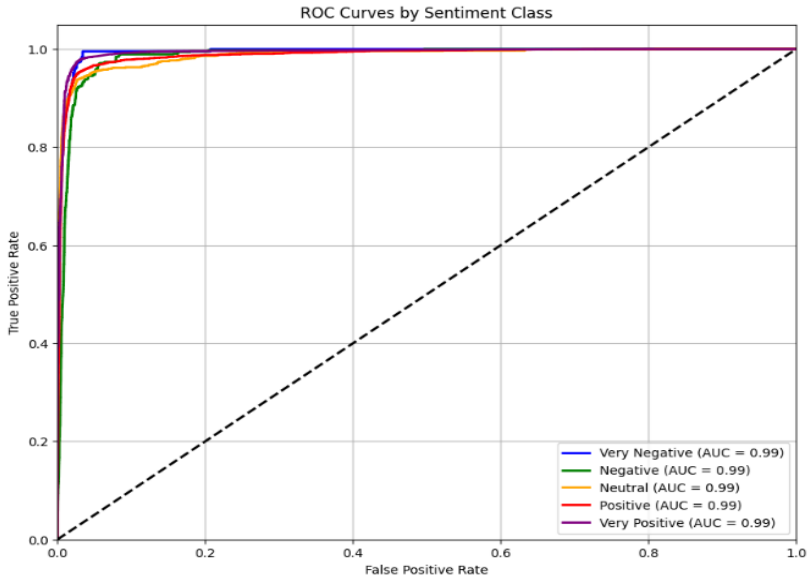

Figure 10. Receiver operating characteristic (ROC) curves for sentiment classification

5. CONCLUSION DAN FUTURE WORK

The originality in our suggested work comes from incorporating a hybrid architecture, which involves combining SBERT and Bidirectional LSTM (BiLSTM) for fine-tuned sentiment classification, used specifically for online course reviews. In contrast to earlier models that generally incorporate SBERT with a basic classification layer, our solution utilizes BiLSTM to process Sentence-BERT's semantic embeddings, enabling the model to identify sequential patterns and long-distance dependencies in text-based data. Another key innovation is the employment of five-class sentiment classification (VN, N, neutral, P, VP), richer than the binary or ternary classifications used in the prior work. In addition, we employ aspect and feature extraction to link sentiment with specific aspects of a course, e.g., content or instructor quality, which further facilitates interpretability and comprehension. Lastly, we present a heuristic aggregation algorithm that can decide the overall sentiment polarity based on the output vectors of SBERT+BiLSTM.

This algorithm is versatile and can be modified to work with binary, three-class, four-class, or five-class classification tasks for various scenarios of sentiment analysis. Additionally, this method could be improved by including interpretability methods such as LIME and SHAP to understand how the model makes its decisions. The application of interpretable methods would not only contribute to the transparency and trustworthiness of the approach but would also help to show whether some features influence sentiment predictions more than others; thus, helping to promote informed and ethical application of AI in practice.

FUNDING INFORMATION

Not Applicable

AUTHOR CONTRIBUTION

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Priyanka Shukla	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Jitendra Nath Singh		✓			✓	✓				✓	✓	✓		
Shachi Mall	✓		✓	✓	✓	✓	✓			✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

INFORMED CONSENT

This research doesn't involve situations where informed consent is applicable; it's purely theoretical or uses publicly available data.

RESEARCH INVOLVING HUMAN AND /OR ANIMALS

This study did not involve research on human participants or animals.

DATA AVAILABILITY STATEMENT

The data supporting of this study are available at Kaggle.com. https://www.kaggle.com/datasets/imuhammad/course-reviews-on-coursera

REFERENCES

[1] H. Khalil and M. Ebner, "MOOCs completion rates and possible methods to improve retention-A literature review," *in In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2014, pp. 1236–1244.

[2] S. L. Schneider and M. L. Council, "Distance learning in the era of COVID-19," *Archives of Dermatological Research*, vol. 313, no. 5, pp. 389–390, 2021.

[3] I. Sindhu, S. M. Daudpota, K. Badar, M. Bakhtyar, J. Baber, and M. Nurunnabi, "Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation," *IEEE Access*, vol. 7, pp. 108729–108741, 2019.

[4] C. A. Sowndarya et al., "Comparative analysis of machine learning and deep learning models for aspect-based sentiment analysis in education," *Journal of Scientific Research and Reports*, vol. 30, no. 12, pp. 567–576, 2024.

[5] A. Kurti and F. Dalipi, "Bridging the gap between academia and industry: Lessons learned from a graduate IT professional development program," *in Proc. 2nd Annual International Conference on Engineering Education & Teaching*, Athens, Greece, Jun. 5–8, 2017, p. 27.

[6] M. J. Gomez, M. Calderón, V. Sánchez, F. J. G. Clemente, and J. A. Ruipérez-Valiente, "Large scale analysis of open MOOC reviews to support learners' course selection," *Expert Systems with Applications*, vol. 210, Art. no. 118400, 2022.

[7] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," *IEEE Access*, vol. 7, pp. 51522–51532, 2019.

[8] S. T. Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," *Array*, vol. 14, Art. no. 100157, 2022.

[9] M. Arbane, R. Benlamri, Y. Brik, and A. D. Alahmar, "Social media-based COVID-19 sentiment classification model using Bi-LSTM," *Expert Systems with Applications*, vol. 212, Art. no. 118710, 2023.

[10] Z. Hameed and B. Garcia-Zapirain, "Sentiment classification using a single-layered BiLSTM model," *IEEE Access*, vol. 8, pp. 73992–74001, 2020.

[11] M. U. Salur and I. Aydin, "A novel hybrid deep learning model for sentiment classification," *IEEE Access*, vol. 8, pp. 58080–58093, 2020.

[12] P. Rasappan, M. Premkumar, G. Sinha, and K. Chandrasekaran, "Transforming sentiment analysis for e-commerce product reviews: Hybrid deep learning model with an innovative term weighting and feature selection," *Information Processing & Management*, vol. 61, no. 3, Art. no. 103654, 2024.

[13] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Information Sciences*, vol. 488, pp. 190–204, 2019.

[14] K. F. Hew, X. Hu, C. Qiao, and Y. Tang, "What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach," *Computers & Education*, vol. 145, Art. no. 103724, 2020.

[15] S. Gupta, N. Singhal, S. Hundekari, K. Upreti, A. Gautam, P. Kumar, and R. Verma, "Aspect based feature extraction in sentiment analysis using Bi-GRU-LSTM model," *Journal of Mobile Multimedia*, vol. 20, no. 4, pp. 935–960, 2024.

[16] M. M. A. Busst, K. S. M. Anbananthen, S. Kannan, J. Krishnan, and S. Subbiah, "Ensemble BiLSTM: A novel approach for aspect extraction from online text," *IEEE Access*, vol. 12, pp. 3528–3539, 2024.

[17] H. T. Phan, N. T. Nguyen, and D. Hwang, "Aspect-level sentiment analysis using CNN over BERT-GCN," *IEEE Access*, vol. 10, pp. 110402–110409, 2022.

[18] I. Setiawan, Y. P. Tjendra, and A. Y. Zakiyyah, "Comparing machine learning algorithms with ensemble model using random oversampling for predicting student's dropouts," *in Proc. 6th Int. Conf. Cybernetics and Intelligent System (ICORIS)*, 2024, pp. 95–100, doi: 10.1109/icoris63540.2024.10903843.

[19] L. Khan, A. Amjad, K. M. Afaq, and H. T. Chang, "Deep sentiment analysis using CNN-LSTM architecture of English and Roman Urdu text shared in social media," *Applied Sciences*, vol. 12, no. 5, Art. no. 2694, 2022.

[20] L. G. Atlas, D. Arockiam, A. Muthusamy, B. Balusamy, S. Selvarajan, T. Al-Shehari, and N. A. Alsadhan, "A modernized approach to sentiment analysis of product reviews using BiGRU and RNN based LSTM deep learning models," *Scientific Reports*, vol. 15, no. 1, Art. no. 16642, 2025.

[21] M. Dragoni, M. Federici, and A. Rexha, "An unsupervised aspect extraction strategy for monitoring real-time reviews stream," *Information Processing & Management*, vol. 56, no. 3, pp. 1103–1118, 2019.

[22] M. M. Agüero-Torales, J. I. A. Salas, and A. G. López-Herrera, "Deep learning and multilingual sentiment analysis on social media data: An overview," *Applied Soft Computing*, vol. 107, Art. no. 107373, 2021.

[23] A. Chauhan, A. Sharma, and R. Mohana, "An enhanced aspect-based sentiment analysis model based on RoBERTa for text sentiment analysis," 2025, doi: 10.31449/inf.v49i14.5423.

[24] C. Zhao et al., "Enhancing aspect-based sentiment analysis with BERT-driven context generation and quality filtering," *Natural Language Processing Journal*, vol. 7, Art. no. 100077, 2024.
[25] N. Wayessa and S. Abas, "Multi-class sentiment analysis from Afaan Oromo text based on supervised machine learning approaches," *International Journal of Research Studies in Science, Engineering and Technology*, vol. 7, no. 7, pp. 10–18, 2020.
[26] C. Wan, Y. Peng, K. Xiao, X. Liu, T. Jiang, and D. Liu, "An association-constrained LDA model for joint extraction of product aspects and opinions," *Information Sciences*, vol. 519, pp. 243–259, 2020.
[27] S. P. Yadav, "Emotion recognition model based on facial expressions," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26357–26379, 2021.
[28] Y. F. Izzaddien, R. Sarno, A. T. Haryono, A. F. Septiyanto, D. Sunaryono, and R. A. H. Winarso, "Multilabel aspect-based sentiment analysis with diverse embedding techniques and finetuned transformers for label detection," *in Proc. 2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C)*, Dec. 2024, pp. 95–100.
[29] X. Li, Y. Lei, and S. Ji, "BERT- and BiLSTM-based sentiment analysis of online Chinese buzzwords," *Future Internet*, vol. 14, no. 11, Art. no. 332, 2022.
[30] Y. Huang and C. Ren, "Research on MOOC reviews oriented sentiment analysis by awareness of emotional distinctions," *IEEE Access*, vol. 14, pp. 154823-154831, 2024, doi: 10.1109/ACCESS.2024.3482232.
[31] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent Twitter sentiment classification," *in Proc. 52nd Annu. Meeting of the Assoc. Comput. Linguistics*, 2014, pp. 49–54.
[32] S. Kiritchenko, X. Zhu, and S. M. Mohammad, "Sentiment analysis of short informal texts," *J. Artif. Intell. Res.*, vol. 50, pp. 723–762, Aug. 2014.

BIOGRAPHIES OF AUTHORS



Priyanka Shukla is Research Scholar in the School of Computing Science and Engineering, Galgotias University. She has 17 years of experience in teaching. She is pursuing a Ph.D. (CSE) at Galgotias University, done M.E. (Computer Engineering) in the Government. S.G.S.I.T.S., Indore (MP) and B.E.(CSE) in SRCEM, Gwalior. In addition, she has published four patents and eight research articles, including book chapters. Her areas of interest includes natural language processing, machine learning, and artificial intelligence. She can be contacted at email: engpriyanka.8@gmail.com.



Dr. Jitendra Nath Singh is working as Professor in School of Computing Science and Engineering, Galgotias University, Greater Noida, India, having 25 years of teaching experience and 2 years of industry Experience. He has completed a master's degree from Madan Mohan Malaviya University of Technology Gorakhpur and Ph.D. in Computer Science from Babasaheb Bhimrao Ambedkar University Lucknow (Central) India. He has served as Chair in many conferences and is affiliated as a member of the program committee of many conferences in India. He has supervised 6 Ph.D. students and many M.Tech. students for their thesis. He is also a Wipro Certified faculty member since 2016. Published several research papers in Conferences, Journals, and book chapters and more than 90 Papers Published in various National, international journals with Galgotias University affiliation. His area of research interest include information retrieval, image processing, software reliability, and natural language processing. He can be contacted at email: singhjn2000@gmail.com.



Dr. Shachi Mall is Professor in the School of Computer Science Engineering, Galgotias University, India. My educational background includes Post-Doctoral Fellow in School of Computer Science, Engineering and Technology, Bennett University, India, Ph.D. and M.Tech in Computer Science and Engineering from Madan Mohan Malaviya University of Technology, Gorakhpur in the state of Uttar Pradesh, India. As a highly motivated and enthusiastic individual with a strong background in explainable AI, machine learning, deep learning and natural language processing, and computer vision. I have published 17 papers in international journals, presented and published 24 papers in international conferences indexed in ESCI, SCI, Web of Science, and Scopus, and authored book chapters and books published by an international publisher. She also holds 11 patents. She can be contacted at email: shachimall@gmail.com.