Check for updates

Systematic review of content analysis algorithms based on deep neural networks

Jalal Rezaeenour [1] · Mahnaz Ahmadi [1] · Hamed Jelodar [2] · Roshan Shahrooei [1]

Abstract

Today according to social media, the internet, Etc. Data is rapidly produced and occupies a large space in systems that have resulted in enormous data warehouses; the progress in information technology has significantly increased the speed and ease of data flow.text mining is one of the most important methods for extracting a useful model through extracting and adapting knowledge from data sets. However, many studies have been conducted based on the usage of deep learning for text processing and text mining issues.The idea and method of text mining are one of the fields that seek to extract useful information from unstructured textual data that is used very today. Deep learning and machine learning techniques in classification and text mining and their type are discussed in this paper as well. Neural networks of various kinds, namely, ANN, RNN, CNN, and LSTM, are the subject of study to select the best technique. In this study, we conducted a Systematic Literature Review to extract and associate the algorithms and features that have been used in this area. Based on our search criteria, we retrieved 130 relevant studies from electronic databases between 1997 and 2021; we have selected 43 studies for further analysis using inclusion and exclusion criteria in Section 3.2. According to this study, hybrid LSTM is the most widely used deep learning algorithm in these studies, and SVM in machine learning method high accuracy in result shown.

✉ Jalal Rezaeenour
    j.rezaee@qom.ac.ir

1 Department of Industrial Engineering, University of Qom, Qom, Iran

2 Faculty of computer science, Dalhousie University, 6050 University Ave, Halifax, NS B3H 1W5, Canada

🖄 Springer

1 Introduction

According to the increasing daily data production, we need a powerful tool to discover valuable information from large volumes of data. An important research area named data mining was developed to extract helpful information and useful knowledge of the text, and data mining serves as a basis for artificial intelligence [5]. AI can learn and train with the help of machine learning and deep learning. This learning and training process helps with the classification and analysis of results. In the 1950s, Alan Turing and Arthur Samuel introduced *machine learning* [62]. Artificial Intelligence (AI) techniques have increased in various domains [47]. This paper discussed and analyzed deep learning because it is an important origin of data science. It is very useful to data scientists who collect, analyze, and interpret large quantities of data; deep learning makes this method quick and simple [51]. In medical science, artificial intelligence has been used for classifying malignant tumours by a non-sequential recurrent ensemble of the deep neural network model. The training and validation accuracy and the ROC-AUC scores have been satisfactory over the existing models [38]. In medical science, artificial intelligence has been used for classifying malignant tumours by a non-sequential recurrent ensemble of the deep neural network model. The training and validation accuracy and the ROC-AUC scores have been satisfactory over the existing models [43].

In the following, the adjective deep in deep learning refers to the use of multiple layers in the network algorithms built in the form of multi-layered models based on various inputs. These models of deep learning are often based on Artificial Neural Networks(ANN), which is why they are called Deep Learning Neural Networks(DLNNs) [39]. Many examples demonstrated that DLNNs have better results than classical models. ANN is one of the most important data mining tools, which has also been used as a powerful learning model in various fields of text mining and data mining since the late 1990s [19].

In fact, in this paper, the application of neural networks in text processing issues concerning our Natural Language Processing (NLP) is considered, which will be reviewed due to the importance of categorization and prediction. It also applies linguistic analysis techniques to process the text. More attention has been paid to text classification due to the significance of assigning a predefined label to many documents automatically [70]. It also uses linguistic analysis techniques to process a text, such as sentiment analysis (SA) (also known as opinion mining) has been a core research topic in artificial intelligence (AI) [78]. A systematic review will guide researchers who are interested in conducting new research. Following an SLR method, all relevant studies from the accessible electronic databases are combined and presented to answer the research questions. An SLR study creates a new viewpoint and helps new researchers learn about advanced technology and scientific discipline [28]. SLR verifies research efforts related to a specific topic [47]. We expect the systematic study to be fully explained at all stages and transparent to other researchers.

This Systematic Literature Review (SLR) follows the Kitchenham and Charters [28] guidelines. It identifies different research papers published from 1997 to 2021 in the context of deep neural networks in programs related to content analysis algorithms. The number of reviewed papers was 130; however, after reviewing the inclusion and exclusion criteria, only 43 articles were in the study. Answered The research questions by extracting the appropriate information from the 43 articles and then forming a statistical representation using tables and figures. The results illustrate the trend of research conducted in this area over the past years and suggest new research topics. The present study systematically demonstrates text processing

processes, including NLP, information detection, text mining, visualization of results, and the integration of neural networks to analyze the collected studies.

The paper is organized as follows: Section 2 contains the literature review, Section 3 includes Materials and methods, Section 4 contains Data, Section 5 contains QualityControl, Section 6 contains Analysis of Results, Section 7 includes Challenges and Recommendations, and Section 8 comparison results and discussion.

2 Review of literature

A massive study has been conducted on data mining to extract relevant and significant information from raw and unstructured text. Much work has been done on extracting useful features using various deep learning models for better results in text classification problems. Some of the related work in text categorization are discussed in this section [44].

In this section, we describe some basic concepts that were used later. AI is a general concept that includes ML and deep learning methods. Over the past decades, due to the importance of data mining, it has been shown that it can display different types of data through text, numbers, image, audio, and video; in this paper, we use textual data; many articles have examined different texts. One of the essential methods in text mining is ANN, which has demonstrated exemplary performance in predicting texts and classifying various problems [72].

The Artificial Neural Network is intended to verify which input is compatible with which output type. This relationship of the Neural Network with the desired output value can be introduced as sentiment analysis, and deep learning models have developed a method for learning vector representation to classify sentiment [64]. Bengio and Mikolov proposed learning techniques for the semantic representation of words in the text. The authors generated embedding word vectors with semantics vectors and used them to transform words into vectors [58]. A systematic study of various methods of neural network classification in the Arabic text shows that text classification is a way to search for textual information and explore the data [65]. Deep Learning techniques have recently attracted the academic community [9]. In the survey, deep learning models have been extensively applied in the field of NLP and show extraordinary possibilities. Deep learning has emerged as a powerful machine learning technique that learns multiple layers of representations or features of the data and produces state-of-the-art prediction results. The several sections in this survey briefly described the main Deep Learning architectures and related techniques that have been applied to NLP tasks and a comprehensive study of its current applications in sentiment analysis [75]. Deep learning methods are useful for high-dimensional data and are becoming widely used in many areas of software engineering [36]. The main criterion of this paper for classification is to examine the model using the confusion matrix, which considers only the recurrent neural network models related to text processing. A deep learning and text mining framework for accident prevention is an example in society [79].

Deep learning model examples are typically designed for text classification (e.g., fast-text and Text-CNN or pre-trained language models, such as BERT(*Bidirectional Encoder Representations from Transformers*) [70]. With the massive progress of the internet, text data has become one of the main formats of big tourism data. Text is an effective and generally existing form of judgment expression and evaluation by users. In the past period, a variety of text mining techniques have been proposed and applied to tourism analysis to improve tourism

value analysis models, build tourism recommendation systems, create tourist profiles, and make policies for supervising tourism markets; Tourism text big data mining methods have made it conceivable to analyze the behaviours of tourists and realize real-time monitoring of the market. As the key text analysis technique, NLP is experiencing a period of strong progress. Both machine learning and present deep learning with high achievements have been greatly applied in NLP. The successes of these techniques have been further boosted by the progress of NLP, Machine Learning, and Deep Learning [34].

In addition, Text mining, a section of synthetic intelligence, is gaining ground nowadays regarding applications in business and analysis. Most of the recent research conducted during this paper focused principally on advanced or the hybrid of deep neural networks to induce economic and higher results. For the same, RNN and LSTM-based models have been used, and the accuracy of the proposed models has been enhanced by varying the hyper parameters and using Glove word embeddings [44].

Deep neural networks (DNNs) have revolutionized the field of NLP; Convolutional Neural networks (CNN) and Recurrent Neural networks (RNN), the two main types of DNN architectures, are widely explored to handle various NLP tasks and have shown that hidden size and batch size can make DNN performance vary dramatically. This suggests that optimizing these two parameters is crucial to the good performance of both CNNs and RNNs [71].

Also, an improved method for link forecast in attributed social networks is presented. One of the newest link prediction methods is embedding methods to generate the feature vector of each node of the graph and find unknown connections. In order to justify the proposal, the authors conducted many experiments on six real-world attributed networks for Comparison with the state-of-the-art network embedding methods. One of the newest link prediction methods is embedding methods to generate the feature vector of each node of the graph and find unknown connections [8].

This article, unlike other articles, examines text processing for emotion analysis using all in-depth learning methods. In this article, we compare the topic of emotion analysis with several methods.

Many studies have been done on text processing and deep neural networks. In this article, all deep learning methods have been studied, some machine learning methods have been studied, and text mining has not been studied only on a specific topic. This subject can be the difference between this article and other articles.The following definitions describe classification techniques:

2.1 Concepts and definitions

Text mining Text mining is an important way to extract meaningful and valuable information from unstructured textual data and helps researchers achieve their goals. Text mining methods, through the identification of the subject, patterns, and keywords, automatically examine a considerable amount of information and increase helpful knowledge and information from them, including various methods (information retrieval) [31, 63].

NLP Natural language processing (NLP) is mentioned as the branch of computer science and, more particularly, the branch of artificial intelligence or briefly (AI) that is concerned with giving computers the ability to understand the text and spoken words in much the same way human beings can. NLP is one of the important superlative technologies of the information

age. The history of NLP dates back to the 1950s [52, 58] NLP is used in classification, clustering, sentiment analysis, text summarization, etc.

Deep learning Deep learning is an extension of ML where deep neural networks are employed for feature extraction and analysis from large-sized datasets [59], and that is a subset of ML in which the computer learns without knowing the human previous characteristics and knowledge. Most Deep Learning models are based on artificial ANNs. The artificial neural network was first published in the 1960s and has strongly emerged since 2006; this network is sorted into layers and works in coordination. Their structure is an input layer, hidden (middle) layers, and an output layer that performs the calculations. During the training process, the weight of the neurons changes to achieve an optimal network. The models applied in NLP include Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). It is the junction of neural networks, graphic modelling, optimization, AI, pattern recognition, and signal processing. Deep learning architecture has emerged as the algorithm of the century because of its ability to generalize, learn, and meet practical expectations [58]. Figure 1 shows summaries and classification of Deep Learning and machine learning and related tasks along with their essential characteristics.

Supervised Learning This type of machine learning is based on using labelled data to train the learning algorithm. The data is labelled since it consists of pairs, an input that a vector can represent, and its corresponding desired output, which can be described as a supervisory signal [39].

Unsupervised Learning Unlike supervised learning, this method uses an input data set without any labelled output to train the learning algorithm. No right or wrong result exists for any input object, and no human intervention to correct or adjust like supervised learning [39].

RNN This type of network, also called RNN, was first created by David Rumelhart in 1986 and developed later. It is a kind of neural network used in speech recognition NLP and
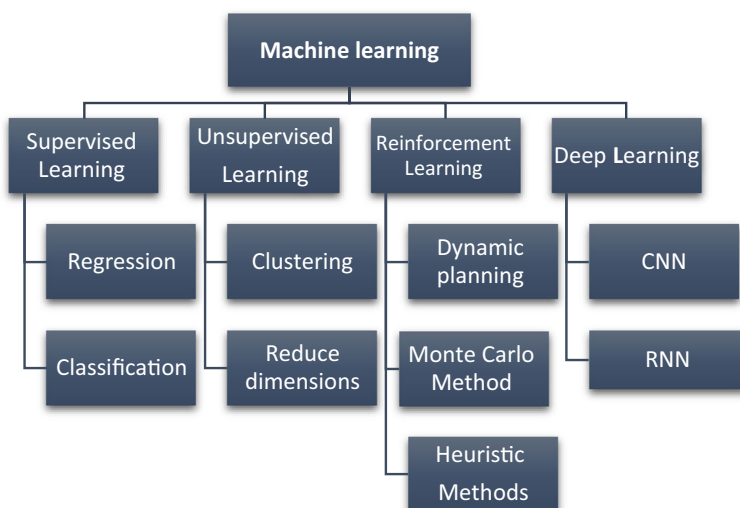

Fig. 1 classification of deep learning and machine learning [39]

sequential data processing. These networks do not transmit information in just one direction (from the input layer to the output one). In RNN, each node acts as a memory cell and carries on operating and calculating. Despite Feed Forward Neural Networks, in RNNs, edges can form circles. These networks can easily remember their previous input because of their memory power and use it to process the following sequence. LSTM is one of the most popular types of these networks. RNN is used for time series, text, and audio data. This network cannot memorize the sentence for long-term; that is the weakness of the networks, which is the so-called Vanishing gradient problem [61]. This network can be combined with convolutional networks; The problem with this network is that it is difficult to train.

The performance of RNN is such that it is necessary to know the previous words to predict the next term. RNN repeats the operations as is evident from its name (implies the same operation on all parts of a sequence (or series) of inputs). It takes the input Xt and uses the output ht. The resulted output depends on the previous calculations. This output indicates how much the output value will be compatible with the input variable. Therefore, the information will be transmitted from one step to the other, enabling us to model a sequence of vectors. We are trying to approximate a probabilistic distribution on all possible next words for the previously given words in a dictionary at each time step. The output layer in an RNN is also a Softmax layer that presents a vector that refers to [59].

LSTM After the RNN encountered gradient disappearance, LSTM networks emerged intending to solve the problem of gradient disappearance or gradient explosion in 1997 Hochreiter and Schmidhuber for the development of the RNN; these networks are for sequential data. In the recurrent network, the content of each step is rewritten. These networks are also the premiers to recurrent networks in assessing accuracy and speed and examining more data [21]. This kind of network always works better than RNN [26]. First, we should convert the list of input sequences into the vectors expected by an LSTM network. That means we must have linguistic representation. Word-embedding methods are used, then the texts converted into vectors are entered into the network one by one, and the amount of memory is updated after each word. Then, prediction or proper labelling will be made.

Convolutional Neural Network (CNN) A CNN is a kind of neural network used in AI, NLP, speech, and images. This network was first introduced by Hubel and Wiesel [23]. These networks are considered deep neural networks consisting of three hidden input and output layers; the convolution neural network center has a convolution layer, which is why this type of network is named. The input layer is an array of numbers. Image Net database data developed this architecture using the Relu activator function. Due to the higher performance speed and reduced training time, these networks have high accuracy in image recognition [56]; and feature recognition without human intervention. CNN no longer has the possibility of long-term dependence, and LSTM is preferred for issues such as language modelling in which dependency is essential. The high number of input data can be a disadvantage of this network.

Neural network The Artificial Neural Network (ANN) is composed of a group of several perceptron's/neurons in each layer. An artificial neural network is called Feed-Forward because the inputs are processed only in the forward direction. ANN consists of information, hidden, and output layers, and each layer tends to learn actual weights. These networks can be deep or shallow. External networks have a hidden layer (i.e., a layer between input and output), but deep networks have more layers called deep neural networks. Can use these networks to

solve problems with tabular data, image data, and text data. The neural network is a part of AI based on the biological model of humans and animals, which is a suitable method for detecting unknown patterns in data with different applications and types [72].

Text classification The text classification process is the method of automatizing a collection of documents into specific groups based on the content of the text itself through the application of particular technologies and algorithms [65].

Word embedding Since NLP is an algorithm, these algorithms must display words numerically as input vectors. In traditional NLP, words are represented as one-hot encoded one-dimensional vectors. For example, the sentence "artificial neural network is the best" comprises six words. The word *Artificial Neural Network is the best*; in this way, each word takes the number one, and the rest becomes zero. The length of this vector is six (100,000, 010000, 001000, 000100, 000010, and 000001). Word embedding is a method that helps to analyze the meanings of words. The embedding description is learned in word embedding using shallow neural networks [56]. A word embedding is a real value vector representing a single word based on the context in which it appears. They represent a dictionary and have a wide range of applications in NLP. There are several ways to learn word embedding [14, 18, 29].

Sentiment analysis In the discussion of NLP, sentiment analysis has a fundamental role. Sentiment analysis includes mental classification that shows a specific text as subjective or objective, and sentiment classification that classifies a mental text as positive, negative, or neutral [45].

3 Materials and methods

A literature review could only be considered a systematic review if it is based on research questions [74]; Fig. 2 follows the main stages of the research. In general, the research questions were identified.

This study uses guidelines for a systematic review of deep neural networks in text processing problems [28]. At first, the keyword search strategy was identified in the research, and then the data sources were identified. Next, the inclusion and exclusion criteria of the paper were determined. Furthermore finally, the quality and data analysis were evaluated. These steps are discussed in detail in the following sections.

3.1 Research strategy

The research was performed by limiting the efficient fundamental concepts related to the subject under study. Numerous studies have been conducted in deep learning, machine learning, different neural networks, RNN, and short-term memory. Many published studies are not reviewed within the scope of the present paper. Following the search strategy, the keywords used to collect the studies contained NLP, RNN, LSTM, deep learning, machine learning, and optimization that briefly and comprehensively showed.

– LSTM
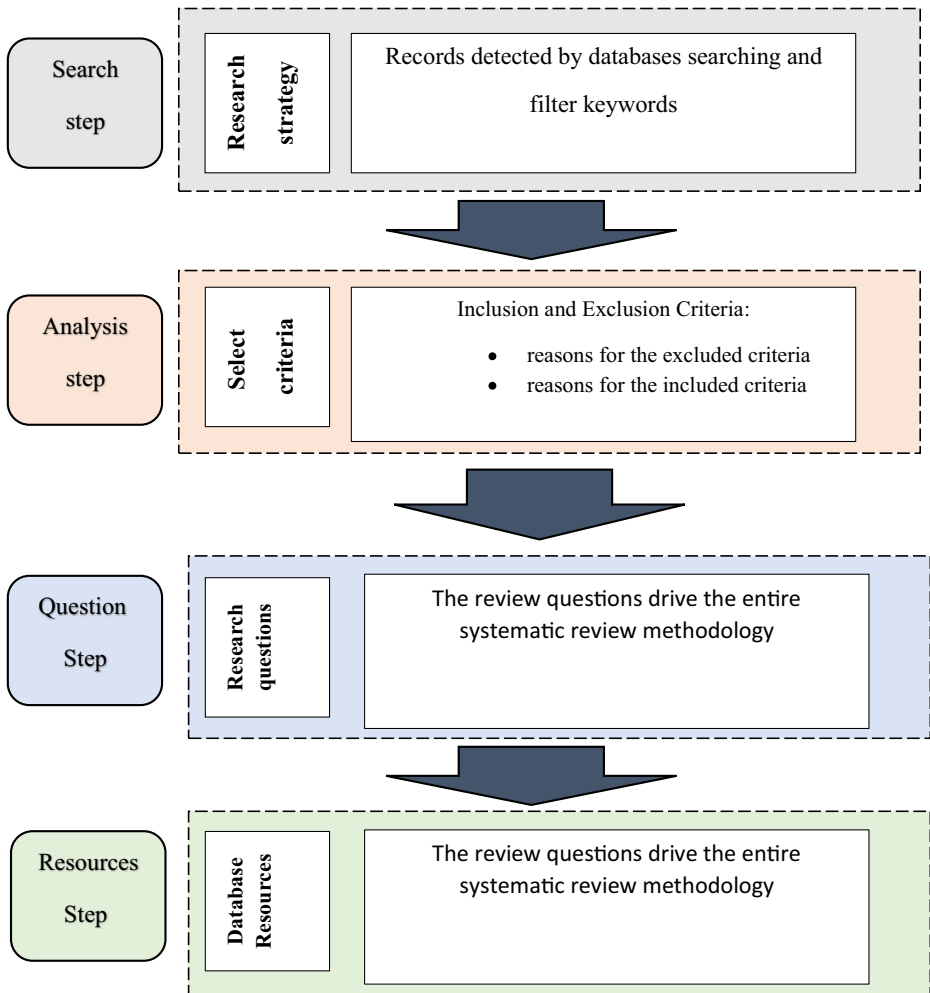– RNN,LSTM

Fig. 2 Research steps flow chart

- TEXT MINING,LSTM
- LSTM,BERT,RNN
- LSTM,DL,ML
- LSTM,NLP
- LSTM,WORD EMBEDDING
- LSTM,BERT
- LSTM,OPTIMIZATION
- BERT,TEXT MINING

In addition to abbreviations, phrases are also thoroughly searched. All the abbreviated words searched in databases show that more than a thousand related studies have been conducted and made available to the public.

3.2 Inclusion and exclusion criteria

+ Inclusion criteria The inclusion criteria have been chosen as follows, and the inclusion criteria are indicated with (+),

  + The study should be of the text classification type.
  + Input data must be textual.
  + There must be a method of embedding words.
  + Reputable journals should be reviewed unless they are part of a text classification study.
  + The presence of NLP words, RNN, short-term memory, text mining, and sentiment analysis in the title or keywords is a preference to enter the study list.

-Exclusion criteria Exclusion criteria include topics that are not valued in the study and are removed from the study list. Exclusion criteria are indicated with (−),

  – Papers that do not have obvious information.
  – Input data is of numerical case.
  – The subject of study is the type of forecasting process of an investigation.
  – The superior method is the kind of machine learning.
  – The papers should only include definitions and reviews of the method.
  – Short papers that are without model description and review.

Figure 3 shows the number of changes in the selection of papers. Out of 130 papers evaluated in the first stage, 63.

articles were removed; consequently, 67 articles remained in the review. Moreover, of those, 43 papers that directly with the issue of text classification and sentiment analysis are shown in Table 7.
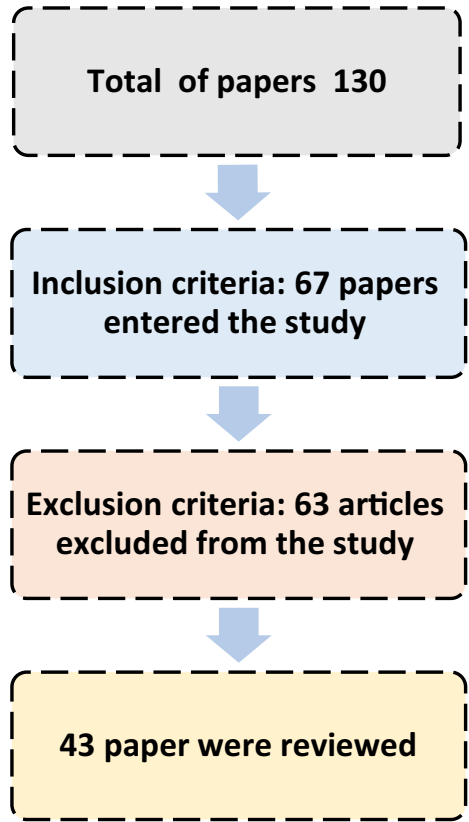
3.3 Research questions

In line with this study, the research questions focus on specific and general aspects of the paper. More specifically, this SLR poses the following four research questions:

1. What are methods of Machine Learning (ML) and deep learning used in classification and word processing?
2. Which methods are more accurate in classifying and processing text?
3. What are the active databases in optimization algorithm studies, data mining in neural networks, word processing, and Long Short Term Memory (LSTM)?
4. What methods are used for word embedding in studies and word processing problems?

3.4 Database resources

Today, digital libraries are suitable for searching for books, magazines, and articles. The present study started in November 2020. Most of the research is from 2016 to 2020; some is

Fig. 3 The process of selecting articles according to inclusion and exclusion criteria

Total of papers 130

Inclusion criteria: 67 papers entered the study

Exclusion criteria: 63 articles excluded from the study

43 paper were reviewed

from previous years. Table 1 shows journals providing papers and the frequency that have been used. In addition, Fig. 4 shows the frequency of articles received from 1997 to 2021. The papers from Google Scholar, Science Direct, Taylor, and Francis digital libraries are searched and stored. Then, the systematic rules necessary to meet the study's aims were applied. Keyword selection is a key step in any systematic review. It determines what articles should be detected.

Table 1 Databases of reviewed papers

DATABASE	Number
Elsevier	34
Springer	7
Taylor and Francis	10
Conference(IEEE)	27
ARXIV	26
MDPI	3
Wiley	1
Research Gate	3
Manuscript	4
Web of Science	2
Mat-aprire-no Information	11
SID	2
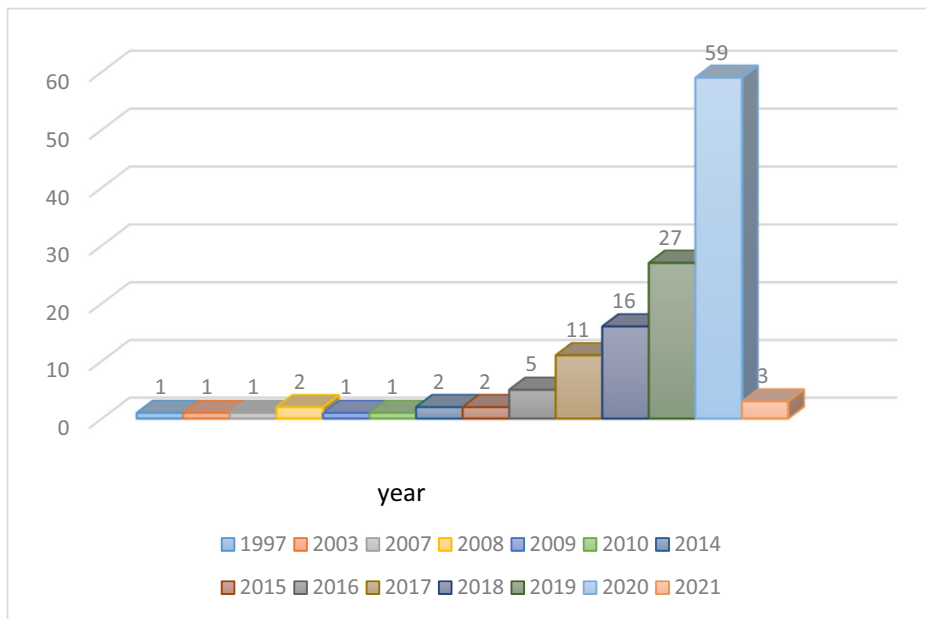
Fig. 4 Frequency of articles from 1997 to 2021

4 Data

In this section, we introduce standard datasets related to text mining to evaluate the proposed models. A variety of data has been used for assessing the proposed models. They have various features, which we discussed in this section, and we will introduce their widely used data.

Book Corpus The dataset consists of 11,038 unpublished books from 16 different types. The pre-training corpus consists of Book Corpus (800 million words) and English Wikipedia (2500 million words).

Internet Movie Database (IMDB) contains actors, movies, and TV series information. It has more than 100,000 movies and various series. It has 83 million registered users as of October 2018, and their opinions and sentiments for several movies [59].

Pima-Indians-diabetes-database (PIDD): This dataset is originally from the National Institute of Diabetes, Digestive, and Kidney Diseases.

Stanford Sentiment Treebank (SST): This dataset includes 11,855 movie reviews and includes train (8544), development (1101), and test (2210) data [67].

SST1 It is an extension of MR(Movie review) but provides five kinds of labels, including very negative, negative, neutral, positive, and very positive [66].

SST2 The set consists of two classes, including negative and positive. The number of training, development, and test sets samples is 6920, 872, and 1821, respectively [15].

Subjectivity (SUBJ) The subjectivity dataset contains sentences labelled concerning their subjectivity status (subjective or objective) [67].

Wikipedia It is a free online encyclopedia created worldwide and is available in 317 languages; it has over 55 million articles.

Twitter Twitter is an American social network where users post messages known as tweets. These datasets can be labelled with three sentiment poles: positive, neutral, and negative [53].

PUBMED It is a free search engine that primarily accesses the MEDLINE database from sources and abstracts of biosciences and biomedical topics.

TREC TREC is a data set for classification questions based on reality. TREC divides all questions into six categories: location, human, entity, abbreviation, description, and numeric. The training dataset contains 5452 labelled questions, while the testing dataset contains 500 questions [80].

Movie Review A set of videos was recovered from imdb.com in the early 2000s by Bo Pang and Lillian Lee. These studies were collected and made available as part of their research on NLP [66].

Amazon This dataset includes reviews of Amazon products and services and contains reviews (ratings, text, and helpful votes), metadata products (descriptions, category information, price, brand, and image features), and links (viewed/also bought graphs).

5 Quality control

Another challenging problem is the evaluation of results. The recommended methods should be evaluated and compared with other methods. We can use machine learning and data mining methods to identify hidden patterns. These methods are available to evaluate the performance of machine algorithms, classification, and regression. The selection criteria must be considered carefully. Also, types of evaluation criteria can be mentioned in Table 2 [18].

Figure 5 pie chart shows the evaluation criteria used in the deep learning paper and Table 3 shows the frequency evaluation criteria, and also the evaluation criteria used in the articles are illustrated in Table 7. The most widely used evaluation of the criterion for the model is accurate.

6 Analysis of results

The papers that were evaluated through inclusion and exclusion criteria were categorized. Table 7 shows that deep learning methods have been reviewed in the papers on sentiment analysis, text classification, and text mining. However, in this research, various models such as RNN, short-term memory, CNN, Bert algorithm, and neural network; were investigated based on hybrid ideas to show that they have achieved high accuracy in the field. Accuracy is a more important factor in evaluation because it shows the proportion of correctly classified numbers to the total number of classifications. A total of 43 articles that have directly discussed sentiment analysis and text classification, along with the methods considered in the papers and the method chosen among the methods, the evaluation criteria, and the type of word

Table 2 Evaluation criteria used in the papers on deep learning and neural network

Formula	Method Name	Evaluation Method
$\frac{TP+TN}{TP+TN+FP+FN}$	Accuracy(ACC)	- The ability of a measurement to match the actual (true) value of the quantity being measured [42].
$\frac{TP}{TP+FP}$	Precision(P)	Precision is the most critical performance measure in this field to evaluate the model that focuses on false positives (FP(or positive) prediction [42].
$\frac{TP}{TP+FN}$	Recall or Sensitivity	- The recall usually utilized to evaluate the model focuses on false negatives (FN). It shows how many samples are truly positive among all the positive examples.
$F1 = 2 * \frac{precision*recall}{precision+recall}$	F1 score or F-measure	Harmonic average of two criteria of precision and recall The best state is one, and the worst condition is zero.
Curve	AUC and ROC(Receiver Operating Characteristic)	- Based on the ROC, the curve area under the curve AUC (Area under the curve) can calculate the overall performance measures, And gives us information about the model's ability to detect classes. The model will be better if the AUC is higher [63].
$MAPE = \frac{1}{N}\sum_{K=1}^{K=N} AP_K$	Map(Mean Average Precision)	A score widely used in ranking problems [44] AP is class K, and N is the number of classes [1].
$MRR = \frac{1}{N}\sum_{I=1}^{N} \frac{1}{rank\ i}$	Mean Reciprocal Rank (MRR)	Mean Reciprocal Rank is a statistical criterion for evaluating each process that provides a list of possible answers to a sample of questions sorted by probability of accuracy [76].
LR	Learning Rate	The learning rate is a configurable hyperparameter used in training neural networks with a positive value, mostly between 0.0 and 1.0 [46].
$MSE = \frac{1}{N}\sum\left(Y-\widehat{Y}\right)^2$	mean squared error (MSE)	- In statistics, the MSE estimate measures the average of the squares of the errors, that is, the average squared difference between the estimated values and the actual value [16, 20].
$RMSE = \sqrt{\frac{1}{N}\sum\left(Y-\widehat{Y}\right)^2}$	root mean squared error (RMSE),	Root Mean Square Error (RMSE) is the standard deviation of the residuals (prediction errors) [1, 16].
$MRR = \frac{1}{N}\sqrt{\left|Y-\widehat{Y}\right|}$	mean absolute error (MAE)	Mean Absolute Error (MAE) is a public metric used to amount accuracy for continuous variables [43].

embedding, have been classified. Other papers classified by systematic method, text mining, explanation of methods, and algorithms were included in the selected articles but not in the analysis category. The analysis of the results showed that the methods related to the deep learning algorithm performed stronger than the machine-learning algorithm in the study related to text classification, sentiment analysis, and text mining methods.

The results showed that the performance of short-term memory alone and the combination of this network with convolutional networks, Bert, as well as attention-based mechanism and BI-LSTM, have mostly been used and have shown a higher accuracy percentage. The superior hybrid methods are categorized in Table 4.

The use of machine learning methods is one of the exclusion criteria in this study, but due to their application in text processing problems, the algorithms mentioned in Table 5 are employed. The methods' definition and utilization are discussed in the [12, 77].

The SVM algorithm has the most potent method among the nominated algorithms that have shown the highest accuracy in classifying problems. In addition, the procedures for word embedding in the studies and text processing problems are shown in the last column of Table 7.
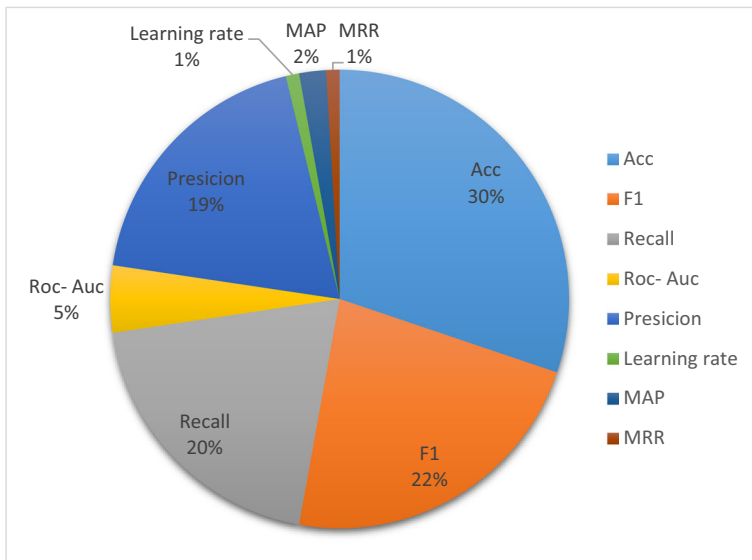
Fig. 5 The pie chart shows the evaluation criteria used in the Deep Learning paper

The methods of word embedding in text processing problems are fully described in the systematic literature review [29] and are shown in Table 6. According to the collected studies, it is found that the process of text processing techniques related to deep learning and machine learning is approximately related to each other.

In both methods of deep learning and machine learning, data preparation, selection, and extraction of algorithm feature selection and learning method, model training based on the train data, model testing based on the test data, evaluation of model accuracy, model application, and final evaluation, and knowledge acquisition from data analysis has been used.

Table 3 Frequency of evaluation criteria used in articles

Evaluation Criteria	Frequency
ACC	30%
F1	22%
Recall	20%
ROC-AUC	5%
Precision	19%
Learning rate	1%
MRR	1%
MAP	2%

Table 4 Hybrid methods with higher accuracy and performance in text classification

Article Citetaion	Hybrid Methods LSTM
Chao et al., 2018 [11]	CONV-LSTM
Ingole et al., 2018 [24]	CNN-LSTM
Song et al., 2020 [55]	BERT-LSTM
Alsentzer et al., 2019 [3]	BI-LSTM
Zhu et al., 2019 [82]	ATT- BI-LSTM
Abdel-Nasser & Mahmoud, 2019 [1]	LSTM-RNN

Table 5 Effective Machine-learning algorithms in classification and text processing

Abbreviate	Machine Learning Method
SVM	Support Vector Machine
NB	Naive Bayes algorithm
KNN	k Nearest Neighbor
LR	Logistic Regression
RF	Random Forest
DT	Decision Tree

From a text processing point of view, these two methods have the same procedure, but the major difference is that in deep learning, more details are considered, the learning process is repeated, and the number of steps is large. It also provides considerable accuracy over machine learning methods and requires more input data. According to hidden layers, deep learning networks do learning in the text, which causes the prediction of the next word, label recognition, questions, and answers (Table 7).

7 Challenges and recommendations

This research focused on the AI methods based on text mining and language processing techniques considering deep neural networks. The problem with these networks is that the concept of test and train is used to achieve the exact number of layers for labelling concepts, and the recurrent of the network helps reach this number, but more time is allocated. We recommend that for future studies, mathematical models to optimize the neural network will examine all number of layers and the layer that has been targeted. The algorithm's speed, computational power, and flexibility of neural networks can be compared. Also, the hybrid models of machine learning with deep learning could be examined in terms of accuracy and quality.

8 Comparison results and discussion

In this section, 4 types of neural networks are compared. Comparisons of these networks are shown in Table 8.

Table 6 Embedding word methods to convert text to vectors

Word Embedding Methods	
One Hot Encoding	
Word2vec	Skip gram
	CBOW
Global Vectors (Glove)	
FAST TEXT	
ELMO(Embedding Language Model)	
BERT	Bio Bert
	Clinical Bert
	SCI BERT
Other methods	ULMFIT (Universal Language Model Fine-Tuning) &GPT
	(Generative Pre-Training) &XLM(Cross-language Model)
	& ERNIE(Enhanced Language Representation with Informative Entities)

Table 7 Text classification and sentiment analysis paper

row	Research work	Classification Algorithm	Algorithm Name	Dataset	Evaluation Criteria	Preferred Method	text representation
1	Aggarwal et al., 2018 [2]	Deep learning	RNN, BERT, LSTM	Book corpus Wikipedia	ACC, F1 R, Roc	Fine-tuned BERT	Word2vec
2	Abdel-Nasser and Mahmoud 2019 [1]	Sentiment analysis Deep learning ML	Sent WordNet,LSTM, BERT, LR	IMDB	ACC, F1 P, R	BERT	BERT
3	Rahman et al., 2020 [46]	Deep learning	CONV-LSTM CNN T-LSTM CNN-LSTM	PIDD	Sensitivity specificity ACC learning rate	CONV-LSTM	–
4	C. Zhou et al., 2015 [80]	Text Classification	CNN, LSTM C-LSTM	SST	ACC	C-LSTM	N-gram Word2vec
5	Du et al., 2019 [15]	Text Classification	GRU, LSTM, BiGRU	SST2, Yelp	ACC	LSTM	BERT ELMO ULMFIT
6	Huang & Feng, 2019 [22]	Text Classification	CRNN, QRNN LSTM, RNN	YELP 2013	ACC	RNN	GLOVE
7	Jurgovsky et al., 2018 [27]	Sequence classification ML Deep learning	LSTM Random forest LR	credit-card transactions march to May 2015	ACC, P R, ROC	LSTM	AS-IS VECTOR EMBEDDI NG
8	Liu et al., 2016 [35]	Text Classification	RNN LSTM	SST1, SST2 IMBD, SUBJ	ACC	Three LSTM	Matrix-Vector NBOW s paragraph vectors
9	Lewis et al., 2017 [33]	Text Modeling	RNN, LSTM C-LSTM	Wikipedia	ACC	C-LSTM	ONE-HOT
10	Song et al., 2019 [53]	Sentiment Analysis	BERT-BASE, BERT-LSTM BERT ATTENTION	ABSA(Aspectbased sentiment analysis) SNLi(The Stanford Natural Language Inference)	ACC, F1	BERT –LSTM	BERT ELMO
11	Reddy & Delen, 2018 [50]	Deep learning Text Classification	ANN,LSTM GRU, RNN	Cerner HealthFacts EMR database 2000to2015	ACC,AUC	LSTM	–
12	F. Zhang et al., 2019 [77]	Text mining ML	NLP,SVM,DT KNN, NB, LR	different causes of accidents	F1 P R	Optimized Ensemble	TF- IDF

Table 7 (continued)

row	Research work	Classification Algorithm	Algorithm Name	Dataset	Evaluation Criteria	Preferred Method	text representation
13	W. Zhang et al., 2019 [76]	Deep learning Semantic Analysis	LSTM,Te LSTM Te LSTM+sc	20newgroup Wiki10 Amazon Semeval07	MRR MAP ACC	Te LSTM(Topic-- Enhanced LSTM neural)+sc(Similarity Constraint)	BOW N-gram
14	Ghourabi et al., 2020 [18]	Deep learning Traditional,ML SMS Classification	LSTM, CNN, SVM KNN LR RF, DT	Smsspam Indian 2730 SMS	ACC, F1 P, R	CNN-LSTM	Word2vec GlOVE TF-IDF
15	X. Wang et al., 2016 [66]	Sentiment Analysis Deep learning	CNN, RNN, LSTM GRU	SST1 SST2 MR	ACC	RNN	Embedding matrix ONE HOT Word2vec
16	X. Zhang & Zhang, 2020 [73]	Sentimental classification	LSTM, NLP RNN	Reuters Twitter	ACC F1	LSTM	Word2vec
17	Bahad et al., 2019 [6]	Sentiment and text classification	CNN,RNN,LSTM BI-LSTM,UNI-LSTM	DS1 and DS2	ACC	BI-LSTM RNN	BOW,TF-IDF GLOVE, word2vec fast text
18	J. H. Wang et al., 2018 [68]	Sentiment Classification, Deep Learning	LSTM ELMO NB	IMDB, Douban Movies., social media platform PTT in Taiwan	ACC, F1 P, R	LSTM	Word2vec CBOW Skip gram
19	Gajendran et al., 2020 [17]	Text processing classification	BLSTM(bidirectional long short term memory) CRF(conditional random field), Text mining	Biomedical NER Corpus	F1 P R	B-LSTM	N-gram
20	Jelodar et al., 2020 [25]	Deep Sentiment Classification	Deep Sentiment Classification	563,079 COVID-19–related comments	ACC	LSTM	GLOVE
21	Ombabi et al., 2020 [41]	CNN LSTM SVM	CNN LSTM SVM	corpus for Arabic text; it contains 63,000 books reviews	ACC, P, R	CNN-LSTM	Fast text Skip gram Word2vec

Table 7 (continued)

row	Research work	Classification Algorithm	Algorithm Name	Dataset	Evaluation Criteria	Preferred Method	text representation
22	H. Chen et al., 2020 [13]	Text Classification	NB, KNN, LSTM. SVM, ATT- BILSTM	Taiwan Hospital website	F1, P R, ACC	ATT-BI LSTM	BOW TF-IDF
23	Banerjee et al., 2019 [7]	Text report classification	SVM, Adaboost CNN, HNN DPA, LSTM	Whole Radiology CORPUS 117816	F1, P R, AUC	DPA-HNN	Word2VEC GLOVE
24	Nowak et al., 2017 [40]	Text and Sentiment Classification	LSTM, GRU BI-LSTM	Amazon dataset	ACC	BI-LSTM	BOW
25	Cai et al., 2020 [10]	Text Classification	LR, CNN BI GRU ATT LSTM HBLA	ARXIV ACADEMIC PAPER DATASET REUTERS CORPUS VOLUME	AUC F1 R p	HBLA(Hybrid BERT model incorporates Label semantics via adjective attention)	Word2vec GLOVE
26	C. Wang et al., 2017 [67]	Sentence Classification,	CNN RNN DNN	MR SST-1 SST-2 SUBJ IMDB	ACC	CONV RNN	GLOVE
27	Rao and Spasojevic, 2016 [48]	Text classification	RNN LSTM CNN	Twitter, Facebook and Google+	ACC	LSTM	GLOVE
28	A. Onan al., 2021 [42]	Sentiment analysis	CNN, RNN.BRNN LSTM	Courps(Massive open online courses (MOOCs)	ACC	LSTM	GLOVE, Fast Text, Word2vec
29	Alsentzer et al., 2019 [3]	Text mining	BERT Bio Bert Clinical Bert	MIMIC-III v1.4 database PubMed	ACC F1	BERT	BERT
30	Sun et al., 2019 [57]	Sentiment Analysis Sentence Classification	BERT LSTM	5215 sentences, 3862 of which contain a single target,	ACC F1 AUC	BERT	–
31	Arase & Tsujii, 2021 [4]	Sentence representation	Bert base NLI PPBERT-base	Twitter URL corpus Wikipedia Para-NMT	ACC	PPBERTbase	–
32	Colón-Ruiz & Segura–Bedmar, 2020 [14]	Sentiment analysis Multi-class text classification Deep learning	CNN LSTM BI LSTM Bert LSTM CNN LSTM	Book corpus	F1	LSTM-CNN	Word2vec
33	Kiran et al., 2020 [30]	Sentiment Classification Deep learning	CNN LSTM	SMS Spam, YouTube Spam, Large Movie Review Corpus, SST,	F1 P R ACC	LSTM-CNN	BOW TF-IDF

Table 7 (continued)

row	Research work	Classification Algorithm	Algorithm Name	Dataset	Evaluation Criteria	Preferred Method	text representation
34	Lee & Hsiang, 2020 [32]	Patent classification	BERT LARGE BERT BASE	Amazon Cellphone & Accessories and Yelp CELF-IP	F1, P R	fine-tuning BERT 1	–
35	Y. Zhang et al., 2021 [78]	Sentiment analysis	sentient, SVM,CNN LSTM,ATT LSTM, icon	3000 multiturn English conversations from several websites (e.g., eslfast.com and focus on English	F1 score P R ACC	interactive LSTM	BOW
36	P. Zhou et al., 2016 [81]	Text Classification	RNN,LSTM B LSTM	SST 1, SST-2 SUBJ, TREC MR	ACC	B- LSTM	GLOVE
37	H. C. Wang et al., 2020 [69]	Text summarization	LSTM, RNN	Science Direct in 6 years	P, R ACC, F-1	RNN	word2vec
38	Song et al., 2019 [54]	Text mining, Abstractive text summarization .	CNN, LSTM LSTM-CNN, LSTM	Stanford natural language process	ROUGE (Recall-Oriented Understudy for Gisting Evaluation	LSTM-CNN	Word2vec
39	Tomihira et al., 2020 [60]	Sentiment analysis	CNN, LSTM BERT	Emoji OF Twitter English Japanese	ACC P, R F1	BERT	Word2vec PCA Fast Text
40	Wahdan et al., 2020 [65]	Text classification- Deep learning, ML	LSTM, CNN RNN, CNN –RNN SVM, LR,NB DT, RF	Arabic text classification	ACC P R F1	LSTM	–
41	Razzaghnoori et al., 2018 [49]	Question classification	RNN LSTM	UTQD.2016 contains 1175 Persian questions	ACC	LSTM	Word2vec TF-IDF
42	Mohasseb et al., 2018 [37]	Question classification Text classification ML	SVM RF	TREC 2007 Question Yahoo Wikipedia	ACC P R F1	LSTM	BOW

Table 8 Comparison of relationships between CNN, RNN ANN networks

Method Features and application	RNN	LSTM	CNN	ANN
Data	Series Data (time series, text, audio)	Series Data (time series, text, audio)	Image, Video	Tabular Data
Feedback relationship	YES	YES	NO	NO
Parameter sharing	YES	YES	YES	NO
Spatial relationships	NO	NO	YES	NO
Vanishing and exploding gradient	YES	NO	YESS	YES
Deep Learning Neural Network	YES	YES	YES	NO
Criteria Evaluation (confusion matrix)	YES	YES	YES	YES

The word yes indicates that these networks are related to the named features. Also, The word data describes what kind of information can be used in any network. That is, digital data, text, audio, and video are done by which type of network. As well as the characteristics of reciprocal relations, Feedback relationships, Parameter sharing, Spatial relationships, Vanishing and exploding gradient, and Deep Learning Neural networks are defined for these networks. Moreover, it shows that the evaluation criteria in the confusion matrix are always used to examine the statistical test.

9 Conclusion_Future

We live in the "Information Age," being exposed to considerable information and data. Due to the production of an enormous amount of information daily in various processes and fields and the problems that come with manually categorizing concepts, we have used concepts related to NLP called text mining. This concept could convert a large amount of information into practical knowledge. Text mining can analyze and process text by machine learning and deep learning algorithms sub-branches of AI. According to the review of articles, this result was obtained from a set of methods used to analyze, predict, label results, and answer questions in text mining. Deep learning, including neural networks, has studied texts in many layers, and this is a survey of the long sequence of the text. The most widely used evaluation criteria for the model is accuracy, and the databases of SST1, SST2, and IMDB data have been used more in papers, respectively.

The neural networks, which achieved the highest accuracy, include the LSTM and RNN network and the hybrid LSTM with BI-LSTM, CNN-LSTM, CONV-LSTM, and BERT-LSTM were studied among neural networks. In addition, after deep learning methods, the support vector machine (SVM) has gained high accuracy of results among machine learning methods.

For activities we can do in the future. Can do sentiment analysis with multivariate decision-making methods. Can be compared MODM, MCDM TOPSIS methods to determine the accuracy of the network against confusion matrix methods. Examine the GRU network and compare its accuracy with the LSTM network.

Appendix

-Softmax activation functions are commonly used in the output layer for categorization problems. Commonly, softmax is used in the last layer of neural networks for final classification in feedforward neural networks.

$\sigma(X)$ of real values in the range (0, 1) that add up to 1. The function definition is as follows [75].

$$\sigma(x)_j = \frac{e^{x_j}}{\sum\limits_{k=1}^{K} e^{x_K}}$$

This function is almost identical to the sigmoid function, and the only difference is that the outputs of this function are normalized so that their sum is one.

-Sigmoid functions are often used in ANN to introduce nonlinear models. With the Using method, it is possible to process the input data linearly and present the result in a sigmoid relation. It is as follows [75].

$$sig(t) = \frac{1}{1 + e^{-t}}$$

Author contribution All authors contributed to the study conception and preparation. All authors read and approved the final manuscript.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest Authors declare that they have no conflict of interest.

References

1. Abdel-Nasser M, Mahmoud K (2019) Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Comput Applic 31(7):2727–2740. https://doi.org/10.1007/s00521-017-3225-z
2. Aggarwal A, Chauhan A, Kumar D, Mittal M, Verma S (2018) Classification of Fake News by Fine-tuning Deep Bidirectional Transformers based Language Model. ICST Trans Scalable Inf Syst April:163973. https://doi.org/10.4108/eai.13-7-2018.163973
3. Alsentzer E et al (2019) Publicly available clinical BERT embeddings, arXiv
4. Arase Y, Tsujii J (2021) Transfer fine-tuning of BERT with phrasal paraphrases. Comput Speech Lang 66: 101164. https://doi.org/10.1016/j.csl.2020.101164
5. Ayre L, T. G. Group (2019) Data Mining for Information Professionals Lori Bowen Ayre LBAyre@galecia.com June 2006, no. July 2006
6. Bahad P, Saxena P, Kamal R (2019) Fake news detection using bi-directional LSTM-recurrent neural network. Procedia Comput Sci 165(2019):74–82. https://doi.org/10.1016/j.procs.2020.01.072
7. Banerjee I et al (2019) Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. Artif Intell Med 97(November 2017):79–88. https://doi.org/10.1016/j.artmed.2018.11.004
8. Berahmand K, Nasiri E, Rostami M, Forouzandeh S (2021) A modified DeepWalk method for link prediction in attributed social network. Computing 103(10):2227–2249. https://doi.org/10.1007/s00607-021-00982-2

9. Bustos O, Pomares-Quimbaya A (2020) Stock market movement forecast: A Systematic review, Expert Syst Appl, vol. 156, https://doi.org/10.1016/j.eswa.2020.113464.

10. Cai L, Song Y, Liu T, Zhang K (2020) A hybrid BERT model that incorporates label semantics via Adjustive attention for multi-label text classification. IEEE Access 8:152183–152192. https://doi.org/10.1109/ACCESS.2020.3017382

11. Chao Z, Pu F, Yin Y, Han B, Chen X (2018) Research on real-time local rainfall prediction based on MEMS sensors. J Sensors 2018:1–9. https://doi.org/10.1155/2018/6184713

12. Chen CW, Tseng SP, Kuan TW, Wang JF (2020) Outpatient text classification using attention-based bidirectional LSTM for robot-assisted servicing in hospital, Inf, vol. 11, no. 2, https://doi.org/10.3390/info11020106.

13. Chen H, Zheng G, Ji Y (2020) Generating Hierarchical Explanations on Text Classification via Feature Interaction Detection, arXiv, no. Cd, https://doi.org/10.18653/v1/2020.acl-main.494

14. Colón-Ruiz C, Segura-Bedmar I (2020) Comparing deep learning architectures for sentiment analysis on drug reviews. J Biomed Inform 110(August):103539. https://doi.org/10.1016/j.jbi.2020.103539

15. Du M, Liu N, Yang F, Ji S, Hu X (2019) On attribution of recurrent neural network predictions via additive decomposition, arXiv, pp. 383–393

16. Fan GF, Yu M, Dong SQ, Yeh YH, Hong WC (2021) Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. Util Policy 73(February): 101294. https://doi.org/10.1016/j.jup.2021.101294

17. Gajendran S, Manjula D, Sugumaran V (2020) Character level and word level embedding with bidirectional LSTM – Dynamic recurrent neural network for biomedical named entity recognition from literature. J Biomed Inform 112:103609. https://doi.org/10.1016/j.jbi.2020.103609

18. Ghourabi A, Mahmood MA, Alzubi QM (2020) A hybrid CNN-LSTM model for SMS spam detection in arabic and english messages. Futur Internet 12(9):1–16. https://doi.org/10.3390/FI12090156

19. Yoav Goldberg, A Primer on Neural Network Models for Natural Language Processing, J Artif Intell Res, vol. 57, pp. 345–420, 2016, [Online]. Available: http://www.jair.org/papers/paper4992.html

20. Hiew JZG, Huang X, Mou H, Li D, Wu Q, Xu Y (2019) BERT-based financial sentiment index and LSTM-based stock return predictability, arXiv, no. 2005

21. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

22. Huang J, Feng Y (2019) Optimization of recurrent neural networks on natural language processing, ACM Int. Conf. Proceeding Ser., no. October 2019, pp. 39–45, https://doi.org/10.1145/3373509.3373573

23. Hubel DH, Wiesel T (1962) and Functional Architecture in the Cat's Visual Cortex From the Neurophysiolojy Laboratory , Department of Pharmacology central nervous system is the great diversity of its cell types and inter- receptive fields of a more complex type (Part I) and to, Most, pp. 106–154

24. Ingole P, Bhoir S, Vidhate AV, (2018) Hybrid Model for Text Classification, Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2018, no. Iaeac, pp. 450–458, https://doi.org/10.1109/ICECA.2018.8474920

25. Jelodar H, Wang Y, Orji R, Huang S (2020) Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. IEEE J Biomed Heal Informatics 24(10):2733–2742. https://doi.org/10.1109/JBHI.2020.3001216

26. Jiang Q, Tang C, Chen C, Wang X, Huang Q (2019) Stock Price Forecast Based on LSTM Neural Network Springer International Publishing. https://doi.org/10.1007/978-3-319-93351-1_32.

27. Jurgovsky J, Granitzer M, Ziegler K, Calabretto S, Portier PE, He-Guelton L, Caelen O (2018) Sequence classification for credit-card fraud detection. Expert Syst Appl 100:234–245. https://doi.org/10.1016/j.eswa.2018.01.037

28. Keele S (2007) Guidelines for performing systematic literature reviews in software engineering, Tech report, Ver. 2.3 EBSE Tech Report EBSE

29. Khattak FK, Jeblee S, Pou-Prom C, Abdalla M, Meaney C, Rudzicz F (2019) A survey of word embeddings for clinical text. J Biomed Informatics X 4(October):100057. https://doi.org/10.1016/j.yjbinx.2019.100057

30. Kiran R, Kumar P, Bhasker B (2020) Oslcfit (organic simultaneous LSTM and CNN fit): a novel deep learning based solution for sentiment polarity classification of reviews. Expert Syst Appl 157:113488. https://doi.org/10.1016/j.eswa.2020.113488

31. Kwartler T (2017) What is text mining?, Text Min Pract with R, pp. 1–15, https://doi.org/10.1002/9781119282105.ch1.

32. Lee JS, Hsiang J (2020) Patent classification by fine-tuning BERT language model. World Patent Inf 61(1): 101965. https://doi.org/10.1016/j.wpi.2020.101965

33. Lewis B, Smith I, Fowler M, Licato J (2017) The robot mafia: A test environment for deceptive robots, 28th Mod. Artif Intell Cogn Sci Conf MAICS 2017, pp. 189–190, https://doi.org/10.1145/1235

34. Li Q, Li S, Zhang S, Hu J, Hu J (2019) A review of text corpus-based tourism big data mining, Appl Sci, vol. 9, no. 16, https://doi.org/10.3390/app9163300.

35. Liu P, Qiu X, Xuanjing H (2016) Recurrent neural network for text classification with multi-task learning, IJCAI Int. Jt. Conf. Artif. Intell., vol. 2016-Janua, pp. 2873–2879

36. Majumder S, Balaji N, Brey K, Fu W, Menzies T (2018) 500+ Times Faster Than Deep Learning, pp. 554–563, https://doi.org/10.1145/3196398.3196424

37. Mohasseb A, Bader-El-Den M, Cocea M (2018) Question categorization and classification using grammar based approach. Inf Process Manag 54(6):1228–1243. https://doi.org/10.1016/j.ipm.2018.05.001

38. Moitra D, Mandal RK (2022) Classification of malignant tumors by a non-sequential recurrent ensemble of deep neural network model. Multimed Tools Appl 81(7):10279–10297. https://doi.org/10.1007/s11042-022-12229-z

39. Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K (2019) Speech recognition using deep neural networks: a systematic review. IEEE Access 7:19143–19165. https://doi.org/10.1109/ACCESS.2019.2896880

40. Nowak J, Taspinar A, Scherer R (2017) LSTM recurrent neural networks for short text and sentiment classification, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10246 LNAI, pp. 553–562, https://doi.org/10.1007/978-3-319-59060-8_50

41. Ombabi AH, Ouarda W, Alimi AM (2020) Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks. Soc Netw Anal Min 10(1):1–13. https://doi.org/10.1007/s13278-020-00668-1

42. Onan A (2021) Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach. Comput Appl Eng Educ 29(3):572–589. https://doi.org/10.1002/cae.22253

43. Ozbayoglu AM, Gudelek MU, Sezer OB (2020) Deep learning for financial applications: a survey. Appl Soft Comput J 93:106384. https://doi.org/10.1016/j.asoc.2020.106384

44. Pandey AC, Garg M, Rajput S (2019) Enhancing Text Mining Using Deep Learning Models, 2019 12th Int. Conf Contemp Comput IC3 2019, pp. 1–5, https://doi.org/10.1109/IC3.2019.8844895

45. Potts C (2014) CS224u: Sentiment analysis [PowerPoint slides], CS224u, p. Natural Language Understanding, [Online]. Available: https://web.stanford.edu/class/cs224u/2014/slides/cs224u-2014-lec15-sentiment.pdf

46. Rahman M, Islam D, Mukti RJ, Saha I (2020) A deep learning approach based on convolutional LSTM for detecting diabetes. Comput Biol Chem 88:107329. https://doi.org/10.1016/j.compbiolchem.2020.107329

47. Rahmani AM et al (2022) Automatic COVID-19 detection mechanisms and approaches from medical images: a systematic review. Multimed Tools Appl 81:28779–28798. https://doi.org/10.1007/s11042-022-12952-7

48. Rao A, Spasojevic N (2016) Actionable and Political Text Classification using Word Embeddings and LSTM, [Online]. Available: http://arxiv.org/abs/1607.02501

49. Razzaghnoori M, Sajedi H, Jazani IK (2018) Question classification in Persian using word vectors and frequencies. Cogn Syst Res 47:16–27. https://doi.org/10.1016/j.cogsys.2017.07.002

50. Reddy BK, Delen D (2018) Predicting hospital readmission for lupus patients: an RNN-LSTM-based deep-learning methodology. Comput Biol Med 101:199–209. https://doi.org/10.1016/j.compbiomed.2018.08.029

51. Sarker IH (2021) Data science and analytics : an overview from data - driven smart computing, decision - making and applications perspective. SN Comput Sci 2:377. https://doi.org/10.1007/s42979-021-00765-8

52. Siddharthan A, Manning CD, Schutze H (2002) Foundations of Statistical Natural Language Processing. MIT Press, 2000. ISBN 0–262–13360-1. 620 pp. $64.95/£44.95 (cloth)., vol. 8, no. 1. https://doi.org/10.1017/s1351324902212851.

53. Song Y, Wang J, Jiang T, Liu Z, Rao Y (2019) Targeted Sentiment Classification with Attentional Encoder Network, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11730 LNCS, pp. 93–103, https://doi.org/10.1007/978-3-030-30490-4_9

54. Song S, Huang H, Ruan T (2019) Abstractive text summarization using LSTM-CNN based deep learning. Multimed Tools Appl 78(1):857–875. https://doi.org/10.1007/s11042-018-5749-3

55. Song Y, Wang J, Liang Z, Liu Z, Jiang T (2020) Utilizing BERT Intermediate Layers for Aspect Based Sentiment Analysis and Natural Language Inference, [Online]. Available: http://arxiv.org/abs/2002.04815

56. Sorin V, Barash Y, Konen E, Klang E (2020) Deep learning for natural language processing in radiology—fundamentals and a systematic review. J Am Coll Radiol 17(5):639–648. https://doi.org/10.1016/j.jacr.2019.12.026

57. Sun C, Huang L, Qiu X (2019) Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence, NAACL HLT 2019–2019 Conf. North Am Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf, vol. 1, pp. 380–385

58. Sur C (2019) Survey of deep learning and architectures for visual captioning—transitioning between media and natural languages. Multimed Tools Appl 78(22):32187–32237. https://doi.org/10.1007/s11042-019-08021-1

59. Tembhurne JV, Diwan T (2021) Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks. Multimed Tools Appl 80(5):6871–6910. https://doi.org/10.1007/s11042-020-10037-x
60. Tomihira T, Otsuka A, Yamashita A, Satoh T (2020) Multilingual emoji prediction using BERT for sentiment analysis. Int J Web Inf Syst 16(3):265–280. https://doi.org/10.1108/IJWIS-09-2019-0042
61. Torfi A, Shirvani RA, Keneshloo Y, Tavaf N, Fox EA (2020) Natural language processing advancements by deep learning: A survey, arXiv, pp. 1–21
62. Turing AM (2012) Computing machinery and intelligence, Mach Intell Perspect Comput Model, pp. 1–28, https://doi.org/10.1525/9780520318267-013.
63. Usai A, Pironti M, Mital M, Aouina Mejri C (2018) Knowledge discovery out of text data: a systematic review via text mining. J Knowl Manag 22(7):1471–1488. https://doi.org/10.1108/JKM-11-2017-0517
64. Vo QH, Nguyen HT, Le B, Le Nguyen M (2017) Multi-channel LSTM-CNN model for Vietnamese sentiment analysis, Proc. - 2017 9th Int. Conf. Knowl. Syst. Eng. KSE 2017, vol. 2017-Janua, pp. 24–29, https://doi.org/10.1109/KSE.2017.8119429
65. Wahdan A, Hantoobi S, Salloum SA, Shaalan K (2020) A systematic review of text classification research based on deep learning models in Arabic language. Int J Electr Comput Eng 10(6):6629–6643. https://doi.org/10.11591/IJECE.V10I6.PP6629-6643
66. Wang X, Jiang W, Luo Z (2016) Combination of convolutional and recurrent neural network for sentiment analysis of short texts, COLING 2016 - 26th Int. Conf. Comput. Linguist. Proc. COLING 2016 Tech. Pap., pp. 2428–2437
67. Wang C, Jiang F, Yang H (2017) A hybrid framework for text modeling with convolutional RNN, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., vol. Part F1296, pp. 2061–2070, https://doi.org/10.1145/3097983.3098140
68. Wang JH, Liu TW, Luo X, Wang L (2018) An LSTM approach to short text sentiment classification with word embeddings, Proc. 30th Conf. Comput. Linguist. Speech Process. ROCLING 2018, pp. 214–223
69. Wang HC, Hsiao WC, Chang SH (2020) Automatic paper writing based on a RNN and the TextRank algorithm. Appl Soft Comput J 97:106767. https://doi.org/10.1016/j.asoc.2020.106767
70. Xu J, Du Q (2020) Learning neural networks for text classification by exploiting label relations. Multimed Tools Appl 79(31–32):22551–22567. https://doi.org/10.1007/s11042-020-09063-6
71. Yin W, Kann K, Yu M, Schütze H (2017) Comparative Study of CNN and RNN for Natural Language Processing, [Online]. Available: http://arxiv.org/abs/1702.01923
72. Zhang PG (2010) Data Mining and Knowledge Discovery Handbook, Data Min Knowl Discov Handb, no., https://doi.org/10.1007/978-0-387-09823-4
73. Zhang X, Zhang L (2020) Topics extraction in incremental short texts based on LSTM. Soc Netw Anal Min 10(1):1–9. https://doi.org/10.1007/s13278-020-00699-8
74. Zhang J, Xie J, Hou W, Tu X, Xu J, Song F, Wang Z, Lu Z (2012) Mapping the knowledge structure of research on patient adherence: knowledge domain visualization based co-word analysis and social network analysis. PLoS One 7(4):1–7. https://doi.org/10.1371/journal.pone.0034497
75. Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: a survey. Wiley Interdiscip Rev Data Min Knowl Discov 8(4):1–25. https://doi.org/10.1002/widm.1253
76. Zhang W, Li Y, Wang S (2019) Learning document representation via topic-enhanced LSTM model. Knowledge-Based Syst 174:194–204. https://doi.org/10.1016/j.knosys.2019.03.007
77. Zhang F, Fleyeh H, Wang X, Lu M (2019) Construction site accident analysis using text mining and natural language processing techniques. Autom Constr 99(June 2018):238–248. https://doi.org/10.1016/j.autcon.2018.12.016
78. Zhang Y, Tiwari P, Song D, Mao X, Wang P, Li X, Pandey HM (2021) Learning interaction dynamics with an interactive LSTM for conversational sentiment analysis. Neural Netw 133:40–56. https://doi.org/10.1016/j.neunet.2020.10.001
79. Zhong B, Pan X, Love PED, Sun J, Tao C (2020) Hazard analysis: A deep learning and text mining framework for accident prevention. Adv Eng Inform 46(August):101152. https://doi.org/10.1016/j.aei.2020.101152
80. Zhou C, Sun C, Liu Z, Lau FCM (2015) A C-LSTM Neural Network for Text Classification, [Online]. Available: http://arxiv.org/abs/1511.08630

81. Zhou P, Qi Z, Zheng S, Xu J, Bao H, Xu B (2016) Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling, COLING 2016 - 26th Int Conf Comput Linguist Proc COLING 2016 Tech Pap, vol. 2, no. 1, pp. 3485–3495

82. Zhu Y, Jiang S (2019) Attention-based densely connected LSTM for video captioning, MM 2019 - Proc. 27th ACM Int. Conf. Multimed., pp. 802–810, https://doi.org/10.1145/3343031.3350932