Sentiment Analysis of Movie Reviews Using Improved Word2Vec and CNN

Azhar Imran
Department of Creative Technologies
Air University
Islamabad, Pakistan
azhar.imran@mail.au.edu.pk

Khattab M Ali Alheeti
Department of Computer Networking Systems
Al Baha University
Ramadi, Iraq
co.khattab.alheeti@uoanbar.edu.iq

Abdulkareem Alzahrani
Faculty of Computer Science&IT
University of Anbar
Al Baha, Saudi Arabia
ao.alzahrani@bu.edu.sa

Muhammad Imran Khan
Faculty of Computing
Riphah International University
Islamabad, Pakistan
imran.fc@riphah.edu.pk

2023 20th International Bhurban Conference on Applied Sciences and Technology (IBCAST) | 979-8-3503-0825-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IBCAST59916.2023.10712797

***Abstract*— The concept of sentiment analysis is spreading across multiple fields including business, politics, movies, religion, academia, etc. Sentiment analysis is a very challenging job, and the accurate prediction of sentiments is very problematic. Convolutional neural networks (CNN) and an enhanced word vector (Word2Vec) were utilized to do sentiment analysis on movie reviews. Word embedding is created in the initial stage using Word2Vec and Latent Dirichlet Allocation (LDA). Then CNN is implemented on the pre-trained word vectors for binary class sentiment analysis of movie reviews. The suggested model's accuracy, sensitivity, and specificity were tested on a publicly available dataset from Stanford University (Movie Review Database), and they were 0.924, 0.932, and 0.924, respectively, And The results indicate that the improved Word2Vec is not only used to gain the word vectors but also improves the feature representation ability of the model by fusing both the local and global connections. The proposed method outperformed all the baseline methods for sentiment analysis of movie reviews.**

Keywords*— *Movie reviews, sentiment analysis, word2vec, text classification.

I. INTRODUCTION

Sentiments may be described as opinions, feelings, emotions, ideas, or judgments aroused by awareness, feelings, and emotions [2]. Sentiment analysis (referred to as sentiment mining or opinion mining) is used to solve the problems of opinions within text, sentiments, and the computational handling of subjective [1]. It may be defined as the procedure of computationally classifying and categorizing views in a text to predict a user's attitude on a particular product and topic. This attitude can be negative, positive, or neutral based on the comments or posts.

The sentiment classification problem has gained worldwide popularity nowadays on various topics. Sentiment analysis is commonly used in business, medical, sports, weather, sociology, psychology, etc. to predict the nature of expressed opinion within the text. Furthermore, the subjective context does not employ to infer that something is wrong [3]. According to the survey, about 90% of information on social media applications is unstructured as shown in Fig.1. The mining of such an application is necessary for decision-making and opinion mining [29].

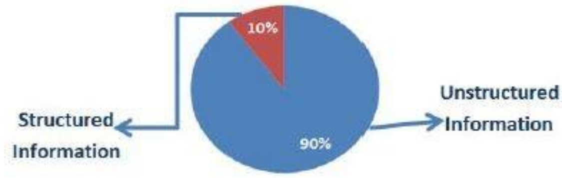

Fig. 1. Status of Social Media Information.

Several techniques are used to convert unstructured text into structured text such as information extraction [10], topic tracking [19], summarization [8] and clustering [13], etc. Existing sentiment analysis approaches can be broadly grouped into four major categories such as dictionary-based method [9], statistical-based method [24], semantic-based method [18] and machine learning method [32]. Most of the baseline methods work in isolation for word-level, topic-level, and document level text classification.

Recently, deep learning has gained worldwide attention for various tasks including natural language processing [6], speech recognition [7] and image classification [12]. CNN is the class of deep learning, which is comprised of input, output and various hidden layers [15]. Word2Vec is an open-source tool for Google. This is commonly used to produce word embedding. The existing Word2Vec only focuses on semantic level features [14]. The latent Dirichlet allocation (LDA) is a probabilistic model for text classification [4].

In this article, we suggested a novel combination of Word2Vec and LDA to effectively represent the matrix model. The improved Word2Vec couldn't only characterize semantic features but also signify the contextual features and improve the feature representation ability of the model. After creating the feature matrix which is fed into the CNN model for binary class sentiment analysis of movie reviews.

The rest of the essay is organized as follows. While Section 3 addresses Materials and Methodology, Section 2 illustrates Related Work. The results and discussion are in Section 4, and the conclusion is in Section 5.

II. RELATED WORK

Although sentiment analysis and opinion mining are the primary sources of business decision-making, still there are a lot of challenges that require further attention. The movie quality can be assessed with the help of movie reviews and their numeric rating. These two features further assist to find qualitative insight into the movie in various facets. The mining of opinions for movie reviews is quite difficult as the English language is itself very ambiguous and complex with different interpretations for one word. Different studies have been proposed earlier to predict whether a movie is worth watching or whether it matches the user's interest or not. The movie rating helps to assess the content of the film. This also gives control to parents of what their kids are viewing.

A. Traditional Approaches to Sentiment Analysis

The four main methods for implementing sentiment analysis like dictionary-based [9], semantic-based [18], statistical-based [24], and machine learning approach [32]. The schematic of these approaches is shown in Fig.2.

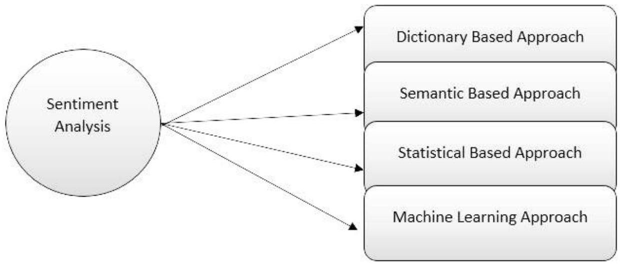

Fig. 2. Based on the notion employed, there are four methods for sentiment analysis, such as (a) Dictionary Based Approach. (b) Semantic-Based Methodology. (b) Using a Statistical Approach (d) The use of machine learning.

Thet et al. proposed a dictionary-based approach to movie reviews using a natural language toolkit (NLTK) [27]. The sentiment strength and orientation have been analyzed by the viewer in several facets of the film. The methods use 32000 words from the SentiWordNet dictionary. Bhoir et al. used SentiWordNet to find out the subjectivity of the sentences [11]. The various aspects of the movie were extracted using the proposed approach to facilitate users. Mahmud et al. used a statistical approach combined with a support vector machine to movie sentiments [21]. The differentiation between words' polarity whether it is negative or positive is performed using SVM, while statistical reasoning has been employed to predict movie success. The proposed technique was assessed using an RBF kernel on the IMDb dataset. Turney et al. suggested using Alta Vista search engine co-occurrence frequencies [28]. The point-wise mutual information (PMI) was constructed for semantic orientation. This technique was categorized into two phases. The phrases were extracted in the first phase while semantic orientation was estimated in the second phase.
Richard Vlach et al [30], devised a tool using this semantic web technology to extract multiple sources of data into a single cluster. The sentiment analysis of Chinese movie reviews was performed using semantic labels in [33]. The emotions were divided into three categories: neutral, positive, and negative. Each review text served as a sentiment phrase sequence from which two sentiment labels were retrieved, one using a semantic analysis method and the other using a machine learning method. Zhang et al. [31] have practiced the word reliance structure to group the sentiments using a semantic-based approach. Zhou et al. [5] analyzed movie reviews using both machine learning and a semantic-based approach [34-38]. The author has used text categorization and supervised classification techniques to segment movie reviews. A corpus was created to test the various classifiers.

B. Neural Network-Based Approaches of Sentiment Analysis

Although, there are a lot of neural network methodologies that have been employed for sentiment analysis of movie reviews, dealing with sentiment challenges including domain independence, and negation rule handling is the least targeted area for the researchers.
For sentence-level classification challenges, Yoon Kim carried out several experiments using CNN-trained over pre-trained word vectors [16]. They demonstrated how well a simple CNN that uses static vectors and little hyperparameter tweaking performs on a variety of benchmarks. CNN was used to complete the sentiment analysis and question categorization tasks. [26] employed a deep CNN model to extract information from movie reviews that ranged from character-level information to sentence-level information. The approaches were evaluated using two datasets: the Stanford Twitter Sentiment corpus (STS) and the Stanford Sentiment Treebank (SSTb). For the SSTB and STS Datasets, the proposed model's accuracy was 85.7% and 86.4%, respectively. An experimental comparison between SVM and ANN was performed for the sentiment categorization task and to obtain better accuracies. The unigram bag of words approach was carried out to evaluate both models. Both models have been evaluated on Pang and Lee movie reviews dataset in which ANN has shown significant results on unbalance data.
Lai et al. [17] have used RCNN without human-designed features for the classification of text. The related information was captured during learning the word illustration using recurrent structure. This method shows a good result with less noise to capture information, unlike existing window based NN methods. Ombabi et al. [22] developed a DL framework incorporated with CNN and Word2Vec for users' interests and opinion classification. The movie review corpus and Sandertwitter sentiment corpus were used. For the sentiment analysis of movie reviews, Ouyang et al. suggested a framework ensemble using Word2Vec and CNN [23]. In the first phase, Word2Vec was employed as an input of CNN to compute word distance and gain the word vector representations. Then a seven-layer CNN model was used for the prediction of sentiment analysis.

There are multiple approaches that have been proposed to deal with movie review sentiment analysis. Most of the existing approaches ignore the semantic feature expression and extract only the word vector matrix of the text granularity level. We proposed a novel approach that is the combination of improved" Wod2Vec with LDA model" and" Convolutional Neural Network". Word2Vec with LDA is designed to create a new text feature representation matrix, which is fed to CNN for movie reviews sentiment classification.

III. MATERIALS AND METHODS

The following steps—dataset, data preparation, data exploration, model construction, and evaluation—have been divided into the assessments of SA's films. Fig. 3 depicts the proposed model schematic

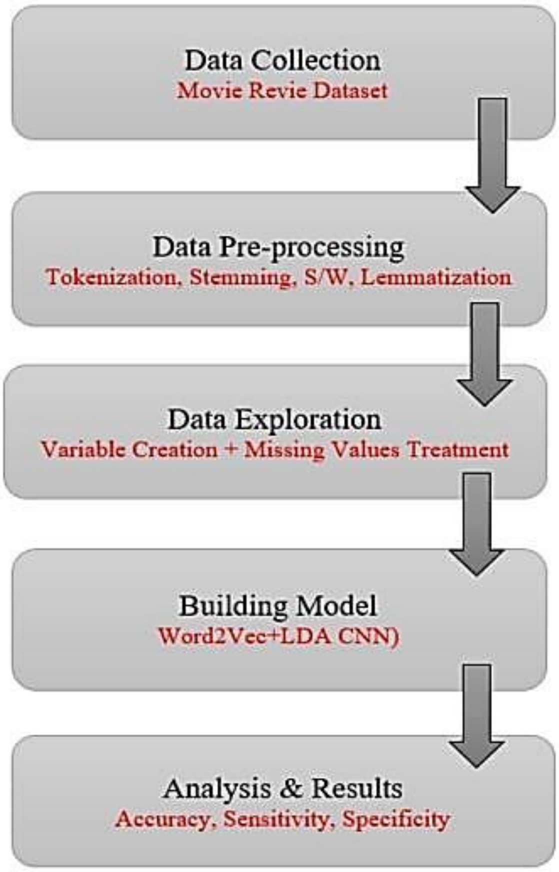

Fig. 3. Proposed methodology of movie review SA using improved Word2Vec and CNN.

In the first phase, the movie review dataset is collected from Stanford University. Then, the preprocessing steps are used to raise the dataset's quality. Manual data exploration is used to address missing values and abnormalities. The proposed model is developed in a combination with Word2Vec and Latent Dirichlet Allocation to gain the word vector and improve feature representation and CNN is employed for binary class sentiment classification of movie reviews. Finally, statistical metrics like accuracy, sensitivity, and specificity are employed to gauge how well the suggested strategy performs.

A. Data

A movie review dataset comprised of 50000 reviews with labelled sentiment polarity was obtained from Stanford University [20]. The dataset was subsequently segmented into training and testing datasets of 25000 polar reviews each for binary class movie review sentiment classification. We also added a long list of positive and negative words (PN Lists). On this list, there are 2006 positive words and 4783 bad terms. Table 1 contains information about the dataset.

TABLE I

DATASET STATISTICS

Dataset	Positive Reviews	Negative Reviews
MRD	25000	25000
PN Lists	2006	4783
Total	27006	29783

B. Preprocessing

Tokenization, stemming, lemmatization, stop word removal, and noise removal are examples of preprocessing steps that can be used to create time-efficient classifiers [25].

- Tokenization: It is the process of dividing the text into tokens. Words, phrases, symbols, and other meaningful text are used as tokens.
- Stemming: It is the process of removing prefixes, affixes, and infixes to reveal the word's root.
- Lemmatization: This is somehow relevant to stemming but differs concerning canonical forms. Stemming fails to give the original form of better/best, and lemmatization gives its canonical form.
- Stop Words Removal: The basic purpose of the removal of English language stop words is to extract the necessary words for sentiment analysis.
- Noise Removal: This is the preliminary step of text preprocessing which includes HTML tags elimination, white spaces, numbering, and punctuation elimination.

C. Data Exploration and Visualization

It is the initial phase in information examination and commonly includes the principal attributes of a dataset. The main steps of data exploration include the treatment of missing values and outliers, variable creation, identification and transformation, and univariate and bivariate analysis of the text.

We have employed a manual technique to treat the missing values because without missing values treatment it often leads to false prediction or classification.

D. Proposed Model

In this study, a novel model for sentiment analysis of movie reviews was proposed. The following section goes through the suggested model's implementation specifics.

1) *Improved Word2Vec:* In our study, we employ the enhanced Word2Vec in conjunction with Latent Dirichlet Allocation (LDA) to obtain the vector as input for the word. The text is trained by utilizing a word vector to build an N-dimensional vector. After the training of the text dataset, the LDA model is applied to obtain a probability model corresponding to the text. The words in the document satisfy the Dirichlet polynomial distribution along with hyperparameters α, and β for document level and word level distribution, respectively. The mathematical expression of probability model ρ with θ document topic distribution containing 'N' words is given in equation (1)

$$\rho(\theta, z, w/\alpha, \beta) = \rho(\theta/\alpha) \prod_{n=1}^{N} (z_n/\theta)\rho(w_n, /z_n, \beta) \quad (1)$$

2) Convolutional Neural Network: On the Word2Vec-acquired trained word vector, we applied a Convolutional Neural Network. In order to implement our model, we employed the input layer, embedding layer, convolutional layers, pooling layers, and fully connected layers. For all convolution layers and max-pooling layers, respectively, we used kernel sizes of 3*3 and 2*2 and stride sizes of 1*1 and 2*2. The proposed technique is composed of three convolution layers, three pooling layers, and two fully connected layers. We employed the SoftMax function on the fully connected layer to obtain a binary classification of the movie reviews. After each convolution layer, we employed the rectified linear unit (ReLU) activation function, and the Dropout layer is used before the final classification layer. The value of dropout is set to 0.5 to overcome the issue of overfitting. The employed hyper-parameter details are given in Table 2.

In the proposed model, we employed a 300-dimensional word embedding model Word2Vec in combination with the document topic model LDA to effectively represent the matrix model. The improved Word2Vec with default parameter settings couldn't only characterize semantic features but also signify the contextual features and improve the feature representation ability of the model. After creating the feature matrix which is fed into the CNN model for binary class sentiment analysis.

The sentence s given in input is decomposed into word vectors $s = w_1, w_2, ..., w_n$. We mapped each word w_i into its embedding vector x_i while allocating input sentences. A matrix is represented in equation (2) which contains sentences of length n and L denotes the concatenation operator.

$$x_{1:n} = x_1 \oplus x_2 \oplus x_3 \oplus .... x_n \quad (2)$$

We used a nonlinear hyperbolic function f and the bias factor b from equation (3) to implement the convolutional filter w to words h.

$$C_i = f(w.x_{i=i+h-1} + b) \quad (3)$$

In equation (4), a feature map is produced for a sentence with length n.

$$C = C_1 + C_2 ..... + C_{n-h+1} \quad (4)$$

To create the maximum value, we used the max pooling on the feature map C $c^\wedge = max(c)$. The important features are extracted with the help of maximum function. The m size feature vector $c^\wedge$ is acquired using multiple filters given in equation 5.

$$C_\wedge = C_{1\wedge}, C_{2\wedge}, ...C_{\wedge m} \quad (5)$$

IV. RESULTS AND DISCUSSIONS

We applied various classifiers to evaluate their performance on the public movie review dataset. The results of these classifiers are recorded under different samples (from 100 to 25000). The parameter settings are varied to get the maximum results, and the final hyper-parameters are selected as per Table 2.

A. Result Analysis

On the Stanford Movie Reviews dataset, the proposed ensemble of improved Word2Vec and CNN is compared to traditional machine learning methods. As shown in Table 3, model accuracy increases as the training set size increases. Table 3 shows the reviews in the training set as well as the accuracy of the support vector machine (SVM), Naive Bays (NB), and the proposed method.

As shown in Table 3, the model's accuracy improves with a larger training set. Table 3 displays the reviews in the training set as well as the precision of the suggested approach, Naive Bays, and the support vector machine (SVM). Positive and negative sentiment prediction is divided into two categories. Figure 4 depicts a graphical representation of the accuracy comparison.

The suggested method's results are compared to those of other standard machine learning approaches, such as baseline CNN and Word2Vec-CNN. The upgraded Word2Vec-CNN method's effectiveness is measured using performance metrics such as recall, accuracy, and F-measure. Table 4 displays the comparison results, and Figure 5 depicts the visual analysis. The Word2Vec-CNN attained precision, recall, and f-measure values of 0.855, 0.867, and 0.845, as shown in Table 4. The performance measures have values that are about 10% higher than the baseline CNN method, with values of 0.836, 0.834, and 0.835, respectively. The proposed method's average precision, recall, and f-measure values are 0.917, 0.932, and 0.924, respectively, which has about 7% better classification rate than the traditional Word2Vec-CNN method. The classification results are improved because our method has strong feature representation ability which may provide both the semantic and syntax level granularity. The experimental findings of this work show that, by combining local and global relationships, our method is effective for sentiment analysis on movie reviews.

TABLE II HYPERPARAMETER CONFIGURATIONS

Dataset	Input Size	Iterations	Momentum	Learning Rate	Hidden Layers	Output Classes
MRD	53*300	14500	0.9	0.001	13	2

TABLE III TRAINING REVIEWS IN THE EXPERIMENT WITH ACCURACY

Sr#	Reviews	NB	SVM	Proposed Method
1	100	47.64	52.48	62.50
2	200	55.77	59.29	67.11
3	500	59.43	64.06	71.70
4	1000	63.15	69.81	75.96
5	2000	67.67	73.22	78.36
6	5000	69.52	76.17	81.28
7	8000	72.88	79.82	84.05
8	10000	74.14	81.24	87.91
9	20000	75.34	83.11	89.11
10	25000	77.68	84.42	92.4

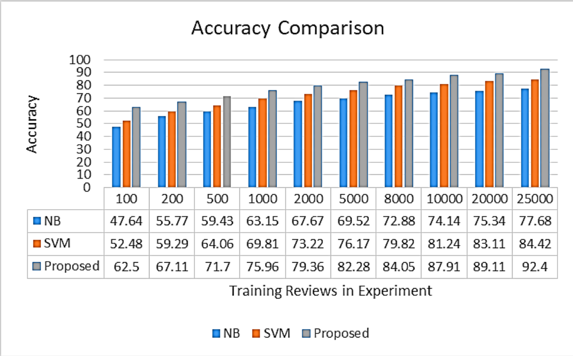

FIG. 4. THE ACCURACY COMPARISON OF BASELINE METHODS SUCH AS THE PROPOSED APPROACH, NAÏVE BAYS (NB), AND SUPPORT VECTOR MACHINE (SVM) IN COMBINATION WITH WORD2VEC, LDA, AND CNN

B. Comparative Analysis with state-of-the-art Methods

The findings of the suggested method are compared to certain state-of-the-art methods presented in the related work section. A convolution-gated recurrent neural network (Conv-GRNN), a randomly initialized convolutional neural network (CNNRand), a recurrent convolutional neural network (RCNN), and a Word2Vec-CNN combined model are compared to the suggested technique. Table 5 shows the performance outcomes of our model and the baseline approaches. Figure 6 depicts a graphical illustration of this comparison.

TABLE IV COMPARISON OF THE PROPOSED METHOD WITH BASELINE METHODS

Method	Precision	Recall	F-measure
SVM	0.755	0.760	0.756
NB	0.768	0.770	0.769
Baseline CNN	0.836	0.834	0.835
Word2Vec-CNN	0.855	0.867	0.845
Proposed Method	0.917	0.932	0.940

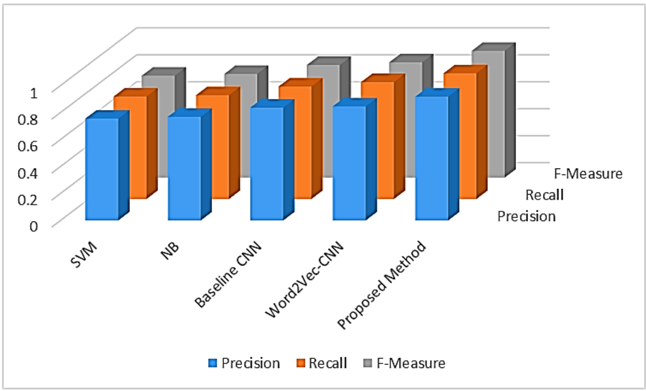

Fig. 5. In terms of precision, recall, and f-measure, the performance of the suggested approach tested on MRD is compared with SVM, NB, Baseline CNN, and Word2Vec-CNN.

TABLE V COMPARATIVE ANALYSIS FOR MOVIE REVIEWS SENTIMENT ANALYSIS

Ref.	Method	Accuracy	Sensitivity	Specificity
[16]	CNN-rand	0.452	0.474	0.455
[26]	Conv-GRNN	0.671	0.651	0.653
[17]	RCNN	0.801	0.782	0.797
[22]	Word2Vec-CNN	0.821	0.798	0.835
Our Method	Word2Vec+LDA-CNN	0.924	0.932	0.924

CONCLUSION

The concept of sentiment analysis is spreading across multiple fields including business, politics, movies, religion, academia, etc. Sentiment analysis is a very challenging job and accurate prediction of sentiment is very difficult. In our research, we performed sentiment analysis on movie reviews by using improved Word2Vec and CNN models. The performance of the suggested model is assessed using a public dataset from Stanford University, which is reviewed in the film. Stemming, tokenization, lemmatization, and stop-word removal are just a few of the preprocessing steps that are used to extract crucial traits that the classifier needs. The upgraded Word2Vec in conjunction with LDA is used to obtain word vectors and increase the model's feature representation ability by combining both local and global relationships. CNN is used for binary class sentiment classification of movie reviews based on this word embedding. The proposed model surpassed all baseline approaches with an accuracy of 92.4%. We plan to test our algorithm on IMDb and movie review polarity datasets in the future.

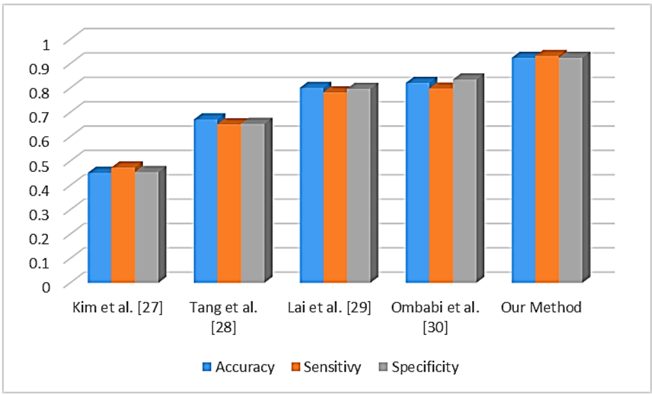

Fig. 6. The proposed method's results are compared to those of state-of-the-art methods such as random CNN, convolutional-gated recurrent neural network, recurrent convolutional neural network, and Word2Vec-CNN

REFERENCES

[1] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J Passonneau. Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)*, pages 30–38, 2011.

[2] Carmen Banea, Rada Mihalcea, and Janyce Wiebe. Multilingual sentiment and subjectivity analysis. *Multilingual natural language processing*, 6:1–19, 2011.

[3] Farah Benamara, Carmine Cesarano, Antonio Picariello, Diego Reforgiato Recupero, and Venkatramana S Subrahmanian. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. *ICWSM*, 7:203– 206, 2007.

[4] Muzafar Rasool Bhat, Majid A Kundroo, Tanveer A Tarray, and Basant Agarwal. Deep lda: A new way to topic model. *Journal of Information and Optimization Sciences*, 41(3):823–834, 2020.

[5] Pimwadee Chaovalit and Lina Zhou. Movie review mining: A comparison between supervised and unsupervised classification

approaches. In *Proceedings of the 38th annual Hawaii international conference on system sciences*, pages 112c–112c. IEEE, 2005.

[6] Long Chen, Yu Gu, Xin Ji, Zhiyong Sun, Haodan Li, Yuan Gao, and Yang Huang. Extracting medications and associated adverse drug events using a natural language processing system combining knowledge base and deep learning. *Journal of the American Medical Informatics Association*, 27(1):56–64, 2020.

[7] Shih-Hau Fang, Yu Tsao, Min-Jing Hsiao, Ji-Ying Chen, Ying-Hui Lai, Feng-Chuan Lin, and Chi-Te Wang. Detection of pathological voice using cepstrum vectors: A deep learning approach. *Journal of Voice*, 33(5):634–641, 2019.

[8] Ibrahim Eldesoky Fattoh, Amal Elsayed Aboutabl, and Mohamed Hassan Haggag. Tapping into the power of automatic question generation. *International Journal of Computer Applications*, 103(1), 2014.

[9] Geli Fei, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. A dictionary-based approach to identifying aspects implied by adjectives for opinion mining. In *Proceedings of COLING 2012: Posters*, pages 309–318, 2012.

[10] Ralph Grishman. Information extraction. *The Oxford handbook of computational linguistics*, pages 545–559, 2003.

[11] A Imran, W Aslam, and MI Ullah. Quantitative prediction of offensiveness using text mining of twitter data. *Sindh University Research Journal-SURJ (Science Series)*, 49(1), 2017.

[12] Azhar Imran, Jianqiang Li, Yan Pei, Ji-Jiang Yang, and Qing Wang. Comparative analysis of vessel segmentation techniques in retinal images. *IEEE Access*, 7:114862–114887, 2019.

[13] Anil K Jain. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666, 2010.

[14] Beakcheol Jang, Inhwan Kim, and Jong Wook Kim. Word2vec convolutional neural networks for classification of news articles and tweets. *PloS one*, 14(8):e0220976, 2019.

[15] Salman Khan, Hossein Rahmani, Syed Afaq Ali Shah, and Mohammed Bennamoun. A guide to convolutional neural networks for computer vision. *Synthesis Lectures on Computer Vision*, 8(1):1–207, 2018.

[16] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[17] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. Recurrent convolutional neural networks for text classification. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[18] Juan Li, Hui Wang, and Samee Ullah Khan. A semantics-based approach to large-scale mobile social networking. *Mobile Networks and Applications*, 17(2):192–205, 2012.

[19] Jing Ma, Fei Wu, and Chi Li. The topic tracking based on semantic similarity of sememe's lexical chain. In *2014 2nd International Conference on Software Engineering, Knowledge Engineering and Information Engineering (SEKEIE 2014))*. Atlantis Press, 2014.

[20] Andrew Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, pages 142– 150, 2011.

[21] Quazi Ishtiaque Mahmud, Asif Mohaimen, Md Saiful Islam, et al. A support vector machine mixed with statistical reasoning approach to predict movie success by analyzing public sentiments. In *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pages 1–6. IEEE, 2017.

[22] Abubakr H Ombabi, Onsa Lazzez, Wael Ouarda, and Adel M Alimi. Deep learning framework based on word2vec and cnnfor users interests classification. In *2017 Sudan Conference on Computer Science and Information Technology (SCCSIT)*, pages 1–7. IEEE, 2017.

[23] Xi Ouyang, Pan Zhou, Cheng Hua Li, and Lijun Liu. Sentiment analysis using convolutional neural network. In *2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing*, pages 2359–2364. IEEE, 2015.

[24] Doug Pender and Harshinie Karunarathna. A statistical process based approach for modelling beach profile variability. *Coastal Engineering*, 81:19–29, 2013.

[25] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47, 2002.

[26] Duyu Tang, Bing Qin, and Ting Liu. Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 1422–1432, 2015.

[27] Tun Thura Thet, Jin-Cheon Na, Christopher SG Khoo, and Subbaraj Shakthikumar. Sentiment analysis of movie reviews on discussion boards using a linguistic approach. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 81–84, 2009.

[28] Peter D Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*, 2002.

[29] Vijay Kumar Verma, Manish Ranjan, and Priyanka Mishra. Text mining and information professionals: Role, issues and challenges. In *2015 4th International symposium on emerging trends and technologies in libraries and information services*, pages 133–137. IEEE, 2015.

[30] Richard Vlach and Wassili Kazakos. Using common schemas for information extraction from heterogeneous web catalogs. In *East European Conference on Advances in Databases and Information Systems*, pages 118–132. Springer, 2003.

[31] Changli Zhang, Daniel Zeng, Jiexun Li, Fei-Yue Wang, and Wanli Zuo. Sentiment analysis of chinese documents: From sentence to document level. *Journal of the American Society for Information Science and Technology*, 60(12):2474–2487, 2009.

[32] Lei Zhang, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, and Bing Liu. Combining lexicon-based and learning-based methods for twitter sentiment analysis. *HP Laboratories, Technical Report HPL2011*, 89, 2011.

[33] Kai Zhao and Yaohong Jin. A hybrid method for sentiment classification in chinese movie reviews based on sentiment labels. In *2015 International Conference on Asian Language Processing (IALP)*, pages 86–89. IEEE, 2015.

[34] *Latif, J., Xiao, C., Imran, A., & Tu, S. (2019, January). Medical imaging using machine learning and deep learning algorithms: a review. In 2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET) (pp. 1-5). IEEE.*

[35] *Rahman, H., Bukht, T. F. N., Imran, A., Tariq, J., Tu, S., & Alzahrani, A. (2022). A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using ResUNet. Bioengineering, 9(8), 368.*

[36] *Imran, A., Aslam, W., & Ullah, M. I. (2017). Quantitative prediction of offensiveness using text mining of twitter data. Sindh University Research Journal-SURJ (Science Series), 49(1).*

[37] *Imran, A., Nasir, A., Bilal, M., Sun, G., Alzahrani, A., & Almuhaimeed, A. (2022). Skin cancer detection using combined decision of deep learners. IEEE Access, 10, 118198-118212.*

[38] *Fahim, S., Imran, A., Alzahrani, A., Fahim, M., Alheeti, K. M. A., & Alfateh, M. (2022, November). Twitter Sentiment Analysis Based Public Emotion Detection using Machine Learning Algorithms. In 2022 17th International Conference on Emerging Technologies (ICET) (pp. 107-112). IEEE.*