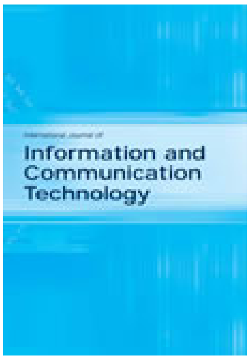
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Data analytics for students' feedback in college education using bi-directional models and fasttext embeddings

Fangbin Song, Di Ma

DOI: 10.1504/IJICT.2025.10071318

Article History:

Received:	17 March 2025
Last revised:	03 April 2025
Accepted:	03 April 2025
Published online:	11 June 2025

Copyright © 2025 Inderscience Enterprises Ltd.

Data analytics for students' feedback in college education using bi-directional models and fasttext embeddings

Fangbin Song and Di Ma*

School of Design Art and Media,
Nanjing University of Science and Technology,
Nanjing, Jiangsu 210094, China
Email: madidi0615@163.com
Email: samuelsong2525@126.com
*Corresponding author

Abstract: This study aims to analyse the students' feedback data for enhancing the educational system. Teachers' feedback serves as a critical tool for assessing educational outcomes and improving teaching strategies. Natural language processing (NLP), an active research area of artificial intelligence (AI), offers novel solutions for analysing and understanding large volumes of feedback data, aiding in the refinement of educational colleges. This paper aims to carry out a comprehensive analysis of students' feedback by classifying content into five classes using advanced AI techniques including machine learning, ensemble methods, and deep learning (DL) combining with both textual features and word embedding features to improve predictive performance. Among all the applied features, the hybrid approach of the latest technique of FastText with DL model of Bi-GRU reveals the highest results with accuracy of 95%. This research confirms that NLP features provide deep insights into content and help us predict the various aspects of students' feedback for improvements in the educational sector.

Keywords: education; natural language processing; NLP; artificial intelligence; AI; sentiment analysis; deep learning; DL; feedback analysis.

Reference to this paper should be made as follows: Song, F. and Ma, D. (2025) 'Data analytics for students' feedback in college education using bi-directional models and fasttext embeddings', *Int. J. Information and Communication Technology*, Vol. 26, No. 17, pp.109–132.

Biographical notes: Fangbin Song is experienced data analytics, natural language processing (NLP), and machine learning (ML) researcher, currently based at the School of Design Art and Media, Nanjing University of Science and Technology. In recent years, his research focus has been on applying advanced computational techniques to analyse textual data in educational settings, experimenting data science approaches for improvement of student feedback analysis and applying data science to support educational outcomes through data literacy.

Di Ma is at the School of Design Art and Media, Nanjing University of Science and Technology, focusing on data science, machine learning and multimedia design. Aside from teaching, his research interest is computational techniques in digital media, interactive systems, and educational technology. At the same time, he teaches how machine learning algorithms can be used to analyse the user behaviour and to improve multimedia content delivery as well as how to develop innovative ideas for interactive learning environments.

Copyright © The Author(s) 2025. Published by Inderscience Publishers Ltd. This is an Open Access Article distributed under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

1 Introduction

In the evolving landscape of higher education, the pursuit of effective teaching strategies and improved learning outcomes has become increasingly vital. As educational institutions strive to meet the diverse needs of students, the continuous assessment and refinement of instructional methods are essential. Education occupies the foundation of individual and organisational improvement and transformation of individuals into productive members of society. In colleges which prepare the students for the job market to acquire necessary skills, the part which educators' plays is of more importance. Teachers are not only the disseminators of knowledge but co-educators, protectors, and facilitators of students' dreams, supporters throughout their school-learners' progress, and trainers for life (Munna and Kalam, 2021). As such, teaching and the experiences that students undergo critically rely on feedback systems that best capture both student learning needs and instructor performance. However, traditional approaches to collecting and analysing feedback do not adequately capture the opinions of students. Structured questionnaires and assessments may offer participants' quality data but usually do not capture the richness of the experiences (Mailool et al., 2020). This limitation fuels the need to look for better ways of analysing feedback to get more complex insight. Here comes, artificial intelligence (AI) (Bardach and Klassen, 2020), innovation in many areas, including healthcare (Saraswat et al., 2022), psychology (Alsini et al., 2024), marketing (Mariani et al., 2022), and education (Urooj et al., 2023). The use of AI technologies within the feedback processes is a powerful way for educational organisations to advance their approaches to measuring teach-learn processes and activities in the given educational context that has no previous parallels by analysing their sentiments (Shaik et al., 2022).

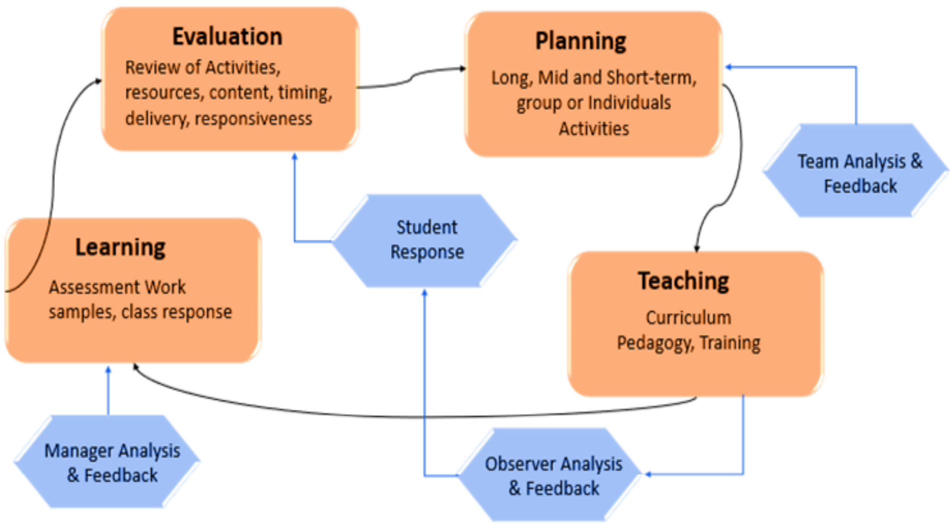
Emerging AI technology particularly natural language processing (NLP) is considered one of the most helpful domains for analysing qualitative data utilised using various tools such as sentiment detection, feedback analysis, grammatical analysis, and summary generation. Sometimes unstructured information such as respondents' comments on surveys, evaluations, and students feedback are hard to analyse with conventional text-mining techniques; It helps institutions go beyond the basic details, as evidenced from teachers' feedback, and gains deeper understanding of feedback trends that can impact instruction and curriculum development (Bhowmik et al., 2023). The consequences of feedback analysis with the help of NLP are quite vast. The findings of the research agree with previous works, where positive feedback is proven to enhance self-efficacy, to promote active involvement in the learning process with the help of surveys, reports, discussions, and questionnaire (Ahmed et al., 2022). Furthermore, as the educational settings at colleges dynamically transform due to the occurring technological and workforce progression, having stable general teachable strategies becomes vital. These innovations, therefore, present a perfect way through which colleges can develop viable spaces such as teaching and learning arenas that support learner achievement and enable instructors to enhance their didactic practices (Wu, 2021). Therefore, the collaboration between sophisticated ML and DL methods and students' feedback analysis reveals a progressive shift in learning paradigms at college level. Due to the use of AI institutions can ensure the development of a more responsive educational environment that can be supportive of learners' needs (Alarfaj et al., 2022). In envisioning the future of education, these technologies have the capability of, not only improving quality of instruction, but more importantly the performance of students in preparation for the promotion of a pool of qualified workforce. When analysing feedback responses from students, sentiments can typically be categorised into five classes: awesome, good, average, poor and awful. All categories offer important information on the performance of colleges. For example, awesome feedback signifies that colleges have exceptional satisfaction with their teaching experiences and institutional support. The positive and optimistic responses by the learners show the level of commitment by the teachers providing excellent teaching environment (Spatiotis et al., 2020).

On the other hand, if the responses by the students are negative such as awful then it provides an opportunity for the students to lessen their frustration. In addition, by sharing their views openly, their disappointment is expressed, and this results in providing the exact reason for the dissatisfaction of the learners whether it is related to course content, the environment in which it is shared or the attitude or lecture delivery style of the teachers. The review of the student can also be about the various institutional level issues which may include lack of administrative support, inadequate resources, etc. The response classes of good and average shows reasonable satisfaction but it also shows that the students are not completely satisfied otherwise they would have opted for better class labels. The class label of neutral shows that their experiences are neither particularly positive nor negative. We can also conclude from such comments, as at basic level the teachers are meeting the learners' expectations, however, there is room for improvements by the teachers. Taking feedback positively, the educational institutes that can use this feedback to identify areas for improvements that could elevate the overall teaching experience (Wang et al., 2025). Considering another class label of good, it depicts that the students feel positive about their learning experiences. For instance, the comments like, "I appreciate the collaborative environment among colleagues", shows that the students satisfied with certain aspects of their roles and have the potential to thrive further with continued support and resources. Having received positive comments may encourage teachers to share best practices (Avila et al., 2020). On the contrary, poor feedback from students often points to specific challenges they face in their teaching practices or interactions. This could include comments about insufficient classroom resources, difficulties in engaging students, or challenges in curriculum delivery. While not as severe as awful feedback, poor responses still indicate areas that require attention and improvement to enhance both teacher satisfaction and student learning outcomes. Addressing these concerns is critical for fostering a positive learning environment and ensuring that the educators' feeling is positive as they think themselves as valued and equipped to succeed (Schles and Robertson, 2019). The breakdown of these feedback categories with the help of NLP techniques describe allows educators to pick up on certain trends and feelings that can be missed when analysing the results using more straight-forward approaches. Such results can be identified the reason that would prompt a discussion of the curriculum used or the teaching methods applied. In this way, by reviewing such trends, also defined in Figure 1, colleges can promote the effectiveness of improvements as well as to meet the requirements of both teachers and the institution.

In this research study textual dataset utilised for the analysis of teachers' feedback of colleges among education and other resource setting. By employing comprehensive study with two features engineering techniques from traditional to advanced approaches such as term-frequency inverse document frequency (TF-IDF) and word embeddings include Fast Text using AI-based techniques such as machine learning (ML), ensemble learning, and deep learning (DL) methods. Models include support vector machine (SVM), decision tree (DT), random forest (RF), gated-recurrent unit (GRU) and Bi-directional GRU)

Bi-GRU), which are evaluated using standard measures of accuracy, precision, recall and f1-score. Using this empirical analysis of teacher's feedback, state-of-the-art Bi-GRU model coupled with FastText embeddings achieves the highest accuracy of 85% showing the roadmap for future research to enhance more predictive model in NLP task.

Figure 1 Flow of teaching and learning strategies (see online version for colours)



The rest of the paper is organised as: Section 2 provides the depth analysis of existing studies based on ML and DL approaches. Section 3 defines the applied methodology. Section 4 provides the details of dataset and performance measures covering experimental setup. Section 5 discusses the results achieved based on methodology. Section 6 presents the conclusion and future work in this domain.

2 Related work

ML and DL techniques have been widely used in automating the analysis of students' feedback in college education. However, SVM and Naïve Bayes have been widely used in previous studies as traditional ML approaches for text classification, sentiment analysis. However, as DL models such as recurrent neural networks (RNNs) and Transformers became more precise with providing feedback analysis, feedback analysis has also come to be more accurate and efficient. Unfortunately, in recent research there are limited successful attempts to incorporate these aspects into model design. This section goes over existing ML and DL based methods for analysing student feedback, its pros and cons, and mentions the requirements of more sophisticated approaches.

2.1 Machine learning

The use of ML geared towards education has recently attracted considerable interest especially in the analysis of teacher feedback or enhancing learning. A systematic review (Hilbert et al., 2021) on the various ML techniques used in educational data analysis and pointed out how specific methods such as DT and SVM of supervised learning have been used in students' performance and learning pattern prediction. This paper highlights the need for designing effective teaching improvements using ML in that they recommend the use of technology to create learning content that matches educational learning needs. Based on these findings, Naïve Bayes and RF models also applied (Song et al., 2024) for analysis of teacher's feedback that was found 95% accurate in identifying sentiments of effectiveness. Other developments in the use of ML were provided by Bernius et al. (2022), a framework made possible using ML, provides feedback on textual student answers in large classrooms. The above elimination also helps improve the efficiency of assessment and offers students meaningful and timely responses to the results that help teachers to maintain their curriculum record and outline. The conclusion generated in this research is like (Dann et al., 2024) who pointed to learning analytics, supported by ML algorithms, for the detection of at-risk teachers' identification that are not supported by institutions to meet their needs, and further analytics allow educators to increase retention levels and improve the overall quality of education. Furthermore, reflective on the vision of the use of ML-based solutions, teacher performance evaluation has seen innovation. For instance, using a video feedback system enhanced by ML algorithms, (Mao, 2022) pointed out how teachers utilised micro-skills to foster reflective practice in education. Likewise, Nyandwi et al. (2023) used ensemble boosting methods including GB, LDA, RF, and SVM classifiers, for the classification of feedback as belonging to which level of Bloom's taxonomy, an improvement of about 5% compared to the single models significantly limit the contextual analysis of diverse feedback. However, these initial works prove that first-order ML methods are suitable for categorised and semi-complex datasets and open the door for the use of higher-order approaches, such as DL. In aggregate, this research points to the fact that although ML techniques are still critically useful in educational feedback analysis, addressing their inherent limitations is best approached by integrating efforts across domains to sustain scalability and inclusiveness.

2.2 Deep learning

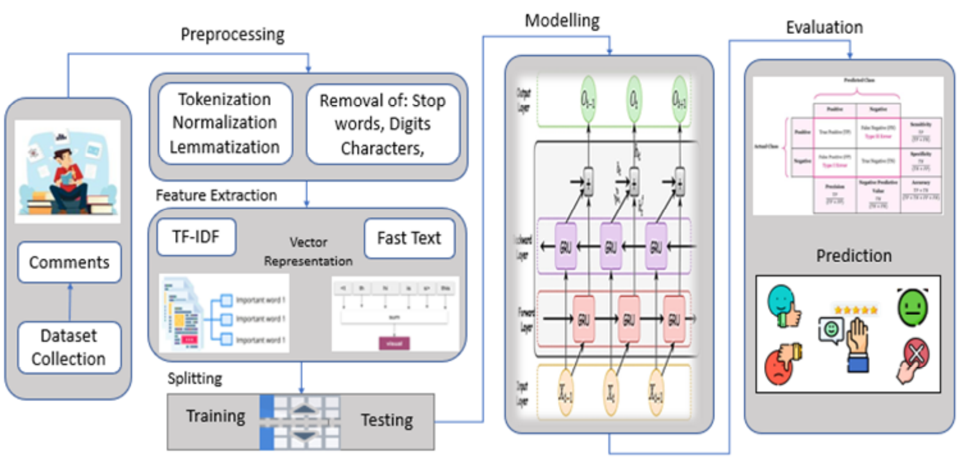
Recently, DL has developed as one of the most effective subsets of ML that has additional possibilities for enhanced learning in the educational database. Some of the previous research have shown positive results of the DL models in handling the teacher's feedback and enhancing the learning practices. For instance, using BERT, which is a transformer-based model, (Topping et al., 2025) automated qualitative feedback on teachers where accuracy was extremely high while identifying features that affect teaching effectiveness. This illustrates the capacity of DL models to learn subtle features of textual data that are not discernible to other methods of ML. Thirdly, Ahmad et al. (2023) assessed the performance of teachers using attention-based mechanism using GRU networks combination a on information gathered sentiments over consecutive polarity based on various services and products. Another study using LSTM networks (Reddy et al., 2022) could capture the temporal dependencies of feedback data to produce better forecasts of teaching quality and of students' satisfaction. As their findings show, this capability is especially useful when feedback is gathered on a consistent basis over a period. This study by Chen et al. (2024) compared the LSTM layers with convolutional neural networks (CNNs) for classroom sentiment analysis; revealing that the DL models used can achieve a very high level of accuracy in sentiment classification and can also highlight sentiment factors that affect student's perceptions of teaching quality. Another paper Ahmed et al. (2023) proposed a spatial-temporal Bi-LSTM model to enhance answer selection in question answering by considering both similarities based on the context. This approach highlights the prediction of relevant answers, addressing the challenge of selecting top-quality responses from a pool of multiple answers, where irrelevant or insufficient replies complicate the process. The process of training DL models entails high computational costs and is complex and demanding in terms of capabilities, which may pose challenges for all institutions of learning, especially universities (Kastrati et al., 2021). More research investigation should be focused on redesigning more efficient DL architectures which would also lead to investigating how transfer learning can be applied to achieve high performance efficiency with less use of resources. There is also a variation of CNN used for text data according to Motevalli et al. (2025) applied GANs based on the tools to analyse text as the spatial data sequence to identify the distinguishing characteristics of critical feedback and analysed that will result moderate results in the feedback between positive and negative. In another extensive study Basiri et al. (2021) used combination of CNNs and RNNs for Sentiment Analysis to take advantage of the strength of both features, spatial and sequential and shown to be effective for fine-grained sentiment classification. Various attention mechanisms have also improved the performance DL models of feedback analysis further using fragile hierarchical attention network (HAN) (Yao, 2024) for teacher feedback classification. The amount of detail was also managed by their method, supported its usefulness for analysing hierarchical structures of feedback text. Despite the high accuracy and scalability of the DL techniques, the present solutions are not perfect. Such work is aligned with Hemmat et al. (2023) since the models indicate that small or imbalance datasets adversely affect the results of DL. Additionally, these models are constrained by the high computational task needed to train and implement, hence not suitable for implementation in resource scant education environment. However, some problems connected with model optimisation and transfer learning have been appearing that DL is increasingly becoming more convenient and effective. These works indicated the effectiveness of DL in the feedback analysis especially for unstructured and complex feedback data. While strengthening the main ML approaches, DL models influence the further development of educational feedback systems, which analyse the effectiveness and students' interest in education.

However, existing studies concerning student feedback analysis have several limitations. However, many traditional ML models like Naïve Bayes and SVM suffer from poor classification accuracy for complex contextual relationships in textual data. Despite such progress, there are still many studies that use static word embeddings, e.g., Word2Vec and GloVe, because they can not capture word semantics in varying contexts. In addition, most prior work does not consider detailed account of the feedback, including perspectives, emotion, or pedagogical relevance. Additionally, these models are not generalisable because of the limited availability of large, annotated datasets and computational complexity is an issue when trying to run these models in real-time in the educational setting.

3 Methods and materials

For this study, the research method was planned in a way that is systematic to address and assess the usefulness of AI-based models for evaluating feedback on teaching and education at colleges. Therefore, employing state-of-the-art NLP, the paper is devoted to textual feedback analysis. Within the present work, a methodical approach was employed as shown in Figure 2, starting with correct data preparation to exclude low-quality data and use only the most meaningful input and feature extraction to obtain valuable representations of the text. Three different techniques of AI-based models including ML, ensemble and DL models were then trained and tested to determine the best methods to classify feedback into different quality categories. This methodology brings to focus the combined use of modern and conventional methods, giving a glimpse of the benefits of using AI in analysing feedback in the teaching education system.

Figure 2 Applied methodology framework (see online version for colours)



3.1 Data cleaning process

This section describes preprocessing phase, as it sets the stage for analysis by tuning up the textual data to get the best from the models. Firstly, all those characters which are extra such as the stop words, digits, special characters, and punctuation marks were not useful for feedback analysis, so substantive content within the feedback was extracted only. Subsequently, lemmatising operation converted the n-grams to equivalent base form to minimise the number of unique words and enhance generalisation capability. The cleaning also involved normalising the text to reduce variability within the dataset; for example, converting all the letters to the lower case or removing multiple spaces. Tokenisation divided the text into substrings so that the models could also analyse it, whether it was by word or by phrase. Moreover, because the feedback contains syntactic information, part-of-speech (POS) tagging was done in order help in analysing the grammatical contextualisation. Collectively, these preprocessing steps represented a process of moving from rough text, which is a raw form of data, into a more refined text that was fit for feature extraction and modelling, as illustrated in Table 1. This made the data highly relevant and representative of the pattern of teachers' feedback since it was strictly collected and analysed.

Table 1 Pre-processing results

Original text	He is terrible! If you have him drop. him IMMEDIATELY! He is mean, and an extremely hard grader, he didn't give 04 GPA ever!!!!
Text normalisation	he is terrible! if you have him drop. him immediately! he is mean, and an extremely hard grader, he didn't give 04 gpa ever!!!!
Tokenisation	['terrible', 'drop', 'immediately', 'mean', 'extremely', 'hard', 'grade', 'didnt', 'give', 'gpa']
Removing punctuations, special character	he is terrible if you have him drop him immediately, he is mean and an extremely hard grader he didnt give 04 gpa ever
Removing digits	he is terrible if you have him drop him immediately, he is mean and an extremely hard grader he didnt give gpa ever
Stopward removals	terrible drop immediately means extremely hard grader didnt give gpa
Lemmatisation	terrible drop immediately means extremely hard grade didnt give gpa

3.2 Feature extraction

Feature extraction is important while text analysis is because feature extraction converts textual data into compact and numerical representations that can be understood and analysed easily by the ML and DL models. Two state-of-the-art methods of feature extraction TF: IDF and FastText were used in this study to extract meaningful features from the text. In TF-IDF approach, different weights for the words are determined depending on their distinct level of importance in the document set, it considers both word frequency in a specific document and its importance in a general document collection, counting terms frequency and document frequency at the same time, computed as in 1.

$$TFIDF(t, d) = TF(t, d) \cdot log\left(\frac{N}{DF(t)+1}\right) \tag{1}$$

where $TFIDF(t, d) = \frac{f_t}{\max(f_t, \ddot{t} \in d)}$ represents the term frequency of t normalised by the maximum frequency of any term in d, N is the total number of documents, and $DF(t)$ is the document frequency of term t.

Besides, TF-IDF technique, FastText, the word embedding model, was used to identify the sematic compatibility between words, where the words were embedded in a dense vector space. FastText generalises over sub-word information, which is especially useful for handling rare or even misspelled words, thus guarantees that the rarer terms have a bigger impact on the final representation of the document. The given form of weighted averaging guarantees that both local (word level) and global (document level) contexts will be corresponding to each other in the feature space, computed as in 2.

$$v_d = \frac{\coprod_{i=1}^{n} \alpha_i, v_{t_i}}{\coprod_{i=1}^{n} \alpha_i} \tag{2}$$

where $\alpha_i, = \frac{1}{\sqrt{DF(t_i)+1}}$ act as an inverse document frequent scaling factor for each word. v_d is an aggregate document embedding for embedding vector as v_{t_i} of word t_i in a document d.

The approach used in the context of feature extraction used in this paper proved to be highly effective in filling the existing gap between raw textual data and computational models for more accurate contextualised prediction.

3.3 Applied models

Model selection is important in text classification and in this study; ML, ensemble learning as well as DL were used in correctly predicting feedback categories. Both models make use of quite different concepts, and they have their advantageous in terms of mining albeit textual features (Onan, 2020). Here, each model is described in detail with a focus on the underlying mathematical equations of each of the models.

3.3.1 Support vector machine

SVM is a strong model of ML that tries to locate the best hyperplane through which data will be separated into different classes. For dataset labels of (x_i, y_i) where $x_i \in \mathbb{R}^d$ are the feature vectors and $y_i \in \{-1, 1\}$ as class labels, SVM optimises the objectives using 3.

$$min_{s,t,\Im} \frac{1}{2}\|s\|^2 + R\prod_{i=1}^{n} \Im_i \text{ subject to} : yi\left(s.x_i + t\right) \geq \forall_i \tag{3}$$

where s, t, $\Im$, R represents as weight vector size, term as bias, slack variable and regularisation function respectively, for controlling the trade-off between margin maximisation and misclassification. In further case of non-linear separability, SVM maps data to higher-dimensional space using kernel function t determines the influence of individual data point say ζ, computed as in 4.

$$K\left(x_i, x_j\right) = \exp\left(\zeta \left\|x_i - x_j\right\|^2\right) \tag{4}$$

3.3.2 Decision tree (DT)

DTs divide the feature space into regions making uses of conditions that bring the most information gain. Tree building is recursive, and the goal is made to try and maximise homogeneity at each node. The impurity at a node is quantified using Gini Index G, defined as in 5.

$$G = 1 - \bigcup_{k=1}^{K} p_k{}^2 \tag{5}$$

where p_k is the proportion of samples belonging to class K at the node. For multi-class problems, tree optimises the split criterion ΔG, as in 6.

$$\Delta G = G_{parent} - \frac{N_{left}}{N_{total}} G_{left} - \frac{N_{right}}{N_{total}} G_{right} \tag{6}$$

Here, N_{left} and N_{right} are the number of samples in the left and right child nodes, respectively.

3.3.3 Random forest

RF is an extension of standard DT, being an ensemble learning model that minimises overfitting by creating a collection of DTs. Given T DTs, the final prediction y_{pred} for a test input x is derived through majority voting, defined as in 7.

$$y_{pred} = mode\left(\left\{f_t(x)\right\}_{t=1}^{T}\right) \tag{7}$$

where $f_t(x)$ is the prediction of the t–th tree. To ensure diversity among trees. RF introduces randomness in two ways: by bootstrapping the dataset and by selecting a random subset of features for each split. The variance reduction in RF described as in 8.

$$Var(\breve{y}) = \frac{1}{T} Var\left(f_t(x)\right) + \left(1\frac{1}{T}\right) Conv\left(f_t(x)\right) \tag{8}$$

where $Conv(f_t(x))$ captures the correlation between predictions from individual trees.

3.3.4 Gated recurrent unit

GRU is a particular type of RNNs developed to cope well with sequential data. It employs gating mechanisms to regulate the passing through of information to overcome the vanishing gradient issue. For an input sequence $\{x_t\}$, the GRU computes the hidden state h_t at each step t as in 9–12:

$$z_t = \propto \left(W_z x_t + U_z h_{t-1} + b_z\right) \tag{9}$$

$$r_t = \propto \left(W_r x_t + U_r h_{t-1} + b_r\right) \tag{10}$$

$$\breve{h}_t = \tanh\left(W_h x_t + U_h \left(r_t \odot h_{t-1}\right) + b_h\right) \tag{11}$$

$$\breve{h}_t = \left(1 - z_t\right) \odot h_{t-1} + z_t \odot \breve{h}_t \tag{12}$$

Here z_t are the updated and reset gates, r_t is the candidate hidden state and $\breve{h}_t$ denotes element-wise multiplication.

3.3.5 Bidirectional GRU

By employing Bi-GRU, GRU is improved because the mechanism can read the input sequence forwards and backwards allowing it to consider the context arising from both prior and subsequent states (Naz et al., 2024). For an input sequence $\{x_t\}$, the forward and backward hidden states, $\vec{h}_t$ and $\overleftarrow{h}_t$, are computed separately as in GRU. The final output o_t is obtained by concatenating the two, as in 13:

$$o_t = \left\{\vec{h}_t, \overleftarrow{h}_t\right\} \tag{13}$$

Ensuring the model leverages the bidirectional dependencies within the text, as working defined using Figure 3.

Figure 3 Architecture of proposed model Bi-GRU (see online version for colours)

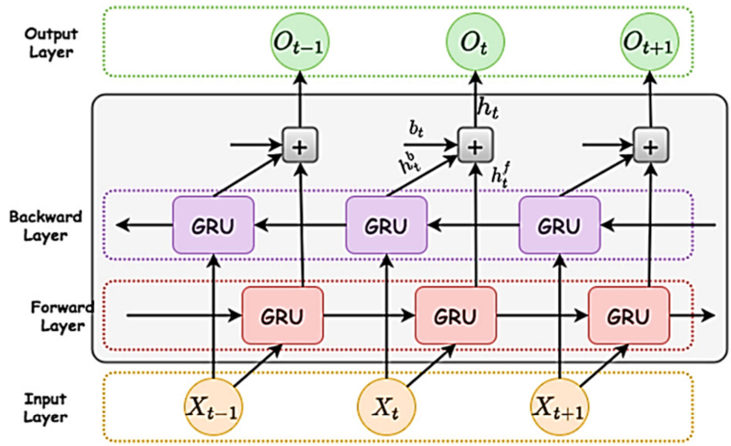


Table 2 Summarising the strength and features of the applied models

Model name	*Nature/working*	*Strength*	*Limitation*
Support vector machine (SVM)	Supervised learning algorithm that finds a hyperplane to separate different classes	Effective in high-dimensional spaces and for text classification tasks	Struggles with large datasets and non-linear relationships
Decision tree (DT)	A tree-like structure that splits data into subsets based on feature values	Easy to interpret and implement; handles both categorical and continuous data	Prone to overfitting and lacks the ability to capture complex patterns
Random forest (RF)	Ensemble method using multiple decision trees to improve accuracy and reduce overfitting	Robust against overfitting; performs well with large datasets	Requires significant computational resources for large datasets
Gated recurrent unit (GRU)	A type of recurrent neural network that focuses on maintaining long-term dependencies in data	Efficient for sequential data and reduces the vanishing gradient problem.	Limited ability to capture complex patterns in long sequences
Bidirectional GRU (Bi-GRU)	A variant of GRU that processes data in both forward and backward directions	Captures richer context by considering past and future information	Computationally more expensive than traditional GRU models

Advanced DL architecture beyond GRU is Bi-GRU, which takes the output data in both the forward and backward directions. By doing this the model gets to cover contextual dependencies better as it better covers information in past and future cases. Hence, in the proposed method, Bi-GRU model is used to classifying student feedback as it is better in understanding sequential data than other models such as SVM, DT and RF. SVM, DT and RF are techniques commonly applied to classification problems, but temporal relationships in sequential data are not incorporated. On the other hand, Bi-GRU takes advantage of the sequential nature of text by processing the data in both directions, but herein it can extract richer contextual information to achieve higher accuracy in sentiment and feedback classification. Because of its capacity to learn long-range dependencies in text, Bi-GRU outperforms the traditional models in a subtle manner that provides great benefit when analysing complex, context dependent feedback. The Table 2 provides a concise comparison of the strengths and limitations of the models applied in the study, helping to highlight the advantages of each approach.

4 Experimental setup

This experimental study selects dataset, processes dataset, implements model and evaluates the model based on performance. We have collected student feedback from different college courses spanning numerous perspectives of teaching methods and the extent of the covered course content as well as overall experience with learning the course. For data preprocessing, the text cleaning, normalisation, tokenisation and embedding of FastText are applied to enhance the feature representation. To implement models, Bi-Directional formats like BiLSTM and architectures based on Transformer are used to capture the contextual dependencies effectively. Standard metrics are used to evaluate the performance of these models on accuracy, precision, recall, and F1-score to determine their efficacy in categorising student feedback as positive or negative.

4.1 Dataset collection

The data used in this study is obtained from online repository used on teaching feedback, data created with the purpose of capturing sentiments and opinions about the different aspects of their teaching process in colleges. It is comprised of tuples as data objects containing textual reports along with related sentiment or quality descriptors. To achieve diverse opinions, the dataset includes an equal distribution of feedback categories as in ‘awesome’, ‘awful’, ‘good’, ‘poor’, ‘average’, as distribution shown in Table 3. Specifically, the textual content of each record of the dataset often includes teacher responses that are wordier or more concise than others. Inclusion of human errors such as typo, use of slang, wrongly spelt words, grammatical errors, and inconsistent use of punctuations in the text data of the dataset makes the data suitable for identifying the real-world strength of NLP models. Due to its nature of textual data and the variety of feedback categories, this dataset can be considered as a reference one for examining the performance of AI-based methods for sentiment analysis in the educational environment.

Table 3 Labels summary statistics

Class	*Text count*
Awesome	629
Awful	489
Good	449
Poor	397
Average	381

4.2 Performance evaluation measures

To assess the models' level of training, the following metrics were selected, as mentioned in table, as major evaluation parameters to be used with ML and DL models, defined in Table 4. Accuracy was applied to estimate the general accuracy of the model and was defined as the ratio of the number of instances correctly predicted out of all instances the model attempted to predict. Accuracy was the concentration on how many of the envisaged positive feedback is indeed positive, giving credence to how the model was pragmatic in minimising on false positive feedback. On the other hand, Recall evaluated how many of the actual positive feedback were correctly identified by the model; thereby reducing the problem of false negatives (Naz et al., 2024). To increase the balance between precision and recall measures, F-Score was computed, which suits cases where the datasets are rather unbalanced. These metric checks that finesse maxima are low in a balanced manner which reduces both false positives and false negatives. Lastly, the area under the receiver operating characteristic curve (AUC-ROC) gave a stable measure of the models' capacity to classify various classes of feedback based on threshold separations with larger AUC values indicating better divided classes. The actual performance metrics provide insights about how accurately the models evaluate the teacher feedback on a scale such as 'awesome,' 'awful,' 'good,' 'poor,' and 'average.'

Table 4 Computation of evaluation measures

Measure	*Equation*
Accuracy	$\frac{TP+TN}{TF+FN+FP+TP}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	$\frac{2(Precision * Recall)}{Precision+Recall}$
AUC-ROC	$\sum_{i=1}^{n=1} \frac{(FP_1+FP_{i=1})(TP_i+TP_{i=1})}{2}$

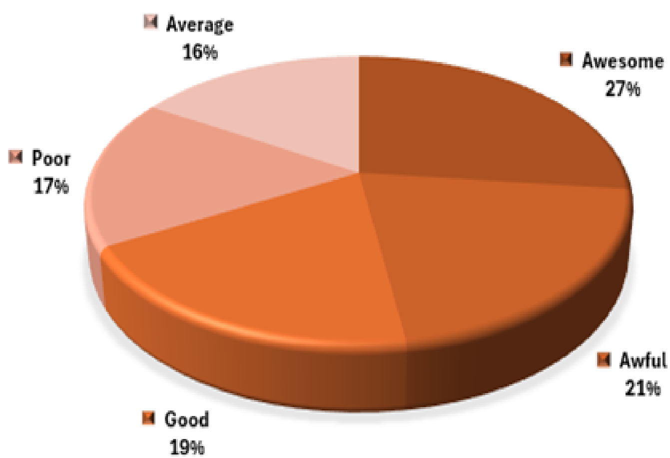
5 Results and discussion

In this section, the evaluation of the proposed approach is presented as the performance of Bi-Directional models with FastText embeddings for classifying student feedback is evaluated. Different performance metrics are used to measure the performance of the models and assess the results based on accuracy, precision, recall and F1 score. As a counterpart, a comparison with the classical ML is also made to emphasize gains in the contextual understanding and classification accuracy. Additionally, the discussion interprets key findings and subsequently discusses potential challenges and the implications of using the DL models for feedback analysis in the college education.

5.1 Visualisation analysis of dataset

Exploratory data analysis gives information on the attributes of the dataset. The EDA offers an overview of the dataset by depicting objective observations of the features. The Word Count Distribution in Figure 4 depicts skewness in the amount of feedback expressed by assembled responses. Simple feedback texts predominantly range from 0–120 words and there is a right-skewed distribution of the number of words in the texts. This means that most of the teaching-based feedback gives somewhat elaborate answers, while some give brief or quite extensive feedback. The density curve is strained over the histogram for clearer revelation of this trend. These analyses lay a good framework for further modelling and feature extraction-oriented work. The aim of the distribution of feedback categories in Figure 5 is to express the amount of feedback with a particular quality label. A similar distribution of words is observed, but the most representative one is the 'awesome' category which implies that overall sentiment of the dataset is positive in Figure 6. The distributions of other categories like 'awful', 'good' 'poor' and 'average' are almost equal but these are less often used than 'awesome'. These analyses form a good basis for further modelling and feature extraction tasks.

Figure 4 Distribution percentage of each class (see online version for colours)



The word clouds present in Figure 7, the main terms of the teacher's feedback and reflect the main topics in relation to different quality levels. The general word cloud shows that practically all the feedback relates to the key words as 'class', 'lecture', 'test', 'professor'. These words mean that teachers often pay attention to issues concerning assignments, assessments, and instructional effectiveness. This gives an indication that the feedback is mainly anchored on academic performance. Such suggestions give a good starting point in building informed approaches to evaluating teacher intentions and feelings with the aim of creating relevant initiatives or improvements in the academic environment.

- Awful: there are seven positive terms: 'good', 'nice', 'well', 'better', 'like', 'best', and 'lest', and eleven negative terms: 'hard', 'funky', 'never', 'nowhere', 'nearly', 'nothing' 'not', 'no', 'never', 'nought', and 'noplace'. Feedback in this category shows with regards to difficulties experienced or inefficient instructional or formative practices.

- Poor: basic words related to study, such as 'test,' 'take,' which are not positive, at least being less rated as the words from the top-rated categories.
- Average: this category combines on-favour and off-favour terms: while there is 'class,' 'test' is a negative term; ''average' is a neutral descriptor that indicates that teachers evaluate experience mediocrity.
- Good: there are some positive terms such as 'really' and the word 'help,' which may be positive remarks made to encourage the recipients to take more positive views about the elements of the class or the teaching style.
- Awesome: the 'awesome' category includes 'easy', 'great' and 'recommend' supporting pleasure of the class or professor.

Figure 5 Distribution of feedback word count analysis (see online version for colours)

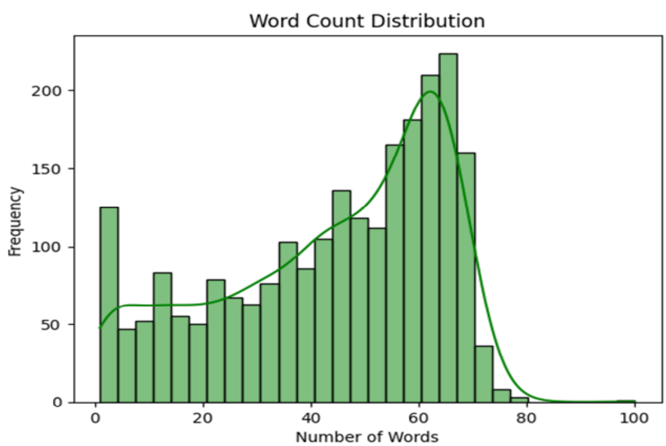


Figure 6 Distribution of labels in the dataset (see online version for colours)

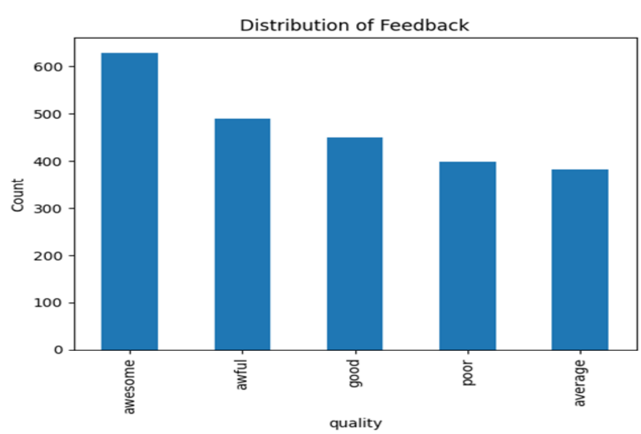


Figure 7 Word cloud of each class label (see online version for colours)

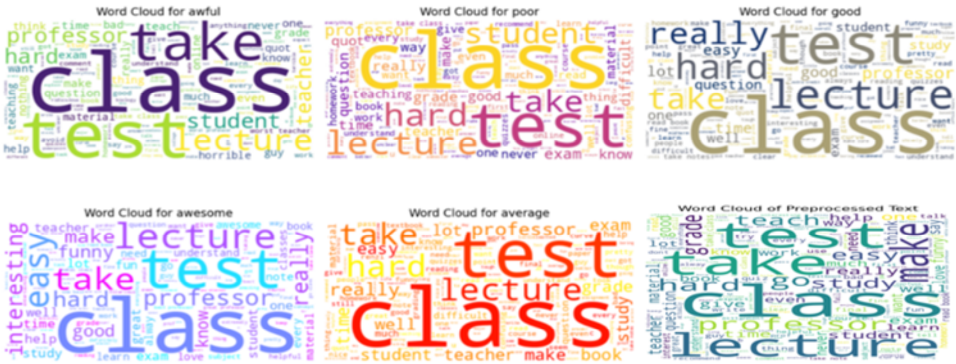
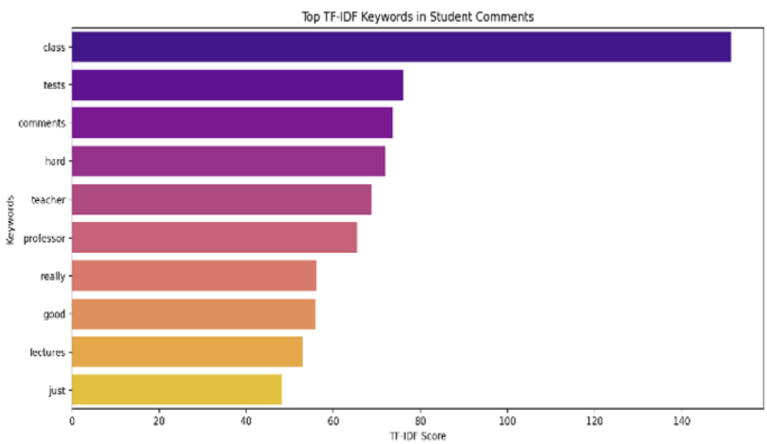


Figure 8 Analysis of TF-IDF based keyword (see online version for colours)



5.2 *Results with machine and ensemble learning models*

The proposed classification, in terms of accuracy for both SVM and DT models shows a decent performance in categorising teachers' feedback according to the features provided (TF-IDF). Out of the relative classifiers, SVM was most accurate with 89% overall accuracy, moderate precision of 78%, but better recall of 89%, resulting in fair F1-score of 89%. This highlights SVM's robustness in handling text-based feedback reviews classification tasks, especially with well-processed TF-IDF features. The DT model achieved an accuracy of 77% and an F1-score of 76. On the other hand, it marginally performed less than SVM although a difference of 0.06 is inconsequential; this means that DT might be more sensitive to over-learning or perhaps less flexible to model the TF-IDF based features. The RF model of the ensemble learning method, as an additional model, possessed a lower accuracy of 74% and F1-score of 67% than those of the SVM and DT. This may explain the performance gap in which RF has been unable to derive accurate insights because of its inadequacy in modelling the feature space of TF-IDF textual data. Nonetheless, a high level of accuracy of 68% shows that RF can provide good predictions for some classes although the problem with generalisation for all labels remains. Therefore, the overall best ML model is SVM, and based on the results, the TF-IDF feature extraction performed well in encoding the text features of the collection, as shown in Figure 8. In contrast, the RF which is generally very resistant was the least accurate for this dataset.

5.3 Results with deep learning algorithms

The proposed DL models, GRU and Bi-GRU demonstrated higher performance than traditional ML techniques since they can model sequential and contextual information in the text. FastText embeddings were used for both models and provided many dense vectors, thus improving the models. The most important hyperparameters that define the model are as follows: embedding dimension – the size of the word embedding; the number of GRU units and the number of layers – the most important factors in terms of model's ability to learn the problematic patterns. To address the problem of overfitting dropout and recurrent dropout rates are used when some neurons are deactivated during the training phase. There are also optimisers like Adam which combined with proper learning rate allow to update the weights of the model effectively. However, some phenomenon such as batch size, sequence length and type of the loss function for the model such as categorical or binary cross-entropy have very crucial impact on the learning process of the model. The inclusion of bidirectionality can enable the model to capture the past and future contexts than when it is used to analyse the feedback given alone since it will enable accurate sentiment predictions. The Bi-GRU model presented the highest overall performance with an accuracy of 95%, precision of 87%, recall of 91%, F1-score of 90. This proves that Bi-GRU can capture the sequential nature of textual feedback as the architecture incorporates bidirectional processing which captures forward and backward contextual dependencies, as results are defined in Table 5. The GRU model had the same meaning accuracy as the SVM, 0.89, and the same F1-score of 0.89. They conclude this by noting that even though it has a minutely higher precision of 78%, GRU made more confident predictions of certain classes than Bi-GRU, but its recall of 79% was less than that of Bi-GRU, as shown in Figures 9 and 10.

Table 5 Analysis of results with applied models and features

Model	*Feature*	*Accuracy*	*Precision*	*Recall*	*F1-score*
		Machine Learning			
SVM	TF-IDF	89	78	89	89
DT		77	76	77	76
		Ensemble learning			
RF	TF-IDF	74	68	74	67
		Deep learning			
GRU	Fast Text	83	78	79	82
Bi-GRU		95	87	91	90

In contrast, DL models outperformed others when dealing with detailed subtleties in the feedback, especially when combining them with pre-trained embeddings such as FastText, as results are shown in Figure 11 based on epochs. Table 6 defines the applied hyperparameter of proposed model, Bi-GRU, as it examines the input sequences in both forward and backward ways. The improved context representation and sequential learning feature of GRU and Bi-GRU make them more effective in the text-based classification than basic ML algorithms.

Figure 9 Accuracy analysis graph over training and validation dataset (see online version for colours)

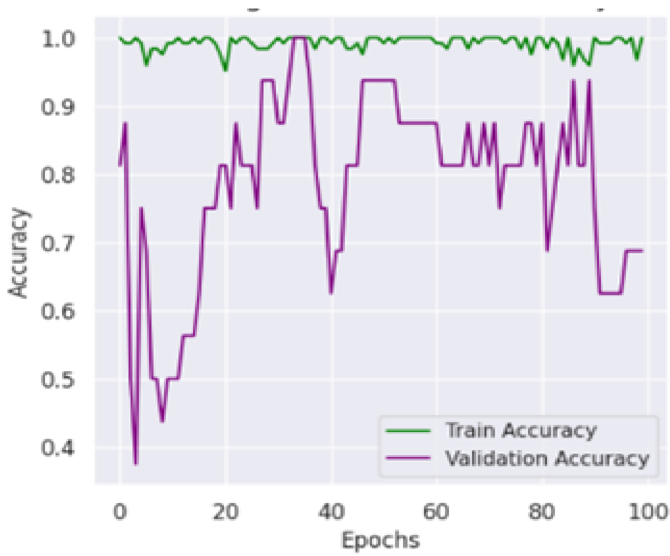


Figure 10 Loss analysis graph over training and validation dataset (see online version for colours)

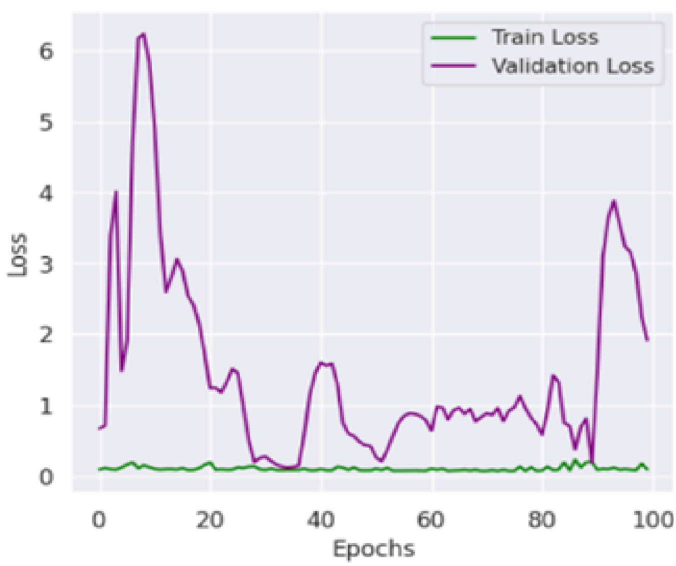


Figure 11 Accuracy analysis of Bi-GRU over epochs (see online version for colours)

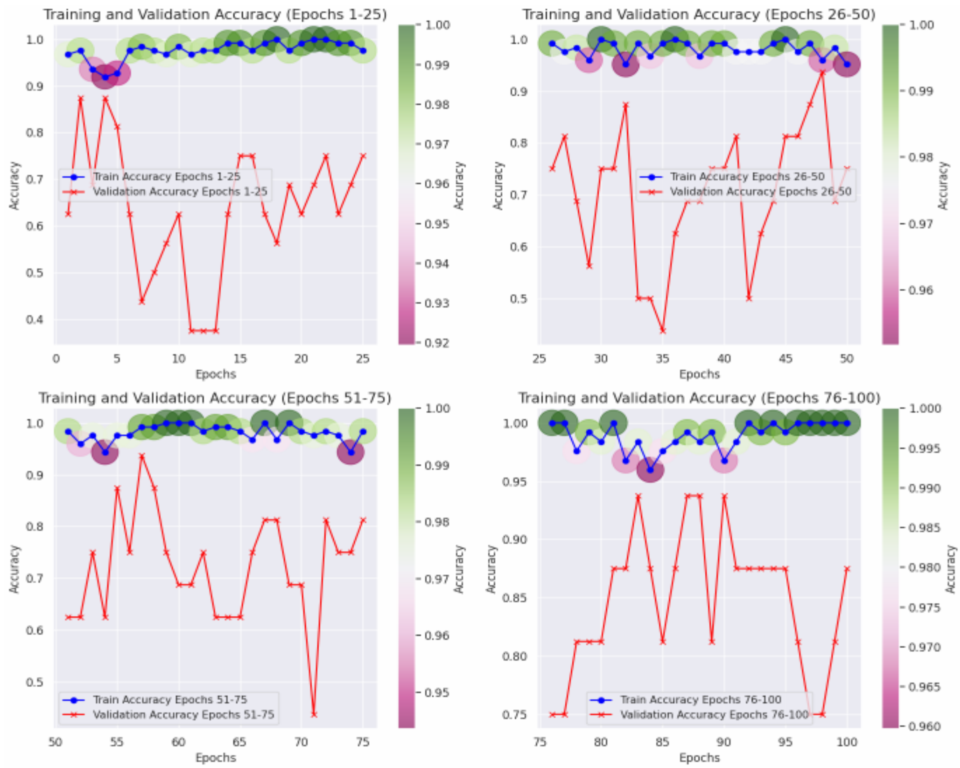
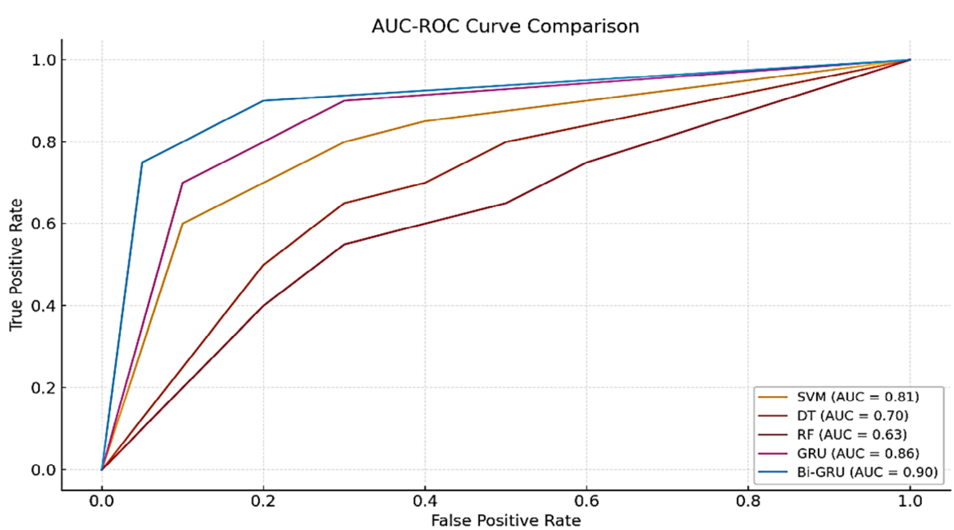


Figure 12 Comparative analysis of ML and DL model using AUC-ROC (see online version for colours)



5.4 Overall observations

When comparing to the other ML approaches, the SVM gave high steady and balanced metrics, but the best model overall was Bi-GRU showing the potential of the DL approach to the text sentiment analysis task. This paper showed that utilisation of pre-trained FastText embeddings helped improve DL models, further proving that good feature representation for text data must be rich. The AUC-ROC curve, as shown in Figure 12, also gives an overall comparison of different models' performance and their capacity in the feedback classification. Out of all the models implemented, Bi-GRU turns out to be the best since it can handle complex features in the text data with an overall AUC of 0.90.

Second, GRU with an AUC of 0.86, which proves the resilience of recurrent structures in understanding temporal relations and context. In the ML field, SVM does surpass many traditional methods; the model does have AUC. 0.8 indicates the ability to handle the text high-dimensional features derived from TF-IDF. Both DT with AUC 0.70 and RF with the minimum AUC 0.63, fit mid-range. This could be due to the weakness inherent in ensemble models like RF to explore complex sparse feature space as developed by the TF-IDF feature extraction process.

Table 6 Hyperparameter setting of proposed models

Hyperparameter	*Description*	*Values*
Embedding dimension	Size of word embeddings	100, 200, 300
GRU units	Number of units in each GRU layer	64, 128, 256
Number of layers	Number of stacked Bi-GRU layers	1, 2, 3
Dropout rate	Fraction of units to drop to prevent overfitting	0.2, 0.3, 0.5
Recurrent dropout	Dropout applied to recurrent connections	0.1, 0.2, 0.3
Batch size	Number of samples per gradient update	32, 64, 128
Learning rate	Step size for the optimiser	0.001, 0.0005, 0.0001
Optimiser	Algorithm for updating weights	Adam
Activation function	Activation for the output layer	Softmax
Sequence length	Length of input text sequences	100, 200, 300
Loss function	Function to minimise during training	Categorical cross entropy
Epochs	Number of times the model sees the entire dataset	100
Bidirectionality	Whether to use bidirectional GRUs	True
Weight initialisation	Method for initialising weights	GlorotUniform
Gradient clipping	Threshold to prevent exploding gradients	1.0, 5.0

The findings revolve around the improvement offered by DL models, especially Bi-GRU and GRU in reducing the mixed nature and complexity of relationships and contexts in the feedback at college levels. This improved performance indicates the appropriateness of these models for such a task, particularly if the sensitivity to the linguistic form and meaning is important. In applying the proposed feedback analysis model in teaching and education at college level, there are alternatives models which reveal a vital performance difference as follows in Table 7. Using student participant data SVM model got 86% accuracy and using student review data LSTM networks got 80% accuracy. Furthermore, in the self-developed set of data applicable to the ELMo model, efficiency measured was 92%. Still, the sentiment analysis Bi-GRU model, proposed in this work, achieved much a higher accuracy of 95% outperforming all the mentioned models.

Table 7 Comparison with existing studies

Ref	*Models*	*Dataset*	*Results (Acc%)*
Nyandwi et al. (2023)	SVM	Student participants	86
Reddy et al. (2022)	LSTM	Students review	80
Bernius et al. (2022)	Elmo	Self-created	92
Proposed	Bi-GRU	Sentiment analysis	95

Hence this shows a significant improvement, and we argue that the Bi-GRU model is better at capturing contextual information and sequential dependence on feedback data that in the previous models fail to handle expression of sentiments. The performance difference implies that previous models could have had a limited capability and bidirectional analysis to process feedback, while the proposed Bi-GRU model achieves better accuracy and resilience in understanding educational feedback.

6 Conclusions and future work

In the field of education, teaching-based feedback plays a pivotal role in assessing teaching quality and improving learning experiences. To assess this accurately, recent development in AI, especially NLP, offer powerful tools for analysing and interpreting feedback on a scale, providing valuable insights to ensure continuous improvement and balance in the education system. AI is equally applied in the education sector to accustom feedback evaluation while enriching the decision-making process and advancing innovations, and individual learning. As stated in the current study, developing models such as Bi-GRU yields a higher accuracy rate of 95% when could with state-of-the-art fast text word embeddings demonstrating the ability of DL in identifying complex feedback patterns. As education continues to evolve, incorporating AI has a high future potential for developing innovative learning systems that could meet individual requirements and mark great progress in learning process technologies and results. Subsequent studies can extend the nature of this work in applying and developing applications of feedback analysis and other AI approaches and tools of a data-driven, teacher-centric perspective in learning.

Declarations

The author declares that he has no conflicts of interest.

Acknowledgements

The Social Science Foundation of Jiangsu Province in 2024, Project Number: 24ZHB011.

References

Ahmad, W., Khan, H.U., Iqbal, T. and Iqbal, S. (2023) 'Attention-based multi-channel gated recurrent neural networks: a novel feature-centric approach for aspect-based sentiment classification', *IEEE Access*, Vol. 11, No. 2, pp.54408–54427.

Ahmed, M., Khan, H.U., Iqbal, S. and Althebyan, Q. (2022) 'Automated question answering based on improved TF-IDF and cosine similarity', in *2022 Ninth International Conference on Social Networks Analysis, Management and Security* (*SNAMS*), , November, pp.1–6, IEEE.

Ahmed, M., Khan, H.U., Khan, M.A., Tariq, U. and Kadry, S. (2023) 'Context-aware answer selection in community question answering exploiting spatial temporal bidirectional long short-term memory', *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 22, No. 9, pp.229–244.

Alarfaj, F.K., Malik, I., Khan, H.U., Almusallam, N., Ramzan, M. and Ahmed, M. (2022) 'Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms', *IEEE Access*, Vol. 10, No. 4, pp.39700–39715.

Alsini, R., Naz, A., Khan, H. U., Bukhari, A., Daud, A. and Ramzan, M. (2024) 'Using deep learning and word embeddings for predicting human agreeableness behavior', *Scientific Reports*, Vol. 14, No. 1, p.29875, https://doi.org/10.1038/s41598-024-81506-8.

Avila, C., Baldiris, S., Fabregat, R. and Graf, S. (2020) 'Evaluation of a learning analytics tool for supporting teachers in the creation and evaluation of accessible and quality open educational resources', *British Journal of Educational Technology*, Vol. 51, No. 4, pp.1019–1038.

Bardach, L. and Klassen, R.M. (2020) 'Smart teachers, successful students? A systematic review of the literature on teachers' cognitive abilities and teacher effectiveness', *Educational Research Review*, Vol. 30, No. 8, p.100312.

Basiri, M.E., Nemati, S., Abdar, M., Cambria, E. and Acharya, U.R. (2021) 'ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis', *Future Generation Computer Systems*, Vol. 115, No. 6, pp.279–294.

Bernius, J.P., Krusche, S. and Bruegge, B. (2022) 'Machine learning based feedback on textual student answers in large courses', *Computers and Education: Artificial Intelligence*, Vol. 3, No. 10, p.100081.

Bhowmik, A., Noor, N.M., Miah, M.S.U. and Karmaker, D. (2023) 'Aspect-based sentiment analysis model for evaluating teachers' performance from students' feedback', *AIUB Journal of Science and Engineering* (*AJSE*), Vol. 22, No. 3, pp.287–294.

Chen, W., Zan, F., Liu, X. and Meng, X. (2024) 'A prediction method for evaluating the quality of Chinese teacher education based on CNN-LSTM', in *Proceedings of the 2024 6th International Conference on Big-data Service and Intelligent Computation*, May, pp.92–98.

Dann, C., O'Neill, S., Getenet, S., Chakraborty, S., Saleh, K. and Yu, K. (2024) 'Improving teaching and learning in higher education through machine learning: proof of concept' of AI's ability to assess the use of key microskills', *Education Sciences*, Vol. 14, No. 8, p.886.

Hemmat, R.A., Pezeshki, M., Bordes, F., Drozdzal, M. and Romero-Soriano, A. (2023) *Feedback-Guided Data Synthesis for Imbalanced Classification*, arXiv preprint arXiv:2310.00158.

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., ... and Stachl, C. (2021) 'Machine learning for the educational sciences', *Review of Education*, Vol. 9, No. 3, p.e3310.

Kastrati, Z., Dalipi, F., Imran, A.S., Pireva Nuci, K. and Wani, M.A. (2021) 'Sentiment analysis of students' feedback with NLP and deep learning: a systematic map-ping study', *Applied Sciences*, Vol. 11, No. 9, p.3986.

Mailool, J., Kartowagiran, B., Retnowati, T.H., Wening, S. and Putranta, H. (2020) 'The effects of principal's decision-making, organizational commitment and school climate on teacher performance in vocational high school based on teacher perceptions', *European Journal of Educational Research*, Vol. 9, No. 4, pp.1675–1687.

Mao, W. (2022) 'Video analysis of intelligent teaching based on machine learning and virtual reality technology', *Neural Computing and Applications*, Vol. 34, No. 9, pp.6603–6614.

Mariani, M.M., Perez-Vega, R., and Wirtz, J. (2022) 'AI in marketing, consumer re-search and psychology: a systematic literature review and research agenda', *Psychology and Marketing*, Vol. 39, No. 4, pp.755–776.

Motevalli, M.M., Sohrabi, M.K. and Yaghmaee, F. (2025) 'Aspect-based sentiment analysis: a dual-task learning architecture using imbalanced maximized-area under the curve proximate support vector machine and reinforcement learning', *Information Sciences*, Vol. 689, No. 21, p.121449.

Munna, A.S. and Kalam, M.A. (2021) 'Teaching and learning process to enhance teaching effectiveness: a literature review', *International Journal of Humanities and Innovation* (*IJHI*), Vol. 4, No. 1, pp.1–4.

Naz, A., Khan, H.U., Alesawi, S., Abouola, O.I., Daud, A. and Ramzan, M. (2024) 'AI knows you: deep learning model for prediction of extroversion personality trait', in *IEEE Access*, Vol. 12, pp.159152–159175, DOI: 10.1109/ACCESS.2024.3486578.

Nyandwi, V., Habimana, O. and Enan, N.M. (2023) 'Ensemble machine learning-based sentiment analysis model for teachers' performance evaluation', *International Journal of Advances in Engineering and Management* (*IJAEM*), Vol. 5, No. 4, pp.1220–1233.

Onan, A. (2020) 'Mining opinions from instructor evaluation reviews: a deep learning approach', *Computer Applications in Engineering Education*, Vol. 28, No. 1, pp.117–138.

Reddy, S.S., Gadiraju, M. and Maheswara Rao, V.V.R. (2022) 'Analyzing student reviews on teacher performance using long short-term memory', in *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA 2021*, pp.539–553, Springer Nature Singapore, Singapore.

Saraswat, D., Bhattacharya, P., Verma, A., Prasad, V.K., Tanwar, S., Sharma, G., ... and Sharma, R. (2022) 'Explainable AI for healthcare 5.0: opportunities and challenges', *IEEE Access*, Vol. 10, No. 6, pp.84486–84517.

Schles, R.A. and Robertson, R.E. (2019) 'The role of performance feedback and implementation of evidence-based practices for preservice special education teachers and student outcomes: a review of the literature', *Teacher Education and Special Education*, Vol. 42, No. 1, pp.36–48.

Shaik, T., Tao, X., Li, Y., Dann, C., McDonald, J., Redmond, P. and Galligan, L. (2022) 'A review of the trends and challenges in adopting natural language processing methods for education feedback analysis', *IEEE Access*, Vol. 10, No. 8, pp.56720–56739.

Song, C., Shin, S.Y. and Shin, K.S. (2024) 'Implementing the dynamic feedback-driven learning optimization framework: a machine learning approach to personalize educational pathways', *Applied Sciences*, Vol. 14, No. 2, p.916.

Spatiotis, N., Perikos, I., Mporas, I. and Paraskevas, M. (2020) 'Sentiment analysis of teachers using social information in educational platform environments', *International Journal on Artificial Intelligence Tools*, Vol. 29, No. 2, p.2040004.

Topping, K.J., Gehringer, E., Khosravi, H., Gudipati, S., Jadhav, K. and Susarla, S. (2025) 'Enhancing peer assessment with artificial intelligence', *International Journal of Educational Technology in Higher Education*, Vol. 22, No. 1, p.3.

Urooj, A., Khan, H.U., Iqbal, S. and Alghobiri, M. (2023) 'Exploring author, article, and venue feature sets for rising star prediction in academic network', *Journal of Scholarly Publishing*, Vol. 54, No. 3, pp.445–473.

Wang, Y., He, P., Geng, J. and Zhu, Z. (2025) 'An innovative teaching practice in analytical chemistry laboratory: merging photoelectric technology with inquiry-based learning through cerium analysis', *Journal of Chemical Education*, Vol. 102, No. 3, pp.1245–1251.

Wu, X. (2021) 'Application of artificial intelligence in modern vocational education technology', in *Journal of Physics: Conference Series*, IOP Publishing, April, Vol. 1881, No. 3, p.32074.

Yao, L. (2024) 'ADHN: sentiment analysis of reviews for MOOCs of dilated convolution neural network and hierarchical attention network based on AL-BERT', *IEEE Access*, Vol. 12, No. 7, pp.98352–98366.