2022 3rd International Conference for Emerging Technology (INCET)
Belgaum, India. May 27-29, 2022

An Analysis and Comparison Of Deep-Learning Techniques and Hybrid Model for Sentiment Analysis for Movie Review

Swapnil Sinha
Electronics and Communication
Government Engineering College Barton Hill.
Trivandrum, India
swapnilsinha139@gmail.com

Abishek Jayan
Civil Engineering
Mar Baselios College of Engineering
Trivandrum, India
abishekjayan98@gmail.com

Rishabh Kumar
Department of Information Technology
Manipal University Jaipur
Jaipur, India
rishabhkumar140999@gmail.com

2022 3rd International Conference for Emerging Technology (INCET) | 978-1-6654-9499-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/INCET54531.2022.9824630

***Abstract*—In today's world, where each day 2.5 quintillion bytes of data is generated, the safe storage and secure processing of such data is a fundamental task. The vastness of the web has brought trillions of data as comments, reviews, blog posts etc. For a service provider the analysis of these are very essential and also finding out the trends in which sentiment analysis plays a huge role. This differs from typical topic-based text classification as it is done by classifying text based on the sentiment it conveys. User generated text is the basis from which opinion and subject knowledge is drawn for sentiment analysis.In this paper, we deal with the sentiment analysis of the IMDB dataset of movie reviews. We explore the effectiveness of different deep learning techniques such as CNN, LSTM, LSTM-CNN, GRU, BERT, BERT-CNN, BERT-LSTM for the sentiment analysis of the IMDB movie review dataset. Data cleaning and the techniques used are explained and the performance of these various algorithms used are measured in terms of recall, precision and accuracy. In this we hope to find the best machine learning model amongst the ones tested for future research.**

***Index Terms*—Sentiment Analysis, CNN, deep learning, LSTM, GRU, BERT, Word Embedding, machine learning**

I. Introduction

In the data world, millions of comments on social networks,financial sites, e-commerce sites are shared by users. This type of data is valuable reference information for customers and suppliers, however, it is difficult for humans to process such a substantial amount of information. So we need to create an automated scoring system to show how the customer is feeling. The system can predict the emotion of the content of praise or criticism. Both the customer's emotional outlook towards the content as well as the polarity must be considered. Using sentiment analysis systems, unstructured data regarding items, companies, movies, or themes on which individuals can express their feelings can be translated into structured data.It enables businesses to have a better understanding of the overall response to their products and their place in the market. In addition, they can better respond to consumer needs and wants while helping their organization grow [4]. It allows the associations to all the more promptly understand the general reaction to their things and their staying watching out. Moreover, they can all the more promptly address the purchaser's issues and needs while simultaneously helping their relationship with creating. In this to compare and contrast various significant learning models In this paper we contemplate the significant learning models and moreover a hybrid of these models like for example CNN, LSTM, LSTM-CNN, GRU, BERT and subsequently we propose one more model of request that joins BERT into two techniques. We will probably show the suitability of BERT when used for assessment examination stood out from various models and a short time later changing BERT with LSTM and CNN and difference them and existing models.

II. Related Works

Analytics of online user content is done using sentiment analysis. Social media faces many challenges in this aspect. This includes the large quantities of data that exist, in addition to their brief nature. All these together contribute to the difficulty of sentiment analysis.

To address this issue, however, a variety of different approaches have been used. Commonly useful approaches for this purpose include the construction of lexical change as well as the implementation of machine learning. Other options include statistical methods and domain knowledge-driven analysis. The task of sentiment analysis were positively influenced due to such techniques. Here we focus on the various machine learning methods adopted for sentiment analysis.

A 7 layer CNN model had been pretrained by Lijun Liu and his team with the help of the word2vec output vectors. This was useful in representing the distance between each word. They chose a 7-layered model to balance heavy computing loads and the accuracy of the output obtained across multiple experiments.[1] Similarily, Jian Wang and team had built a model to perform sentiment analysis on Chinese text using Long Short Term Memory Network (LSTM). [2] Salma Akter Lima and team, however, designed a system for analysing IMDB movie reviews using CNN, LSTM and LSTM-CNN models, achieving high performance results from all three models. [3] Chien Cheng Lee and team used a pre trained BERT Model to analyse text messages pertaining to the stock

Authorized licensed use limited to: Humber College. Downloaded on April 05,2025 at 00:58:52 UTC from IEEE Xplore. Restrictions apply.

market to view public opinion on US stocks. [4] Similarily, Manish Munikar and team also used a BERT model for performing sentiment analysis on a Stanford Sentiment Treebank(SST) dataset. [5] However, Junchao Dong and his team has taken it one step further, performing a sentiment analysis on commodity reviews by utilising a hybrid BERT-CNN Model. The hybrid model was custom designed to complement the individuals and as such, was a step up from the performance show by the individual models tested. [6]

Feature abstraction abilities were demonstrated in recent NER research investigations via deep learning, with remarkable potential. In NLP, the GRU and LSTM are the most commonly employed models. Furthermore, Bidirectional RNNs (BRNNs) have been proven to be beneficial in performing a wide range of NLP problems as the input sequence are processed from both sides, resulting in better representations and detecting the patterns which unidirectional RNNs may have missed.[19]

Charu Nanda and team designed a system that performs sentiment analysis on Hindi movie reviews using a classification algorithm. Similar works had been done for movies reviews in Thai, Bengali and English. [7] However, it is observed that while there are many ways to perform sentiment analysis, we have seen little amount of progress in this field to compare the different methods to discover the most effective efficient algorithm. Furthermore, to the best of our knowledge, There has been very little work reported regarding finding the most efficient algorithm for sentiment analysis for the IMDB movie reviews. The ones that have been observed make conclusions among the comparison of a select few models. This is where our work comes in. Our project is to perform sentiment analysis of IMDB movie reviews using a plethora of deep learning models as an example CNN, LSTM, BERT to name a few. Then, the outputs shall be compared and contrasted to discover the most efficient algorithm that can be applied for future studies.

III. Background

A. Sentiment Analysis

Machine learning and lexicon based techniques are the two techniques used in sentiment analysis. Finding whether the polarity of texts are negative or positive by utilizing NLP and machine learning is the purpose of sentiment analysis. By inputting corpus, the analysis is done. Bag of words modelling, n-grams, Parts of Speech tagging and bi-grams are some of the machine learning techniques used for this matter. For analysing complex data, Natural Language Processing Tool Kit (NLTK) is the base of the open-source library Textblob which is used here. As order is a vital component in a sentence, document term matrices are not used in sentiment analyis. A database of linked English words is used to score sentences on two scoring methods. They are the overall polarity score and the overall subjective score. -1 to +1 is the range for the first method, the polarity score. A positive value is expected for this scoring method for optimum efficiency. The second method, the subjectivity score ranges from 1 for subjective sentiments to 0 for objective sentiments. The Textblob library is built from the Natural Language Tool Kit (NLTK) of Python which allows it access various tools such as Parts of Speech Tagging which tags words based on the context and their respective definitions.

B. Models

1) CNN: Consists of neurons in multiple layers which connect to neurons in subsequent layers which as a whole form an improved version of the traditional neural network. A rudimentary neural network,it's specialty is multi layer convolution. Convolutional layers, pool layers and result layers work in parallel for conventional neural nets. The CNN applies channels and learns the size of the channels in isolation depending upon the endeavor that ought to be done, which for the present circumstance is the portrayal of words using the channels.The area designated to the outcome as well as the information neurons are calculated by the CNN. A many layer neural network is made after summarizes the pool results that arrangements with similar concepts like that of the neurons of the frontal cortex.

2) LSTM: It is a kind of Recurrent Neural Network. Its capacity to utilize its inward state memory to deal with successions is unrivaled. It can be utilized for an assorted arrangement of displaying undertakings. It is prepared to do putting away a state throughout a significant stretch of time. solves the drawbacks of recurrent neural networks, such as the disappearance or decline of the gradient vector over time.[8] The middle layer in a LSTM is a forgot gate. The forgot gate is used to determine which data needs to be memorised and which data has to be discarded for efficient long-term learning.[19] The data is taken from the input layer by the middle layer, and the result is displayed by the output layer.[12]

3) BERT: BERT (Bidirectional Encoder Representations from Transformers) a technique for NLP that employs unsupervised language representation and bidirectional Transformer models (Dynamically set weights use the attention mechanism between them based on their relationship and a deep learning method in which each input element is linked to each output element).[14] Using the bidirectional capabilities, BERT has been pre-trained in two distinct but related NLP tasks: Next Sentence Prediction and Masked Language Modeling. BERT can handle ambiguity, which is the most challenging component of interpreting natural language, as well as accurately analyse languages that are similar to human languages.[13]

IV. Our work and Analysis

A. Data Exploratory Analysis

By getting a complete grasp of the data, Exploratory Data Analysis' primary objective is attained.The identification of various trends and abnormalities within a dataset using graphical representations of the aforementioned data is called Exploratory Data Analysis.

Authorized licensed use limited to: Humber College. Downloaded on April 05,2025 at 00:58:52 UTC from IEEE Xplore. Restrictions apply.

3 stages comprise the technique. Firstly, the data is analysed for trends and patterns. The visualizing and modelling using Python libraries are done after goal setting the investigating data. Null values and incomplete data is considered during the analysis as well as variations.[11] Finally, Python libraries are used for proper visualisation processing.

The second stage involves modelling of the data. This is done using Python libraries such as Boxplot; which creates visuals using the IQR method. Outliers were removed for improving accuracy.This study's major goal is to observe and analyse IMDB movie reviews.[3] Before undertaking a sentiment analysis, it can be useful to visualise and analyse the distribution of variables.

B. Data Preprocessing

Unstructured data and useless texts are abundant on social media platforms. The identification and formatting of valid information from structureless data via techniques such as text mining is called preprocessing. Corpus matrix and document-term matrix are the two types of data obtained during pre-processing. Corpus matrix consists of various useless text such as punctuation and URLs within its content. However, it is compatible with Numpy, having been built using Pandas, the data analytics tool. Cleaning the data allows for the removal of unnecessary content from the corpus. Removal of such irrelevant words is highly beneficial for the accuracy of the corpus texts. Prior to or successive to the Natural language text processing, stop words could be filtered out. The removal of drivel such as stop words, symbols, URLs as well as punctuation gives us processable tweets. Furthermore, all letters were lowercased in order to increase data simplicity. While doing so we made sure to check that the meaning and semantics of the sentences would not change in the process. Through the deconstruction of textual matter into components called tokens, lexical analysis is done. This process is called tokenization. A clean data output is obtained through processing wherein raw tweets are taken as input and noise is removed. For this matter, both bigrams and unigrams were used. In order to separate the phrases, both n-grams and bi-grams are used as features for this. However, sentiment classification is crucial for our method, which involved the analysis of tweets involving thoughts and feelings. Context is considered in bi-grammical classification and analysis of sentiments. "Not" is a word that when coming with words such as "bad" and "hate" changes the meaning of sentences into something positive, rather than negative. Word by word analysis of the corpus is done by N-grams. Order of the words are analysed by unigrams. The division of data into bigrams and unigrams succeeds specific words. Another functionality of the Natural Language Tool Kit is the ability to reduce words to its basal form, a process known as Stemming. Finally, the document term matrix is obtained for further analysis.

C. Word Embeddings

Word embeddings is a text mining method for constructing word associations in Corpus or the textual data. A combination of Machine Learning, Computer Linguistics, Artificial Intelligence and overall Computer Science birther the Natural Language Proecessing field which in turn gave rise to Word Embedding. The context in which words are employed reveals their syntactic and semantic meanings. The distributional hypothesis proposes that the words which that appear in similar settings have semantic similarities. Word embedding can also be divided into two types: count-based and prediction-based. Count-based embedding, like the standard bag-of-words paradigm, fails to preserve context in textual data. The researchers at Stanford have developed a vector depiction of words that has shown very good performance for context preservation and is done by utilizing an algorithm that identifies the vector depiction of words. The GloVe or also known as Global Vector for Words Representation performs well at this task.[10]Target word prediction based on context is done by prediction-based embedding.

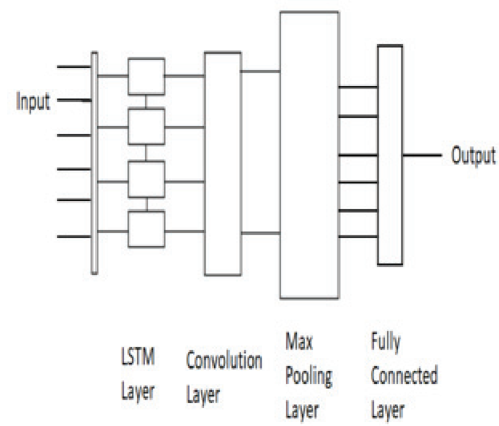

Fig. 1. CNN-LSTM architecture

D. Models and results

Here, we will discuss the hybrid models which we have made for the comparison and about the precision, recall, and F1 scores obtained by the respective models.

1) CNN-LSTM: In this, sequence prediction is improved by linking the CNN layer which is utilized for the extraction of features of the input data to the LSTMs.The Convolution Layer, which receives words as inputs, is the first layer of the CNN-LSTM Model. After convolution, the result is pooled via Max Pooling. After that, the output is sent to the LSTM layer. Finally, the positive or the negative sentiment are assigned to the textual data.

2) LSTM-CNN: In the LSTM-CNN Model, the input is fed to the LSTM layer. The LSTM layer and CNN layer are in sequence.The LSTM layer's output is given to the CNN layer, in which convolution occurs and output is pooled & then outputted as one of two types positive or negative sentiment.[11]

Authorized licensed use limited to: Humber College. Downloaded on April 05,2025 at 00:58:52 UTC from IEEE Xplore. Restrictions apply.

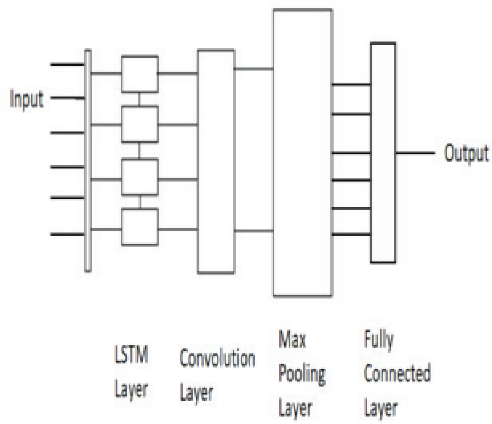

Fig. 2. LSTM-CNN architecture

3) BERT-CNN and BERT-LSTM: There are two ways BERT is used for sentiment analysis:

- By extracting the features as output for successive classification models to work on, the overall BERT model architecture is preserved. This process is called feature extraction.
- By modifying the model design by adding extra layers to the model, this allows for retraining of the model as well as bench-marking multiple tasks. This process is called fine-tuning.[15].

Within the fine tuning process, there exist two more sub methods that may be followed.

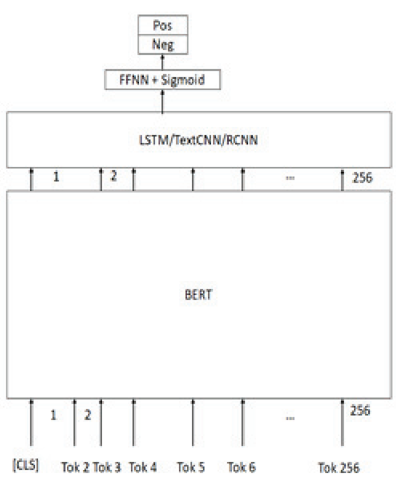

Fig. 3. BERT-LSTM architecture

- In the first, Devlin and his team added a special CLS token at the beginning of each sentence. This allows for sequential fine tuning of BERT via sending the output of this token into the neural network as an input. This process is called Fine tuning BERT using CLS.[14]
- Second, The entire output (CLS token inclusive) can be used as an input for the neural net. A SEQ LEN xh matrix is formed by the resulting combination with SEQ LEN being the input length and h being the hidden vector length. CLS token as well as the BERT output is used here. This process is called fine-tuning BERT using all tokens. This in turn becomes the input for further classifications.[16] With h as the length of hidden vectors and SEQ LEN as the maximum length of the input sequence,an SEQ LEN ×h matrix, .

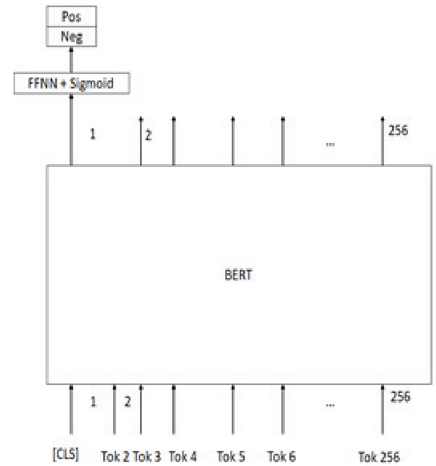

Fig. 4. BERT architecture using all tokens

V. Results and Observations

TABLE I
BERT-LSTM ARCHITECTURE

Model	Precision	Recall	F1
CNN	80.38	84.98	82.61
LSTM	83.25	82.33	82.78
CNN-LSTM	81.12	81.45	81.29
LSTM-CNN	82.39	78.21	80.25
GRU	84.89	84.02	84.45
BERT	87.52	88.39	87.95
BERT-CNN	89.61	88.39	88.98
BERT-LSTM	89.13	89.32	89.22

The dataset comparison results are displayed as follows. As such, the following observations are made:

- When the IMDB movie dataset is juxtaposed, it is clear that the prediction scores for the CNN model is much lower when compared with other models, followed closely by the LSTM and GRU models. However, when the BERT model is brought to the equation the performance is much better in contrast to the above-mentioned models. The training period of the BERT based approaches was reported to be longer than the training period of the CNN, LSTM, GRU, and hybrid approaches; however, the overall outputs fair differently.
- The CNN-LSTM and the LSTM-CNN models have not shown promising results when compared to LSTM, GRU and BERT and it's hybrid counterparts but report better

Authorized licensed use limited to: Humber College. Downloaded on April 05,2025 at 00:58:52 UTC from IEEE Xplore. Restrictions apply.

results than the CNN model. All three BERT models are observed to have good performance.

- However, only a marginal difference could be observed amoung the results of the models under consideration in comparison to BERT. There are two possible reasons for this outcome:
- The reduced length of the comments allow for the elimination of potential confusions during pre-processing, making the comment-classification process simpler.
- All BERT based approaches are seen to require an exorbitant amount of computational resources. Due to a lack of resources on our end however, we are unable to meet the demand of requiring 512 for the input string of the sub split BERT. The acquisition of more computational resources for meeting the demand put forward by such highly taxing architectures is under consideration for further research

CONCLUSION

Of all the deep learning models under consideration, it is observed that CNN has shown the worst performance amongst all the models under consideration. Furthermore, hybrid deep learning models like CNN-LSTM and LSTM-CNN which had also been implemented, were reported to have similar results to the above mentioned models.

The fine-tuning of the BERT model using pre-trained dual approach BERT models has shown to exhibit better performance in comparison to other all other models considered in this paper.

Lastly, the experimental results have shown that the good accuracy is observed in the BERT-CNN model for our proposed fine tuning method. While computationally taxing, it exhibits potential for further enhancements. In contrast, BERT CNN and BERT LSTM yielded no considerable performance bonuses. We intend to extend the suggested method or architecture for sentiment analysis based on aspects in the future, and more thorough experiments on different datasets might be undertaken.

REFERENCES

[1] X. Ouyang, P. Zhou, C. H. Li and L. Liu, ”Sentiment Analysis Using Convolutional Neural Network,” 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015, pp. 2359-2364, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.349.

[2] J. Wang and Z. Cao, ”Chinese text sentiment analysis using LSTM network based on L2 and Nadam,” 2017 IEEE 17th International Conference on Communication Technology (ICCT), 2017, pp. 1891-1895, doi: 10.1109/ICCT.2017.8359958.

[3] M. R. Haque, S. Akter Lima and S. Z. Mishu, ”Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews,” 2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), 2019, pp. 161-164, doi: 10.1109/ICECTE48615.2019.9303573.

[4] C. -C. Lee, Z. Gao and C. -L. Tsai, ”BERT-Based Stock Market Sentiment Analysis,” 2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), 2020, pp. 1-2, doi: 10.1109/ICCE-Taiwan49838.2020.9258102.

[5] M. Munikar, S. Shakya and A. Shrestha, ”Fine-grained Sentiment Classification using BERT,” 2019 Artificial Intelligence for Transforming Business and Society (AITB), 2019, pp. 1-5, doi: 10.1109/AITB48515.2019.8947435.

[6] J. Dong, F. He, Y. Guo and H. Zhang, ”A Commodity Review Sentiment Analysis Based on BERT-CNN Model,” 2020 5th International Conference on Computer and Communication Systems (ICCCS), 2020, pp. 143-147, doi: 10.1109/ICCCS49078.2020.9118434.

[7] C. Nanda, M. Dua and G. Nanda, ”Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning,” 2018 International Conference on Communication and Signal Processing (ICCSP), 2018, pp. 1069-1072, doi: 10.1109/ICCSP.2018.8524223.

[8] D. Li and J. Qian, ”Text sentiment analysis based on long short-term memory,” 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI), 2016, pp. 471-475, doi: 10.1109/CCI.2016.7778967.

[9] J. Zhang, Y. Li, J. Tian and T. Li, ”LSTM-CNN Hybrid Model for Text Classification,” 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2018, pp. 1675-1680, doi: 10.1109/IAEAC.2018.8577620.

[10] B. Oscar Deho, A. William Agangiba, L. Felix Aryeh and A. Jeffery Ansah, ”Sentiment Analysis with Word Embedding,” 2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST), 2018, pp. 1-4, doi: 10.1109/ICASTECH.2018.8506717.

[11] N. Chen and P. Wang, ”Advanced Combined LSTM-CNN Model for Twitter Sentiment Analysis,” 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), 2018, pp. 684-687, doi: 10.1109/CCIS.2018.8691381.

[12] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, ”Sentiment Analysis of Comment Texts Based on BiLSTM,” in IEEE Access, vol. 7, pp. 51522-51532, 2019, doi: 10.1109/ACCESS.2019.2909919.

[13] T. -L. Truong, H. -L. Le and T. -P. Le-Dang, ”Sentiment Analysis Implementing BERT-based Pre-trained Language Model for Vietnamese,” 2020 7th NAFOSTED Conference on Information and Computer Science (NICS), 2020, pp. 362-367, doi: 10.1109/NICS51282.2020.9335912.

[14] J. Devlin, M. Chang, W. M and K. Lee, ”Bert: Pre-training of deep bidirectional transformers for language understanding”, arXiv preprint, vol. 10, no. 04805, 2018.

[15] R. Man and K. Lin, ”Sentiment Analysis Algorithm Based on BERT and Convolutional Neural Network,” 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2021, pp. 769-772, doi: 10.1109/IPEC51340.2021.9421110.

[16] K. Zhang, M. Hu, F. Ren and P. Hu, ”Sentiment Analysis of Chinese Product Reviews Based on BERT Word Vector and Hierarchical Bidirectional LSTM,” 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE), 2021, pp. 9-14, doi: 10.1109/CSAIEE54046.2021.9543231.

[17] S. Agrawal, S. Dutta and B. K. Patra, ”Sentiment Analysis of Short Informal Text by Tuning BERT - Bi-LSTM Model,” IEEE EUROCON 2021 - 19th International Conference on Smart Technologies, 2021, pp. 98-102, doi: 10.1109/EUROCON52738.2021.9535535.

[18] Y. Pan and M. Liang, ”Chinese Text Sentiment Analysis Based on BI-GRU and Self-attention,” 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020, pp. 1983-1988, doi: 10.1109/ITNEC48623.2020.9084784.

[19] J. Li, S. Zhao, J. Yang, Z. Huang, B. Liu, S. Chen, et al., ”WCP-RNN: A novel RNN-based approach for bio-NER in Chinese EMRs”, J. Supercomput., vol. 76, no. 3, pp. 1450-1467, Mar. 2020.

Authorized licensed use limited to: Humber College. Downloaded on April 05,2025 at 00:58:52 UTC from IEEE Xplore. Restrictions apply.