A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents

PRAPHULA KUMAR JAIN, Indian Institute of Technology (Indian School of Mines)
VIJAYALAKSHMI SARAVANAN, Rochester Institute of Technology
RAJENDRA PAMULA, Indian Institute of Technology (Indian School of Mines)

With the fastest growth of information and communication technology (ICT), the availability of web content on social media platforms is increasing day by day. Sentiment analysis from online reviews drawing researchers' attention from various organizations such as academics, government, and private industries. Sentiment analysis has been a hot research topic in Machine Learning (ML) and Natural Language Processing (NLP). Currently, Deep Learning (DL) techniques are implemented in sentiment analysis to get excellent results. This study proposed a hybrid convolutional neural network-long short-term memory (CNN-LSTM) model for sentiment analysis. Our proposed model is being applied with dropout, max pooling, and batch normalization to get results. Experimental analysis carried out on Airlinequality and Twitter airline sentiment datasets. We employed the Keras word embedding approach, which converts texts into vectors of numeric values, where similar words have small vector distances between them. We calculated various parameters, such as accuracy, precision, recall, and F1-measure, to measure the model's performance. These parameters for the proposed model are better than the classical ML models in sentiment analysis. Our results analysis demonstrates that the proposed model outperforms with 91.3% accuracy in sentiment analysis.

CCS Concepts: • **Computing methodologies → Natural language processing**; **Machine learning**; • **Information systems → World Wide Web;**

Additional Key Words and Phrases: Convolutional neural network, long short-term memory, deep learning, sentiment analysis, social media, word embedding

84

ACM Reference format:
Praphula Kumar Jain, Vijayalakshmi Saravanan, and Rajendra Pamula. 2021. A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 20, 5, Article 84 (July 2021), 15 pages.
https://doi.org/10.1145/3457206

Authors' addresses: P. K. Jain, Department of Computer Science & Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, JH, 826004, India; email: praphulajn1@gmail.com; V. Saravanan, Department of Software Engineering, Rochester Institute of Technology, USA; email: vijayalakshmi.saravanan@gmail.com; R. Pamula, Department of Computer Science & Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, JH, 826004, India; email: rajendra@iitism.ac.in.

2375-4699/2021/07-ART84 $15.00
https://doi.org/10.1145/3457206

1 INTRODUCTION

With the rapid growth of information and communication technology, Internet users are increasing daily; due to that drastic increase, the Web 2.0 trend has come into existence [36]. Web 2.0 is mainly described by social media engagements in terms of user-generated content. The main features of Web 2.0 are collaborating information sharing and user interactions on the online platform, which leads to a tremendous amount of text information on various topics. Hence, the social media platform becomes an essential part of information sharing and social networking. The social media appearance and popularity like Twitter, Facebook, and online review websites, have found the attraction of Web 2.0 generation [15].

With the exponential increase of social media information and messaging data, sentiment analysis using online reviews is getting massive attention from various organizations (e.g., private, government, and academia) [26]. There is a lot of social media mining going on online review; a few examples are connection, content, and user data [16]. Evaluation and analysis of qualitative content of social media information to get essential aspects is a valuable contribution in the field of travel and tourism management [43], scientific research [19], and consumer behavior analysis [2]. In the existing literature, many scholars focused their study on various application areas, such as news analysis [9], prediction election results [28], and healthcare [1]. The study of sentimental analysis of qualitative content based on emotional detection or opinion mining has become an attractive research field for many scholars [32, 42].

In the existing literature, scholars are using various computational models in sentiment analysis to obtain insights from qualitative information, **NLP**, and **text mining (TM)** to detect individual emotions or opinions [4, 44]. Proposing an accurate model in improving sentiment or emotion classification performance is the primary constraint of such studies. In the current scenario, sentiment analysis is the most popular strategy in textual content to get consumer opinions or sentiments and formulate a new model. Furthermore, the study on sentiment analysis encompasses different aspects of the domain and covers various technical information carvings. Due to this fact, sentiment analysis is a new area of research in knowledge fusion using **machine learning (ML)** and an attractive topic of research in the field of artificial intelligence [11, 33].

In the recent research era, online reviews provide an opportunity in business growth by exploring consumer reviews in understanding perception and how they can use these perceptions for consumer service improvement in this cutting-edge competition. Online review data are unstructured; there are possibilities of irregularities, ambiguities, and a massive volume of data that becomes challenging in nature while analyzing manually [10, 22]. We have used the proposed hybrid **convolutional neural network–long short-term memory (CNN-LSTM)** model for sentiment analysis in extracting opinions from the qualitative content to make it faster and efficient. Sentiment analysis is one of NLP's critical areas of research. Sentiment analysis is the method by which views are identified and extracted from a piece of text. Sentiment analysis predicts whether a given text is either positive, negative, or neutral. Sentiment analysis takes unstructured reviews of users and offers insights about the review.

This article implements a hybrid model that combines CNN and LSTM. We used the convolution layer to extract features from the text and the LSTM model to determine long-term dependence between word sequences. CNN recognizes patterns across space, and LSTM learns patterns across time. We used a combination of these two models, which will exploit both neural network models' properties. There are different ways of performing sentiment analysis, such as document-wise, sentence-wise, and aspect-wise sentiment analysis. We have focused on document view sentiment analysis in which we take a review and classify it as positive or negative. In addition, classification can be supervised and unsupervised. In this study, we used supervised learning because collected reviews are already labeled previously. We also compared the proposed model in terms of accuracy with various ML models, such as **Naive Bayes (NB)**, **Logistic Regression (LR)**, **Decision Tree (DT)**, **Support Vector Machine (SVM)**, CNN, and LSTM.

The rest of the article is structured as follows. Section 2 presents a literature review on this study's related work, with the beauty of online reviews in classifying consumer sentiment. In this clime, the theoretical model related to our work and research questions are also presented. Section 3 presents the suggested research techniques, and Section 4 presents the proposed hybrid model in sentiment classification. Section 5 shows the results of our implementation. Section 6 presents a discussion of our implementation strategy and findings. Finally, Section 7 concludes our research work and provides future directions with this study's limitations.

2 EXISTING LITERATURE AND RESEARCH QUESTIONS

2.1 Sentiment Analysis

Sentiment analysis, also known as opinion mining or emotional extraction, uses TM and NLP techniques to draw consumer attention [7]. Sentiment analysis provides various benefits like upselling opportunities, agent monitoring, and live insights [21]. Using sentiment analysis, service providers get to know where they are lacking and the aspects with which consumers are satisfied [6]. Sentiment analysis works by identifying the positive and negative feelings within text reviews, and this can be very hard to know within subtle wordplays [30]. Sentiment analysis can be used to get accurate and more in-depth analysis. Some related research work on sentiment analysis is mentioned in Table 1.

In this study, the goal is to find whether the given review is positive or negative; in this context, we have implemented various techniques to get the results. Many of the researchers have carried out their research in this area [31, 45, 47]. A **deep learning (DL)** framework is a powerful tool for supervised and unsupervised learning. Many researchers are trying to use DL for sentiment analysis because of better results as well [14, 38]. In recent years, there has been a lot of work on analyzing emotions using DL models. The deep neural network has achieved great success in ML, especially in computer vision and NLP. A deep neural network for sentiment analysis using different filters was proposed by Kim [25]. Pang and Lee [37] first introduced the concept of document-level sentiment analysis. A massive amount of data generates through reviews on the web daily, which need to be utilized and extract meaning from text.

2.2 LSTM- and CNN-Based Sentiment Analysis Research

The CNN model has been applied in the existing literature and shown to be useful for sentiment analysis problems. The dynamic CNN for semantic modeling of sentences presented by Kalchbrenner et al. [24]. Dynamic kmax pooling is used for the dynamic CNN, which performs max pooling over a linear sequence. CNN with Word2Vec was proposed by Jang et al. [23], which operates seven layers model to improve the proposed model's accuracy on the IMDB dataset. In an implementation, the authors used Word2Vec and the parametric rectified linear unit with the dropout and CNN model [37]; this proposed model gains better accuracy compared to other classification models.

Ouyang et al. [35] used CNN for sentiment analysis on the rotten tomato dataset. Instead of only having a short-term memory, LSTM networks also have a long-term memory and can thus handle long-term dependencies required for sentences. The meaning of words depends upon the order in which they appear. The position of a word changes the sentiment of the sentence. Different CNN and LSTM combinations have been applied on datasets to get better results [34]. Various experiments proved that word vectors are an essential part of NLP in DL. A character-level architecture for NLP proposed by Al-Smadi et al. [3], and a high-level representation of a sentence,

Table 1. Related Recent Literature on Sentiment Analysis Using Online Reviews in Hospitality and Tourism

S. No.	Authors	Domain	Data Source	ML Techniques
1	Sezgen et al. (2019) [39]	Airline	TripAdvisor	Latent semantic analysis
2	Banerjee and Chua (2016) [5]	Airline	TripAdvisor	Statistical analysis
3	Kumar and Zymbler (2019) [27]	Airline	Twitter	SVM, ANN, CNN
4	Ye et al. (2014) [48]	Hotel	Daodao	LR
5	Chen et al. (2019) [8]	Hotel	MilitaryLife PTT	LSTM
6	Dey et al. (2016) [12]	Hotel	UCI	NB, KNN
7	Lee et al.(2018) [29]	Hotel	TripAdvisor	DT, NB, SVM
8	Guo et al. (2017) [18]	Hotel	TripAdvisor	Latent Dirichlet analysis
9	Siering et al. (2018) [40]	Airline	Airlinequality	NB, SVM, NN
10	Our study (2020)	Airline	Twitter, Airlinequality	LR, DT, NB, SVM, LST, CNN, proposed CNN-LSTM

also achieved significant results with lesser convolution layers on movie reviews and the Stanford sentiment tree bank dataset.

2.3 Basic LSTM Model

LSTM is a **recurrent neural network** (RNN) for sequential data, designed to generate better long-term dependencies than normal RNN. RNN is a neural network designed to capture temporal dependence. In RNN, output depends only on the previous state's results, making them applicable to tasks such as handwriting recognition, speech recognition, document generation, or chatbots. Like RNNs, at each step, the LSTM network also gets its input from the present timestep and the output from the last time stage. The output of the current stage is provided as input to the next stage. The layer from the previous timestep or sometimes all of them are used for classification. The vanilla RNN often experiences an exploding or vanishing gradient [41]. The LSTM network tries to conquered such a problem by utilizing a few internal gates in each node.

There are three gates in the LSTM model: the input gates, output gate, and forget gate. These gates help avoid the long-term dependency problem. The cell recognizes values over a random time interval, and the three gates monitor the flow of data in and out of the cell. A softmax layer is used to apply the softmax function. Figure 1 shows an LSTM module's internal design, and Figure 2 represents the basic LSTM model.

The association between input, hidden states, and different gates in a LSTM cell is mentioned in Equations (1) through (5):

$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}), \tag{1}$$

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}), \tag{2}$$

$$o_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)}), \tag{3}$$

$$c_t = \sigma(f_t \bullet c_{t-1} + i_t \bullet tanh(W^c x_t + U^{(c)}h_{t-1} + b^{(c_t)})), \tag{4}$$

$$h_t = o_t \bullet tanh(c_t), \tag{5}$$

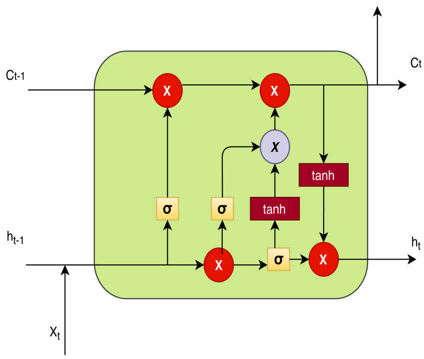

Fig. 1. The internal architecture of a standard LSTM module.

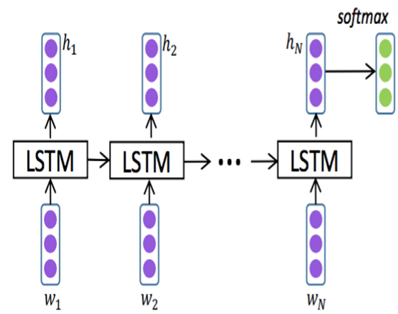

Fig. 2. The architecture of a standard LSTM model.

where W and U represents the input weights and recurrent connections, respectively. All small letters represents vectors, at time t, x is input ($x_t \in R^d$), and the feature dimension of each word is indicated using d, σ shows the sigmoid function, and $\bullet$ means the element-wise product. The memory cell is created to reduce the vanishing, and the exploding gradient is indicated by c_t ($c_t \in R^h$), which helps learn of dependencies over a long time. This was not feasible with the conventional recurrent network; i_t ($i_t \in R^h$) denotes the input gate, o_t ($o_t \in R^h$) represents the output gate, and these control input and the cell's output. The forget gate indicated by f_t ($f_t \in R^h$) is used for resetting the memory cell, and the hidden state vector is described by h_t ($h_t \in R^h$).

2.4 Basic CNN Model

CNN is a specific type of neural network that works well with spatial data. CNN is useful with images and videos for classifications and segmentation. In recent years, CNN has been very successful for NLP tasks as well. One of the early works that used CNN for text classification [13], which used CNN on various text classification tasks, showed excellent performance. The basic CNN model is shown in Figure 3.

2.5 Research Questions

In this study, we have presented theoretical aspects that drive our strategy in classifying consumer decisions as positive or negative based on the online reviews provided by the valuable consumer. Existing literature on sentiment analysis focuses on consumer content; similarly, our research has focused on the research queries mentioned next:

Research query 1: What is the research objective of this study?
Research query 2: Can the opinion expressed by the reviewer in online reviews be classified as positive or negative?
Research query 3: What is the power of the proposed model compared to an existing model?

3 PROPOSED CNN-LSTM MODEL FOR SENTIMENT ANALYSIS

Our proposed model contains a CNN and an LSTM system (refer to Figure 4). The word embedding applied to flight reviews is given as input to the model. We used the word embedding technique with LSTM and CNN models because representing words into vectors provides better performance for NLP problems [17]. As DL needs numbers as input, we use Keras to generate word embedding that converts the text data into vector values. The embedding assumes that two words sharing

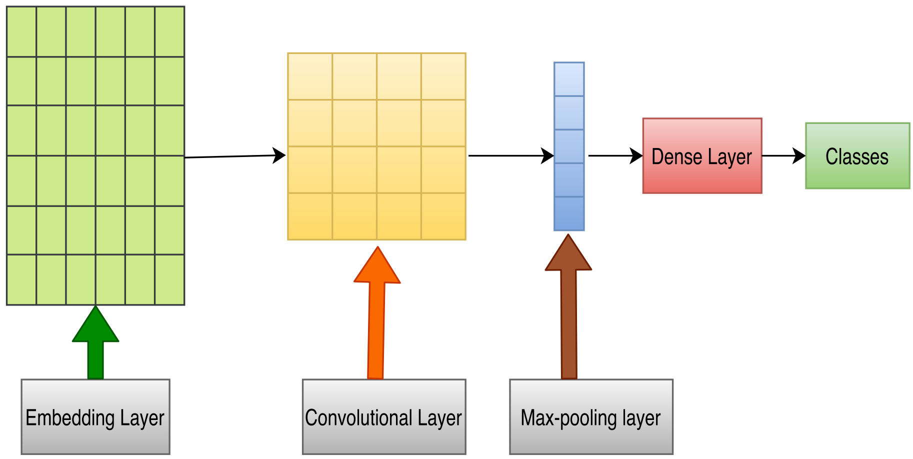

Fig. 3. The architecture of CNN-based text classification models.

similar contexts also have a similar meaning and consequently have a similar vector representation. For example, “dog,” “puppy,” and “pub” are often used in a similar context, and with similar words such as “cute,” “fluffy,” or “bite,” they will share a similar vector representation. These vectors become the input to the proposed CNN and LSTM.

3.1 The Proposed CNN Model

In this segment, We defined the CNN model, which uses word embedding as input parameters. The embedding layer translates the text into numeric vectors, and then we have applied it to the CNN model to train these vectors. We have used one pooling, one convolution layer, and two fully connected layers in the proposed CNN. We have used Keras’s open-source python library for numerical computation. In this study, we have also used dropout layers and **Rectified Linear activation unit (ReLu)** Relu for getting better results because neurons can get overfit quite easily, to regularize neural network Dropout can implement [46]. Dropout prevents neural networks from overfitting by randomly skipping neurons by a given probability on each epoch. The proposed CNN model is described as:

Convolution layer. The embedding layer takes the tokenized vector from text and converts it into a vector of the desired size. Our CNN model consists of one convolution layer. CNN convolves this layer’s input with pooling layers, and pooling reduces the complexity of computation and controls overfitting. The convolution filter size is 4; the max pooling is applied after the convolution layer, and dropout is applied after the dense layers.

Max pooling. Max pooling reduces the representation’s spatial size to reduce the number of parameters and the time taken for training. It helps reduce overfitting and computational costs. In this study, we use one-dimensional pooling for CNN; in this technique, the max of all given values is considered in the max-pooling process for the given pooling size.

Dense layer. A dense layer in a neural network feeds all outputs from the previous layer to all of its neurons, with each neuron providing one output to each neuron of the successive layer. We use two dense layers: one in the hidden layer with 10 neurons and one in the output layer with 1 neuron for the sentiment classification.

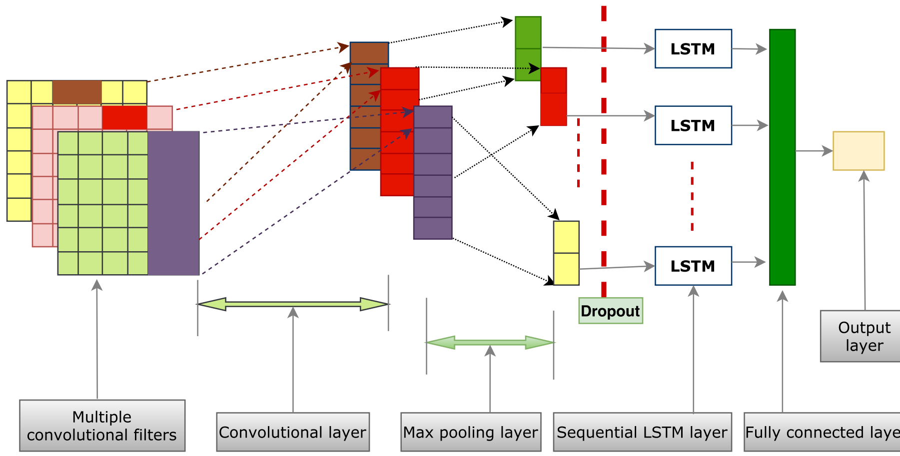

Fig. 4. Architecture of the proposed ensemble CNN-LSTM model.

Loss function and optimizer. Loss functions indicate how well a network performs the task it was intended to do; we used binary cross entropy as a loss function, which generally is used as a loss function with binary class output. In this study, we consider zero for the negative sentiment and one for the positive sentiment. The loss is calculated over the hidden dense layer and output layer. *Activation function.* ReLu is used for our model in both LSTM and CNN. To add non-linearity into the network, we use the ReLu activation function. The output of this layer is the same shape as the input shape. In the output layer for the binary classification tasks, we use sigmoid as the activation function.

3.2 The Proposed LSTM Model

This section describes the proposed LSTM model that uses the proposed CNN model's output as input parameters. In the proposed LSTM, we use a sequential LSTM layer, a fully connected layer, and an output layer. The proposed LSTM model is described as follows.

Activation function. This study used the ReLu function for both of the proposed models to classify the airline quality dataset and the Twitter dataset in the hidden layer. In the output layer, the sigmoidal function is used in binary classification for the output layer.

Dropout. We use dropout to prevent the network from overfitting. Dropout randomly sets the outgoing edges of a hidden unit to zero at each update of the training phase. We use a dropout of 0.3 to reduce overfitting while training in both proposed models (LSTM and CNN).

Batch normalization. This layer normalizes the distribution for all batch after dropout. Batch normalization reduces internal covariate shift and hence guides to convergence at a faster rate. It reduces overfitting because of its regularization effects. We use batch normalization in the LSTM network.

Dense layer. We use two dense layers: one in the hidden layer with 10 neurons and one in the output layer with 1 neuron for classification.

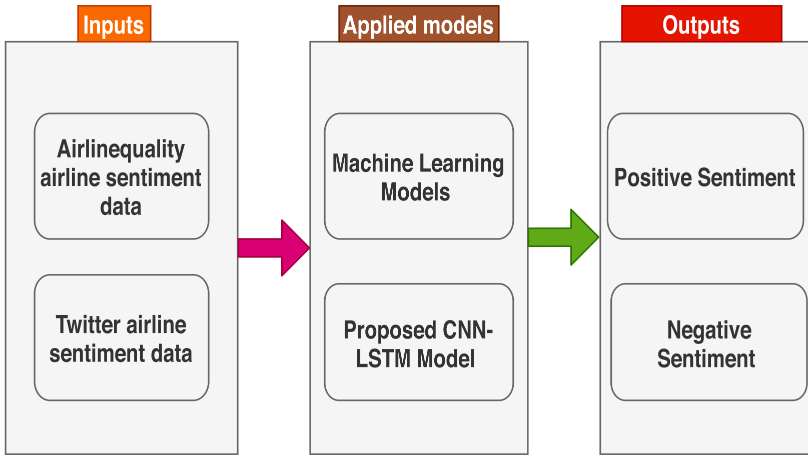

Fig. 5. Proposed research work flow diagram.

4 EXPERIMENT

4.1 Research Process

The suggested research methodology is shown in Figure 5. To explain the research queries mentioned earlier, we focus on the airline sentiment and Twitter data from different sources and classifying the consumer sentiment as positive and negative using classical ML and the proposed model (research queries 1 and 2). Following this, we introduce a classification model based on CNN and LSTM. We form the analysis of the results and show the proposed model's power compared to classical ML models (research query 3). The proposed sentiment analysis is implemented with the help of the Python programming language.

4.2 Dataset Description

4.2.1 Airlinequality Airline Sentiment Data. For our experimental analysis, the airline reviews dataset has been collected randomly from www.airlinequality.com. This online platform acts as a vast source of information over which travelers can evaluate their experience for the offered services. The randomly collected data consists of reviews provided by the airline consumer's experience during their travel. The considered data is balanced with 41, 421 recommendations and 41, 421, not recommendations. Consumer recommendation about services or products shows that they were satisfied or unsatisfied with the offered travel facilities. In this study, we consider recommendations as positive sentiment and not recommendations as negative sentiment, respectively. The collected data samples have reviews of the travelers from 85 various countries who have referred to 186 various airlines. The maximum number of total words used in this dataset is 68, 374, and the maximum number of words in a single review is 3,618.

4.2.2 Twitter Airline Sentiment Data. This dataset is taken from CrowdFlower's data for everyone's library (www.data.world/crowdflower/). The collected data includes 15 columns. We only use the reviews and the sentiment label from the dataset for all rows and utilize it for the sentiment classification job on each significant U.S. airline. Twitter airline sentiment data is scraped from February 2015, and contributors were asked first to classify positive, negative, and neutral tweets, followed by categorizing negative reasons (e.g., "late flight" or "rude service"). There 14,640 tweets abeled as positive, negative, and neutral. We took positive and negative tweets into consideration for our purpose. The maximum number of total words used in this dataset is 12,195, and the maximum number of words in a single review is 130.

4.3 Data Preprocessing

For getting hidden sentiment from the text data, text preprocessing is an essential step in NLP tasks. In text preprocessing, gathered raw text data needs few changes to convert it into a form in which ML algorithms could perform better. We cannot use text data directly for classification, so we have cleaned it by removing unwanted words, unknown words, and symbols that will not help in training. Tokenization and padding are applied to clean text data to get a sequence of numbers for each review and a word represented by a number. In this study, we consider positive sentiment as one and negative emotion as zero. We discard neutral sentiment data from our dataset because of its neutrality. If some review is more significant than the maximum length defined, it will be truncated from the back to the desired length. The dataset was preprocessed to a dictionary size of 4,000 words with a zero-padded maximum sequence of 100 words per review; any more data became insignificant to the network's objective. We used pooling here because pooling helps in reducing the complexity of computation and also controlling overfitting.

4.4 Sentiment Analysis

In assessing the power of online reviews posted by a consumer over the online platform to classify in the positive and negative class accurately, we propose a hybrid CNN-LSTM model and evaluated its performance. Toward that approach, we also implement various classical ML models over the data collected from different online platforms, such as airlinequality and Twitter. Further, we evaluate applied traditional ML models with the proposed hybrid model. Finally, we validate our obtained results with a suggested evaluation methodology in ML to confirm that the obtained results are accurate and do not influence overfitting.

4.5 Evaluation Measures

For the validation of the obtained results on training data, we have used accuracy, precision, recall, and F-measure [20]. All of these calculative performance measurements are shown in Equations (1) through (4). Table 2 represents the relationships among these variables and all of the samples.

Accuracy. Accuracy gives the measure of the correct number of predictions from the entire predictions. It provides general performance measures. However, we need other means to get a better view of the result.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

Precision. Precision is the ratio of all positives examples that are correctly classified to all of the positives classified samples.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

Recall. Recall is the ratio of the positive samples that are correctly classified to the total number of positive samples.

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

Table 2. Relation Among Evaluatory Variables

Predicted	**Actual**	
	Positive	**Negative**
Positive	TP	FP
Negative	FN	TN

TP, true positive; FP, false positive; FN, false negative; TN, true negative.

F-measure. F-measure is the harmonic mean of precision and recall. It is an indicator of the relationship between recall and precision.

$$F1{-}measure = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{9}$$

5 RESULTS ANALYSIS

The sentiment analysis results in the case of airlinequality airline review samples are shown in Table 3. The obtained results show that the classifier builds upon the consumer sentiment using various ML approaches (Classifiers A-D), as well as the different DL models (E and F), have a comparable classification performance in terms of accuracy (up to 85.2% in DL). A further classifier based upon the proposed hybrid CNN-LSTM model (Classifier G) exhibits higher efficiency when classifying the consumer sentiment as positive or negative hidden into the raw data collected from the online platform (up to 90.2%). These obtained results show that the proposed hybrid model based on CNN and LSTM performs better in sentiment analysis.

The sentiment analysis results in the case of Twitter airline review samples are shown in Table 4. The results analysis represents the classification done based on consumer sentiment using the classical ML approach in Classifiers A-D. Simultaneously, Classifier E and F's various DL models may have a comparable classification accuracy (up to 88.2% in the case of DL). A different classifier based upon the proposed hybrid CNN-LSTM model (Classifier G) exhibits higher efficiency when classifying the consumer sentiment as positive or negative hidden into the raw data collected from the online platform (up to 91.3%). These obtained results analyses represent that the proposed hybrid model based on CNN and LSTM performs better in sentiment analysis.

Regarding the various classical ML models, our findings are that LR performs slightly higher than all other classifiers such as DT, SVM, and NB in sentiment analysis from the unstructured data collected from the online platform. Interestingly, classifiers considered that the proposed DL-based hybrid model outperforms other single DL models or classical ML models. Furthermore, precision and recall for the proposed model are also better. Finally, we may conclude that the online reviews' qualitative content is essential in accessing considerations that do not know consumer sentiment regarding their satisfaction. Such a proposed model may be utilized to classify the consumer content into a positive or negative attitude accurately.

6 DISCUSSION

6.1 Discussion of the Findings

A significant metric is consumer sentiment analysis, the attitude indicator of customer satisfaction, and their behavioral measure's definitive outcome. Organizations have to take care of consideration for a long- or short-term goal in consumer policy making. This study has tried to classify positive sentiment and negative sentiment based on the user-generated feedback data. In

Table 3. Comparisons of Accuracy, Precision, Recall, and F-measure on Airlinequality Airline Sentiment Data

Model	Algorithm	Accuracy	Precision	Recall	F-measure
A	LR	84.5	84.3	86.8	86.0
B	NB	83.2	83.2	83.2	83.2
C	DT	78.1	82.3	76.5	79.4
D	SVM	84.3	84.5	84.6	83.7
E	CNN	85.0	83.2	81.1	82.3
F	LSTM	85.2	86.5	86.1	86.3
G	Proposed CNN-LSTM	90.2	87.8	87.0	87.6

Table 4. Comparisons of Accuracy, Precision, Recall, and F-measure on Twitter Airline Sentiment Data

Model	Algorithm	Accuracy	Precision	Recall	F-measure
A	LR	88.2	87.1	89.3	88.0
B	NB	87.1	84.6	89.6	82.7
C	DT	78.1	82.3	76.5	77.8
D	SVM	80.2	76.4	82.4	81.5
E	CNN	87.1	83.0	81.1	82.5
F	LSTM	88.2	87.1	86.0	86.2
G	Proposed CNN-LSTM	91.3	87.8	87.0	87.5

our implementation, we considered qualitative measures provided inside the online reviews. We used various classical ML techniques and NLP techniques to get insights from the user-generated qualitative review content. We also proposed a hybrid model based on DL techniques such as CNN and LSTM. In our implementation, we first extracted consumer sentiment insights from online reviews. We used such ideas generated from online reviews in classifying the sentiment into positive or negative content. Finally, we compared classical ML models with the proposed model in sentiment analysis.

Research query 1: What is the research objective of this study?
Findings. Consumer sentiment analysis is the primary objective of this study; we collected user-generated reviews from the Airlinequality and Twitter online platforms in the field of travel and tourism, particularly airline reviews. We implemented TM and NLP techniques over qualitative content; these insights were utilized to classify the sentiment into positive or negative.

Research query 2: Can the opinion expressed by the reviewer in online reviews be classified as positive or negative?
Findings. The results obtained indicate that consumer sentiment is correlated with qualitative information, and we can extract such secret sentiment from online reviews. Extraction of the consumer sentiment can be possible with TM and NLP processing techniques. Derived opinion can be classified as positive or negative with the help of classical ML models. During this study, we found that consumer sentiment may benefit from the recommendation of various services based on online reviews.

Research query 3: What is the power of the proposed model compared to an existing model?

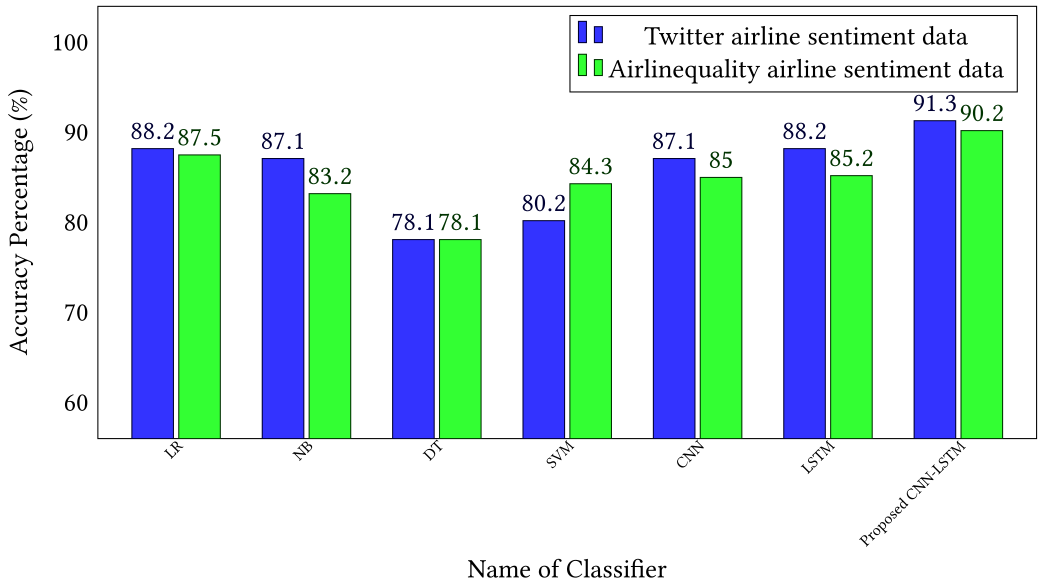

Fig. 6. Accuracy comparison of existing models with the proposed model for the considered datasets.

Findings. In the proposed research work, qualitative features included consumer sentiment analysis, and consumer sentiment regarding various offerings depends on these factors. In this study, we evaluated consumer sentiment based on the textual information present in the online reviews. For the implementation and evaluation of the research goal, we proposed a hybrid CNN-LSTM model. We also presented a performance comparison of the proposed model with the classical ML models shown in Tables 3 and 4. Our result analysis concluded that the proposed model outperforms other ML models in the process of consumer sentiment analysis shown in Figure 6.

6.2 Theoretical and Managerial Contribution

This study identified the various importance of the information available in consumer feedback while explaining consumer sentiment as an attitudinal construct and sentiment analysis. This study proposed a hybrid model based on CNN and LSTM DL models in sentiment classification as positive or negative. In this context, we implemented our model on two different datasets collected from various online sources. We applied qualitative review content as input to the ML model and proposed model. Our experiment results were evaluated based on the evaluation measures mentioned in the literature. We can conclude from the results that the proposed model outperforms other ML models in sentiment analysis.

This study gives a managerial direction, which will be significant for the services provider organizations in travel and tourism to serve their consumers better. This study finds that managers have to concentrate on consumers' qualitative content in their online feedback. This research work is also drawing the intention toward the power of qualitative content in building consumer sentiment. This study also points toward the method of feedback collection and getting insights to understand consumer opinion. This study can help bring ideas from online reviews and a quick understanding of consumer sentiment. The proposed model can be considered in any service provider organization, which can help classify consumer review and business growth and profitability based on positive feedback.

7 CONCLUSION AND FUTURE WORK

The pervasiveness of the ICT has provided us with a large amount of textual information that can be utilized to find great and valuable insights. Nowadays, many organizations understand the benefits of sentiment analysis, which helps them in detecting the textual emotional tone and offers critical insights. Also, consumer sentiment plays a vital role in organizational growth. In evaluating consumer sentiment using online feedback, DL techniques have been implemented in many research works. This paper has suggested an ensemble model for sentiment analysis, using CNN and LSTM models. In this research, the word in the online reviews converted into vectors of numbers with Keras embedding; the obtained embedding feeds to the CNN as an input vector. Further, the output from the CNN model has been applied as an input to the LSTM model. Our experimental analysis demonstrates high accuracy compared with other ML models. We may also conclude that online reviews may be utilized in consumer recommendation decisions for the forthcoming consumer because both are highly related in nature.

This study provides many future research pathways; as this study only focused on the sentiment analysis of online reviews of airline data, many travel and tourism industry areas can be considered to the future disentangling consumer sentiment. Further, as a particular aspect may also influence consumer sentiment that does not evaluate this research, a specific aspect may be considered in the subsequent investigation. In future studies, various TM algorithms can be implemented on online review data, and decent performance can be evaluated. Future research could be carried out on survey-based data and online data, and a comparison of the results obtained in both cases can be compared. In this study, we have only considered data obtained from an online platform that is in the form of English sentences. Thus, the consumer reviews written in other languages does not include in sentiment analysis; further study may consist of textual reviews provided in other languages or a combination of various languages. Consumer sentiment may vary based on many features such as staying, spoken language, airline logo, airline facilities, etc. In a further study, consumer sentiment can also be concluded based on these parameters. We have proposed a CNN-LSTM hybrid classifier for sentiment classification; other hybrid models can better solve this problem. Finally, future research can implement various optimization techniques with ML models in sentiment analysis.

REFERENCES

[1] Laith Abualigah, Hamza Essam Alfar, Mohammad Shehab, and Alhareth Mohammed Abu Hussein. 2020. Sentiment analysis in healthcare: A brief review. In *Recent Advances in NLP: The Case of Arabic Language*. Springer, 129–141.

[2] Ibrahim Said Ahmad, Azuraliza Abu Bakar, and Mohd Ridzwan Yaakub. 2020. Movie revenue prediction based on purchase intention mining using YouTube trailer reviews. *Information Processing & Management* 57, 5 (2020), 102278.

[3] Mohammad Al-Smadi, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh. 2019. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *International Journal of Machine Learning and Cybernetics* 10, 8 (2019), 2163–2175.

[4] Oscar Araque, Ignacio Corcuera-Platas, J. Fernando Sánchez-Rada, and Carlos A. Iglesias. 2017. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications* 77 (2017), 236–246.

[5] Snehasish Banerjee and Alton Y. K. Chua. 2016. In search of patterns among travellers' hotel ratings in TripAdvisor. *Tourism Management* 53 (2016), 125–131.

[6] Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems* 31, 2 (2016), 102–107.

[7] Iti Chaturvedi, Erik Cambria, Roy E. Welsch, and Francisco Herrera. 2018. Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion* 44 (2018), 65–77.

[8] Liang-Chu Chen, Chia-Meng Lee, and Mu-Yen Chen. 2019. Exploration of social media for sentiment analysis using deep learning. *Soft Computing* 24 (2019), 8187–8197.

[9] Mu-Yen Chen and Ting-Hsuan Chen. 2019. Modeling public mood and emotion: Blog and news sentiment and socio-economic phenomena. *Future Generation Computer Systems* 96 (2019), 692–699.

[10] Mu-Yen Chen, Hsiu-Sen Chiang, Edwin Lughofer, and Erol Egrioglu. 2020. Deep learning: Emerging trends, applications and research challenges. *Soft Computing* 24, 11 (2020), 7835–7838.

[11] Po-Jen Chen, Jian-Jiun Ding, Hung-Wei Hsu, Chien-Yao Wang, and Jia-Ching Wang. 2017. Improved convolutional neural network based scene classification using long short-term memory and label relations. In *Proceedings of the 2017 IEEE International Conference on Multimedia and Expo Workshops (ICMEW'17)*. IEEE, Los Alamitos, CA, 429–434.

[12] Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, and Sweta Tiwari. 2016. Sentiment analysis of review datasets using naive Bayes and K-NN classifier. arXiv:1610.09982.

[13] Jiachen Du, Lin Gui, Ruifeng Xu, and Yulan He. 2017. A convolutional attention model for text classification. In *Proceedings of the National CCF Conference on Natural Language Processing and Chinese Computing*. 183–195.

[14] Ashraf Elnagar, Ridhwan Al-Debsi, and Omar Einea. 2020. Arabic text classification using deep learning models. *Information Processing & Management* 57, 1 (2020), 102121.

[15] Maryam Esmaeili and Alberto Vancheri. 2010. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. IEEE/WIC/ACM.

[16] Imene Guellil and Kamel Boukhalfa. 2015. Social big data mining: A survey focused on opinion mining and sentiments analysis. In *Proceedings of the 2015 12th International Symposium on Programming and Systems (ISPS'15)*. IEEE, Los Alamitos, CA, 1–10.

[17] Antonio Gulli and Sujit Pal. 2017. *Deep Learning with Keras*. Packt Publishing Ltd.

[18] Yue Guo, Stuart J. Barnes, and Qiong Jia. 2017. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Management* 59 (2017), 467–483.

[19] Viktor Hangya and Richárd Farkas. 2017. A comparative empirical study on social media sentiment analysis over various genres and languages. *Artificial Intelligence Review* 47, 4 (2017), 485–505.

[20] Fatemeh Hemmatian and Mohammad Karim Sohrabi. 2019. A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review* 52 (2019), 1495–1545.

[21] Charles F. Hofacker, Edward Carl Malthouse, and Fareena Sultan. 2016. Big data and consumer behavior: Imminent opportunities. *Journal of Consumer Marketing* (2016).

[22] Stefan Jackson and John Tozer. 2020. A vision for data science at British Airways. *Impact* 2020, 1 (2020), 15–19.

[23] Beakcheol Jang, Inhwan Kim, and Jong Wook Kim. 2019. Word2Vec convolutional neural networks for classification of news articles and tweets. *PLoS One* 14, 8 (2019), e0220976.

[24] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. arXiv:1404.2188.

[25] Yoon Kim. 2014. Convolutional neural networks for sentence classification. arXiv:1408.5882.

[26] Akshi Kumar and Geetanjali Garg. 2019. Systematic literature review on context-based sentiment analysis in social multimedia. *Multimedia Tools and Applications* 79 (2019), 15349–15380.

[27] Sachin Kumar and Mikhail Zymbler. 2019. A machine learning approach to analyze customer satisfaction from airline tweets. *Journal of Big Data* 6, 1 (2019), 62.

[28] Jin-Ah Kwak and Sung Kyum Cho. 2018. Analyzing public opinion with social media data during election periods: A selective literature review. *Asian Journal for Public Opinion Research* 5, 4 (2018), 285–301.

[29] Pei-Ju Lee, Ya-Han Hu, and Kuan-Ting Lu. 2018. Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics* 35, 2 (2018), 436–445.

[30] Xiao-Qin Liu, Qiu-Lin Wu, and Wen-Tsao Pan. 2019. Sentiment classification of micro-blog comments based on randomforest algorithm. *Concurrency and Computation: Practice and Experience* 31, 10 (2019), e4746.

[31] Kigon Lyu and Hyeoncheol Kim. 2016. Sentiment analysis using word polarity of social media. *Wireless Personal Communications* 89, 3 (2016), 941–958.

[32] Trupthi Mandhula, Suresh Pabboju, and Narsimha Gugulotu. 2019. Predicting the customer's opinion on Amazon products using selective memory architecture-based convolutional neural network. *Journal of Supercomputing* 76 (2019), 5923–5947.

[33] Mika V. Mäntylä, Daniel Graziotin, and Miikka Kuutila. 2018. The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review* 27 (2018), 16–32.

[34] Fenna Miedema. 2018. *Sentiment Analysis with Long Short-Term Memory Networks*. Vrije Universiteit Amsterdam.

[35] Xi Ouyang, Pan Zhou, Cheng Hua Li, and Lijun Liu. 2015. Sentiment analysis using convolutional neural network. In *Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. IEEE, Los Alamitos, CA, 2359–2364.

[36] Tim O'Reilly and John Battelle. 2004. Opening welcome: State of the internet industry. San Francisco, California, October 5 (2004).

[37] B. Pang and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2, 1–2 (2008), 1–135.

[38] Seungwan Seo, Czangyeob Kim, Haedong Kim, Kyounghyun Mo, and Pilsung Kang. 2020. Comparative study of deep learning-based sentiment classification. *IEEE Access* 8 (2020), 6861–6875.

[39] Eren Sezgen, Keith J. Mason, and Robert Mayer. 2019. Voice of airline passenger: A text mining approach to understand customer satisfaction. *Journal of Air Transport Management* 77 (2019), 65–74.

[40] Michael Siering, Amit V. Deokar, and Christian Janze. 2018. Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews. *Decision Support Systems* 107 (2018), 52–63.

[41] Tristan Stérin, Nicolas Farrugia, and Vincent Gripon. 2017. An intrinsic difference between vanilla RNNs and GRU models. In *Proceedings of the 9th International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE'17)*. 76–81.

[42] Farideh Tavazoee, Claudio Conversano, and Francesco Mola. 2019. Recurrent random forest for the assessment of popularity in social media. *Knowledge and Information Systems* 62 (2019), 1847–1879.

[43] Chih-Fong Tsai, Kuanchin Chen, Ya-Han Hu, and Wei-Kai Chen. 2020. Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management* 80 (2020), 104122.

[44] Mikalai Tsytsarau and Themis Palpanas. 2012. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery* 24, 3 (2012), 478–514.

[45] Ge Wang, Pengbo Pu, and Yongquan Liang. 2018. Topic and sentiment words extraction in cross-domain product reviews. *Wireless Personal Communications* 102, 2 (2018), 1773–1783.

[46] Haibing Wu and Xiaodong Gu. 2015. Towards dropout training for convolutional neural networks. *Neural Networks* 71 (2015), 1–10.

[47] Feng Xu, Zhenchun Pan, and Rui Xia. 2020. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Information Processing & Management* 57, 5 (2020), 102221.

[48] Qiang Ye, Huiying Li, Zhisheng Wang, and Rob Law. 2014. The influence of hotel price on perceived service quality and value in e-tourism: An empirical investigation based on online traveler reviews. *Journal of Hospitality & Tourism Research* 38, 1 (2014), 23–39.

Received October 2020; revised January 2021; accepted March 2021