An Insight on Sentiment Analysis Research from Text using Deep Learning Methods

D. Christy Daniel, L. Shyamala

Abstract: ***Nowadays, Deep Learning (DL) is a fast growing and most attractive research field in the area of image processing and natural language processing (NLP), which is being adopted across several sectors like medicine, agriculture, commerce and so many other areas as well. This is mainly because of the greater advantages in using DL like automatic feature extraction, capability to process more number of parameters and capacity to generate more accuracy in results. In this paper, we have examined the research works which have used the DL based Sentiment Analysis (SA) for the social network data. This paper provides the brief explanation about the SA, the necessities of the pre-processing of text, performance metrics and the roles of DL models in SA. The main focus of this paper is to explore how the DL algorithms can enhance the performance of SA than the traditional machine learning algorithms for text based analysis. Since DL models are more effective for NLP research, the text classification can be applied on the complex sentences in which there are two inverse emotions which produces the two different emotions about an event. Through this literature appraisal we conclude that by using the Convolutional Neural Network (CNN) technique we can obtain more accuracy than others. The paper also brings to the light that there is no major focus on mixed emotions by using DL methods, which eventually increases the scope for future researches.***

Index Terms: Sentiment Analysis, Deep Learning, Machine Learning, Neural Networks.

I. INTRODUCTION

Sentiment Analysis (SA) also known as Opinion Mining (OM) is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine the writer's attitude towards a particular topic or a product. Both the terms SA and OM are interchangeable and they express a mutual meaning. However, some researchers stated that OM and SA have slightly different notions. OM extracts and analyzes people's opinion about an entity, while SA identifies and analyses the sentiment expressed in a text [1]. These days, people are sharing their emotions to friends, relatives and the society through the social network applications such as Twitter, Facebook, blogs, etc., This has resulted in the availability of huge volume of unstructured data across many social platforms in the internet. Mining these social network data available in unstructured format and changing that into structured data format using the pre-processing techniques is the first step in SA for getting better results.

In general SA is being investigated at three levels namely:

- Document-level SA
- Sentence-level SA
- Aspect-level SA

The document-level SA, classifies a whole document as a positive or negative sentiment [2], [3]. It considers the whole document pertaining to one topic as a basic information unit. This level of analysis assumes that each document expresses emotions/opinions on a single event/entity (e.g., a trip to hill station/mobile charger) and it is not applicable to documents which review multiple events /entities. Whereas, the sentence-level SA, targets to classify sentiments expressed in each sentence. Here, the first step is to identify whether the sentence is subjective i.e. expressing people's views and opinions or objective i.e. expressing factual information [4]. If the sentence is subjective, then the sentence-level SA will determine whether the sentence expresses positive or negative opinions. However, there is no fundamental difference between a document level SA and a sentence level SA, because the sentence level SA also comes with a similar assumption that one sentence should only contain sentiment about one topic. But, in reality, many a times, within the same sentence, multiple entities are compared leading to the expressions of contrasting sentiments. On the other hand, an aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities. Often, a document-level SA or a sentence-level SA cannot provide the specific opinions on all aspects of the entity when a reviewer gives different opinion for different aspects of the same entity or product, as in the case of this sentence "the car looks very good but it gives very poor mileage". Hence, aspect-level SA is also called as feature-based SA [5]. Practically speaking, in any social network platform, or blogs or online reviews, the sentiments expressed regarding an event or a product will seldom be explicitly positive or negative, rather people tend to express a mixed opinion about various features of the product or an event, with some being positive and others as negative. This process of expressing more than one emotion at a time is called as Mixed Emotions. Thus, mixed emotion can be broadly defined as the co-occurrence of both the positive and negative effects [6].

Nowadays, the approaches in doing the SA are also evolving very rapidly with lot of researches being done in this area. Broadly, there are three different approaches in SA (Fig.1) namely, machine learning approach, lexicon based approach and hybrid approach (which is a combination of machine learning and lexicon based approaches). The machine learning approach can be further classified into supervised learning, unsupervised learning and DL approaches. The supervised learning techniques use huge number

Revised Manuscript Received on August 05, 2019.

D. Christy Daniel, Research Scholar, School of Computing Science and Engineering, VIT Chennai Campus, Chennai, 600127, India.

L. Shyamala, Associate Professor, School of Computing Science and Engineering, VIT Chennai Campus, Chennai, 600127, India.

of labeled dataset where annotated data are used for training the classifier which results in getting high levels of accuracy while performing SA. Some of the supervised learning techniques are Maximum Entropy (ME), Naive Based (NB), Support Vector Machines (SVM) [7]. On the other hand, the unsupervised learning approach is used when the labeled dataset are not available. Some clustering algorithms for this approach are k-means, matrix factorization, principal component analysis (PCA) [8]. The DL, which is a third approach under the machine learning method uses many hidden layers for analyzing the data deeply for producing more accurate results. It is found that the DL algorithms can do the feature extraction automatically without the human intervention [9] such as convolutional neural networks (CNN), recurrent neural networks (RNN), recursive neural networks (RvNN), deep belief network (DBN). The lexicon based approach, as the name implies it relies on the sentiment lexicon for SA and it can be either corpus based or dictionary based approach by using the statistical methods to find the sentiment polarity [10]. In this paper, we have reviewed multiple research articles published recently towards DL approach in SA. The motivation for writing this paper in SA focusing on text with DL approach was mainly due its recent advancements in SA research, as it is being considered as a promising technique for the analysis of human emotions with higher precision. In this literature evaluation, the authors have examined 28 journal articles and international conference papers from ScienceDirect, ACM digital library, Google Scholar and IEEE Xplore.

The remaining part of this paper is organized as follows: section2 describes the literature review on data acquisition, pre-processing, DL networks. Section 3 highlights the functions of the DL models and the performance metrics. Discussion and analysis part is provided in section4 and finally the section 5captures the concluding remarks of the paper and future work.

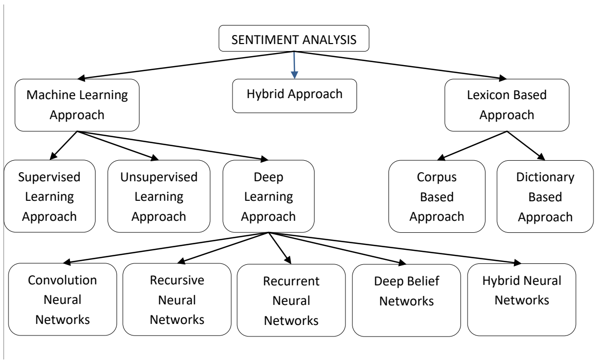

Fig.1. Approaches in sentiment analysis

II. LITERATURE REVIEW

This section reviews 28 research papers focusing on DL based SA researches (presented in Table 1 to 5),which got published recently in various peer reviewed journals. The insights gained from this review has been categorized into various sub headings and presented under this section.

A. Types of Data Acquisition

The different types of data acquisition followed by the researchers in DL based SA are presented in Table1. It was found that all of the datasets used for DL based SA are of benchmark datasets. Generally, in the research papers that were reviewed for this work, it was observed that three types of data acquisition were followed namely, self-reported data acquisition, social network data acquisition and other language data acquisition for SA. A self-reported data is typically a data collected from the respondent or user through an interview questionnaire or poll in which respondent read the question and select a response by themselves without any external interference. Whereas, the social network data are the data which were collected from social networking platforms like Twitter, Facebook etc., and other language data are the data collected from other than English language sites like online Hindi movie review, Tibetan micro-blog.

B. Data Pre-Processing

All the reviewed research works on SA has done text pre-processing steps, before the dataset were fed into the DL models as the input. The type of pre-processing varied from one research paper to another which were primarily based on the kind of data found in the dataset and there is no strict thumb rule that all dataset have to go through a fixed pre-processing procedure. The researcher can determine the pre-processing needs based on the quality of the available dataset. Xu et al. [11] reported that the pre-processing steps are mainly the conversion from natural language text into numerical values. Araque et al. [12] and Al-Sallab et al. [13] in their research have applied pre-processing to corresponding dataset which must be done before converting the texts into word vectors which includes sentence splitting as well as data cleaning. Majumder et al. [14] and Li et al. [15] have done unification, such as reduction to lowercase. Hash tags, user name mentions, links, emoticons, extra spaces removal are the cleanings for word level processing [16]. Jianqiang et al. [17] reported that pre-processing operations are involved to avoid the noisy and unstructured data such as removal of non-English characters, removal of URL expansion of acronyms and slang to their full form. Chatterjee et al. [18] has performed negation conversion like "won't" into "will not", replacing the emoticons and emoji into their original text.

C. Deep Learning Networks

In this section, we have discussed the different models of DL like Convolutional Neural Networks (CNN), Recursive Neural Networks (RvNN), Recurrent Neural Networks (RNN) and Deep Belief Networks (DBN). It was found in the review that more than one DL models have been used by many researchers for their study and this has been discussed under the hybrid neural network section.

Convolutional Neural Networks (CNN)

The CNN was popularly used in image processing earlier, but nowadays it is used in NLP as well. In this, the text data is moved forward into the hidden layers which comprises of three layers such as 1) Convolutional layers 2) Pooling layer 3) Fully connected layers. Akhtar et al. [19] have reported that the CNN architecture has shown the best performance for different domains and different languages while using the twitter dataset. In this research, they have analyzed for both sentence and aspect levels by the combination of the CNN-SVM (Convolutional Neural Networks- Support Vector Machine) by using word embedding method for getting the input data, which converted the word into vector values using word2vec tool. Kim [20] in his research paper had shown the model architecture (Fig: 2) with two channels in CNN through a clear example sentence.

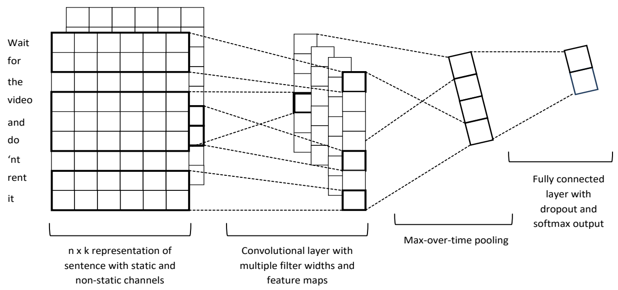

Fig.2.A model framework of CNN with two channels [20]

Araque et al. [12] in their research, have enhanced the SA using DL based classifier which required a baseline classifier such as word embedding method and some machine learning models. To get better improvement for the DL technique, they had used six different classifier [sentiment140 Go et al. [21], Stanford Core NLP Manning et al. [22], Sentiment WSD Kathuria [23], Vivekn Narayanan et al. [24], pattern.en Smedt and Daelemans [25], TextBlob Sentiment Classifier Loria [26]]. In this research, the researchers have conducted the test on convolutional functions of maximum, minimum and average by using the sentiment140 twitter dataset which got the good results. Majumder et al. [14] have developed a DL based document modeling for personality detection from text. The researchers have taken four different personalities for doing the personality detection, which describes an attitude of each and every persons such as Extroversion (EXT), Neuroticism (NEU), Agreeableness (AGR), Conscientiousness (CON) and Openness (OPN). They have used Deep-CNN (DCNN) to develop a document modeling with a help of two features in document and word level, such as stylistic features and semantic features. These two features were aggregated as the input variable representation and send it to the DCNN for the final classification of document modeling for personality detection. Guimaraes et al. [27] in their research, have classified the age groups from the user's profile in social network using DCNN based on their writing style, from the write-ups on the same topic and their different mental attitude among people of the similar age group. Ren et al. [28] have found a text finding technique established on a CNN for Chinese text features which consists of a text component detector layer for learning the Chinese linguistic structure, a spatial pyramid layer and a deep belief network (DBN) based on the multi-input-layer. The DBN multi-input-layers were changed as the fully connected layer in which that the hidden layers got the different scale features which were comparable. Finally, the CNN technique was already trained which was capable to extract the features from the Chinese complex text structure. Jianqiang et al. [17] have introduced the Global Vectors (GloVe) which is for vector formation model of word embedding to extract the sentiment from large twitter corpora using DCNN which is unsupervised based learning on five datasets such as the Stanford Twitter Sentiment Test (STSTd), SemEval2014 Task9, the Stanford Twitter Sentiment Gold (STSGd), the Sentiment Evaluation Dataset (SED), the Sentiment Strength Twitter dataset (SSTd). They have formed the feature set by n-gram features and polarity scores for the sentiment words. They have proven that the result of the datasets has shown the best performance than the baseline model. Chen and Zhang [29] have proposed the text emotion classification model using the combined features of the CNN and SVM model. In CNN, the pooling layer was producing the output which was comprised by the distributed features as the input to the SVM. In this, CNN has the ability to automatically learn the features and the SVM was act as the emotional classifier which were used to deal for the text emotion classification.

Yiming Li et al. [30] have proposed a model of text concept vector in which the input data must be conceptualized initially and a huge knowledge database was created by the input data along with the concept. The neural network converted the conceptualized data into the vector forms for both sentence and document level tasks with the meaning and the concept of the original text which provided good results for both the tasks. For sentence level experiment, they have tested the semantic relations between the two sentences using Sentences Involving Compositional Knowledge (SICK) dataset. For document level, they have used both the dataset for sentiment analysis. Rezaeinia et al. [31] have proposed a Improved Word Vectors (IWV) model which were done on the various lexicon methods, position of the emotion word, along with the tag and word to vector conversion, which were tested by the benchmark dataset and the DL algorithms. The word2vec embedding model consists of Continuous Bag-of-Words (CBOW) and Skip-gram algorithms which produced the best results for the conversion from the texts to word vectors. Alharbi and de Doncker [32] have proposed a method that used the tweet as the input for integrating the human behavioral by using CNN which produced good performance than the base line model. Zhao and Mao [33] have introduced the topic models such as TopCNN and TopLSTMs for sentence representations, which were used to improve the vector conversion for multiple feelings in Chinese text data. In this, the topic-aware of CNN and LSTM neural networks were formed using topic method with different semantic group processes. They have also developed a model based on the word sense which was a collection of general feelings and target feeling words. The researchers have used standard datasets such as Movie Review dataset (MR), Subjectivity dataset (Subj), Customer Reviews (CR), the MPQA (Opinion on a phrase level polarity detection subtask) dataset and Question classification dataset (TREC). The results have shown that the TopCNN and TopLSTMs models were providing the best performance than the other topic models. Table 2, captures additional information from the research articles.

Recurrent Neural Networks (RNN / ReNN)

Cho et al. [34] in their paper have described that the RNN is the most popular model for linguistic evaluations in NLP. RNN is essential for sequential information for

NLP. It is like a short-term memory units which consists three layers such as input, output and hidden layers. In the next step is to extend to the deep RNN which solves the difficulties of learning in deep networks. The paper reports that, since RNN is a short-term unit, the network was not able to remember the initial value of the sequence. To solve this issue, the researchers have used the Long Short Term Memory (LSTM) which produced the memory block for the sequence connections.

Dragoni and Giulio [35] in their research have implemented a tool called NeuroSent which used to analysis the sentiment on multi-domains. NeuroSent tool were developed by RNN with LSTM which was repeatedly tested by the Dranziera dataset and compared the results between the NeuroSent model and the baseline model using Average (Precision, Recall, F1) and Average (Deviation, Min.F1, Max.F1) in which Avg (MAX.F1) was produced best result. Finally, the proposed NeuroSent was shown the best results than the baseline model for multi-domain analysis. Li et al. [30] in their research paper reported that they have found a method called a Hierarchical User Attention Network model (HUAN) for classifying user's review based on the user information by using Bi-LSTM. They have extracted three user preferences as features, at word level, aspect level and polarity level. HUAN also has obtained the different features of the products of different users. Lee et al. [36] have collected the movie review dataset from the Internet Movie Script Database (IMSDB) site for predicting the emotions using DL network. They have found the similarity between target and the predicted emotions using cosine similarity methods. Meisheri et al.[16] have proposed a method for analyzing the sentiment from a short text using Bidirectional LSTM (BiLSTM) . For deriving the short text , they have applied the embedding techniques on word and character part, which can effectively predict and classify the emotion from the short text. Kratzwald et al. [37] have analyzed the emotion state of the human beings and how it has affected their decision making with the help of RNN and transfer learning to improve the performance. In addition to this, Table3 highlights the additional information on RNN / ReNN.

Recursive Neural Networks (RNN / RvNN)

RvNN is one of the deep neural networks which used to produce structure prediction when applying the same set of weight on the input data. Al-Sallab et al. [13] in their research have developed AROMA in which they have evaluated the opinion mining on three Arabic corpora with different writing styles such as Arabic Treebank (ATB), the Qatar Arabic Language Bank (QALB) corpus and Twitter datasets. An AROMA has solved the issues of the morphological complexity and language sparsity which has improved the performance of the opinion mining than the baseline method called Recursive Auto Encoder (RAE). Huang et al. [38] in their research have analyzed the sentiment on the Standford Sentiment Treebank (SST) and the movie review datasets using RNN with LSTM. They contribute to the idea of finding the syntactic information in which they have found how to learn the part of speech (POS) tag with embedding vectors in the composition process using RNN, Recursive Neural Tensor Network (RNTN), and LSTM networks. Table 4, captures additional information from the researches.

Deep Belief Networks

DBN is a hybrid probabilistic framework. It has multiple layers of both directed and undirected connections with hidden units. All the layers are connected with each other but not the units. There are two top layers which used to form the Restricted Boltzmann Machine (RBM) with undirected connection. The lower layer is used to receive the input from the above layer because it has the direct connection. Hinton [39] has stated in his research paper that DBN has to use an unsupervised learning approach to reconstruct the input. Xu et al. [11] have proposed a method to apply the Deep Belief Network (DBN) based dimensionality reduction on 150,000 tweets from Chinese microblog platform such as Sina Weibo. For training the data, the semantic similarity was retrieved by the retweet and hashtags which was not for the training purposes but also used in modifying the process of fine tuning for retrieving the best dimensionality reduction. An additional research information on the reviewed papers are presented in Table 5.

Hybrid Neural Network

Hybrid neural network is the combination of two or more DL algorithms for processing the data. Sun et al. [40] in their research have developed a content extension method to solve the feature sparse problem generated in the Chinese micro blog due to the restriction of using more number of words (length) for short text using CNN in sentiment classification. To extract the contextual information, a new convolutional auto encoder was proposed by the researcher. In that, they have included DBN, that consists of several layers of RBM which used to extract some higher level features. The final classification of SA was achieved by a ClassRBM (Classification RBM) layer. The researchers have proven that the convolutional auto encoder and DBN were the best model for short-length document classification with the proposed extension of the feature method. Chaturvedi et al. [41] in their research have analyzed the ways to learn the dependencies between words using Variable-order Belief Network (VBN) framework which were applied on the real world dynamic dataset of the basketball games as well as the processed dataset of 20 Newsgroups. The research has proved that the VBN has showed an improved accuracy level on the real-world dataset than the baseline work to an extent of 30 per cent. Hassan and Mahmood [42] have shown that CNN can have the capability to learn higher level features that cannot be changed for local translation. They have combined the CNN and RNN to capture the long-term dependencies when the length of the input sequence was developed. Recurrent layers have helped to reduce the loss of information and get the long-term dependencies more effectively. Xia and Zhang [43] have analyzed the emotions from text using DL techniques such as CNN, RNN, and LSTM on IMDB movie review dataset. DL technique used keras and TensorFlow as the back end. CNN, RNN, and LSTM has shown the best results in the emotion classification task and produced higher accuracy than the traditional learning. The researchers have proven in their paper that these three models were very effective.

Chen et al. [44] in their research have improved the SA based on the sentence type classification by CNN and BiLSTM-CRF (Bidirectional Long Short-Term Memory with Conditional Random Fields) to obtain the specific words from the sentences which were opinionated. Further, they have also classified the sentences into three group based on the target words like nothing-targeted, only one word targeted and more than one word targeted, which were applied to train the One-dimensional CNN for performing the classification of SA on all individual groups and predicting the polarity for sentiment words. It was found from their research that BiLSTM-CRF has achieved the better results from the benchmark datasets. Sun et al.[45] have analyzed the sentiment on Tibetan micro-blog using hybrid neural networks models such as CNN-LSTM which has produced higher accuracy. Lin et al.[46] have solved the sentiment prediction problem in a cross media of SA field. For that, they have proposed a method called explicit emotion signals using LSTM which has produced the best results. Wu et al. [47] have developed a model with three modules such as to encode the text, predict the sentiment polarity, decoding the module based on the different auto encoder for analyzing the sentiment. Chatterjee et al. [18] have introduced a novel approach for extracting the emotions from the dialogues which was the combination of the semantic and sentiment emotion detector. It consists of two LSTM layer such as LSTM-SSWE(LSTM-Sentiment Specific Word Embedding) and LSTM-GloVe for acquiring best results than the traditional machine learning technique. An additional information on this topics are reviewed in Table 6.

III. FUNCTION OF THE DL MODELS AND PERFORMANCE METRICS

A. Function of the DL

The functions of the DL models used in the NLP processing, which were presented by the researchers are presented in Table 7

B. Performance Metrics

The performance metrics from the appraised paper, which were presented by the researchers are presented in Table 8

IV. DISCUSSION AND ANALYSIS

In this section, we have briefly analyzed and discussed the recent trends and techniques of different DL models for text data, the researcher’s preferences towards these models, different type of analysis like SA, OM etc., and SA using text from other languages. Fig.3, describes about the percentage of research articles reviewed in this paper for the SA from the beginning of 2015 to March 2019 in the area of sentiment, opinion and other text related works. It was found that the number of research work related to the text analysis was in an increasing trend from the year 2015 to 2019 March. Though the year 2019 has less than 15 per cent, the count will definitely increase as there is hardly two quarter have got completed in this year so far.

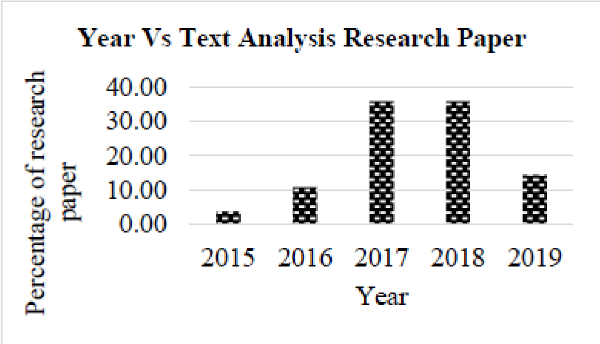

Fig. 3: Percentage of research articles using for the SA research over years

Fig.4, captures the percentage of reviewed research articles using different type of analysis over the past five years. From the graph it is evident that there is a huge growth of research in the SA field than the opinion mining and other text analysis related works starting from the beginning of 2015 to March 2019. It is evident from this data that SA has great scope in this space when compared to opinion mining. When we compare the opinion mining and other text related analysis, it was found that the areas like text extraction, word dependencies in text, topic models have significant research focus among the researchers.

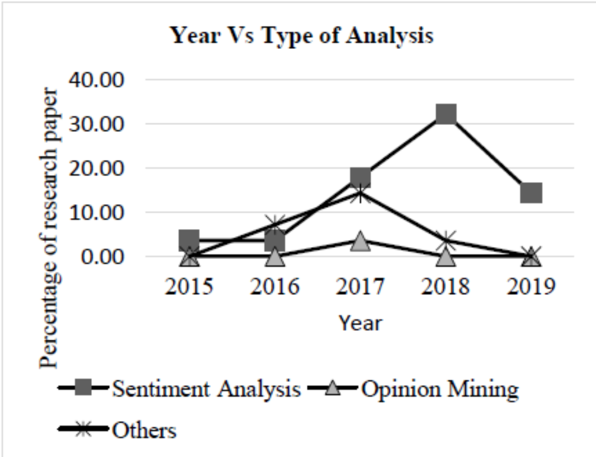

Fig. 4: Percentage of research articles using different analysis over years (2015 to 2019)

Fig.5, highlights the percentage of research articles using different DL models over the years starting from the beginning of 2015 to March 2019. It was found from the review of 28 research papers that the researchers have used different DL models such as CNN, RNN, RvNN, DBN, and Hybrid models for their research works. From this analysis, we can infer that CNN model was the predominantly used model over these years by many researchers. In 2016, DBN, CNN, Hybrid models were used in the same level. RNN have started coming into the picture from 2017 onwards. RvNN was used only in 2017. From this usage pattern we infer that CNN is the most preferred model because of its more efficient hidden layers.

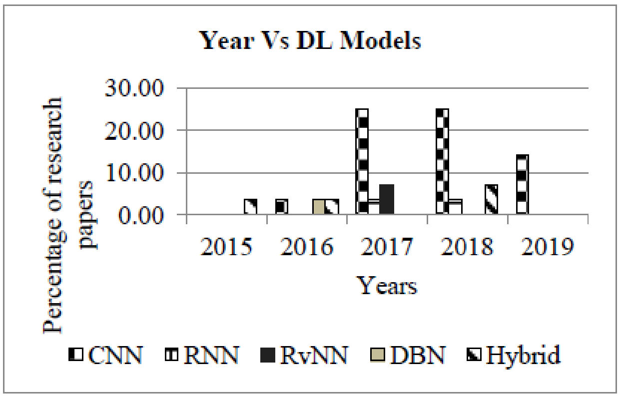

Fig.5: Percentage of research articles using different DL models

Fig.6, describes the percentage of research articles using different languages over years. Recently, researchers are using non-English languages for SA alongside with English. In 2015, Chinese language was used for SA. English, Chinese, and Hindi were equally used in the year 2016. English language is used for all the years except in the year 2015. From the analysis, it can be observed that most of the researchers wanted to use the English language because of the availability of the lot of data in social network domain than the other languages. But nowadays, the researchers are getting the challenges to build corpus for other languages also.

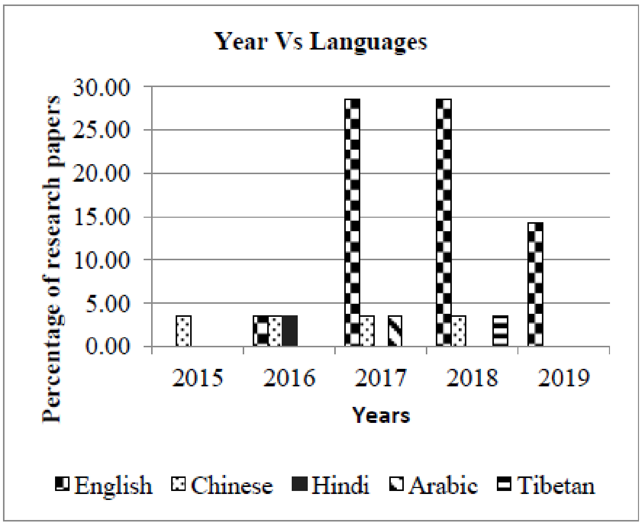

Fig. 6: Percentage of research articles using different languages

The table 9a to 9d compares the results of various DL models with the baseline works done in the machine learning by various researchers. It was observed from our review that all the DL models excelled very well over the baseline comparison with machine learning scores.

V. CONCLUSION AND FUTURE WORK

The analysis of social network data has been done to discover the human's emotions and attitudes towards an event or a subject. The reviewed papers taken for this study talks about the SA, OM, and emotion detection from the social network text data using different DL models and also captures the recent research focuses in SA. Though the machine learning is found to perform well in SA, the accuracy levels of detecting sentiments, increases when DL models are employed. This is mainly because of the more number of hidden layers in DL models which analyze the data very deeply and produce more precise results than the machine learning models. But the presence of more hidden layers have also created the 'overfit' of the data which may leads to errors by predicting irregular features. All the surveyed papers have explained about various models of DL and also the hybrid models in giving more accurate results than the other traditional models. The examined literature have shown that the accuracy level has increased to a maximum extent of 94.49 per cent while using DL (CNN model), on the other hand the traditional learning method like Classifier Ensemble Model could only able to achieve a maximum accuracy of only 86.48 per cent for a same study. Hence the DL models can be considered to produce greater accuracy than the other models. From the studied papers, it was also found that most of the English language data has produced more accuracy than the other language data due to the lack of data resources.

It was found from the surveyed DL researches, that there is no major work done with reference to complex sentences. Hence there is a desperate need to focus on the complex sentences in which we can do the research on mixed emotions which describes the positive and negative emotions along with the context of the sentence. This brings to the light that the future research works can focus in the direction of mixed emotions for SA using DL methods.

REFERENCES

1. Tsytsarau, Mikalai, Themis Palpanas. Survey on mining subjective data on the web. Data Mining and Knowledge Discovery, 24.3 (2012): 478-514.
2. Pang, Bo, Lillian Lee, ShivakumarVaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics, 2002.
3. Turney, Peter D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th annual meeting on association for computational linguistics, Association for Computational Linguistics, 2002.
4. Wiebe, Janyce M, Rebecca F. Bruce, Thomas P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, Association for Computational Linguistics, 1999.
5. Hu, Minqing, Bing Liu. Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2004.
6. Larsen JT, McGraw AP. Further evidence for mixed emotions. J Pers Soc Psychol, 2011 Jun; 100(6):1095-110.
7. Annett M, Kondrak G. A comparison of sentiment analysis techniques: Polarizing movie blogs. In Conference of the Canadian Society for Computational Studies of Intelligence, 2008 May 28 (pp. 25-35), Springer, Berlin, Heidelberg.
8. Shukla A, Misra R. Sentiment Classification and Analysis Using Modified K-Means and Naïve Bayes Algorithm. International Journal of Advanced Research in Computer Science and Software Engineering, 2015 Aug;5(8):80-5.
9. Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyu ML, Chen SC, Iyengar SS. A Survey on Deep Learning: Algorithms, Techniques, and Applications. ACM Computing Surveys (CSUR), 2018 Sep 18;51(5):92.
10. MedhatW, Hassan A, Korashy H. Sentiment analysis algorithms and applications:

11. Xu L, Jiang C, Ren Y, Chen HH. Microblog dimensionality reduction—a deep learning approach. IEEE Transactions on Knowledge and Data Engineering, 2016 Jul 1;28(7):1779-89.
12. Araque O, Corcuera-Platas I, Sanchez-Rada JF, Iglesias CA. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems with Applications, 2017 Jul 1;77:236-46.
13. Al-Sallab A, Baly R, Hajj H, Shaban KB, El-Hajj W, Badaro G. AROMA: a recursive deep learning model for opinion mining in Arabic as a low resource language. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 2017 Jul 13;16(4):25.
14. Majumder N, Poria S, Gelbukh A, Cambria E. Deep learning-based document modeling for personality detection from text. IEEE Intelligent Systems, 2017 Mar;32(2):74-9.
15. Li J, Li H, Kang X, Yang H, Zong C. Incorporating Multi-Level User Preference into Document-Level Sentiment Classification. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 2018 Nov 19;18(1):7.
16. Meisheri H, Ranjan K, Dey L. Sentiment extraction from Consumer-generated noisy short texts. In Data Mining Workshops (ICDMW), 2017 IEEE International Conference on 2017 Nov 18 (pp. 399-406), IEEE.
17. Jianqiang Z, Xiaolin G, Xuejun Z. Deep Convolution Neural Networks for Twitter Sentiment Analysis. IEEE Access, 2018;6:23253-60.
18. Chatterjee A, Gupta U, Chinnakotla MK, Srikanth R, Galley M, Agrawal P. Understanding Emotions in Text Using Deep Learning and Big Data. Computers in Human Behavior, 2019 Apr 1;93:309-17.
19. Akhtar MS, Kumar A, Ekbal A, Bhattacharyya P. A hybrid deep learning architecture for sentiment analysis. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers 2016 (pp. 482-493).
20. Kim Y. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882, 2014 Aug 25.
21. Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 2009 Dec;1(12).
22. Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D. The Stanford CoreNLP natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations 2014 (pp. 55-60).
23. Kathuria P.(2015). Sentiment wsdgithub repository . https://github.com/kevincobain20 0 0/sentiment _ classifier/ . Accessed on May 30, 2016.
24. Narayanan V, Arora I, Bhatia A. Fast and accurate sentiment classification using an enhanced Naive Bayes model. In International Conference on Intelligent Data Engineering and Automated Learning 2013 Oct 20 (pp. 194-201), Springer, Berlin, Heidelberg.
25. Smedt TD, Daelemans W. Pattern for python. Journal of Machine Learning Research. 2012;13(Jun):2063-7.
26. Loria, S. (2016). Textblob documentation page . https://textblob.readthedocs.org/en/ dev/index.html . Accessed on May 30, 2016.
27. Guimaraes RG, Rosa RL, De Gaetano D, Rodriguez DZ, Bressan G. Age Groups Classification in Social Network Using Deep Learning. IEEE Access, 2017;5:10805-16.
28. Ren X, Zhou Y, He J, Chen K, Yang X, Sun J. A convolutional neural network-based chinese text detection algorithm via text structure modeling. IEEE Transactions on Multimedia, 2017 Mar;19(3):506-18.
29. Chen Y, Zhang Z. Research on text sentiment analysis based on CNNs and SVM. In2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA) 2018 May 31 (pp. 2731-2734). IEEE.
30. Li Y, Wei B, Liu Y, Yao L, Chen H, Yu J, Zhu W. Incorporating knowledge into neural network for text representation. Expert Systems with Applications, 2018 Apr 15;96:103-14.
31. Rezaeinia SM, Rahmani R, Ghodsi A, Veisi H. Sentiment analysis based on improved pre-trained word embeddings. Expert Systems with Applications, 2019 Mar 1;117:139-47.
32. Alharbi AS, de Doncker E. Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information. Cognitive Systems Research, 2019 May 1;54:50-61.
33. Zhao R, Mao K. Topic-aware deep compositional models for sentence classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017 Feb;25(2):248-60.
34. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, 2014 Jun 3.
35. Dragoni M, Petrucci G. A neural word embeddings approach for multi-domain sentiment analysis. IEEE Transactions on Affective Computing, 2017 Oct 1;8(4):457-70.
36. Lee SH, Kim DM, Cheong YG. Predicting Emotion in Movie Scripts Using Deep Learning. In Big Data and Smart Computing (BigComp), 2018 IEEE International Conference on 2018 Jan 15 (pp. 530-532). IEEE.
37. Kratzwald B, Ilić S, Kraus M, Feuerriegel S, Prendinger H. Deep learning for affective computing: Text-based emotion recognition in decision support. Decision Support Systems, 2018 Nov 1;115:24-35.
38. Huang M, Qian Q, Zhu X. Encoding syntactic knowledge in neural networks for sentiment classification. ACM Transactions on Information Systems (TOIS), 2017 Jun 9;35(3):26.
39. Hinton GE. Deep belief networks. Scholarpedia, 2009 May 31;4(5):5947.
40. Sun X, Gao F, Li C, Ren F. Chinese microblog sentiment classification based on convolution neural network with content extension method. In 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) 2015 Sep 1 (pp. 408-414). IEEE.
41. Chaturvedi I, Ong YS, Tsang IW, Welsch RE, Cambria E. Learning word dependencies in text by means of a deep recurrent belief network. Knowledge-Based Systems, 2016 Sep 15;108:144-54.
42. Hassan A, Mahmood A. Convolutional Recurrent Deep Learning Model for Sentence Classification. IEEE Access, 2018;6:13949-57.
43. Xia F, Zhang Z. Study of text emotion analysis based on deep learning. In 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA) 2018 May 31 (pp. 2716-2720), IEEE.
44. Chen T, Xu R, He Y, Wang X. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Systems with Applications, 2017 Apr 15;72:221-30.
45. Sun B, Tian F, Liang L. Tibetan Micro-Blog Sentiment Analysis Based on Mixed Deep Learning. In 2018 International Conference on Audio, Language and Image Processing (ICALIP) 2018 Jul 16 (pp. 109-112), IEEE.
46. Lin D, Li L, Cao D, Lv Y, Ke X. Multi-modelity weakly labeled sentiment learning based on Explicit Emotion Signal for Chinese microblog. Neurocomputing, 2018 Jan 10;272:258-69.
47. Wu C, Wu F, Wu S, Yuan Z, Liu J, Huang Y. Semi-supervised dimensional sentiment analysis with variational autoencoder. Knowledge-Based Systems, 2019 Feb 1;165:30-9.
48. Socher R, Lin CC, Manning C, Ng AY. Parsing natural scenes and natural language with recursive neural networks. In Proceedings of the 28th international conference on machine learning (ICML-11) 2011 (pp. 129-136).

Table 1. Types of data acquisition

S No	Type of data acquisition	Used by the researcher for DL based SA research
1	Self-reported data	Kratzwald et al.[37] .
2	Social network data	Chaturvedi et al.[41] ; Dragoni M and Petrucci G [35] ; Araque et al. [12] ; Chen et al. [44] ; Huang et al. [38] ; Majumder et al. [14] ; Guimaraes et al. [27] ; Zhao R and Mao K [33] ; Meisheri et al. [16] ; Jianqiang et al. [17] ; Chen Y and Zhang Z [29] ; Li J et al. [15] ; Yiming Li et al. [30] ; Kratzwald et al. [37] ; Lee et al. [36] ; Hassan A and Mahmood A [42] ; Xia et al. [43] ;Wu et al. [47] ; Rezaeinia et al. [31] ; Alharbi and de Doncker [32] ; Chatterjee et al. [18].

3	Other language data	Sun et al. [45] ; Xu et al. [11] ; Akhtar et al. [19] ; Ren et al. [28] ; Al-Sallab et al. [13] ; Sun et al. [45] ; Lin et al. [46] ;

Table 2. Analysis of Convolutional Neural Network Method

S.No	Paper	DL Models	Task in NLP	Datasets	Results	Performance
1	Akhtar et al. [19]	CNN	Sentiment Analysis	1)TwitterH 2)ReviewAH 3)ReviewS 4)MovieH	The authors have applied a sentiment word vector using CNN framework for their experiment and finally predicted the sentiment which was done with the help of softmax method using best classifier. It produced better results than the baseline work for all datasets across the domains and different languages.	Accuracy: 1) 62.52 2) 65.96 3) 57.34 4) 44.88
2	Araque et al. [12]	Convolutional Functions (Max,Min ,Avg)	Sentiment Analysis	1)SemEval 2013 2)SemEval 2014 3)Vader 4)STS-Gold 5)IMDB 6)PL04	In this paper, the proposed model has yielded better classification performance from the simpler form of features on six public datasets. Meta learning approach was produced better results than the baseline model.	F1-Score: 1)87 .87 2)88 .19 3)89 .52 4)89 .24 5)90.93 6)94.49
3	Majumder et al. [14]	DCNN	Personality detection from text	James Pennebaker and Laura Kings (Essay Dataset)	In this, to receive the document vector from the sentence vector, they have used the convolution filters with 1-max pooling layers to developed a document model.	Accuracy for five personality traits: 1) Extroversion 58.09 2)Neuroticism 59.38 3) Agreeableness 56.71 4)Conscientiousness (57.30) 5)Openness 62.68
4	Guimaraes et al.[27]	DCNN	Age group classification from sentences	Twitter	This paper proved that the classification has performed well using deep CNN for the two groups such as adult age and teenager .The precisionwas0.95 in the validation part.	F1-Score : 0.940
5	Ren et al. [28]	CNN	Chinese text detection	ICDAR 2011 and 2013	In this paper, the CNN was extracting the Chinese complex text in which the text structure component detector (TSCD) layer was the best layer to recognize the text regions than other layers.	F1-Score: 0.77

6	Jianqiang et al .[17]	DCNN	Sentiment Analysis	1) SSTd 2) SED 3) STSGd 4) SE2014 5) STSTd	In this paper, the GloVe-DCNN model along with word embedding method(n-grams and polarity scores) were produced (87.62%) higher accuracy than the baseline model on the STSTD and also Precisian, Recall, and Accuracy values were higher than the baseline model for all the dataset.	Accuracy: 1) 81.36 2) 87.39 3) 85.97 4) 85.82 5) 87.62
7	Chen and Zhang [29]	CNN	Sentiment Analysis	NLPCC2014	In this paper, the proposed CNNs-SVM model has developed to analyze the sentiment from text data. CNNs-SVM model was produced good performance than the traditional model.	F1-Score: 88.8(Positive) 88.9(Negative
8	Yiming Li et al.[30]	CNN	Text Concept Vector	1) SICK 2) Yelp 2013 Yelp 2014 Yelp 2015 3) IMDB	In this research, they have proposed the Text Concept Vector(TCV) method for both sentence and document level task. It has obtained good results on three dataset by using external knowledge base to get more understanding about the text which has helped to solve the semantic problems between two sentences.	1) SICK: Pearsons 0.8691 Spearmans 0.8044 2)Yelp for TVC: 67.8 69.2 71.5 3) IMBD for TVC: 50.5
9	Rezaeinia et al. [31]	CNN (Word2vec,Glove)	Sentiment Analysis	1) MR 2) CR 3) SST 4) TR 5) SST-1	In this paper, the proposed model called Improved Word Vectors (IWV) has shown the highest performance when they have combined the four different methods, such as word to vector conversion, Part-of-speech technique, location of the words and the lexicon approach, which improved the accuracy of word embedding model.	Average Accuracy: 1) 81.275 2) 84.55 3) 85.95 4) 81.6 5) 45.275
10	Alharbi and de Doncker [32]	CNN	Sentiment Analysis	SemEval-2016	In this research paper, the CNN and LSTM have produced the more accuracy than the other classifier.	Accuracy: 88.71
11	Zhao and Mao [33]	CNN LSTM	Topic models	1) MR 2) Subj 3) CR 4) MPQA 5) TREC	In this, the TopicCNN (TopCNN) with word and sentence and TopCNN ensemble model was produced the best performance than the compositional models	Accuracy: TopCNNens 1) 83.0 2) 95.0 3) 86.4 4) 91.8 5) 94.1

					(CNN and RNN based) on the five dataset.	TopLSTMsens 1) 81.9 2) 94.5 3) 82.9 4) 90.8 5) 91.9

Table 3. Analysis of Recurrent Neural Network Method

S.No	Paper	DL Models	Task in NLP	Datasets	Results	Performance
1	Dragoni and Giulio [35]	RNN LSTM	Multi-Domain Sentiment Analysis	Dranziera dataset	In this research, the proposed model called NeuroSent for multi-domains was obtained the best results than the baseline model on the Dranziera dataset.	Avg.Max. F1-Score: 87.30
2.	Li Yiming et al. [30]	Bi-LSTM	Sentiment Classification	1)Tripadvisor, 2)Yelp2014	This paper has shown that a) An aspect layer was helpful for document representation. b) Considering user information can boost sentiment classification performance by a large margin. c)HUAN has outperformed state-of-the-art methods significantly.	Accuracy: User-agnostic models. 1)0.684 2)0.684 User-aware models. 1)71.5 2)67.2
3.	Lee et al. [36]	RNN LSTM	Predicting Emotion	IMSDB	In this research, the movie review dataset as the input to the DL model which is evaluated the dataset to predict the emotions in the best way and they have found that the higher level cosine similarity between the predicted and target emotions.	Average Cosine Similarity in Ratings (3.5~4.49) = 0.98205
4	Meisheri et al. [16]	Bi-LSTM	Sentiment extraction	1)WASSA-2017 Shared Task 2)SemEval-2016 Task 4 SubtaskC	In this research work, the word, character level and attention, character level BiLSTMs was produced the higher performance on the SemEval dataset. For WASSA, the word and dropout(avoid overfitting) with bias initializer were worked together and produced best result than the baseline model.	1)WASSA Pearson 0.701 Spearman 0.688 2)SEMEVAL MAEM(Mean Absolute Error over each class) = (0.9175)
5	Kratzwald et al. [37]	RNN Bi-LSTM	Emotion recognition	1) Literacy tales 2) Election tweets 3) ISEAR 4)Headlines 5) General tweets	In this paper, the proposed sent2affect (transfer learning)method was used for affect computing which was effectively improved the performance of the sentiment analysis.	F1-Score 1) 91.4 2) 82.9 3) 92.8 4) 85.6 5) 85.9

Table 4. Analysis of Recursive Neural Network Method

S.No	Paper	DL Models	Task in NLP	Datasets	Results	Performance
1	Al-Sallab et al. [13]	A Recursive Deep Learning Model for Opinion Mining in Arabic (AROMA)	Opinion Mining	1) ATB 2) Tweets 3) QALB	This research has shown that the AROMA introduced for evaluating the Arabic data using recursive model which was produced better accuracy than the other models.	1)ATB Accuracy 86.5 F1-Score 84.9 2) Tweets Accuracy 79.2 F1-Score 75.5 3) QALB Accuracy 76.9 F1-Score 68.9
2	Huang et al. [38]	RNN LSTM	Sentiment Classification	1) SST 2) Movie Review (MR)	This research work has tagged with the part-of-speech to increase the representation of sentence level. The proposed model was produced more accuracy for sentiment classification.	Accuracy 1) 48.8 2) 76.6

Table 5. Analysis of Deep Belief Network Method

S.No	Paper	DL Models	Task in NLP	Datasets	Results	Performance
1	Xu et al. [11]	DBN	Microblog Dimensionality Reduction	Twitter from Sina Weibo,	The research has applied deep networks to perform dimensionality reduction on microblog texts	-

Table 6 Analysis of Hybrid Neural Network Method

S.No	Paper	DL Models	Task in NLP	Datasets	Results	Performance
1	Sun et al. [45]	CNN DBN	Sentiment Classification	1)Twitter 2) Sina Microblog	This research has developed a DBN model which was working effectively for extending the content of the post only for the short text of feature extraction and fine tuned parameter with the proper structure were produced the best sentiment classification in Chinese microblog.	F1-Score 1) 0.719 2) 0.723
2	Chaturvedi et al. [41]	Deep recurrent belief network (DRBN)	word dependencies in text	1)20Newsgroups 2)BasketballGames.	The researchers have used the combination of the deep belief and recurrent networks to learn the long time delay from the short	F-Measure 1) 0 .94 2) 0.81

					time delay which can be used dynamically in the real time problems using multivariate Gaussians.	
3	Hassan et al. [42]	CNN and RNN	Sentence Classification(Sentiment Analysis)	1) IMDB 2) SSTb	The framework developed in this research was to analyze the sentence classification of short text by the combined model of CNN and RNN which was produced more accuracy than the other model. To reduce the more number of CNN layers, LSTM has to place on top of the CNN model.	Accuracy 1) 93.2 2) 89.2
4	Xia and Zhang [43]	CNN RNN LSTMNN	Text Emotion Analysis	IMDB	This paper has shown that the combination of CNN and word2vec produced more accuracy than the other models for emotion analysis from text data.	F-Measure 88.7
5	Chen et al. [44]	BiLSTM-CRF, CNN.	Sentence type classification	1) Movie review 2) SST-1 3) SST-2 4)Customer reviews	The researchers have applied the BiLSTM-CRF on all the dataset using various languages which was produced good performance in 5 out of 6 different languages. BI-LSTM has reached the 20.73 F1 score.	Accuracy 1) 82 .3 2) 48 .5 3) 88 .3 4) 85 .4
6	Sun et al. [45]	CNN, CNN-LSTM, SVM.	Sentiment analysis	Tibetan micro-blogs	In this research work, they have used CNN-LSTM model. This hybrid model has worked very well and produced more accuracy for sentiment classification than CNN,SVM models.	Accuracy 86.21
7	Lin et al. [46]	CNN LSTM.	Sentiment Learning on cross media content.	Sina Weibo	These researchers have investigated Explicit Emotion Signal (EES) for an image, text and the combination of both which has worked on the weakly labeled dataset. EES was produced best results and it is needed the less samples for the experiments.	Accuracy a) Image =(50.9%) b) Text =(56.4%) c) Image and Text = 60.1%

Retrieval Number J93160881019/2019©BEIESP
DOI: 10.35940/ijitee.J9316.0881019

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

8	Wu et al. [47]	CNN,LSTM ,Bi-LSTM, ST-SLSTM, CT-SLSTM, SV-SLSTM	Sentiment Analysis	1) CVAT 2) Facebook 3)Emobank	In this paper, the Bi-LSTM has the capacity to hold the contextual information which was used to predict the polarity values of the sentiment accurately.	1)Valence =0.676 Arousal =0.387 2) Valence =0.620 Arousal =0.883 3) Valence =0.605 Arousal =0.490 Dimension =0.307
9	Chatterjee et al. [18]	CNN LSTM	Emotion Analysis	Twitter Firehose	The researchers have proposed a model to detect the following three emotions like angry, happy and sad from the textual dialogue of human beings based on the combination of the semantic and sentiment related sentences which has produced more precise results.	F1(MACRO) = 71.34 F1(MICRO) =71.4

Table 7.The functions of the DL models

DL Network	Uses of the DL network	Used in a Journal Papers
CNN	CNN is used for NLP as well as the image. It provides more precise results and require less pre-processing when using more hidden layers.	Akhtar et al. [19] Li Yiming et al. [30]
RvNN	It is also used for NLP in a tree structure format	Socher et al. [48]
RNN	RNN utilizes the sequential information in the network.	Xia et al. [43]
DBN	It is the probabilistic model also called RBM, which has the capability to learn the features layer by layer, which holds the input of the next layer.	Sun et al. [45] Xu et al.[11]

Table 8. Collection of Performance Metrics from the appraised paper

S.No	Performance Metric	Symbols	Description
1.	Precision	P	The combination of total number of true positive and false positive are producing the fraction of true positive. The formula is P = True Positive /(True Positive + False Positive)
2.	Recall	R	The combination of total number of true positive and false negative are producing the fraction of true positive. The formula is R = True Positive / (True Positive +False Negative)
3.	F1 Score	F1	It is the average value of Precision and Recall. The formula is F-measure = 2*(Recall * Precision) / (Recall + Precision)
4.	Classification Accuracy	CA	It is the correct predictions which are divided by the total number of input samples.
5.	Spearmansρ	Spearmansρ	The performance test is for either ordinal variables or for continuous data.
6.	Pearsons$r$	r	It is used to calculate the weightage between two variables X and Y for linear correlation. The values ranges from +1 to -1.
7.	Mean Square Error	MSE	It lies between predicted and observed values

8.	Root Mean Square Error	RMSE	The deviations between the predicted values and the observed values. A normalized RMSE (N-RMSE) has been used in Guimaraes et al. [27]

Table 9a. Results of CNN model vs. Machine learning for SA

Article	Dataset	CNN Score	Baseline Score
Araque et al. [12]		**F1-Score:**	**F1-Score:**
	SemEval 2013	87 .87	85.34
	SemEval 2014	88 .19	84 .16
	Vader	89 .52	87.71
	STS-Gold	89 .24	83 .43
	IMDB	90.93	84.06
	PL04	94.49	86 .48
Jianqiang et al.[17]		**Accuracy:**	**Accuracy:**
	SSTd	81.36	59.11
	SED	87.39	62.51
	STSGd	85.97	68.79
	SE2014	85.82	73.21
	STSTd	87.62	68.81
Chen and Zhang[29]		**F1-Score:**	**F1-Score:**
	NLPCC2014	88.8 (Positive)	86.1 (Positive)
		88.9 (Negative)	86.0 (Negative)
Rezaeinia et al. [31]		**Average Accuracy:**	**Average Accuracy:**
	MR	81.275	79.9
	CR	84.55	83.15
	SST	85.95	84.225
	TR	81.6	80.525
	SST1	45.275	43.75
Alharbiand de Doncker [32]		**Accuracy:**	**Accuracy:**
	SemEval-2016	88.71	48.31

Table 9b. Results of RNN model vs. Machine learning for SA

Article	Dataset	RNN Score	Baseline Score
Dragoni and Giulio [35]	Dranziera dataset	**Avg.Max. F1-Score:**	**Avg.Max. F1-Score:**
		87.3	73.95
Li Yiming et al. [30]		**Accuracy:**	**Accuracy:**
		User-agnostic models.	User-agnostic models.
	Trip advisor	0.684	0.413
	Yelp 2014	0.644	0.392
		User-aware models.	User-aware models.
	Trip advisor	0.715	0.693

	Yelp 2014	67.2	0.653
Meisheri et al. [16]	WASSA	**Pearson** - 0.701	**Pearson** - 0.648
		Spearman - 0.688	**Spearman** - 0.641
	SEMEVAL	**MAEM** (Mean Absolute Error over each class) = (0.9175)	**MAEM** (Mean Absolute Error over each class)=1.13
Kratzwald et al.[37]		**F1-Score:**	**F1-Score:**
	Literacy tales	91.4	87.4
	Election tweets	82.9	78.7
	ISEAR	92.8	92.4
	Headlines	85.6	84.2
	General tweets	85.9	84.6

Table 9c. Results of ReNN model vs. Machine learning for SA

Article	Dataset	ReNN Score	Baseline Score
Huang et al. [38]		**Accuracy:**	**Accuracy:**
	SST	48.8	40.7
	Movie Review	76.6	75.9

Table 9d. Results of Hybrid model vs. Machine Learning for SA

Article	Dataset	ReNN Score	Baseline Score
Hassan et al. [42]		**Accuracy:**	**Accuracy:**
	IMDB	93.2	83.5
	SSTb	89.2	82.4
Xia and Zhang [43]		**F-Measure:**	**F-Measure:**
	IMDB	88.7	81.3
Chen T et al. [44]		**Accuracy:**	**Accuracy:**
	MR	82 .3	79
	SST-1	48 .5	43.2
	SST-2	88 .3	82.4
	Customer reviews	85 .4	80
Sun et al. [45]		**Accuracy:**	**Accuracy:**
	Tibetan micro-blogs	86.21	80.83
Wu et al. [47]	CVAT	Valence = 0.676	Valence = 0.547
		Arousal= 0.387	Arousal = 0.296
	Facebook	Valence = 0.620	Valence = 0.466
		Arousal = 0.883	Arousal = 0.823
	Emobank	Valence = 0.605	Valence = 0.485
		Arousal = 0.490	Arousal = 0.358
		Dimension = 0.307	Dimension = 0.212
Chatterjee et al .[18]	Twitter Firehose	F1(MACRO) = 71.34	F1(MACRO) = 49.97
		F1(MICRO) = 71.4	F1(MICRO) = 50.81

AUTHORS PROFILE

Christy Daniel D is currently a full time research scholar in School of Computing Science and Engineering, VIT, Chennai, Tamilnadu, India. Her research interests are Text Analysis, Neural Networks, Deep Learning, Big data analytics. She holds a Master of Engineering in Computer Science and Engineering and has six years of teaching experience as Assistant Professor.

L. Shyamala has completed her Bachelors in Engineering (Electronic and Communication Engineering) in 1998 under Madras University, Master of Technology in Information Technology in 2006 from College of Engineering under Anna University and Ph.D in Engineering under Anna University in 2016. She got ACRF scholarship to do her Ph.D in full time. She is having 16 years of teaching experience and Published around 10 papers in journals, Currently working as Associate Professor in VIT University, Chennai.