Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network: A Comprehensive Survey

PM. Lavanya
Department of Information Technology
Easwari Engineering College
Ramapuram, India
pm.lavanya92@gmail.com

E. Sasikala
Department of Computer Science and Engineering
SRMIST
Kattankulathur, India
sasikale@srmist.edu.in

2021 3rd International Conference on Signal Processing and Communication (ICPSC) | 978-1-6654-2864-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICSPC51351.2021.9451752

Abstract— **The social media is becoming an increasing trend for sharing the thoughts, ideas, opinions, etc. based on online reviews which generates a tremendous amount of unstructured data (ie. User posts). For processing those unstructured data supervised learning algorithms are preferred which helps for better performance optimization. Few years ago, Deep Learning (DL) techniques (ie. Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)) models has become popular in healthcare applications by giving the rise in complicacy of the healthcare data. Deep Learning (DL) Techniques provides an effective and efficient model for data analysis by uncovering the masked patterns and find the meaningful information from the significant amount of health data whereas the traditional analytics does not able to produce within a stipulated period. Specifically, Deep Learning (DL) techniques consist of yielding good results by using the models of pattern recognition for social healthcare networks. The study of this paper focuses on by investigating the models of deep learning (DL) techniques applied to classify the text in social media healthcare networks. The main intention of this review provides an insight for training the data and to classify the text by analyzing and extracting the raw input and produce the output with the help of Natural language processing (NLP). Overall, the purpose of this review is to enhance the performance of the text classifier based on effectiveness to improve accuracy and text processing speed by using a suitable methodology in order produce the promising results in the future.**

*Keywords— **Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Deep Learning (DL), Natural Language Processing (NLP), Classifier, social healthcare network.***

I. INTRODUCTION

Recently, accessing the healthcare is turning out to be a major challenge. The awareness has to be disseminated among the people about the diseases and symptoms for the prevention of the being infected in prior. In such cases a social media makes it possible for disseminating the information where users communicate through it. Social media helps the users to share their opinions and ideas from different regions all over the world through online. The people's opinion are posted through questions and answers given by some experts to educate them. In particular, social healthcare network is created for the purpose to bridge the communication among the patients and healthcare professionals in order to interact and educate the patients in health-related information and helps to protect and improve the people's health. On day-to-day basis the data produced is increasing tremendously from the social networks which is of both structured and unstructured. In order to process the unstructured data since social networks does not follow any rules from patients' postings, the best techniques have to be used. For this case an algorithm from supervised learning can be adopted in order to find the patterns in data and for classifying them by collecting the data and analyzing it from the previous learning. There are two types of supervised learning. One is classification and other one is regression. In this study, the main task is classification. The useful text has to be extracted from the social healthcare network by the patient's where the questions are being posted by the them and the answers are being posted by healthcare professionals. Deep learning (DL) algorithms were used previously for classifying the text. The classifiers are evaluated based on the performance. The sole purpose of classifier is to analyze the text automatically and to assign it with the pre-defined tags. Fig 1, shows the stages of supervised learning.

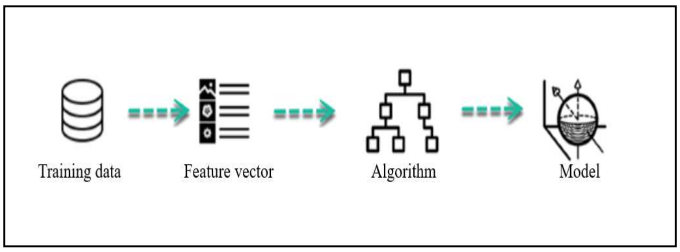

Fig1. Learning stage of a supervised learning

The process of assigning the labels to texted areas (ie. queries, documents, sentences and paragraphs) is one such problem in NLP which is called as text classification or text categorization (TC). Some of the common applications include, sentiment analysis, question-answering (QA), categorizing news, user intent-based classification. Text data can be obtained from various sources from mails, chats, social media, web data, customer reviews, and feedbacks. Text basically contains an information which is of richly sourced. But extracting accurate text might be a challenging and time-consuming, since it contains unstructured data. Classifying the text can be done either by automatic labeling or manual explanation. Due to increased volume in text data, classifying the text automatically becomes very important. The performance for classifying the text is evaluated from the algorithms used in DL by based on its accuracy Classifying the text automatically is divided into 2 types:

• Rule based method
• Machine learning (ML) (data-driven) based method

Rule-based method helps to categorize the text into various categories with the help of pre-defined rule set, and it requires a thorough domain understanding. The other type, machine learning (ML) methods are used to categorize the text by observing the data. As a recent trends Deep learning (DL) is gaining importance due to the terms of accuracy when it is trained with large amount of data. Shah, A. M et al. [3] It is a one of the class of machine learning where it performs much better on unstructured data. Now-a-days DL plays a vital role in healthcare applications. Z. Liao et al. [1] Advancement in modern technologies have caused a transformation towards digital health in healthcare, where diagnostic and organizational activities can be assisted by computer-generated analytics, and with the use of electronic patient medical records. S. Amin et al. [4] Now-a-days, there is an increase in using social media applications which gives a valuable information for analyzing and interpreting the user emotions, thoughts, and opinions on different various topics such as healthcare, politics, education, sports, research, arts, etc. Social network users frequently post their information by updating about their present circumstances based on their status. The information about their day today life circumstances are shared by them, and also if an epidemic situation occurs in a region which is growing rapidly and also if infected by any disease.

J. Li, A. Sun et al. [5] One of the algorithm in NLP is Named entity recognition (NER) which is used to identify the designators from text and map those designators to the predefined types semantically such as person, organization, location, etc. Text summarization, question-answering (QA) and machine translation are some of the NLP applications which is served by NER. Recently, machine learning (ML) approaches is gaining more attention, by analyzing the patterns particularly from raw data or images. M. Bates et al. [7] addressed how the machine learning progress allows the epidemiologists by mining through broad sets of data digitally. A. Mike et al. [8] reviewed that the conjunction of machine learning and natural language processing (NLP) with social media networks to support by analyzing the huge dataset for the purpose of mental health research in population.

In spite of various methods, modulation in ML, few types of architecture have its familiarity. For e.g, When compared to various machine learning (ML) classifiers such as NB, NBM Modal, and SVM K. Nargund et al. [9] the k Nearest Neighbor (kNN) classifier precision was seemed to be superior, by classifying the tweets among the 2 classes (i.e., awareness or real occurrence of tweets). Similarly, K. Lee et al. [10] proved that performance for text categorizing was attained using Multinomial Naive Bayes Modal (MNBM) with F-measure of 0. 811 when comparing with the other text classifiers such as Random Forest (RF), NB and SVM. K. Lee et al. [11] explored a backpropagation algorithm for twitter by using the multilayer perceptron and predicts the data status by US population who are infected with ILI. The personal health record tweets are detected and analyzed by various supervised ML algorithms R.A. Calix et al. [12] and deep gramulator approach is used for improving the precision when it is applied to independent set of data.

A. *Tasks Involed In Text Classification*

The process of grouping the text (e.g., articles, tweets, customer reviews) into ordered categories is called as text classification. Text classification process includes the categorizing the news, analyzing the text based on user opinion and classifying the text based on topic. Later researchers found that by using DL based text classifiers it is efficient and effective by casting multiple Natural Language Understanding (NLU) tasks (For e.g., natural language inference, QA system) as text classification. The following are the five text classification tasks in recent studies of Deep Learning.

Classification based on user opinion: The technique used to analyze and determine the opinion of people from textual data (e.g., tweets, customer reviews, etc) and extracting the viewpoint and polarity. The task can emit either multi-class or binary problem. Classifying the texts into fine grained labels or multiple level intensities is called as Multiple class opinion analysis. Classifying the text based on positive and negative classes is called as binary sentiment analysis.

Categorizing news: The articles related to news is one of the major and noticeable information sources. Based on the interest it helps the user to obtain information in real-time e.g., Based on user interest it emerges the new topics and recommend the relevant news.

Topic Analysis: The process of identifying the topic or context from a text (For e.g., Either a feedback of a product given by the user can be about "ease of use" or "customer support").

Question-Answering (QA): The two kinds of QA tasks are: QA which is extractive and QA which is generative. First task includes, a process where a question is given and with a set of answers by the candidate (e.g., P. Rajpurkar et al. [21] SQUAD, existence of a text in a document), a classification is performed by the system based on the candidate answer whether it is correct or not. Other type of QA, is a generating the text task as it required the answers to be generated in real time.

Natural language inference (NLI): It is otherwise called as Recognizing Textual Entailment (RTE), used to predict whether the hypothesis meaning is entailed from another. M. Marelli et al. [22] The RTE system is needed to be assigned with a text pair unit such as contradiction, inference and neutral. Generalized form of NLI is the paraphrasing, which is otherwise called as pairing text comparison, the similarity of a text pair is measured semantically by indicating that how frequently a sentence is paraphrased to other sentence or text.

B. Corpora Used For Text Classification

a) Corpora used for Sentiment Analysis

Text Classification of Reviews using Yelp: Yelp dataset consists of data about reviews for the task by classifying on two sentiments. One is for detecting the fine-tuned sentiment tasks which is known as Yelp 5. Second, it is used to predict the positive and negative opinions, which is called as Yelp 2 or Yelp Polarity on reviews. Yelp 5 dataset has 60K test sample and 750K training sample for individual class, and Yelp 2 includes 38K test sample and 560K training sample for positive and negative class.

Text Classification on movie reviews using IMDb: is one of the dataset labels used for classifying the user opinion based on reviews by movies. The dataset is divided equally between test sets and training sets with 25K reviews for each set and it has equal number of negative reviews and positive reviews.

Text Classification on Question answering using MPQA: The Multiple Perspective (MP) Question Answering (QA) dataset is developed for sentiment dataset by containing with two class tasks. MPQA is obtained from articles produced by news which is related to a broad variety of sources based on news and it consists of 10,606 sentences. Here the dataset is imbalanced with 7,293 negative and 3,311 positive opinions.

b) Corpora used for Topic Classification

Text Classification on topic using DBpedia: Is an enormous dataset, which is created from the commonly used information boxes containing within Wikipedia and it multilingual knowledge based. In this dataset, the attributes like properties and classes are removed and added in every release and it is published every month. The most popular version of DBpedia contains 70K test sample and 560K training sample, each of containing 14 class tasks

Text Classification on topic using Ohsumed: Is a subdivision of the Medical line repository. It contains 7.5K documents where single document is labeled by uni or multiple classes which are selected from 24 categories of cardiovascular diseases in medical abstract.

Text Classification on topic using PubMed: Created by the National Library of Medicine for biological and medical documents, consisting of category of documents. Every document is labelled with the class of MeSH set called as set label which is used in PubMed. The sentence in a document is labelled as abstract by using one of the following classes such as review, aim, methodology, outcome, or summarization. Other types of datasets for topic classification consists of Irony (composed with explained comment described from the social news website reddit, The dataset from twitter is used for classifying the topic based on tweets and arXiv collection).

c) Corpora used for QA

Text Classification on QA using (SQUAD): Stanford Question Answering Dataset P. Rajpurkar et al. [21] a group of questioning and answering pairs attained by various articles. The question which contain the correct answers in the given text can be of any sequence of tokens. Because the QA is generated by human through the crowdsourcing, is different from some other QA datasets. The version SQUAD 1.1 pairs on 536 articles by containing 107,785 question-answer. Presented SQUAD 2.0, the upgraded version, which combines 100K queries to SQUAD 1.1 contain almost 50K un-answerable questions which are written in adversarial by crowd workers and those are same as answerable.

Text Classification on QA using Trec Question Answering: Is the studied and most familiar dataset for question answer research. It consists of two types, such as Trec 6 & Trec 50. Trec 6 is splitted into 6 categories by containing the questions while Trec-50 contains in 50 categories. In both the versions there is test datset and trained dataset by containing 500 and 5,452 questions respectively.

Text Classification used for QA using QUORA: Is designed for identifying the paraphrase (i.e. For detecting replicated questions). To identify the duplicate questions, a subset of QUORA data is presented by authors that contains of over 400K pairs of question. A question pair results in binary value which indicates that the weather two queries are identical or not. The other types of datasets includes the WikiQA.

Text Classification used for QA using WIKIQA: It contains the question answers pairs, those are grouped and explained for research in open domain. Even researchers are allowed to evaluate the answer for question which has no answers in dataset by using triggering models. It also includes the questions which contain no correct answers, and allows researchers by using triggering models for answers by evaluating them.

II. RELATED WORKS

A. Feed Forward Neural Networks

One of the simplest methods used for text representation is Feed forward neural networks (FFNN). A high accuracy is achieved on classifying the text by using this model. Text is viewed as a bag of words. It learns how to use embedding model for representing the vector which is termed as Glove J. Pennington et al. [24] or word2vec T. Mikolov et al.[23] for each word, and obtain the sum vector or by calculating the mean of word embeddings as text representation, and pass the sum vector through single or more layers of feed forward, which is also otherwise called as Multiple Layer Perceptron (MLP), and later it performs the classification by using the techniques such as Naïve Bayes, logistic regression, or Support Vector Machine on the representation of final layer M. Iyyer et al.[25]. To illustrate these models the (DAN) Deep Average Network is used for its simplicity, it performs better on huge models whose design is explicit for learning about the text composition. Eg. DAN performs better than the syntactic approach with increased syntactic variance on datasets. A. Joulin et al. [26] A simple text classifier which is called as fastText is proposed. fastText is similar to DAN which is used to represent the text by describing the number of times the words are being occurred in a document. Similarly, fastText also used additional features such as bag of n grams for capturing the order for local word information. S. Wang et al. [27] the results produced are efficient by achieving a comparable result by using the word order explicitly. T. Mikolov et al. [28] proposed doc2vec model, by using an unsupervised algorithm and learn about the representation of fixed length features of variable segments of information, such as paragraph, sentence, and whole document. The doc2vec model is close to Continuous Bag of Words (CBOW) model [28, 23].

B. Recurrent Neural Networks (RNN) Models

The approach of RNNs views the text as words in sequence, that are planned for capturing the text structures and dependencies on words for classifying the text. When comparing with Feed forward model, vanilla RNN model does not performed well. Though RNNs consists of many variants, the most popular architecture Long Short-Term Memory (LSTM), is modelled for capturing the dependencies which are long termed. LSTM introduced a memory cell for remembering the values in a random time intervals and gates for regulating the information flows inside and outside of each cell by addressing the gradient vanishing problems which is occurred by vanilla RNNs. K. S. Tai et al. [29] In order to learn rich semantic representations, it builds a tree based LSTM model ie. tree structured topologies of generalized LSTM. It is shown that that for NLP tasks Tree-LSTM approach is better than to use chain structured LSTM because natural language exploited the properties syntactically by naturally combining the words with phrases. The performance of Tree based LSTM is validated upon two tasks: predicting the relationship semantically between two sentences and sentiment classification. To design a word relation with a long-span for reading system, J. Cheng et al. [30] instead of single memory in LSTM structure a large memory network is augmented. By using this model, it has achieved a better result on sentiment analysis, NLI and modelling language.

It's noted that RNN is a subset of DNN, which is called as recurrent neural networks. RNN, an equal weight is applied in recursive manner over a input for producing a processed representation of a vector or structured prediction over the fixed variable sized input. R. Socher et al. [31] RNN with a linear chain designed input, also operate on tree based design, ie.parse trees of sentences in a natural language, which combines the child representation into parent representation. RNN is considered as most emerging model for classification of text because they are easy and effective to use – since the text is viewed as a sequence of tokens and sentences. L. Pang et al. [37] used multiple layer CNN for identifying the n gram patterns and casted matching text as recognizing image. J. Wang et al. [38] proposed a CNN model by combining implicit and explicit depiction of short text for classification. It shows a rapid increase by using CNN to for text classification in the field of biomedical applications [39–42].

it does not require structure label in additional (ie. parse trees).

C. Convolutional Neural Networks (CNN) Models

Y. LeCun [32] Recognizing the patterns based on time is trained by RNN, whereas recognizing the patterns based on space is trained by using CNN. The models of RNN works well for the applications of tasks involved in NLP such as QA or Parts of Speech (POS) tagging where it requires a long range semantics, whereas the models of CNN is used for detecting the position invariant patterns. To express a particular sentiment, the patterns can be of key or phrases (e.g. "We love" or a context like" animals"). N. Kalchbrenner et al. [33] Therefore, CNN have also become one of the most familiar approach for categorizing the text). The model which uses a k-max-pooling dynamically, is termed as Dynamic CNN (DCNN). The top layer of DCNN uses the word embedding for a word in the sentences by constructing the matrix. Next, convolutional design consists of convolutional layers widely with dynamic pooling layers built by dynamic k maximum-pooling and generates a feature map over the sentence and explicitly capture the short and long-distance relations between words and phrases. The pooling parameter k can be chosen dynamically which depends on sentence level and size in the hierarchical convolution.

Later, Y. Kim et al. [34] proposed a CNN model for text classification (TC). In unsupervised neural language (i.e., word2vec) model, it uses one convolution layer at the top of words vector. Y. Kim et al. [34] compares the four various models for learning word embedding: (i) rand, all the word embeddings are initialized in random and later it is altered during training; (ii) static, the word2vec embeddings those are pre-defined is used and they are stayed fixed at training model; (iii) non static, in each task the word2vec embeddings are fine tuned (iv) multiple-channel, two sets of word embedding vectors are used, where both the sets are assigned by word2vec, where one is fixed while the other is updated during model training. These CNN models are reported for improving the art on analyzing the sentiment and classifying the question. J. D. Prusa et al. [35] presented an approach for encoding the text using CNN models which is used to reduce the memory space and time and is required for learning representation of text at character level. The model allows to preserve large amount of data from the actual text to improve performance in classification. There are other CNN-based models. L. Mou et al. [36] presented a tree structured CNN for capturing the semantic level of

D. Attention Models

The model attention has become an useful tool and familiar concept in deep learning models (DL) for NLP M.T. Luong et al. [43]. In short, vector of importance weights is interpreted by using attention in the language models. For predicting a word in a sentence, it is estimated by how strongly it correlates with other words by taking the total of their values as the approximation of the target weighted by attention vector. T. Shen et al. [44] used a directional attention network in self for CNN/RNN language understanding, by which the attention of intermediate elements of input sequence are directional and multi-dimensional. Based on the proposed attention a lightly weighted neural network which is used to learn embedding the sentence by without the inclusion of CNN/RNN structure. T. Shen et al. [44] proposed a LSTM approach for NLI with inner-attention. The model uses a two-staged process for sentence encoding. First, Bi-LSTM word level uses the average pooling for representing a first stage sentence. Secondly, for better representations, attention is applied on the same sentence by replacing the average pooling. Z. Lin et al. [49] to extract interpretable sentence embeddings self-attention is used. S. Wang et al. [50] To produce variable n grams features, densely connected CNN is proposed with multiple scale feature attention. I. Yamada et al. [45] performs to classify text by using some attributes to a base, and neural attention bag of attributes mechanism is used. A. P. Parikh et al. [46] To break down a problem into sub divisions attention is used for solving it separately. Q. Chen et al. [47] To enhance sentence embedding generalized pooling method is developed, and proposed a multiple head vector-based attention model. M. E. Basiri et al. [48] for sentiment analysis an attention method based bidirectional deep model of CNN-RNN is proposed.

E. *Hybrid Models*

G. Chen et al. [51] performed multiple label text classification by using a Convolutional and Recurrent model for capturing both local and global semantic text and therefore, have a tractable computational complexity by modelling higher-order label correlations. D. Tang et al. [52] a gated RNN is used to learn representing a document that is used to encode the intrinsic relations between sentences and used CNN model to learn for representing the sentences. It views the sentences as a character sequence in a document, instead of words, and proposed a model that both recurrent layers for document encoding and character-based convolution. By using this model, it shows comparable performances by using with a smaller number of parameters, while comparing with word level models. In order for capturing long range of contextual dependence it applies a recurrent structure for learning word representations. Maximum pooling is derived automatically to reduce the noise, by selecting only the obvious words which are critical for the task of classifying text. A Hierarchical Deep Learning technique is developed for classifying the text (HDLTex). HDLTex employed a stack of combining the model of deep learning (DL) architectures, including RNN, NLP and CNN, which provides a deep understanding at every stage of the hierarchical document.

The following Table I shows the performance and limitations those are tabulated based on existing algorithms used for text classification (TC) techniques in DL.

TABLE 1: RESULTS OF DIFFERENT STUDIES USING NATURAL LANGUAGE PROCESSING (NLP) APPROACHES WITH DL IN TERMS OF ACCURACY AND ITS LIMITATIONS.

Ref	**Authors**	**Approach**	**Attributes**	**Techniques & Tools**	**Accuracy**	**Merits**	**Limitations**
[13]	V. Agarwal et al.	-For better performance noisy words are removed. - Bag of words, n grams and TF-IDF methods are used.	-Text data	- Logistic Regression & Support Vector Machine (SVM)	-F1 measure=62%	-In each and every step it gains data by utilizing a stacked model. -Performance is compared by letting the features into multiple classifiers	-It requires better feature selection for accurate results.
[15]	S. Deepak et al.	- LSTM, RNN, LSTM and GRU overcomes the issues posed by NLP. -Based on previous results RNN can get the entire meaning of the news.	-Text data	-GRU -LSTM	- LSTM=91.2% -GRU=84.4%	-Better results is produced by using the data mining approach. -It utilizes live mining by incorporating the features for better performance from news articles	-BY using the pure NLP approach results can be improved.
[18]	S. A. Cammel et al.	-Frequency and sentiment values are measured to improve the progress.	- Stemming - Sentiment analysis -Stop word removal	-N gram analysis -TF-IDF	-Accuracy is not measured quantitatively	-In spite of capturing the same set of topics of data from different hospital, the model used is transferrable.	-Multi labelling can be used for capturing multiple topics to provide accurate opinion measures from patient response.
[17]	M. U. Salur et al.	-RNN with Fast text and Word2Vec	-Text data	-LSTM -CNN	-80.44% of accuracy	-It utilizes roadmap for textual embedding tweets. -For better accuracy representing the	-Models have no other methods to handle jargon. -To capture the morphological features

						different methods for text is used.	attention mechanism can be utilized.
[19]	C. Zhang et al.	-Sentiment analysis, Statistical analysis and linguistic cues analysis.	- Similarity value is found between the events and cluster	-Affinity Propagation -K means	-Precision-90%	Trustworthy and legitimate news is employed into topic clusters and compare the event authenticity.	-Opinions are not discussed. -Validation is performed by using fake new sources.
[14]	W. Haitao et al.	- To improve short text classification by using CNN with n grams and non linear sliding	-Features produced by neural network -Text data	-CNN -N gram	-The precision calculated for CNN and N gram is higher than by using the traditional ML algorithms	- It is useful in classifying the short text data. -For efficient feature extraction and classification attention mechanism and pooling methods are used.	-Different text representation and word embeddings are explored for optimal solutions when applied to various problems.
[16]	Z. Li, F. Yang et al.	-Disambiguation -NLP of biomedical text	-Text data	-Bi-LSTM	-96.71% of accuracy	-It contains information of both prefix and suffix words at a same time. -Used to predict the ambiguous word in surrounding context.	-Results can be improved if biomedical data loss occurs it uses unsupervised learning and unlabeled data
[20]	K. Liu et al.	-MSMTC model	-Medical Text data	CNN -LSTM -Bi-LSTM -Naïve Bayesian Model	-CNN Accuracy-87.65% Precision-87.95% Recall-87.39% F1-Measure-87.30%	-structure of dual channel design produces more accurate classification than learning the sequence of words directly.	-lack of normative expressions from original social media text.

III. ENDOWMENT

The structure of this study was concise in the following points.

- Provides the sole purpose of why to use deep learning algorithms in social healthcare networks.
- Discussing about the deep learning (DL) approaches by using NLP on text classification (TC).
- Providing insights of performance based on accuracy by using the existing algorithms in DL.
- Exposing the open challenges and limitations by current deep learning models in social network.

IV. CONCLUSION

As an emanate technique, deep learning with Natural Language Processing (NLP) has a colossal capability in tackling the problems in text classification. As a whole, the performance based on text classification techniques has been focused and has been projected by using deep learning (DL) techniques for producing successive results. DL based techniques are considered to be a powerful tool for dealing with text classification such as preprocessing, feature extraction and selection, classification, and clustering steps. The performance is analyzed based on accuracy and it is discussed. It is summarized and concluded that the hybrid methods based on DL equipped with NLP produces better performance when comparing with a single approach. It can be concluded that, the combination of two or more methods can improve the performance and might yield better results. Therefore, depicting and applying suitable methods for social healthcare network is a major challenge.

REFERENCES

[1] Z. Liao et al., "Medical Data Inquiry Using a Question Answering Model," 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 2020, pp. 1490-1493, doi: 10.1109/ISBI45749.2020.9098531.

[2] Jeyaraj, P., & Nadar, E. R. S. (2019). Deep Boltzmann Machine Algorithm for Accurate Medical Image Analysis for Classification of Cancerous Region. Cognitive Computation and Systems. doi:10.1049/ccs.2019.0004

[3] Shah, A. M., Yan, X., Shah, S. A. A., & Mamirkulova, G. (2019). Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach. Journal of Ambient Intelligence and Humanized Computing. doi:10.1007/s12652-019-01434-8

[4] S. Amin et al., "Recurrent Neural Networks With TF-IDF Embedding Technique for Detection and Classification in Tweets of Dengue Disease," in IEEE Access, vol. 8, pp. 131522-131533, 2020, doi: 10.1109/ACCESS.2020.3009058.

[5] J. Li, A. Sun, J. Han and C. Li, "A Survey on Deep Learning for Named Entity Recognition," in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2020.2981314.

[6] Gupta, A., & Katarya, R. (2020). Social Media based Surveillance Systems for Healthcare using Machine Learning: A Systematic Review. Journal of Biomedical Informatics, 103500.

[7] M. Bates, Tracking Disease: Digital Epidemiology Offers New Promise in Predicting Outbreaks, IEEE Pulse. 8 (2017) 18–22. doi:10.1109/MPUL.2016.2627238.

[8] A. Mike, C. Daniel, Social media, big data, and mental health: current advances and ethical implications, Curr. Opin. Psychol. (2016). doi:10.1016/j.copsyc.2016.01.004

[9] K. Nargund, S. Natarajan, Public health allergy surveillance using micro-blogs, 2016 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2016. (2016) 1429–1433. doi:10.1109/ICACCI.2016.7732248

[10] K. Lee, A. Agrawal, A. Choudhary, Datasets, Mining Social Media Streams to Improve Public Health Allergy Surveillance, (2015) 815–822.

[11] K. Lee, A. Agrawal, A. Choudhary, Forecasting Influenza Levels Using Real-Time Social Media Streams, Proc. - 2017 IEEE Int. Conf. Healthc. Informatics, ICHI 2017. (2017) 409–414. doi:10.1109/ICHI.2017.68.

[12] R.A. Calix, R. Gupta, M. Gupta, K. Jiang, Deep gramulator: Improving precision in the classification of personal health-experience tweets with deep learning, Proc. - 2017 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2017. 2017-Janua (2017) 1154–1159. doi:10.1109/BIBM.2017.8217820.

[13] V. Agarwal, H. P. Sultana, S. Malhotra, and A. Sarkar, "Analysis of Classifiers for Fake News Detection," Procedia Comput. Sci., vol. 165, no. 2019, pp. 377–383, 2019, doi: 10.1016/j.procs.2020.01.035.

[14] W. Haitao, H. Jie, Z. Xiaohong, and L. Shufen, "A short text classification method based on n-gram and cnn," Chinese J. Electron., vol. 29, no. 2, pp. 248– 254, 2020, doi: 10.1049/cje.2020.01.001.

[15] S. Deepak and B. Chitturi, "Deep neural approach to Fake-News identification," Procedia Comput. Sci., vol. 167, no. 2019, pp. 2236–2243, 2020, doi: 10.1016/j.procs.2020.03.276.

[16] Z. Li, F. Yang, and Y. Luo, "Context Embedding Based on Bi-LSTM in Semi-Supervised Biomedical Word Sense Disambiguation," IEEE Access, vol. 7, pp. 72928–72935, 2019, doi: 10.1109/ACCESS.2019.2912584.

[17] M. U. Salur and I. Aydin, "A Novel Hybrid Deep Learning Model for Sentiment Classification," IEEE Access, vol. 8, pp. 58080–58093, 2020, doi: 10.1109/ACCESS.2020.2982538.

[18] S. A. Cammel et al., "How to automatically turn patient experience free-text responses into actionable insights: A natural language programming (NLP) approach," BMC Med. Inform. Decis. Mak., vol. 20, no. 1, pp. 1–10, 2020, doi: 10.1186/s12911-020- 1104-5.

[19] C. Zhang, A. Gupta, C. Kauten, A. V. Deokar, and X. Qin, "Detecting fake news for reducing misinformation risks using analytics approaches," Eur. J. Oper. Res., vol. 279, no. 3, pp. 1036–1052, 2019, doi: 10.1016/j.ejor.2019.06.022.

[20] K. Liu and L. Chen, "Medical Social Media Text Classification Integrating Consumer Health Terminology," in IEEE Access, vol. 7, pp. 78185-78193, 2019, doi: 10.1109/ACCESS.2019.2921938.

[21] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, “Squad: 100,000+ questions for machine comprehension of text,” arXiv preprint arXiv:1606.05250, 2016.

[22] M. Marelli, L. Bentivogli, M. Baroni, R. Bernardi, S. Menini, and R. Zamparelli, “Semevaltask 2014 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment,” in Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014), 2014, pp. 1–8.

[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

[24] J. Pennington, R. Socher, and C. Manning, “Glove: Global vectors for word representation,” in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543

[25] M. Iyyer, V. Manjunatha, J. Boyd-Graber, and H. Daumé III, “Deep unordered composition rivals syntactic methods for text classification,” in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015, pp. 1681–1691.

[26] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, “Fasttext. zip: Compressing text classification models,” arXiv preprint arXiv:1612.03651, 2016.

[27] S. Wang and C. D. Manning, “Baselines and bigrams: Simple, good sentiment and topic classification,” in Proceedings of the 50th annual meeting of the association for computational linguistics: Short papers-volume 2. Association for Computational Linguistics, 2012, pp. 90–94.

[28] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in Advances in neural information processing systems, 2013, pp. 3111–3119

[29] K. S. Tai, R. Socher, and C. D. Manning, “Improved semantic representations from tree-structured long short-term memory networks,” arXiv preprint arXiv:1503.00075, 2015

[30] J. Cheng, L. Dong, and M. Lapata, “Long short-term memory-networks for machine reading,” arXiv preprint arXiv:1601.06733, 2016

[31] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank,” in Proceedings of the 2013 conference on empirical methods in natural language processing, 2013, pp. 1631–1642.

[32] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[33] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, “A convolutional neural network for modelling sentences,” in 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference, 2017.

[34] Y. Kim, “Convolutional neural networks for sentence classification,” in EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, 2018.

[35] J. D. Prusa and T. M. Khoshgoftaar, “Designing a better data representation for deep neural networks and text classification,” in Proceedings - 2016 IEEE 17th International Conference on Information Reuse and Integration, IRI 2016, 2016.

[36] L. Mou, R. Men, G. Li, Y. Xu, L. Zhang, R. Yan, and Z. Jin, “Natural language inference by tree-based convolution and heuristic matching,” arXiv preprint arXiv:1512.08422, 2015.

[37] L. Pang, Y. Lan, J. Guo, J. Xu, S. Wan, and X. Cheng, “Text matching as image recognition,” in 30th AAAI Conference on Artificial Intelligence, AAAI 2016, 2016.

[38] J. Wang, Z. Wang, D. Zhang, and J. Yan, “Combining knowledge with deep convolutional neural networks for short text classification,” in IJCAI International Joint Conference on Artificial Intelligence, 2017.

[39] S. Karimi, X. Dai, H. Hassanzadeh, and A. Nguyen, “Automatic Diagnosis Coding of Radiology Reports: A Comparison of Deep Learning and Conventional Classification Methods,” 2017.

[40] S. Peng, R. You, H. Wang, C. Zhai, H. Mamitsuka, and S. Zhu, “DeepMeSH: Deep semantic representation for improving large-scale MeSH indexing,” Bioinformatics, 2016.

[41] A. Rios and R. Kavuluru, “Convolutional neural networks for biomedical text classification: Application in indexing biomedical articles,” in BCB 2015 - 6th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2015.

[42] M. Hughes, I. Li, S. Kotoulas, and T. Suzumura, “Medical Text Classification Using Convolutional Neural Networks,” Studies in Health Technology and Informatics, 2017.

[43] M.-T. Luong, H. Pham, and C. D. Manning, “Effective approaches to attention-based neural machine translation,” arXiv preprint arXiv:1508.04025, 2015.

[44] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, “Disan: Directional self-attention network for rnn/cnn-free language understanding,” in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[45] I. Yamada and H. Shindo, “Neural attentive bag-of-entities model for text classification,” arXiv preprint arXiv:1909.01259, 2019.

[46] A. P. Parikh, O. Tackstrom, D. Das, and J. Uszkoreit, “A decomposable attention model for natural language inference,” arXiv preprint arXiv:1606.01933, 2016.

[47] Q. Chen, Z.-H. Ling, and X. Zhu, “Enhancing sentence embedding with generalized pooling,” arXiv preprint arXiv:1806.09828, 2018.

[48] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, “Abcdm: An attention-based bidirectional cnn-rnn deep model for sentiment analysis,” Future Generation Computer Systems, vol. 115, pp. 279–294, 2020.

[49] Z. Lin, M. Feng, C. N. d. Santos, M. Yu, B. Xiang, B. Zhou, and Y. Bengio, “A structured self-attentive sentence embedding,” arXiv preprint arXiv:1703.03130, 2017.

[50] S. Wang, M. Huang, and Z. Deng, “Densely connected cnn with multi-scale feature attention for text classification.” in IJCAI, 2018, pp. 4468–4474.

[51] G. Chen, D. Ye, E. Cambria, J. Chen, and Z. Xing, “Ensemble application of convolutional and recurrent neural networks for multi-label text categorization,” in IJCNN, 2017, pp. 2377–2383.

[52] D. Tang, B. Qin, and T. Liu, “Document modeling with gated recurrent neural network for sentiment classification,” in Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 1422–1432.