Chapter 5

Aspect Category Detection for Hindi

In ABSA, the finer-level aspect terms are extracted and categorized. Aspect category detection(ACD) allocates a subset of predefined categories to a particular review. For example, consider the established list of restaurant aspect categories - *food*, *price*, *ambience*, *service* and *general*, and the following review-
"*Highly praise that as great value for excellent pizza and service.*"
The two phrases "*pizza*" and "*service*" belong to existing aspect categories "*food*" and "*service*". A review may have multiple aspect categories. Aspect category detection helps to achieve multi-label text classification. It is helpful to handle reviews with multiple-aspect terms. The following are the major contributions of this chapter:

- Aspect category detection is formulated as a multi-label problem, and
- Proposed a deep learning-based method for aspect category detection. Experiment results of our supervised model on IITP-II demonstrate that our method outperforms strong baselines significantly.

5.1 Methodology

The proposed model for aspect category detection includes one BiLSTM unit and one dense layer. The multi-label assignment module assigns different aspect categories to each sentence in the dataset. Pre-processing is performed on all sentences in the dataset. The dataset is split into three sets. These sets are train set, test set and validation set. The model is trained on the train set and it uses the test set for accuracy analysis. A validation set is employed to fine-tune the hyper-parameter [121] of the model. For aspect category detection, the sigmoid function [78] is used by output layer. Sigmoid function enables independent probability estimation for each category. BiLSTM layers allow the model to capture the input data's intricate dependencies and contextual information. The BiLSTM unit is able to learn more about the situation because it can process sequences both forward and backward. That helps a lot with jobs like finding the aspect category, where the meaning and relationship between words in a sentence are very important. A brief discussion of the BiLSTM unit is provided in Section 4.2. An example sentence with the predicted category *Electronics*, is shown in figure 5-1.

5.2 Experimental Setup

Hindi sentences are taken from IITP-II. Description of IITP-II is presented in section 2.6.

5.2.1 Data pre-processing

Hindi review sentences in the IITP-II belong to the following categories:- design, ease of use, GUI, hardware, miscellaneous, music, performance, place, price, reachability, scenery, software and story. These categories are distributed unevenly in

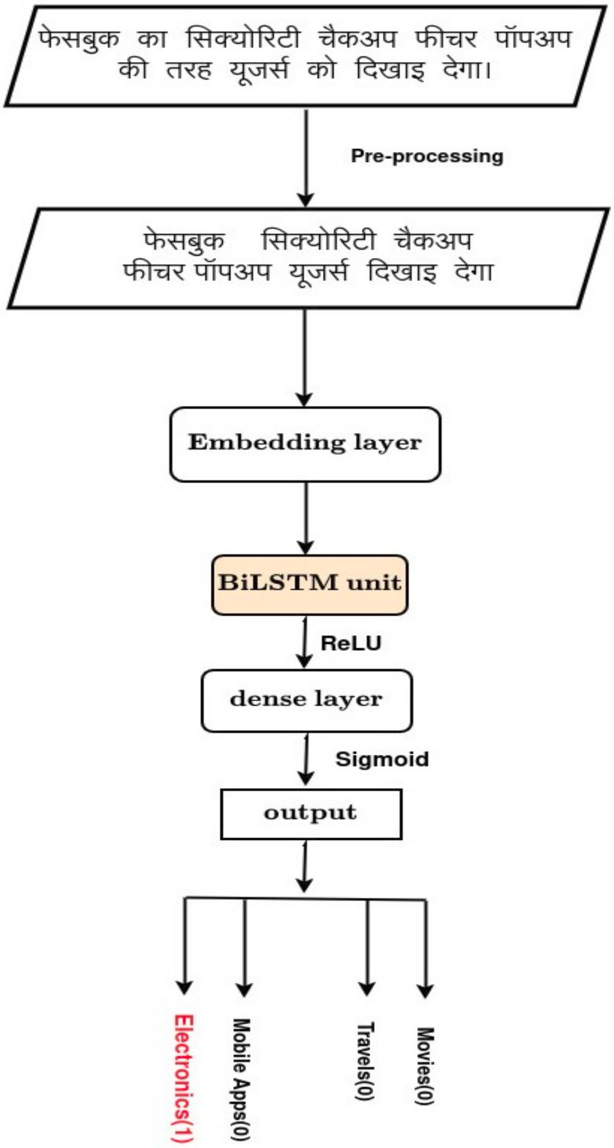


Figure 5-1: Aspect category detection model

the data set. Some categories are rare in the data set. Bar-graph in figure 5-2 represents it.

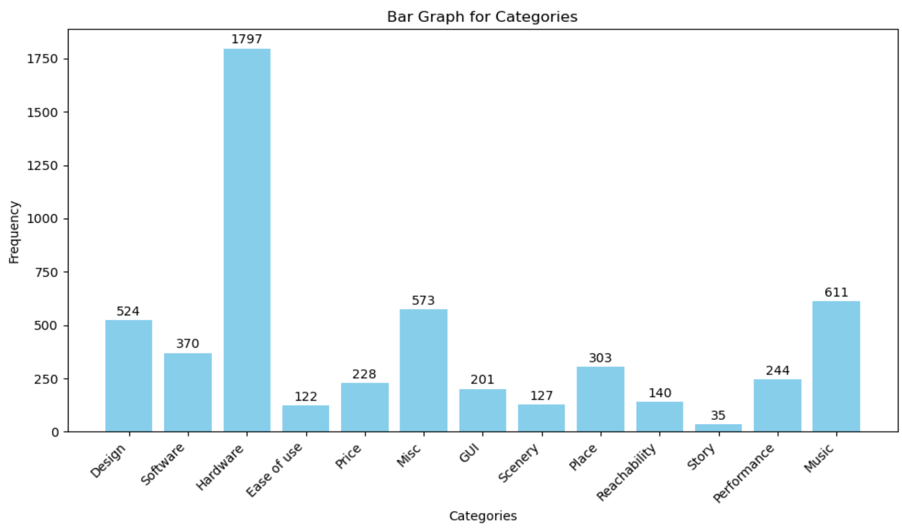


Figure 5-2: Hindi sentence distribution between categories

These aspect categories are grouped into four domains. Table 5.1 presents the aspect category distribution in these four domains.

Table 5.1: Aspect categories by domain

Domain	**Aspect Categories**
Electronics	Design(524), Software(370), Hardware(1797), Ease of use(122), Price(228), Misc(573)
Mobile apps	GUI(27), Ease of use(29), Price(11), Misc(134)
Travels	Scenery(127), Place(303), Reachability(35), Misc(105)
Movies	Story(35), Performance(244), Music(38), Misc(573)

There are 5275 aspect categories with their polarity among these 4930 review sentences in the dataset. 4601 Hindi reviews have one aspect category. 313 Hindi reviews have two aspect categories. Sixteen Hindi reviews have three aspect categories. Aspect category distribution among different domains is presented in figure 5-3.

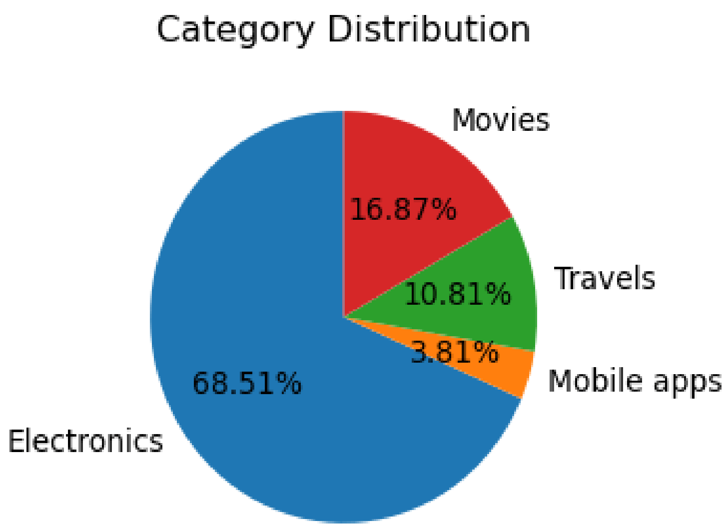


Figure 5-3: Hindi sentence distribution between categories

- **Multi-label assignment** The following steps are performed using python [71] for multi-label assignment to the Hindi sentences:-

  I. Initialize an empty list to store Hindi sentence data and aspect category labels

  II. Store each sentence in one row

  III. Initialize with 0 to all aspect categories for each sentence

  IV. For each sentence:

     (a) Identify and collect aspect categories

     (b) Update the set of aspect categories with 1 for the present aspect categories

  Two Hindi reviews from the IITP-II are presented in table 2.9. Figure-5-4 shows the multi-label assignments for these two Hindi reviews.

ID	Text	mobile_apps	electronics	movie	travel
app_2	फेसबुक का सिक्योरिटी चैकअप फीचर पॉपअप की तरह यूजर्स को दिखाइ देगा।	1	0	0	0
app_94	इसमे बहुत उच्च स्तर के ग्राफ़िक्स नहीं हैं लेकिन यह गेम बहुत आकर्षित करने वाला है।	1	0	0	0

Figure 5-4: Multi-label assignment

Data pre-processing is vital for effective machine learning. For aspect category detection in Hindi,it involves cleaning, standardising, removing duplicates and normalising text data, which includes eliminating punctuation, tokenisation, stemming for simpler words, and removing non-essential "stop words". An example of pre-processing on Hindi review is illustrated below.

Raw sentence:

फेसबुक का सिक्योरिटी चैकअप फीचर पॉपअप की तरह यूजर्स को दिखाइ देगा।

After pre-processing:

फेसबुक सिक्योरिटी चैकअप फीचर पॉपअप यूजर्स दिखाइ देगा

5.2.2 Fully connected layer

A fully connected layer [25] is a dense layer, where each neuron or node is connected to every neuron in the previous and subsequent layers. A fully connected layer usually performs an activation function after calculating the weighted sum of the input data. The weights associated with each connection are learnable parameters that the model adjusts during training to optimise performance on a specific task. The biases are introduced in the model to account for the offset or shift in the data.

Common activation functions in fully connected layers include ReLU [14], sigmoid, and hyperbolic tangent ($\tanh(x)$). These activation functions introduce non-linearities, allowing the neural network to model complex relationships in the data. As for "फेसबुक" in above review dense layer allows the model to learn an optimal combination of forward state h_T and backward state h'_T.

$$h_T^{(\text{Combined})} = \text{ReLU}(W[h_T, h'_T] + b)$$

where W and b are learnable parameters and bias.

A final dense layer with four units and sigmoid activation is used. Sigmoid activation is used here to predict probabilities of each of the four classes independently.

5.3 Evaluation Results and Analysis

The model is trained and assessed using the dataset after pre-processing. The dataset is divided into training set, validation set and test set in ratio 70-20-10.

5.3.1 Experimental Setup

Python [73] is used to construct the model. Keras [60], Scikit-learn [62], and TensorFlow [42] libraries are utilized within the development environment. This environment is made available via the Google Colab [18] virtual platform throughout the process. A GPU that is publicly available and 12 gigabytes of RAM make up the hardware setup. Visualizing and analyzing the results is accomplished through the matplotlib library [53].

Each training iteration comprises the processing of 32 samples, which allows for the training process to be fine-tuned. For optimization of the training process and to avoid over-fitting, an early-stopping technique [74] is incorporated. This strategy allows the model to stop training if the validation loss does not improve for three consecutive epochs. This prevents the model from being trained for an extended period of time, which is necessary when additional performance improvements are unlikely to occur.

The binary cross-entropy method [77] is utilized for loss estimation for the model. It is possible to prevent over-fitting by using dropout [13], which is an effective regularization strategy [94]. At each successive layer of training, dropout causes a random skewing of a few neurons.

5.3.2 Model Evaluation

We model the detection of aspect categories as a multi-label classification exercise; the averaging function can be used as a micro-average to calculate precision [89], recall and F1-score. Precision is the number of correctly predicted aspect categories divided by the total number of aspect category predictions. The number of aspect categories a system successfully predicted divided by the total number of aspect categories is termed as recall. F1-score is the harmonic mean of precision and recall. F1-score attempts to balance precision and recall. Accuracy is the ratio of correctly classified aspect categories (true positives and negatives) out of the classifier's total number of aspect categories. Note that the calculation ignores instances of the aspect categories that appear more than once in a single sentence.

5.3.3 Result Analysis

Accuracy and F1-score is measured to analyze effectiveness of proposed model. The model shows an impressive train accuracy of 93.91% and 0.8345 F1-score. The model accurately labels the data, as evidenced by his high accuracy and F1-score. Output for two reviews are shown as follows.

Hindi review – I (Before pre-processing)

फेसबुक का सिक्योरिटी चैकअप फीचर पॉपअप की तरह यूजर्स को दिखाइ देगा।

Hindi review – I (After pre-processing)

फेसबुक सिक्योरिटी चैकअप फीचर पॉपअप यूजर्स दिखाइ देगा

Predicted multi-label assignment

$$1\ 0\ 0\ 0$$

First label is 1 and other labels are 0, it indicates that predicted aspect category for Hindi review is **Electronics**. Review does not belongs to aspect categories

Mobile apps, **Travels** and **Movies**.

Hindi review – II (Before pre-processing)

इसमें बहुत उच्च स्तर के ग्राफ़िक्स नहीं हैं लेकिन यह गेम बहुत आकर्षित करने वाला है।

Hindi review – II (After pre-processing)

इसमें बहुत उच्च स्तर के ग्राफ़िक्स नहीं हैं लेकिन यह गेम बहुत आकर्षित करने वाला है

Predicted multi-label assignment

$$1\ 1\ 0\ 0$$

Predicted labels are 1 1 0 0. It means that review is predicted under two aspect categories, **Electronics** and **Mobile Apps**, out of four possible aspect categories.

5.3.4 Accuracy Analysis

We assess the performance of the suggested model in terms of its accuracy. In figure 5-5 the training accuracy and the validation accuracy are plotted against the number of epochs. In order to complete an epoch, the full training dataset must be processed by the model just once. In other words, each and every data point makes a contribution to the parameter update. There are slight differences in accuracy between training and validation. At first, the model's accuracy rises rapidly as the epochs increase. Later on, it becomes better marginally. It affirms that the model is suitable for the given problem.

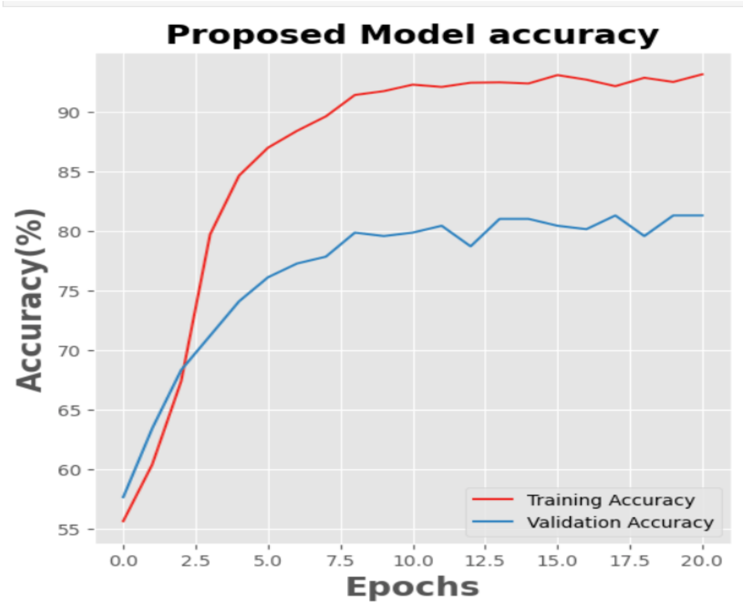


Figure 5-5: Accuracy versus epochs

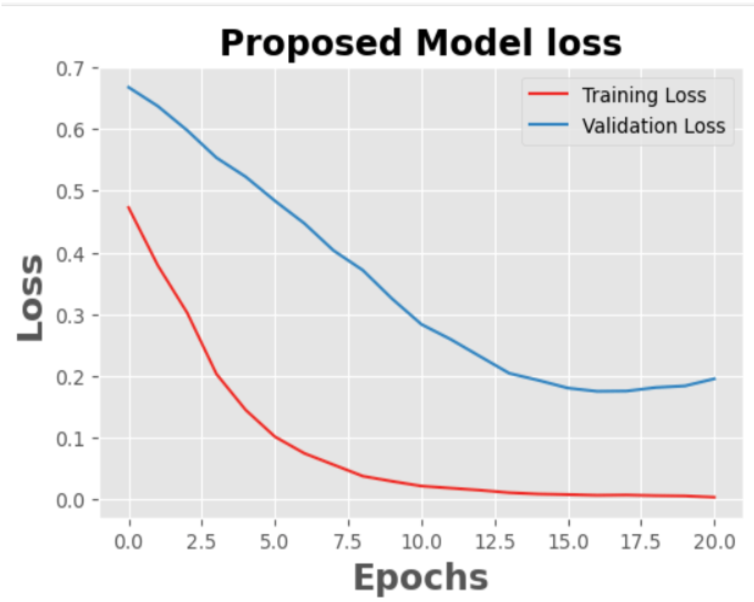


Figure 5-6: Loss versus epochs

Figure 5-6, is a graph for the loss function of the model, which also provide information that is consistent with the previous figure. In the initial stages of the learning process, the validation loss is greater than the training loss. When the model acquires a new aspect category domain feature, the accuracy of the model improves while the validation loss reduces. This is because the model is learning new features. Figure 5-7 demonstrates that the model accurately detects each and every aspect category.

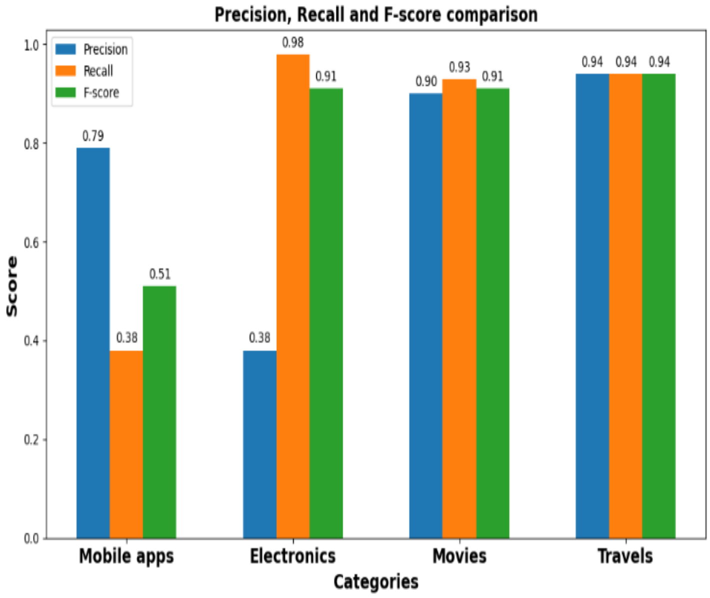


Figure 5-7: Aspect category-wise performance concerning precision, recall and F1-score

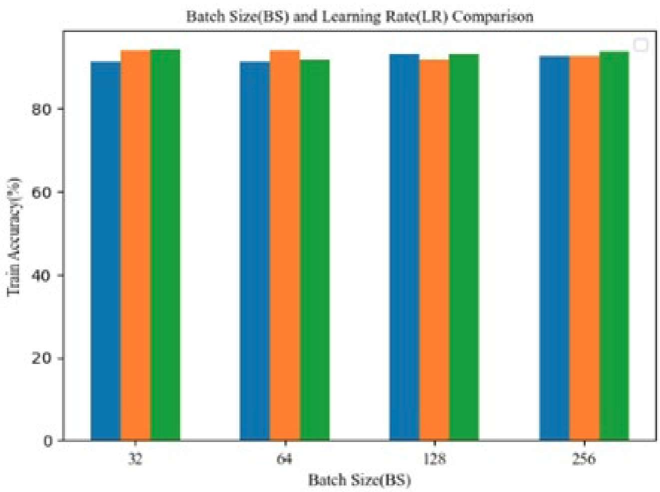
The model's overall train accuracy is 93.91%, validation accuracy is 84.39% and the F1-score is 0.8345. The hyper-parameters are specified in table 5.2.

Table 5.2: Dimensions and optimal hyper-parameters for the aspect category detection model

Parameter	**Value**
Input's maximum length	80
LSTM units	25
Dropout	0.5
Number of neuron units in dense layer	50
Activation function	ReLU
Number of neuron units in output layer	04
Epochs	33
Mini-batch size	32
Rate of learning	0.0001
Function of activation	ReLU
Rate of dropout	5.0×10^{-1}

5.3.5 Influence of batch size and learning rate

We observe that there is a considerable connection between the hyper-parameter batch size and the learning rate of the aspect category detection model. The behavior of the model is observed over the duration of this experiment, and it is done so across a wide range of mini-batch sizes and learning rates. The results obtained with batch sizes of 32, 64, 128 and 256 are presented in figure 5-7. Additionally, the learning rates were 0.01, 0.001, and 0.0001 respectively. With learning rate(LR) of 0.0001 and 32 mini-batch size, the model recommended is able to obtain the best possible validation accuracy. When applied to a smaller batch size, more accuracy is achieved with a lower learning rate.



LR = 0.01	90.99	91.12	92.89	92.51
LR = 0.001	93.65	93.78	91.62	92.39
LR = 0.0001	93.91	91.75	93.02	93.40

Figure 5-8: Accuracy obtained across different mini-batch sizes and learning rates

5.3.6 Comparison of aspect Category Detection Results

Here, we examine the similarities and differences between the outcomes of our proposed models' aspect category identification and the findings of the current state of the art. When compared to the findings of Akhtar et al. [6] and Hetal et al. [51], the outcomes demonstrate a notable improvement. The comparison of the F1-scores can be seen in figure 5-9. With the BiLSTM model suggested, we are able to achieve a maximum F1-score value of 0.94 for the Travels. When compared to the findings considered to be state-of-the-art for Hindi, the F1-scores for all three domains, except the Mobile apps domain, our results exhibit an improvement ranging from 0.19 to 0.64. It is important to note that the Mobile apps domain deviates from this trend. The F1-score value of 0.51 indicates that the Mobile apps aspect category did not demonstrate any significant growth throughout the course of the study. The fact that there are fewer occurrences (201) of the Mobile apps domain, as shown in table 5.1, could be one of the reasons for this particular phenomenon. A comparison with state-of-art method is shown in table 5.3. Within the Electronics, Movies and Travels aspect category, the model obtains a notable advancement in its F1-score in contrast to its previous score.

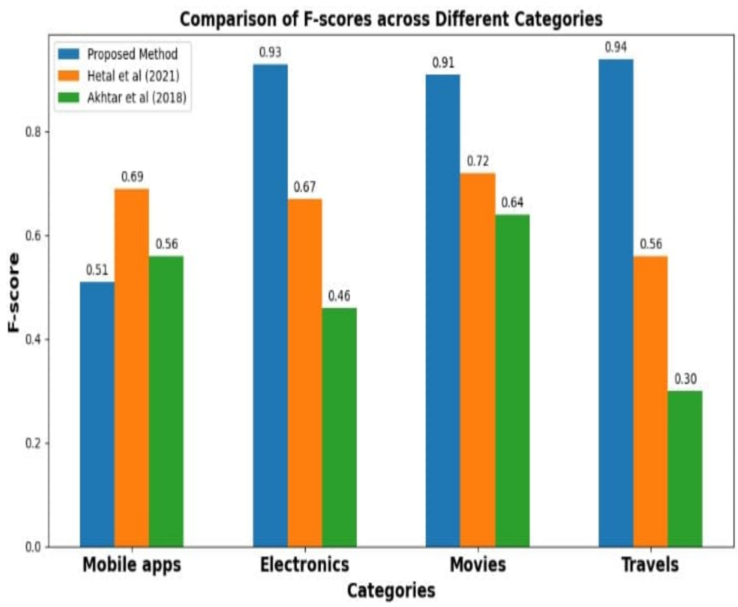


Figure 5-9: F1-score comparison of the proposed model with state-of-art results

Table 5.3: Comparison of F1-scores of Aspect Category Detection methods for Hindi

Domain	References	F1-score
Electronics	Akhtar et al. [5]	0.46
	Hetal V et al. [51]	0.67
	Our model	**0.93**
Mobile apps	Akhtar et al. [5]	0.56
	Hetal V et al. [51]	**0.69**
	Our model	0.51
Travels	Akhtar et al. [5]	0.30
	Hetal V et al. [51]	0.56
	Our model	**0.94**
Movies	Akhtar et al. [5]	0.64
	Hetal V et al. [51]	0.72
	Our model	**0.91**

5.4 Summary

We have conducted a study on recognizing aspect types with a framework based on deep BiLSTM. An aspect category is a generalization of the characteristics that are addressed in a review, which presents numerous features. Specifically, the aspect category detection can be seen as a problem of multi-label categorization. The model has an outstanding F1-score of 0.8345 and an accuracy of 93.91% across four domains, which is higher than the results that are considered to be state-of-the-art for Hindi. In addition, the model has a high accuracy rate. It achieves an acceptable balance between the F1-score and the accuracy measurements. BiLSTM is able to successfully handle sequences of variable lengths, which is a significant advantage given the broad range of lengths that are present in Hindi phrases. The complex relationships that existed between the words in the sentences are successfully captured by the BiLSTM units. With a dataset that is larger and contains a greater number of domains involving combination of hand-crafted qualities, it would be able to do more in-depth testing of the proposed approach. We believe that this will result in an improvement in the performance of the model. In addition, we are looking forward to improving performance by using techniques that are founded on the concept of transfer learning.