Nowadays, a lot of data like raw facts, reviews, or opinions has stored on social media or e-services websites. In recent years, people have preferred to purchase products in online, so, the customers choose the best products and a lot of information is gathered in the form of customer feedback. Customer feedback helps with product development and boosts sales of the product.Reviews posted by users on e-commerce websites offer useful details about the product.Sentiment analysis on the text review helps to analysing the user sentiment toward product and predict sales of the product.The Long Short Term Memory (LSTM) and Convolutional Neural Network (CNN)-based methods utilized in current methodologies for sentiment analysis involve the vanishing gradient problem and overfitting problem, respectively. For the purpose ofenhance classification execution, the Word Embedding Attention (WEA) strategy is introduced in the Bi-directional Long Short Term Memory (Bi-LSTM) model. From the input dataset for sentiment analysis, the CNN model is used to retrieve the features. To preserve the gradient in the network and resolve the vanishing gradient problem in the network, the balanced cross-entropy is presented.The WEA technique gives terms with a strong connection to class higher weight, this method helps to improve the model's execution with regards to classes, which increases an accuracy and recall value.The CNN function helps to improved performance with less training data.While the existing CNN model has 97.1% accuracy and 85.4% precision in sentiment analysis, the WEA-BiLSTM model has 97.4% accuracy and 86.8% precision respectively.

4.1 Introduction

The COVID-19 pandemic and recent developments in internet technologies have advanced digital transformation and increased usage of social media and e-commerce platforms. Online consumer reviews are a common and expanding form of User-Generated Content (UGC) and are maintained by a number of review websites, including Amazon, Yelp, and TripAdvisor. Before making a purchase, customer’s check the feedback to evaluate the high standard of an organization, item, or provision. Online customer reviews generally have a significant influence on many prospective consumers' shopping decisions [76]. Due to the fact that consumers prefer to buy things online, a lot of information is gathered in the form of customer feedback to help potential buyers select the best items. These studies provide optional languages that helps the e-commerce industry identify areas that require development.

The information is very useful for businesses looking to comprehend customer feedback on their goods or services. The emotional analysis data is really useful for companies seeking to comprehend customer feedback on their goods or services, as well as experience [77]. Sentiment analysis is the technique of identifying the intended sentiments that are expressed in text-based resources like social media postings, product reviews, and online communities. The majoraim of sentiment analysis is to make it easy for machine learning-based algorithms to recognise these intended emotions when they are conveyed in textual materials. The application of sentiment analysis has performed highly effective in a variety of fields. Similar to this, companies typically use the sentiments collected from client reviews and comments to polish and enhance the USPs of their goods [78].

One of the most essential concepts in Natural Language Processing (NLP) and text mining for sentiment analysis is text categorization. Associating pertinent material with labels that already exist is the description of the text classification problems. In this rank of labelling, set of data form is important. Each text can be illustrated by a single label or several labels, based on the issue's current state. Stock trading firms also collect information and assess their view of various facts using sentiment analysis technologies [79]. Using sentiment analysis, characteristics of software products may be extracted from user evaluations together with a summary of what users think of each feature. Product managers could find this information quite helpful in prioritising their work for upcoming releases. It is also possible to utilise the user assessment at the feature level to sporadically check on the features and general health of the product. Review classification therefore emerges as the fundamental tool for textual data organisation [80]. Based on user feedback, product characteristics may be categorised into many groups. The quick advancement of machine learning and artificial intelligence opens the door to the computer analysis of customer evaluations for opinion mining. Customer review data offers various benefits, starting with the fact that there are lot of reviews that are publicly available online. Second, there are several extremely useful qualities that may be used to characterise customer review data. Third, customer reviews are raw, open data that are collected from regular people and made available to all manufacturers [81]. In essence, consumer feed back display how the customer feels about the enterprise that is important to understanding what the consumer thoughts. Theseestimation significantly influence other customers' choices and serve as the cornerstone of business development [82] without any client feedback on their goods or services. Customers can offer feedback by way of an online survey, a social media website, or handwritten evaluations. There are a tonne of thoughts and evaluations available, but it might be challenging to classify them as favourable or unfavourable. Customer satisfaction is crucial to the hotel industry's ability to deliver higher-quality services and improve customer relationships [83]. However, the fast growth of online purchasing has resulting in an excess of product reviews, making it challenging for customers to choose the most useful ones. Therefore, it becomes crucial for platforms to effectively identify useful reviews in order to help consumers make intelligent purchasing decisions. As a result, large reviews sometimes include a lot of pointless or boring material, which lowers the categorization performance of review length. [84]. As a result, the purpose of this study is to determine whether customer feedback from applications and online reviews is related. Customer sentiment is the term used to describe the sentiments that consumers express through text reviews. These feelings can be either good, negative, or neutral. The research investigates consumer feelings and describes them in terms of the polarity of those ratings [85].

4.2 Problem Statement

- The LSTM model has vanishing gradient problem and HNN model has overfitting problem in classification.
- The fusion of features creates overfitting technique and irrelevant features affects the classification performance.
- The learning capacity of the model was less and the feature selection process was not effective.

4.3 Contributions

- The WEA technique is proposed to provide higher weight value to the words having strong relation to the classes. This helps to increases the learning rate of the method and expands the class wise execution of the method.
- The balanced cross entropy is applied to maintain the gradient in the network to reduce the vanishing gradient problem. The gradient is maintained in the network based on the loss calculation in the network.

- The CNN based feature extraction technique is applied to generate more features in the convolutional layer and helps the model to provide higher performance for less number of training data.
- The WEA-BiLSTM model has higher performance in sentiment analysis compared to existing techniques. The WEA-BiLSTM model solves vanishing gradient problem and reduces the overfitting problem in the classification.

4.4 Proposed Method

The CNN model is used to retrieve features from the input dataset of Amazon reviews. The Bi-LSTM model's attention layer applies weight values to the features in order to classify sentiment. The network's gradient is maintained by applying the balanced cross entropy. The WEA-BiLSTM model's process for categorising sentiment is shown in Figure 4.1.

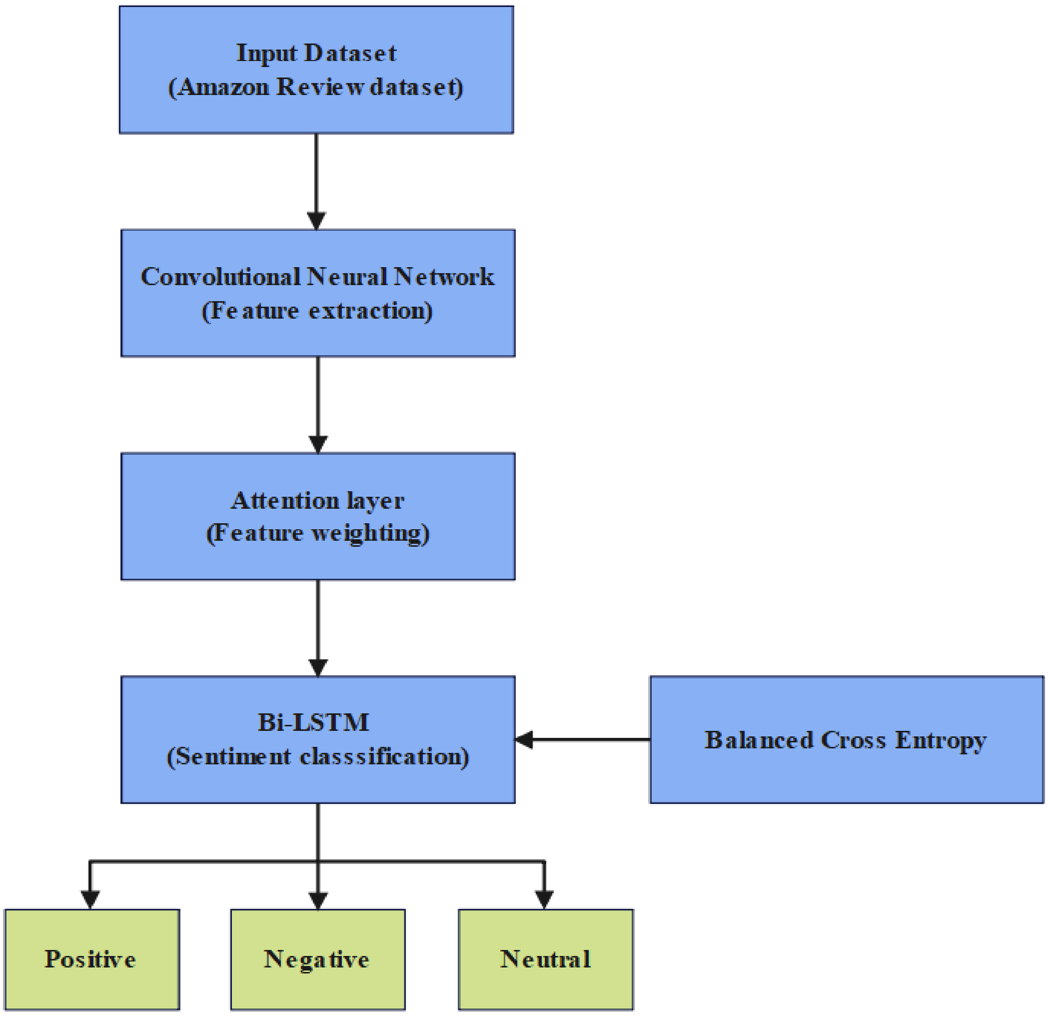

Fig. 4.1 :The flow of WEA-Bi-LSTM model for sentiment classification

4.4.1 Convolutional Neural Network

Humans find the process of understanding pictures intriguing, and this is a straightforward task for them. The machine that interprets images has many hidden intricacies. The CNN model is a deep learning system that seeks to mimic animal visual processing and is motivated by the visual cortex of the brain [86]. In the area of image processing, which includes detection, localization, segmentation, and classification, among other things, CNN's represents a giant step forward. The model is widely used primarily due to the great CNN efficiency in image categorization. The CNN model uses a convolutional layer with trainable weights and biases that are modelled after animal neurons. As demonstrated in Figure 4.2,Convolutional layers, fully linked layers, and activation functions are the core building elements of the CNN model. This paper provides a succinct explanation of the CNN concept, and the research includes a thorough description of CNN. [87, 88]

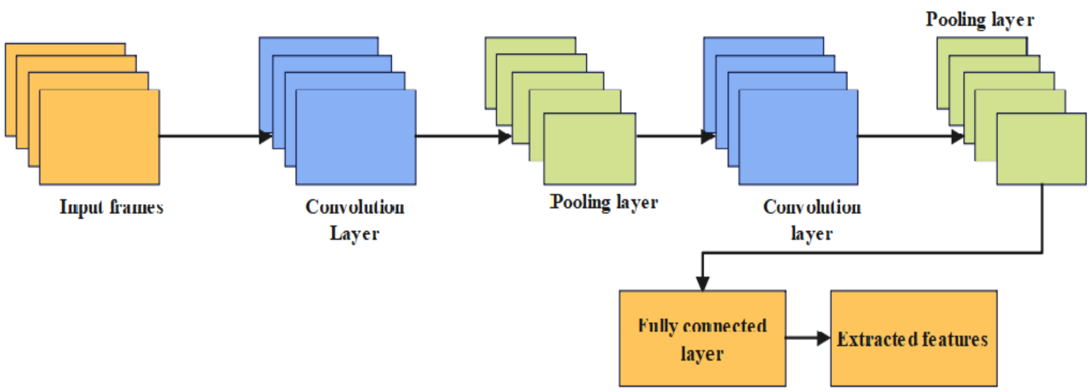

Fig. 4.2 :CNN architecture model for feature extraction

4.4.1.1 Convolutional Layer:

The visual cortex's neuronal cells are involved in extracting characteristics from pictures in animal brains. Each neural cell extracts different characteristics that aid in the comprehension of a picture. Convolutional layers are used to simulate neuronal cells, and this allows for the extraction of properties including gradient direction, texture, colours, and edges. In convolutional layers and size, convolutional filter or kernels are learning filters is $n \times m \times d$, where image depth is d.During the forward pass, the Kernels are twisted over the input volume's height and width, andInput and filter entries are generated using the dot product. CNN learns which filters to use for texture, colour, edge, etc. An activation function layer is applied using the convolution layer's output.

4.4.1.2 Activation Function:

Since most real-world data is non-linear,Utilizing non-linear data transformation's activation functions. This guarantees that input space representation is translated to various output spaces in accordance with requirements.

This requires real-value number x and converting it into a range of 0 and 1. Large positive and negative inputs in particular are placed near 0 and unity, respectively. This is expressed in equation (4.1).

$$f(x) = \frac{1}{1+e^{-x}} \tag{4.1}$$

A number with real value x is taken into account in non-linear functions and converters x to 0 if x is negative. A common non-linear function that takes less time to compute and is better than the sigmoid is the ReLU activation function and tan h functions, as shown in Equation (4.2).

$$f(x) = \max(0, x) \tag{4.2}$$

4.4.1.3 Pooling

The pooling layer performs a non-linear down sampling's converged feature. As a result, dimensional reduction data processing requires less computer resources. To lower the spatial size, data are aggregated based on feature type or space; rotational variation in pictures overcomes translation; and overfitting is controlled. A rectangle patch set is created when inputs are partitioned using a pooling method. Based on the kind of pooling procedure, a single value is computed to replace each patch. Maximum pooling and average pooling are the two pooling types that are most often utilised.

4.4.1.4 Fully Connected Layer:

According to an artificial neural network, the inputs are linked to each node in the preceding layer, and weight values are given to each node this is known as a fully connected layer. The model's output is the total of the inputs multiplied by the appropriate weights. A completely linked layer is coupled to the sigmoid activation function to carry out the classification task. In Figure 4.3, can see CNN's completely linked layer.

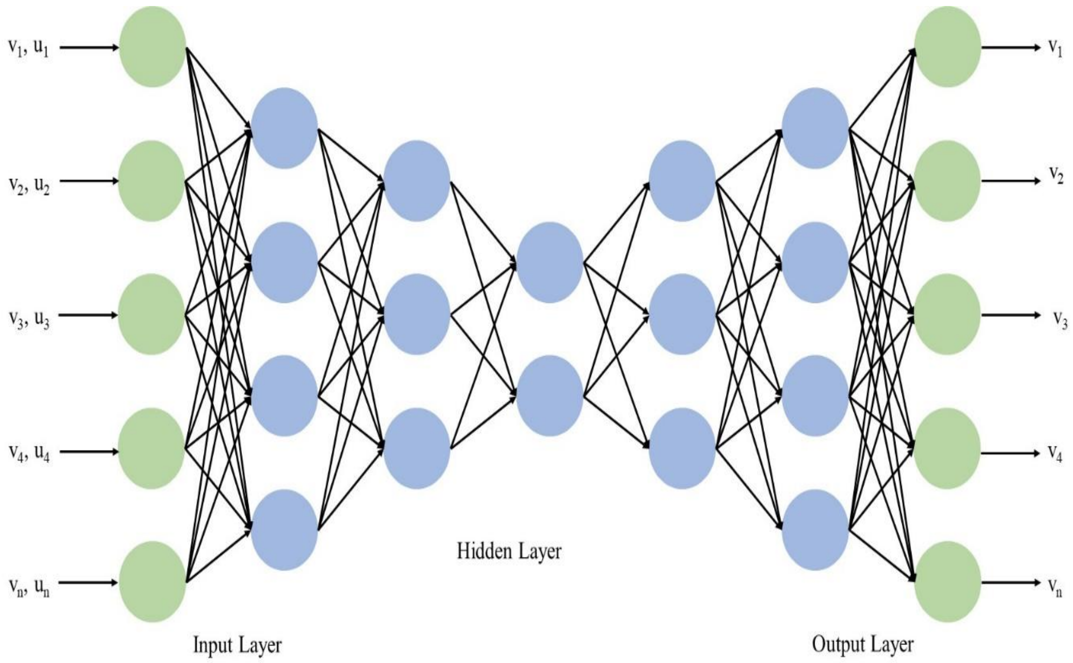

Fig. 4.3 :CNN model's fully connected layer

4.4.2 Attention Layer

Third region in the weighted word representation layer, CNN networks are used to utilise textual information. Each tweet is used as the basis for applying the word representation to a convolutional layer. The F_n Convolutional word vector matrix is computed using weighted matrix and is defined as $w \in R^{t \times m}$ a word vector is widely used to quantify local and fundamental features t is selected in F_n matrix, as in equation (4.3)

$$h_i = f(V_{i:j+t-1} \times W[i] + b_i) \qquad (4.3)$$

Word vector when t design a feature map of $h \in R_{n-t+1}$, bias is b, The weight of the matrix is expressed as W, and non-linearity activation function is represented as f. Convolutional layers are used to create feature maps, while the max-pooling layer is utilized to lower the dataset's dimension and extract essential components, as indicated in equation (4.4).

$$p_i = Max[h_i] \qquad (4.4)$$

where the feature map's max-pooling layer was applied to indicate $P_i \in R^{n-t+\frac{1}{2}}$.Maxpooling is used to extract equivalent essential components from the characteristics of different CNN layer filters without focusing on polarity or semantic relevance. An attention layer is used to give each feature on CNN-generated features value was shown in the equation (4.5).

$$A_i = \frac{\exp(p_i)}{\sum_i \exp(p_i)} \tag{4.5}$$

The above equation provides the max-pooling attention calculation and produces an attention score A_i is employed in every feature contextp_i.

To understand the feature context, the Bi-LSTM use the attention score output. Using a sequential map, the Bi-LSTM model creates the final features. The final feature context p_i is utilized for CNN's greatest feature map, and attention scores are shown as A_i.

4.4.3 Bi-LSTM Model

Due to its higher performance on sequences of data, the LSTM model is commonly employed in sentiment prediction. The input x_k at the current time step k and concealed layer result is denoted as h_{k-1} as earlier time step of concealed layer output. The LSTM cell state c_k and LSTM unit network architecture is shown in Figure 4.4.Since the gradient is propagated by the cell state, LSTM preserves the pertinent characteristics for a longer period of time than an RNN model.

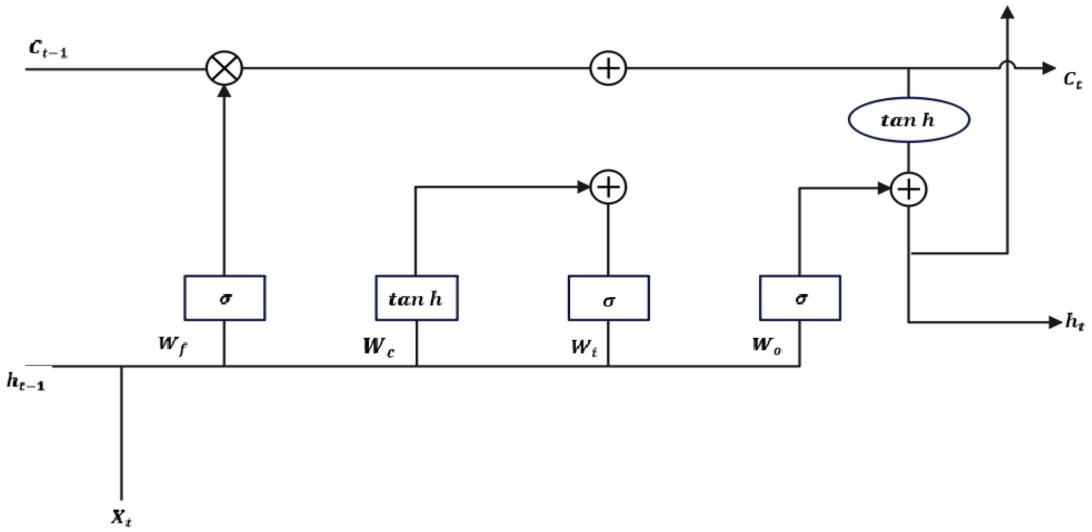

Fig. 4.4 :The LSTM unit comprises of input, output and forget gate in Bi-LSTM

Forget gate f_k isFirst unit of the LSTM model, and the input gate is provided with information to update. Temporarily save the g_k new candidate value prior to updating a new cell state value. Input is received as $x_k$and h_{k-1},At each stage, the output is computed using biases, the weight parameter, and the ReLU activation function.. Equation (4.6-4.8) is used to measure gates and the ReLU activation function, which solves the vanishing gradient problem.

$$f_k = \sigma(W_x^f x_k + W_h^f h_{k-1} + b^f) \tag{4.6}$$

$$i_k = \sigma(W_x^i x_k + W_h^i h_{k-1} + b^i) \tag{4.7}$$

$$g_k = ReLU\left(W_x^g x_k + W_h^g h_{k-1} + b^g\right) \tag{4.8}$$

The forget gate is used to apply the element-wise product of the cell state f_k at previous time step c_{k-1}, and new candidate value g_k isapplied to an input gate using the elementwise product i_k. For determining cell state, two determines sum c_k, as in equation (4.9).

$$c_k = f_k \odot c_{k-1} + i_k \odot g_k \tag{4.9}$$

The output gate o_k is last unit.This gate determines the output information Similar to the preceding procedures, the output gate is determined as shown in equation (4.10).

$$o_k = \sigma(W_x^o x_k + W_h^o h_{k-1} + b^o) \tag{4.10}$$

Equation (11) is used to compute the hidden state h_t.

$$h_k = o_k \odot ReLU(c_k) \tag{4.11}$$

Utilizing the LSTM model, which takes input from historical capacity data, the prediction is carried out $\{C_k\}$, size of input and output are same.Both the input and output sizes are the same, and the input and output data types are the same. Input and output formats are the foundation for the LSTM prediction.

Two order of each token in relation to its past and future context, two LSTM layers are used. The LSTM model analyses the left-to-right sequence and a separate sequence from right-to-left, as shown in Figure. 4.5. a front layer that is concealed $\vec{h}$ every time step, a hidden unit function t is depending on the input data step to calculate x_t and previous hidden state $\vec{h}_{t-1}$. The data entered at this stage x_t and the future hidden state $\overleftarrow{h}_{t+1}$ are to calculate the hidden unit function $\overleftarrow{h}$ in a hidden backward layer. The forward and backward context representations are used to produce $\vec{h}_t$and $\overleftarrow{h}_t$,that are joined together by a long vector, respectively. On combined results, forecasts' target signals are based.

The fully connected dense layer transforms the bi-directional network into a high-level sentiment representation to estimate text sentiment polarity. Equation (4.12) provide the output.

$$h_i = Relu(w_i h_p + b_i) \qquad (4.12)$$

The feature map that the Bi-LSTM network gives h_p, the obtained features is h_i, and learning new parameters are w_i, and b_i. The output layer's sentiment categorization is applied by the merge feature layer, as indicated in Figure 4.5.Application of sigmoid and sigmoid classifiers to binary and multi-class data is done in turn. Disparity between real and anticipated text sentiment is performed via balanced cross entropy.

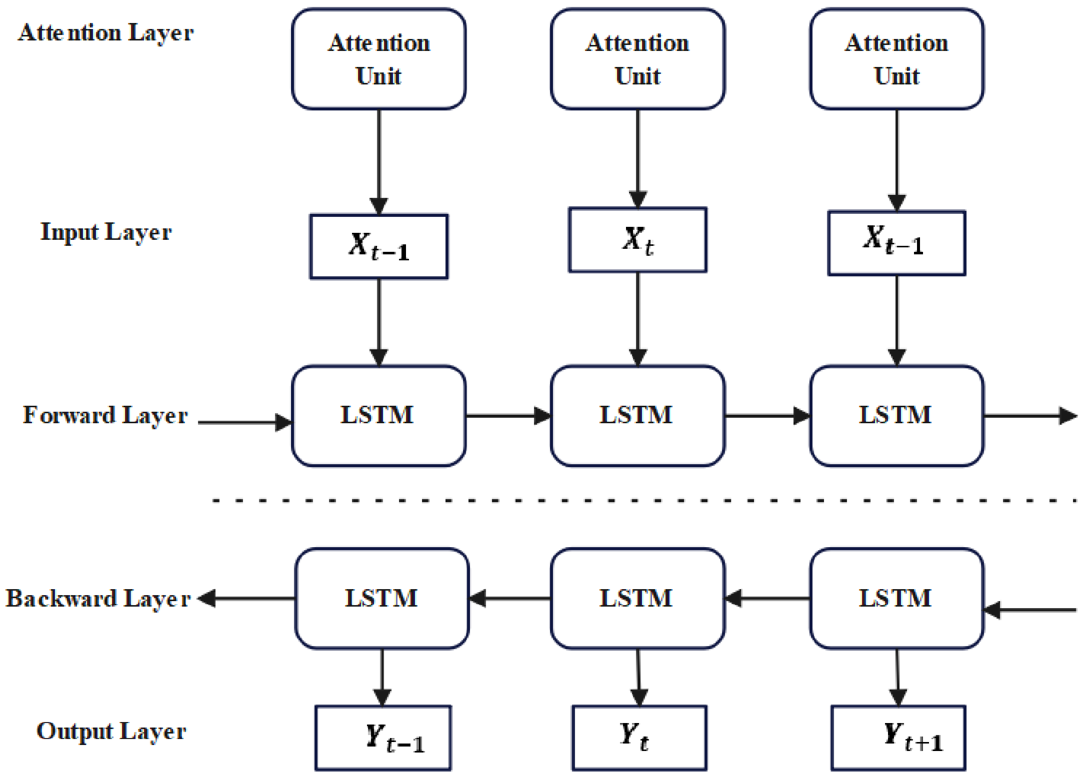

Fig. 4.5 :Attention layer in Bi-LSTM model for sentiment classification

4.4.4 Balanced Cross-Entropy

Two different kinds of optimization are paired with a tiny replay memory and the deep neural network baselinethe typical cross-entropy of softmax and distillation loss.

Total loss is used to train the model. L, to define weighted distillation loss L_d and average softmax cross-entropy loss L_c, as in equation (4.13)

$$L = \rho L_d + (1 - \rho) L_c \qquad (4.13)$$

Where ρ is applied as N_{t-1}/N_t with incremental step t and N_t asamount of courses overall, to make up for two losses of significance.The amount of new classes utilized for learning is compared to the total number of classes used for learning at each training stage, and the relevance of distillation loss is changed.

Each time a new step is added t, the new parameters θ_t are utilized for initialization and the parameters from the previous phase. θ_{t-1}which, by use of distillation loss, is employed to preserve previously knowledge learned. The distillation loss L_d for each training sample $(x, y) \in X_t \cup X_M$, as in equation (4.14 & 4.15):

$$L_d(x) = \sum_{k=1}^{N_{t-1}} -\hat{p}_k(x) \log\big(p_k(x)\big) T^2, \qquad (4.14)$$

$$\hat{p}_k(x) = \frac{e^{\frac{\hat{z}_k(x)}{T}}}{\sum_{j=1}^{N_{t-1}} e^{\frac{\hat{z}_j(x)}{T}}}, \quad p_k(x) = \frac{e^{\frac{z_k(x)}{T}}}{\sum_{j=1}^{N_{t-1}} e^{\frac{z_j(x)}{T}}} \qquad (4.15)$$

Whenever y is the related ground truth label, the input data is x, the current model θ_t output logits is $z(x) = [z_1(x), \dots, z_{N_t}(x)]$, and the previous incremental step θ_{t-1} output logits of model is $z(x) = [\hat{z}_1(x), \dots, \hat{z}_{N_t} - 1\ (x)]$.

4.5 Results and Discussion

The Amazon dataset is used to apply the WEA-BiLSTM model for sentiment analysis, and it is contrasted with alternative approaches. For comparison, the WEA-BiLSTM model outputs were tested for performance and error metrics.

Table 4.1 :Performance of WEA-Bi-LSTM model for various iterations

Iteration	Precision (%)	Accuracy (%)	F-measure (%)	Recall (%)
0	0	0	0	0
50	81.4	91.7	76	72.3
100	82.1	92.5	76.5	73.4
150	82.3	92.8	77.2	73.6
200	82.5	93.4	77.8	73.8
250	83.3	93.7	78.1	74.1
300	84	94	78.3	75.2
350	85.7	94.9	78.7	76
400	86	95.6	80.4	76.1
450	86.8	97	80.4	76.3
500	86.8	97.4	81.2	76.7

The WEA-BiLSTM model is used for sentiment analysis, and Table 4.1 displays performance metrics for several iterations. This demonstrates that the model's ability to detect emotion gives it more accuracy. Word embedding is used in the attention technique's convolution process to improve feature representation. Because the model contains textual information on sarcasm and distinguishing features, their recall value is low. The WEA-BiLSTM model's accuracy is 95.6%, precision is 86%, recall is 76%, and F-Measure is 78.7% in the 400th iteration.

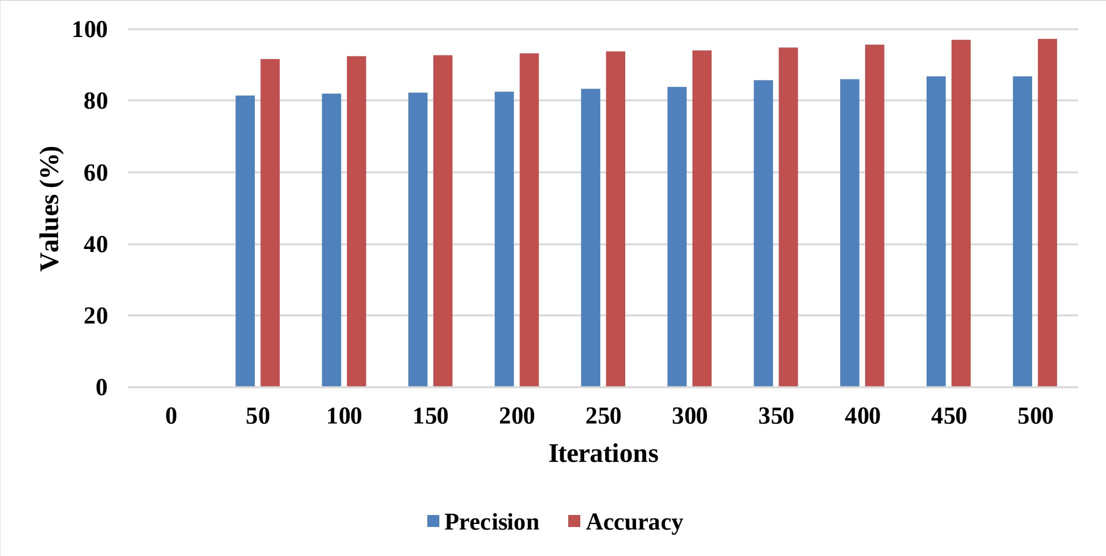

Fig. 4.6 :Accuracy and F-Measure of WEA-BiLSTM model for various iterations

Figure 4.6 displays the accuracy of the WEA-BiLSTM model and the F-measure for several rounds of sentiment analysis. This demonstrates that the WEA-BiLSTM model performs much worse than F-Measure while having better accuracy. The WEA-BiLSTM model performs less well in the negative class, which lowers the model's F-measure. Due to the user's caustic knowledge, the model performs significantly lower in the negative class.

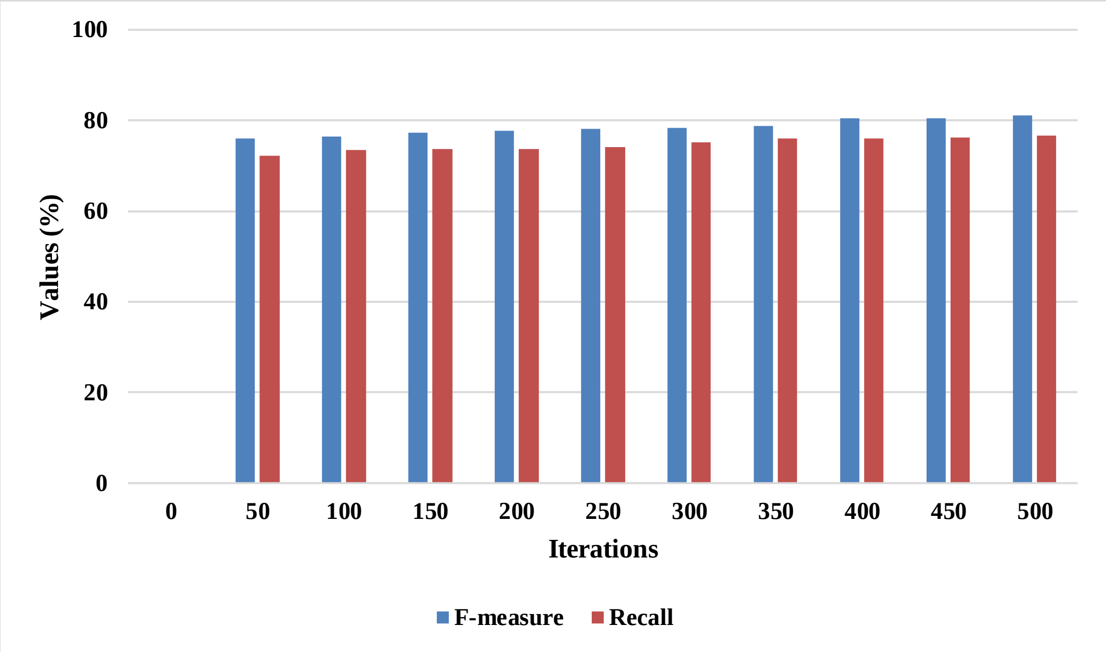

Fig. 4.7 :accuracy of the WEA-BiLSTM model

Figure 4.7 shows the accuracy and recall measurements for the WEA-BiLSTM model for several iterations. After 50 iterations, the WEA-BiLSTM model reaches the higher accuracy and recall levels. Based on the weight values supplied to the input text, the WEA approach accelerates the BiLSTM model's learning rate. The WEA approach gives terms with a strong connection to a class more weight, which boosts the model's effectiveness.

Table 4.2 :The WEA-BiLSTM model classifier comparison on sentiment analysis

Comparison	Precision (%)	Accuracy (%)	F-measure (%)	Recall (%)
SVM	81	91.8	75.6	73
ANN	71.9	81.9	76.8	72
LSTM	83	94.1	78.57	74.5
APSO-LSTM	85.28	96.8	80.04	76.08
CNN	85.6	97.1	80.64	76.33
Word Embedding CNN	86.8	97.4	81.2	76.7

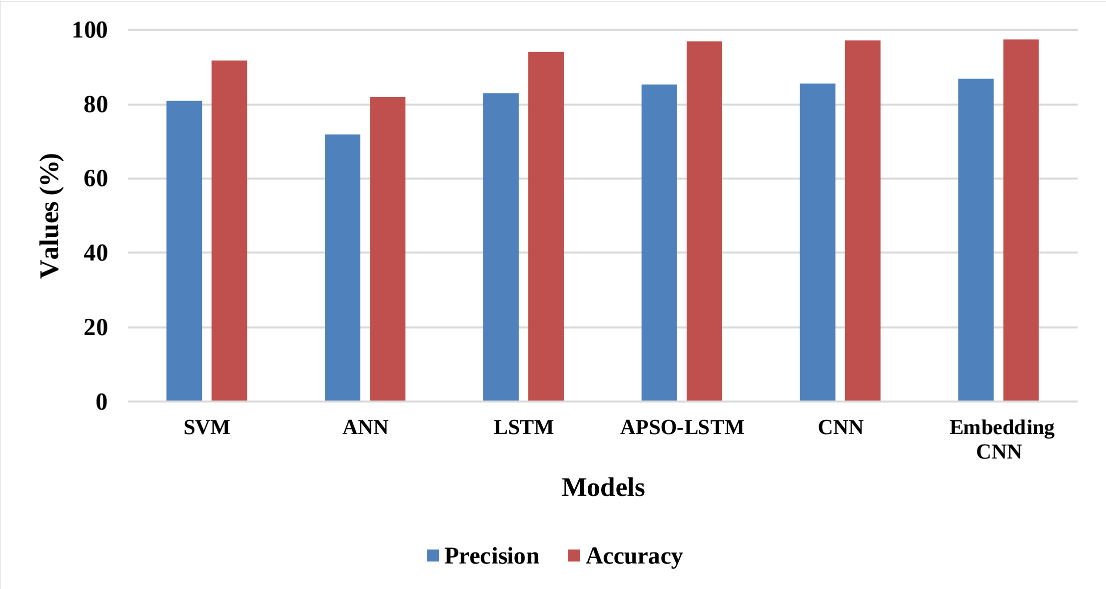

Fig. 4.8 :The WEA-BiLSTM model classifier comparison on sentiment analysis

Table 4.2 and Figure 4.8 compare the WEA-BiLSTM performance of the model to those of other existing classifiers. The WEA with CNN has the capacity to produce in-depth reports and directed outward with a strong connection to class greater weight. Existing SVM models have issues with imbalanced data, and LSTM models have issues with vanishing gradients. The APSO-LSTM factor represents certain ability to be flexible for classification and has a local optima trap. Due to the feature extraction procedure producing additional features, the CNN model suffers an overfitting issue. While the current CNN model has 97.1% accuracy and 85.4% precision in sentiment analysis, the WEA-BiLSTM model has 97.4% accuracy and 86.8% precision.

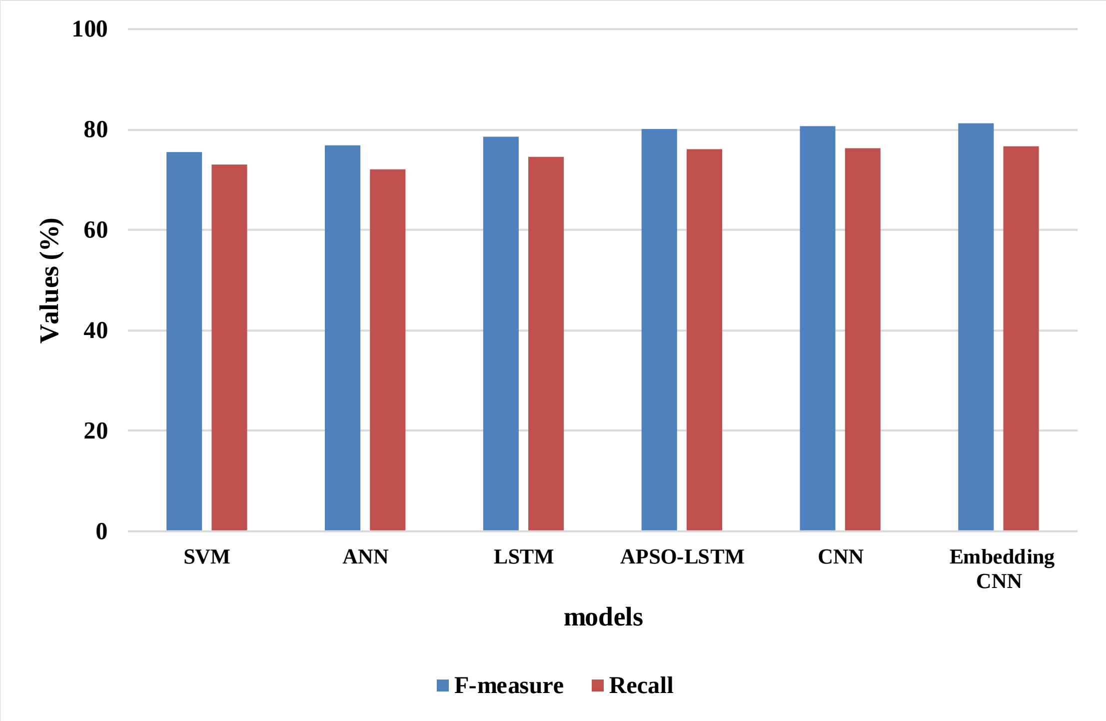

Fig.4.9 :WEA-BiLSTM precision and recall on sentiment analysis

Figure 4.9 shows the calculation of the WEA-BiLSTM model precision and recall value and a comparison with current sentiment analysis classifiers. The WEA-BiLSTM model employs embedding to provide phrases with strong relationships to classes a greater weight value. This improves the class-related performance of the WEA-BiLSTM model and raises its accuracy and recall values. Due to the addition of extra features in the convolutional layer, the existing CNN model performs less efficiently. The LSTM model's performance is decreased because to the vanishing gradient issue.

Table 4.3 :The WEA-BiLSTM model performance for various categories

Category	Methods	AUC	Accuracy (%)	Recall (%)	F-measure (%)	Precision (%)
Books	IG	0.93	88	88	88	88
	CNN	0.94	90	90	90	90
	Proposed CNN	0.96	93	92	92	92
DVD	IG	0.94	90.25	90	90	90
	CNN	0.95	91.4	90	90	90
	Proposed CNN	0.97	93.5	93	93	93

Category	Methods	AUC	Accuracy (%)	Recall (%)	F-measure (%)	Precision (%)
Electronics	IG	0.95	91.63	92	92	92
	CNN	0.96	92.3	93	93	93
	Proposed CNN	0.97	93.7	95	95	95
Kitchen	IG	0.96	93.09	93	93	93
	CNN	0.97	94.1	93	93	93
	Proposed CNN	0.98	95.2	94	94	94
AVG	IG	0.94	90.74	90.75	90.75	90.75
	CNN	0.95	91.95	91.5	91.5	91.5
	Proposed CNN	0.97	93.85	93.5	93.5	93.5

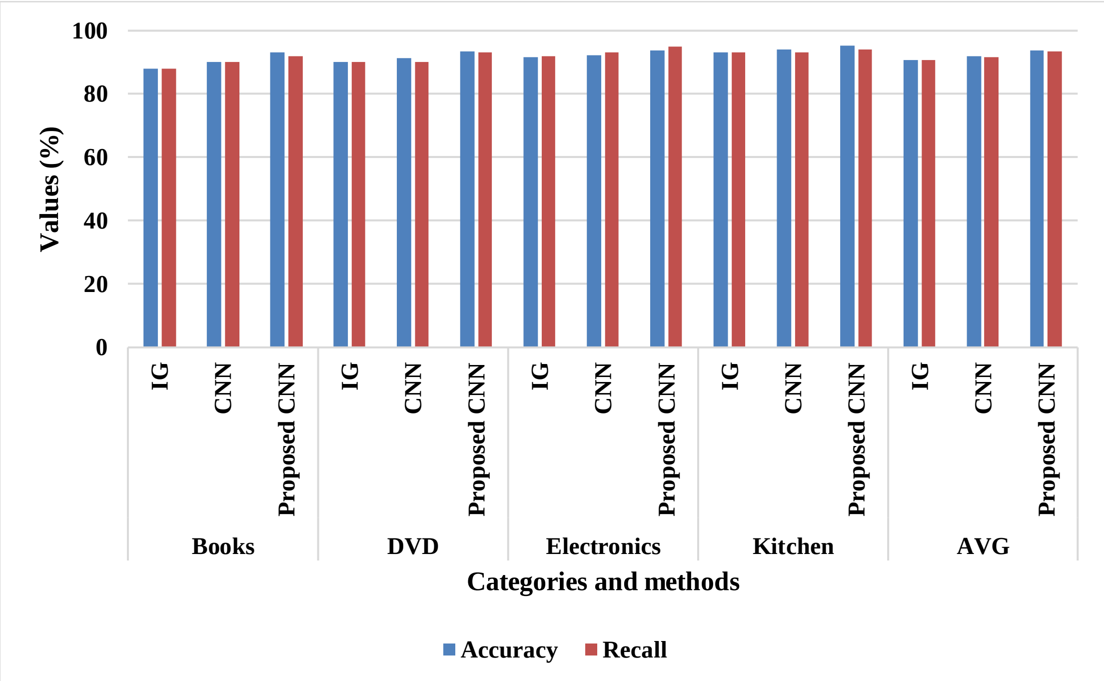

Fig. 4.10 :Accuracy and F-measure of the WEA-BiLSTM model for different categories

According to Table 4.3 and Figure 4.10, the WEA-BiLSTM model's effectiveness is evaluated for a number of categories in the dataset. Every dataset category has a better performance for the WEA-BiLSTM model. The WEA-BiLSTM model uses word embedding method to provide terms with strong relationships to classes a greater weight value. The learning rate of the LSTM model in classifications is accelerated by the CNN features extracted and weight value. To solve the vanishing gradient problem, the balanced cross-entropy maintains the gradient values in the model. Both the WEA-BiLSTM and CNN models have accuracy and precision ratings of 93.85% and 93.5%, respectively.

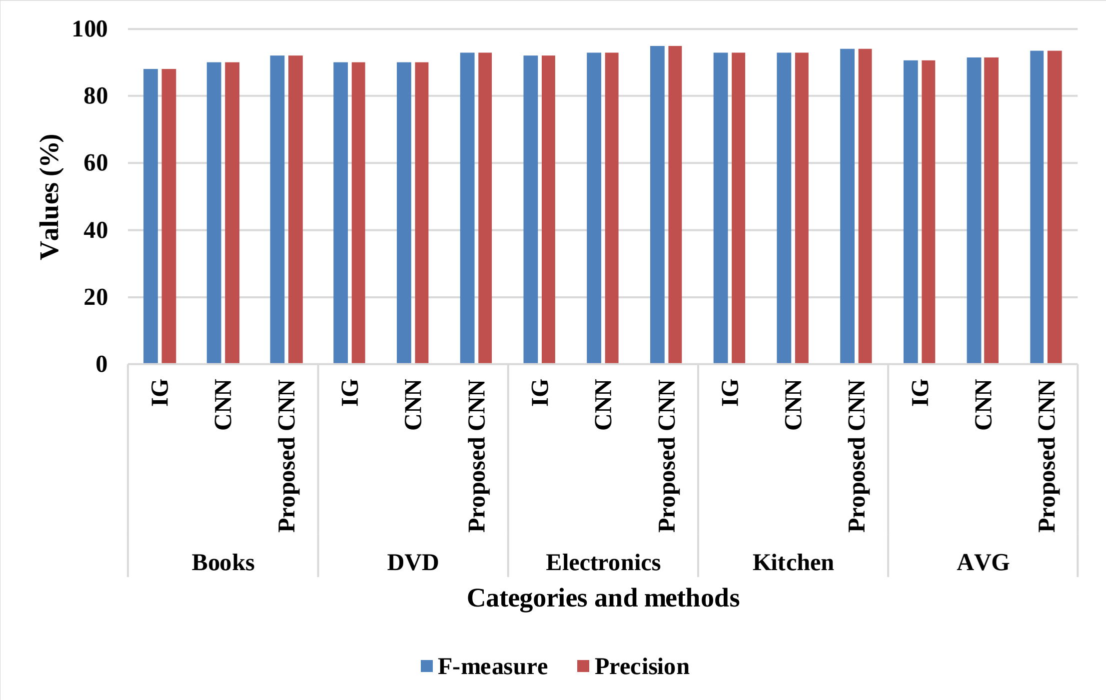

Fig. 4.11 :The WEA-BiLSTM model Precision and Recall for various categories

Figure 4.11 shows the accuracy and recall values for the WEA-BiLSTM model across the dataset's various categories. The WEA-BiLSTM model performs sentiment analysis with more precision and recall than other methods. The WEA-BiLSTM model gives terms with a strong connection to classes a greater weight value. This improves the model's performance class-wise and increases its accuracy and recall value.

Table 4.4 :WEA-BiLSTM model MAE on sentiment analysis

Datasets	ItemKNN	SlopeOne	A3NCF	SVD	CNN	UTER	Proposed CNN
Baby	0.8536	0.8557	0.8075	0.8123	0.7854	0.7966	0.7704
Sports and Outdoor	0.6771	0.663	0.633	0.6374	0.6087	0.6169	0.5843

Datasets	ItemKNN	SlopeOne	A3NCF	SVD	CNN	UTER	Proposed CNN
Pet Supplies	0.9202	0.8917	0.8746	0.88	0.8542	0.855	0.8412
Video Games	0.8482	0.8251	0.7916	0.7935	0.7891	0.7907	0.77

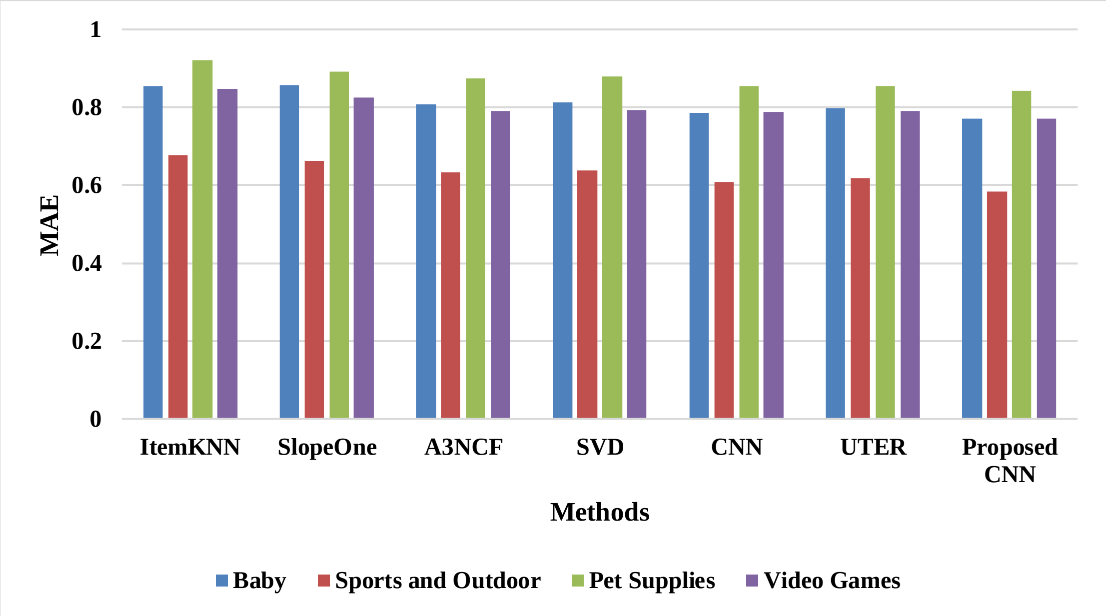

Fig. 4.12 :MAE of WEA-BiLSTM model on sentiment analysis

Table 4.4 and Figure 4.12 illustrate how the MAE value of the WEA-BiLSTM model for several categories on the Amazon dataset is determined and evaluated with modern techniques. The WEA-BiLSTM model performs better than the current approach. The WEA approach gives words with a strong connection to classes a greater weight value. The network's gradient value is sustained by balanced cross-entropy, which also eliminates the vanishing gradient problem.

Table 4.5 :WEA-BiLSTM RMSE on sentiment analysis

Datasets	ItemKNN	SlopeOne	A3NCF	SVD	CNN	UTER	Proposed CNN
Baby	1.1854	1.2012	1.0324	1.0704	0.9987	1.0146	0.9862
Sports and Outdoor	0.9959	1	0.8548	0.9065	0.8421	0.8501	0.8369

Datasets	ItemKNN	SlopeOne	A3NCF	SVD	CNN	UTER	Proposed CNN
Pet Supplies	1.2658	1.2454	1.1058	1.1448	0.9965	1.0627	0.9921
Video Games	1.1634	1.1421	1.0269	1.0511	0.9988	1.0164	0.9908

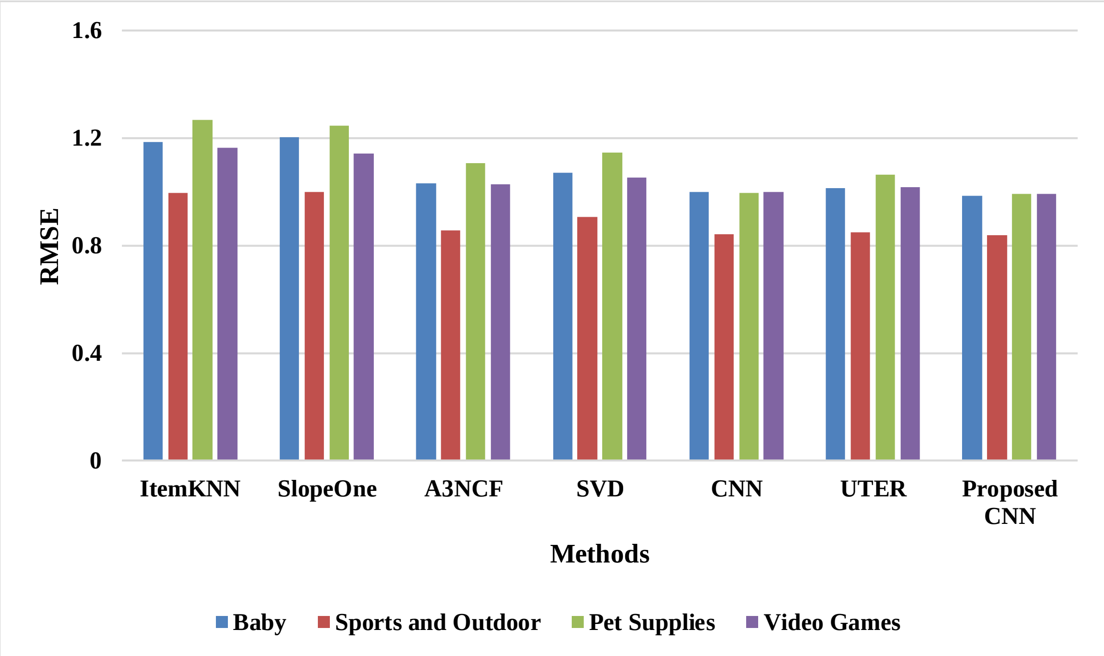

Fig. 4.13 :RMSE of WEA-BiLSTM model on sentiment analysis

Table 4.5 and Figure 4.13 illustrate how the WEA-BiLSTM model produces the RMSE value for several categories on the dataset. Due to the weight values that are provided in connection to classes, the WEA-BiLSTM model performs better in terms of classes. Due to the overfitting issue caused by the addition of extra features in convolutional layers, existing CNN-based models have limitations.

Table 4.6 :WEA-BiLSTM model for sentiment analysis

Algorithm	User Cold Start		Item Cold Start		User-Item Cold start	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
UTER	0.8123	1.2036	0.8106	1.2033	0.8112	1.2089
CNN	0.8108	1.1984	0.8065	1.1982	0.8057	1.2008
SVD	0.8409	1.2235	0.8446	1.2252	0.848	1.2276

Algorithm	User Cold Start		Item Cold Start		User-Item Cold start	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
A3NCF	0.8274	1.2137	0.8313	1.221	0.8336	1.2227
SlopeOne	0.8726	1.2626	0.8805	1.2639	0.8876	1.2683
ItemKNN	0.8653	1.251	0.8714	1.2501	0.8752	1.2532
Word Embedding CNN	0.8044	1.1920	0.7996	1.1905	0.8003	1.1941

Table 4.6 shows the MAE and RMSE values for the WEA-BiLSTM model evaluation for various cold start problems. Items with cold starts have less data about the products and less details about the users. The CNN-based feature extraction used by the WEA-BiLSTM model helps in extracting deep features and gives terms with strong relationships greater weight values.

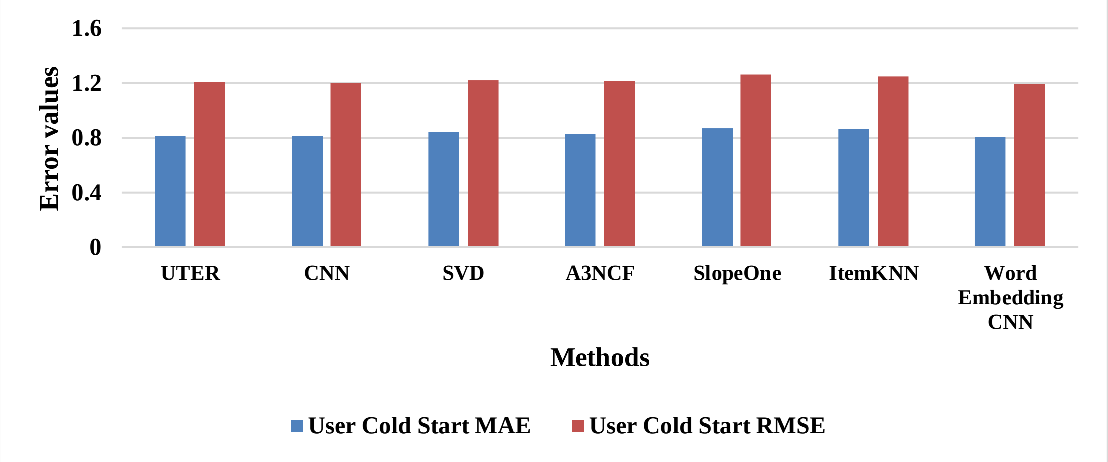

Fig.4.14 :The WEA-BiLSTM model for user cold start

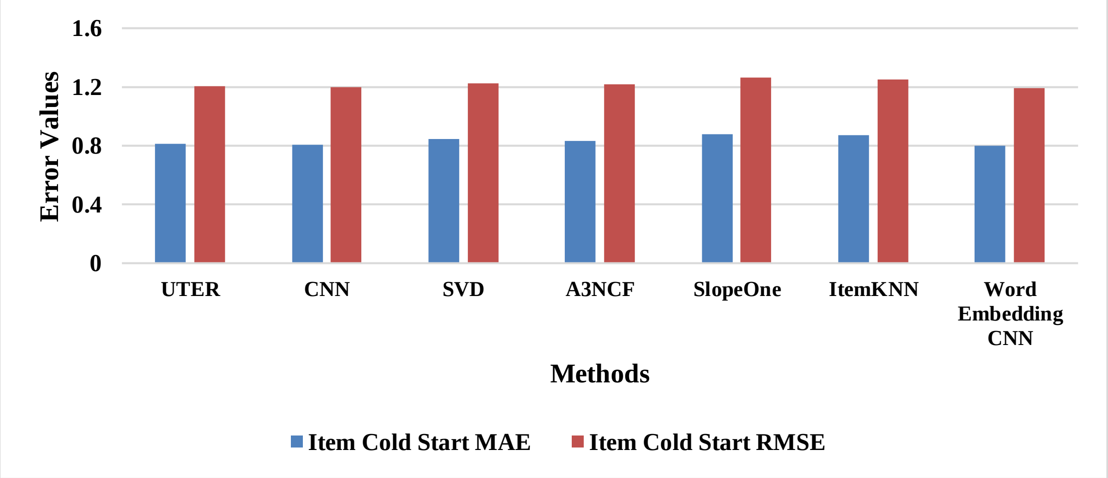

Fig. 4.15 :The WEA-BiLSTM model for Item cold start

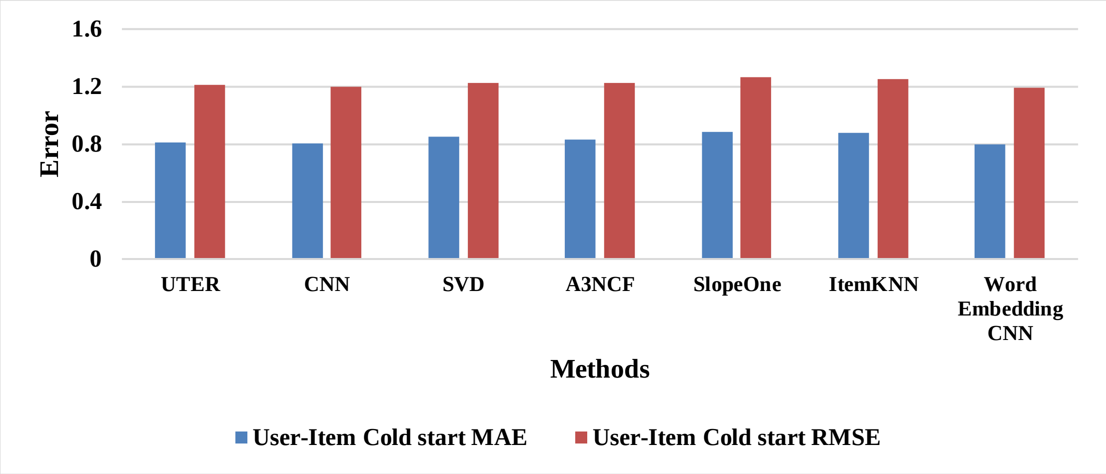

Fig. 4.16 :WEA-BiLSTM model for user Item cold start

According to Figures 4.14, 4.15, and 4.16, the MAE and RMSE metrics are used to analyze the WEA-BiLSTM model performance for user cold start, product cold start, and user-item cold start, respectively. The WEA-BiLSTM model performs sentiment analysis categorization more accurately than other models. Due to the model's utilization of CNN feature extraction for classification, the WEA-BiLSTM model performs better with lesser training data.

4.6 Summary

Sentiment is the process of identifying the intended sentiments that are conveyed in text-based materials like social media postings, product reviews, and online discussion environments. Similar to how sentiments derived from customer reviews and comments are often applied by businesses to develop and improve their goods and postal service. The user's sentiment about the product is provided through sentiment analysis on reviews, which is important for product development. The vanishing gradient problem and the overfitting issue in classification were constraints of the existing LSTM and CNN-based models used in sentiment analysis. This study suggested the WEA approach to give words with a strong relationship to classes a greater weight value. Due to the convolutional layer generating more features, the CNN model is used for feature extraction. To maintain the gradient in the system and solve the vanishing gradientproblem, balanced cross entropy is used. With less training data required for classification, improved performance is achieved by CNN feature extraction. Future work on this methodology will focus on using the feature selection strategy to choose pertinent features and improve performance in imbalance data.