Deep Learning Methods for Sentiment Analysis of Drug Reviews: A Novel Approach

1st Jay Rabari
Chandubhai S Patel Institute Of Technology
Charotar University of Science and Technology (CHARUSAT)
Gujarat, India
21cs053@charusat.edu.in

2nd Vraj Prajapati
Chandubhai S Patel Institute Of Technology
Charotar University of Science and Technology (CHARUSAT)
Gujarat, India
21cs052@charusat.edu.in

3rd Dharmendrasinh Rathod
Chandubhai S Patel Institute Of Technology
Charotar University of Science and Technology (CHARUSAT)
Gujarat, India
dharmendrarathod.cse@charusat.ac.in

***Abstract*—Diverse data sets with important insights into a wide range of health issues have proliferated in the medical industry in recent years.It takes a sophisticated study to evaluate medical items; costs, side effects, user experiences, and appropriate dose must all be taken into account.Drug reviews also include detailed information about health issues and the chemical makeup of medications, which further complicates the evaluation process.This research suggests an enhanced deep learning-based method to address this issue. We have suggested a hybrid method for it that uses BI-LSTM to combine CNN and RNN. The proposed model achieved the 93% accuracy rate.**

***Index Terms*—Sentimental Analysis, Deep Learning, CNN, RNN, LSTM, Bi-LSTM,Random Forest, GRU**

I. Introduction

As technology continues to evolve, people increasingly rely on online reviews to inform their choices, autonomously deciding on their preferred treatments. People have become habituated to first garnering information by exploring these reviews and blog posts, before making any sort of purchase. This development indicates the importance of reviews for the advancement of any product. Over the years, the number of reviews increased, leading to an increase in the amount of data to explore through. To deal with this large amount of information sentimental analysis or opinion mining is a very beneficial technique.

With the progression in the web, the clinical area also now has a wealth of assorted information gathered throughout the long term, accessible for dissecting different medical issue. This information should be handled, assessed and broke down to help people in looking for data. For this, opinion investigation is the fitting arrangement. Feeling examination is a NLP strategy that includes the utilization of computational techniques to decide and extricate opinions, sentiments, and feelings communicated in message information. Subsequently, feeling examination is utilized to extricate and classify sentiments on different items and administrations.

II. Related Work

This paper introduces a novel approach to Sentiment Analysis through deep learning techniques. To address the vanishing gradient issue in commonly used RNNs, the study explores the efficacy of three different Recurrent Neural Networks: LSTM, GRU and also CNN.The study focuses on the analysis of sentiments in various contexts, employing different deep learning models for sentence-level and aspect/target-level sentiment analysis, which is a crucial component of Natural Language Processing (NLP), has witnessed significant advancements with the emergence of deep learning techniques.

In [1] Rathod et al.(2023) offer research on the sentiment analysis of medication reviews using machine learning approaches, and they achieve good accuracy rates in their findings. This work can be found in this article.

In [2] Colón-Ruiz and Segura-Bedmar (2020) compared several deep learning models for sentiment analysis on medication reviews. Their study showed that Long Short-Term Memory (LSTM) recurrent neural networks and convolutional neural networks (CNNs) are both useful for analyzing drug sentiments. They also introduced the use of a Bi-LSTM in conjunction with Bidirectional Encoder Representations from Transformers (BERT). The study draws attention to the possible loss of data in CNN as a result of pooling layers and suggests using RNN—more especially, LSTM—as an alternative. LSTM is selected because it can lessen the vanishing gradient issue that arises with a standard RNN [2].

In a concentrate by Vikas Goel et al. [3], feeling examination of multilingual tweets is led utilizing a Google Interpreter Programming interface. Subsequent to preprocessing and interpretation into English, the creators utilize Innocent Bayes and Recursive Brain Organization (RNN) for arrangement. RNN exhibits altogether higher precision contrasted with Gullible Bayes in grouping multilingual tweets. [3]

Inspecting on the web item audits and client created content,

TABLE I
COMPARISON METHODS WITH MODELS AND ACCURACY

Author	Year	Model used	Accuracy
Ali et al.(2013). [10]	2013	Logistic Regression	68.5%
Aditya et al.(2019). [11]	2019	Neural Network	90.28%
Theres B. et al.(2020). [4]	2020	RNN-Bi LSTM	83.9%
Colón-Ruiz et al.(2020). [2]	2020	BERT + LSTM	90.46%
Md Nazim U. et al.(2022). [12]	2022	Random Forest	94.06%

the paper utilizes Bidirectional Long Momentary Memory (Bi-LSTM) to address the constraints of customary component based strategies. Bi-LSTM, a brain network extricating text qualities for handling and expectations, is joined with Repetitive Brain Organization (RNN). The outcomes feature the upgraded message qualities and expanded order precision accomplished by the Bi-LSTM model, particularly in catching feelings significant for opinion examination. [4]

Johnson and Zhang (2016) proposed a technique using LSTM for locale embeddings in regulated and semi-managed text classification. The application of LSTM demonstrates its potential for sentiment analysis, particularly in capturing contextual information over word sequences, despite the fact that their study focused on general text categorization. [5], [6]

Wiley et al. (2014) investigated drug jabber on internet based informal communities, giving experiences into the sorts of conversations and opinions communicated by clients. In spite of the fact that their attention was on examining conversations as opposed to opinion examination strategies, their discoveries could advise the plan regarding feeling investigation frameworks for drug reviews. [7]

In a tweet suggestion study, different models are looked at utilizing certifiable Twitter information. When it comes to classifying hashtags based on the content of tweets, LSTM and RNN, particularly LSTM-RNN, perform better than other algorithms. LSTM-RNN shows predominant exactness (28.6%) contrasted with elective models, underlining the adequacy of LSTM in catching tweet semantics. [3], [8], [9]

Ali et al. (2013) investigated opinion examination on clinical discussions, displaying the relevance of feeling examination methods in medical care areas. While their review didn't explicitly target drug surveys, the philosophies and difficulties recognized are relevant for feeling examination in comparative contexts. [10]

In summary, as shown in Table-I existing literature demonstrates the effectiveness of deep learning techniques, such as CNNs, LSTMs, and GRUs, for sentiment analysis in various domains, including pharmaceuticals. However, there is a need for further research to tailor these approaches specifically for analyzing sentiments in drug reviews, considering the unique language and context inherent to this domain.

III. METHODOLOGY

This paper proposes a sentiment analysis system, which takes customer reviews as the input and gives sentiment of review(either positive or negative) as output. This system uses a Bi-directional LSTM model.

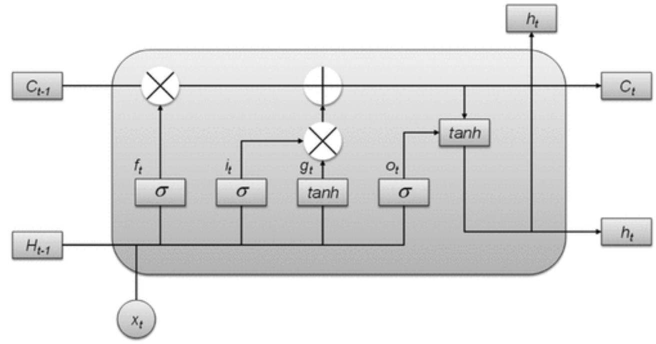

Fig. 1. LSTM Architecture

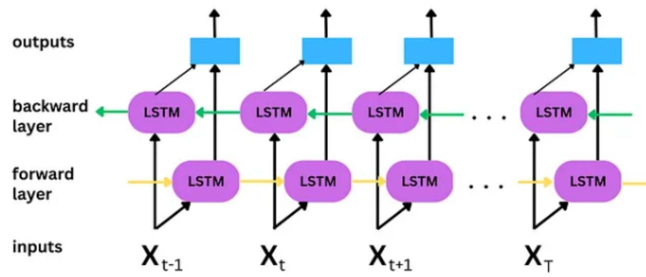

Fig. 2. LSTM Architecture

A. Forward Long Short Term Memory (LSTM)

LSTM models, inspired by RNNs tailored for long-term dependencies in sequential data, address RNNs' limitation in capturing continuous signals due to aging information. They excel in discerning crucial data within sequences, crucial for tasks like signal processing and speech analysis. By extracting interval data and forecasting porosity, LSTMs effectively handle long-term dependency in various sequential datasets.The diagram that follows depicts the structure of the LSTM network. [13]

Figure-1 elaborate the LSTM Architecture where xt is the current input, ht–1 is the status value output at the previous moment, ht is the output of the current moment state value, ct–1 is the memory unit of the previous moment, ct is the memory unit of the current moment, ft is the forgetting gate, it is the input gate and ot is the output gate. [13]

LSTM possibly removes forward data highlights while handling grouping data, so LSTM has weaknesses. The BiLSTM can extensively utilize information setting data. The structure of the BiLSTM is depicted in figure below. BiLSTM is made out of a forward and a converse LSTM. Consequently, contrasted and the LSTM model, the elements separated by BiLSTM are more extensive. [13]

B. Bi-directional Long Short Term Memory(Bi-LSTM)

In the Figure-2, we observe two layers of LSTM, operating sequentially and facilitating data progression in reverse. This

TABLE II
OTHER ALGORITHMS USED

Algorithm Name	Description
GRU	Gated Recurrent Unit (GRU), a type of recurrent neural network architecture, is suitable for capturing sequential dependencies in drug review texts. It can be utilized as a potential deep learning model for sentiment analysis. [15]
CNN	Convolutional Neural Network (CNN), renowned for its effectiveness in capturing local patterns, is applicable in analyzing drug reviews. Its ability to extract features from text data makes it a valuable deep learning architecture for sentiment analysis. [16]
Logistic Regression	Logistic Regression, a simple yet powerful linear model, serves as a baseline for sentiment analysis in drug reviews. Its straightforward interpretation and computational efficiency make it suitable for comparison with more complex models. [17]
Random Forest	Random Forest, an ensemble learning method, is adept at handling non-linear relationships in text data. Applied to sentiment analysis in drug reviews, it offers robustness in capturing intricate sentiment patterns. [18]

prompts a discussion on the primary utility of BiLSTM and raises questions about future timestamp values. The additional LSTM layer, functioning in reverse, serves as a representation of future timestamps. These reversed values, flowing oppositely, are termed future timestamp values in contrast to past timestamps. Essentially, the input sequence reverses in the supplementary LSTM layer, as depicted in the diagram, simplifying backtracking when implemented in models. Presently, BiLSTM is gaining popularity for solving real-world problems efficiently, offering robust methods that can be validated and integrated within shorter timeframes, thereby reducing the overall complexity of AI models. [14]

C. Other Algorithms used

In Table-II Various algorithms, including Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), Logistic Regression, and Random Forest, are employed for sentiment analysis in drug reviews, each offering distinct advantages in capturing sequential dependencies, local patterns, baseline comparisons, and handling non-linear relationships in text data, respectively.

IV. PROPOSED SYSTEM

The process of examining reviews written about various medications and classifying the tone that is conveyed in each review as either favourable, negative, or neutral is referred to as sentiment analysis. The Figure-3 depicts a broad framework for dealing with and stalling a dataset of drug reviews using a Bidirectional Long Short Term Memory (BiLSTM) model. It begins with the rough dataset containing drug reviews, go on through essential preprocessing steps encompassing data cleaning and course of action to ensure the dataset's accessibility for examination. Following this hidden stage, the data goes through tokenization, a critical cycle that destroys the printed information into indisputable units like words

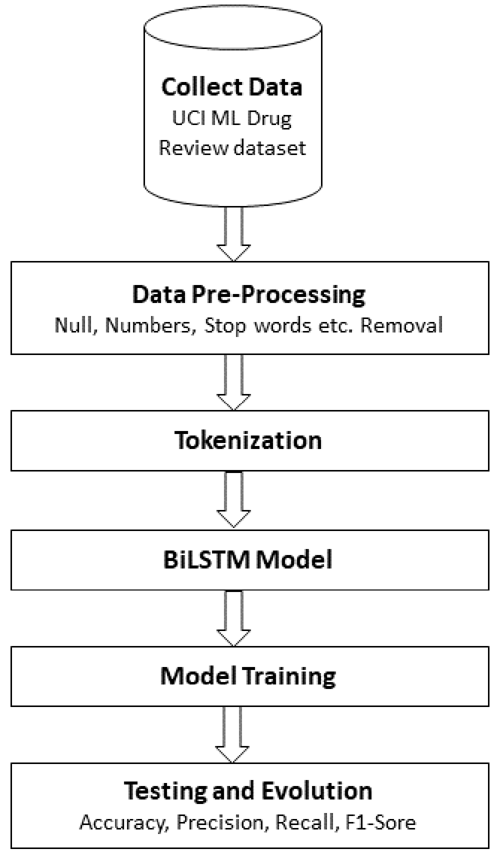

Fig. 3. Proposed System Flow-chart

TABLE III
DATA DESCRIPTION

Attribute Name	Type	Description
drugName	Categorical	Name of the drug
Condition	Categorical	Name of condition
Review	Text	Patient review
Rating	Numerical	10 star patient rating
Date	Date	Date of review entry
usefulCount	Numerical	Number of users who found review useful

or tokens. The preprocessed dataset is then exposed to the BiLSTM model and then model is trained on training data. Finally, the flowchart wraps up in the gathering and appraisal of the results got from the pre-arranged model.

A. Dataset gathering

This Research paper contains a dataset from UCI Machine Learning Repository, named as 'Drug Review Dataset (Drugs.com) Data Set'. There are a total of 215063 instances and 6 attributes. Dataset is already divided into Training and Test dataset.Training dataset contains a total of 1612987 instances and Testing dataset contains 53766 instances. 6 attributes are Drugname, condition, review, rating, date and usefulcount as shown in Table-III.

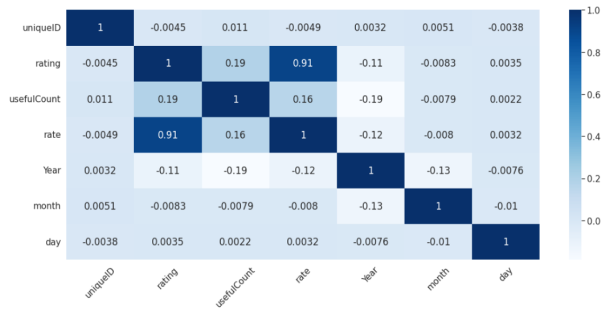

Fig. 4. Correlation Matrix

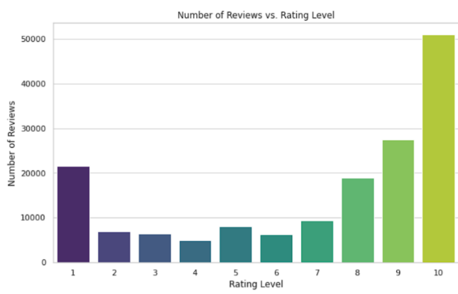

Fig. 5. Rating Count

B. Exploratory Data Analysis

Figure-4 shows the correlation between all attributes and gives the impacting or importance of attributes on the label.
As displayed in Figure-5 contains the rating vs no of reviews for each rating graph.from that we can analyze how many reviews are above 5 and how many are below 5.
Additionally showed in Figure-6, we have found the best 10 conditions for which the patient has given reviews. We can deduce that the majority of the surveys focus on conception prevention, followed by dejection, which has the second highest number of surveys. The circumstances like insomnia, obesity and weight loss has basically a similar count and furthermore collect the information about the most well known drugs for conception prevention.

C. Data Pre-processing

The drug reviews are cleaned by eliminating all the whitespace, converting to lower case letters, collecting the stopwords and all the other general processing techniques. To improve the accuracy and reduce the risk of overfitting, feature extraction we used tokenization. [4]

1) Feature Extraction: It is used to reduce the number of features in dataset by creating new features from existing ones. The new reduced feature is used to summarise the most of the information present in the previous ones. [19]

2) Undersampling: Undersampling involves reduction of observations which are not impacting the data heavily. It is used to reduce our dataset to only the necessary portion containing the majority of data.
As Figure-7 shows the pie presentation of sentiments we divided the rating above 5 as a positive rating and below 5 as a negative rating.According to the data visualized by the chart we have 77.91% positive rating and 22.09% negative rating.

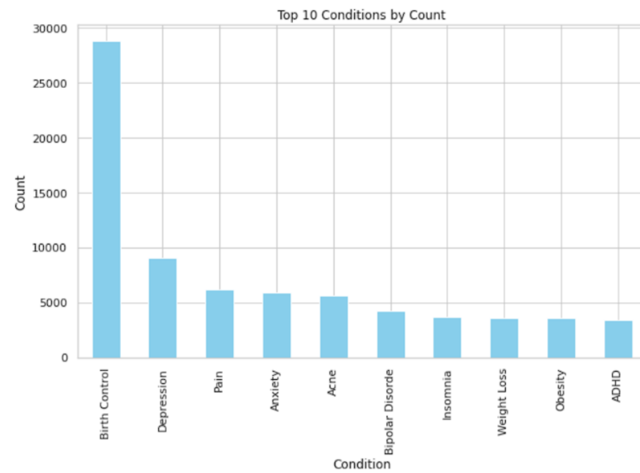

Fig. 6. Top 10 health conditions

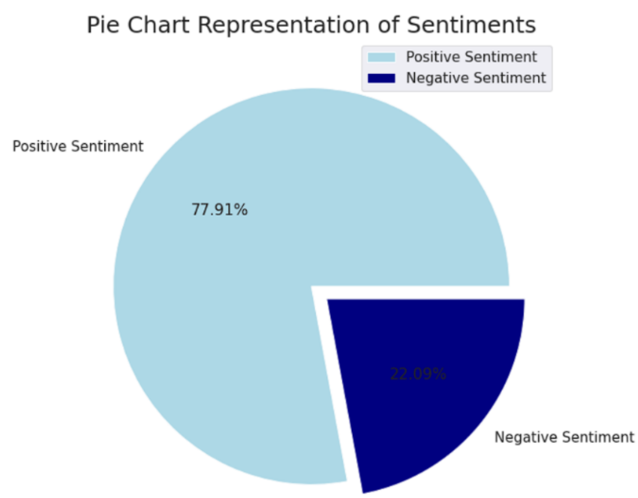

Fig. 7. Pie Chart Representation of Sentiments

The natural language toolkit is used to analyse the texts and then wordclouds are created to understand the important words in reviews.
Figure-8 and Figure-9 are the wordcloud which means common words in the positive reviews and the negative reviews. As shown in the Figure-10 it is a yearly reviews count and we got the maximum reviews between the year 2015-2016 and avg reviews are 10,000 per year.

3) Tokenization: Tokenization is an essential step in natural language processing, where text undergoes a series of well-defined processes for thorough analysis. Essentially, tokenization involves segmenting text into smaller, meaningful units known as tokens. These tokens typically represent individual words or sequences of words, aiding in the extraction of semantic meaning from the text. By breaking down text into its

Fig. 8. Worcloud of positive reviews

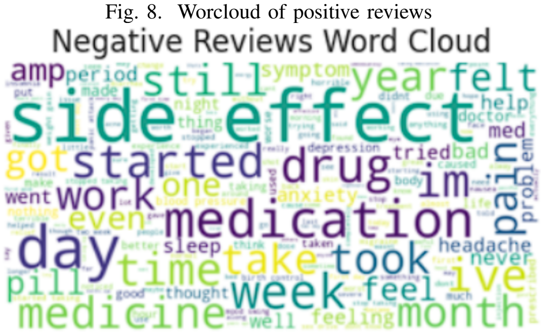

Fig. 9. Worcloud of negative reviews

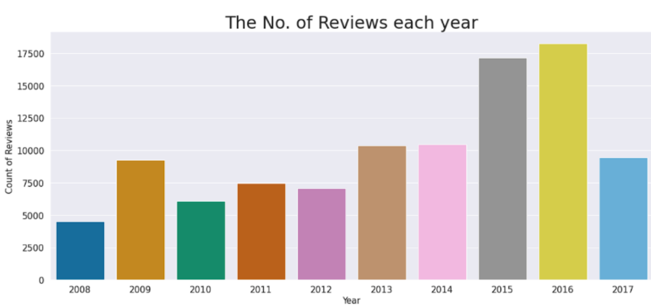

Fig. 10. Number of reviews per year

constituent tokens, tokenization enables machines to interpret and understand human language more effectively. [11], [12], [20]

V. RESULTS

According to all our algorithms, we got the result given below. We have used 5 algorithms and compared accuracy and F1-score(micro,macro). To evaluate the sentiments of the reviews, the positive sentiments are considered as value 1 and negative sentiments as value 0. We are evaluating the results using CNN, Random Forest, GRU, Logistic Regression and BiLSTM models. As Table-IV And Figure-11 shows we got the best accuracy for BI-LSTM algorithm. This model gives us the best sentiment output with an accuracy of 93%. Figure-12 and Figure-13 above shows the decrease in training loss and increase in accuracy over an increasing number of epochs for our Bi-LSTM model.

TABLE IV
COMPARISON OF DIFFERENT ALGORITHM

Algorithms(Model)	Macro F1	Micro F1	Accuracy(%)
GRU	0.4369	0.7760	0.7761
CNN	0.8898	0.9280	0.9237
Logistic Regression	0.4388	0.7761	0.7759
Random Forest	0.8613	0.9170	0.9170
Bi-LSTM	0.9028	0.9339	0.9323

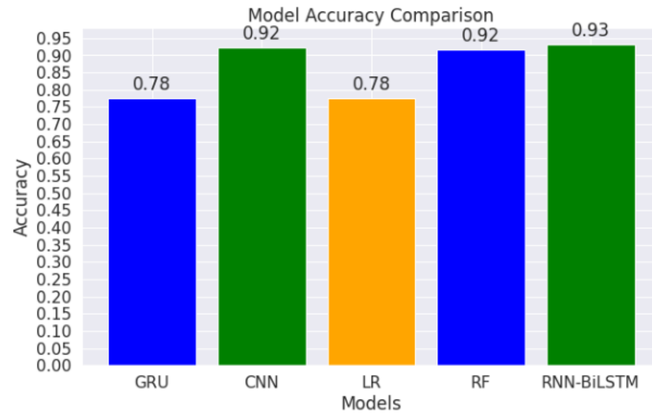

Fig. 11. Graph of comparison of different models

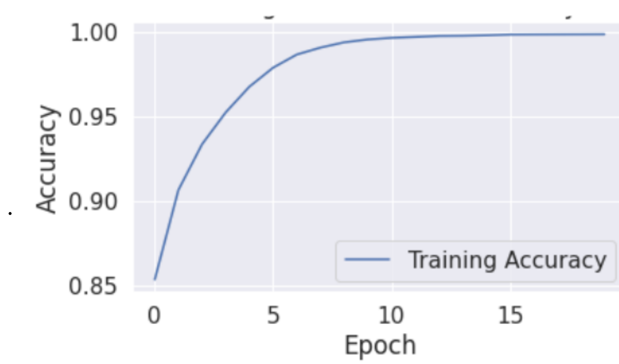

Fig. 12. Training Accuracy

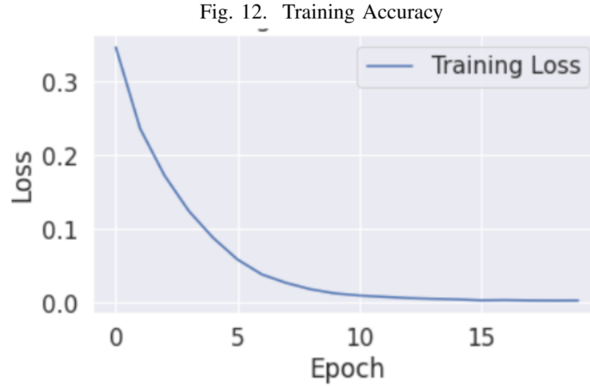

Fig. 13. Training Loss

VI. CONCLUSION

People now have the ability to voice their opinions about a wide range of products on the market thanks to significant technological advancements, particularly the internet. One such field is evaluating drugs for clinical conditions. Because so many people rely on these reviews, finding out whether a particular drug is working and what might make customers angry is made easier by extracting information from these reviews. We have utilized Bi-LSTM calculation to analyze the sentiments of drugs which gives a precision of 93%. As a piece of our future work, we might likewise want to utilize more granular client data. For example, client age, orientation and treatment length to additionally further develop results and get to the next level bits of knowledge.

REFERENCES

[1] D. Rathod, K. Patel, A. J. Goswami, S. Degadwala, and D. Vyas, "Exploring drug sentiment analysis with machine learning techniques," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, 2023, pp. 9–12.

[2] C. Colón-Ruiz and I. Segura-Bedmar, "Comparing deep learning architectures for sentiment analysis on drug reviews," *Journal of Biomedical Informatics*, vol. 110, p. 103539, 08 2020.

[3] Q. Li, S. Shah, X. Liu, and A. Nourbakhsh, "Data sets: Word embeddings learned from tweets and general data," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 11, no. 1, 2017, pp. 428–436.

[4] T. Bemila, I. Kadam, A. Sidana, and S. Zemse, "An approach to sentimental analysis of drug reviews using rnn-bilstm model," *Library & Librarianship Types eJournal*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:235022752

[5] R. Johnson and T. Zhang, "Supervised and semi-supervised text categorization using lstm for region embeddings," in *International Conference on Machine Learning*. PMLR, 2016, pp. 526–534.

[6] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *arXiv preprint arXiv:1801.06146*, 2018.

[7] M. T. Wiley, C. Jin, V. Hristidis, and K. M. Esterling, "Pharmaceutical drugs chatter on online social networks," *Journal of biomedical informatics*, vol. 49, pp. 245–254, 2014.

[8] V. Bobicev, M. Sokolova, Y. Jafer, and D. Schramm, "Learning sentiments from tweets with personal health information," in *Advances in Artificial Intelligence: 25th Canadian Conference on Artificial Intelligence, Canadian AI 2012, Toronto, ON, Canada, May 28-30, 2012. Proceedings 25*. Springer, 2012, pp. 37–48.

[9] D. Rathod and M. H. Bohara, "Sentiment analysis for feature extraction using dependency tree and named entities," in *2017 International Conference on Innovations in information Embedded and Communication Systems (ICIIECS)*, vol. 1, 2017, pp. 587–592.

[10] T. Ali, D. Schramm, M. Sokolova, and D. Inkpen, "Can i hear you? sentiment analysis on medical forums," in *Proceedings of the sixth international joint conference on natural language processing*, 2013, pp. 667–673.

[11] A. Aditya, "Review sentiment analysis and rating prediction on drugreview dataset," Ph.D. dissertation, 11 2019.

[12] M. N. Uddin, M. F. B. Hafiz, S. Hossain, and S. M. M. Islam, "Drug sentiment analysis using machine learning classifiers," *International Journal of Advanced Computer Science and Applications*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:246543392

[13] Y. Sun, J. Zhang, Z. Yu, Y. Zhang, and Z. Liu, "Bidirectional long short-term neural network based on the attention mechanism of the residual neural network (resnet–bilstm–attention) predicts porosity through well logging parameters," *ACS Omega*, vol. 8, no. 26, pp. 24 083–24 092, 2023. [Online]. Available: https://doi.org/10.1021/acsomega.3c03247

[14] S. Nath, "Why is bilstm better than lstm ?" Apr 2023. [Online]. Available: https://medium.com/@souro400.nath/why-is-bilstm-better-than-lstm-a7eb0090c1e4

[15] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *International conference on machine learning*. Pmlr, 2013, pp. 1310–1318.

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

[17] Y. So, "A tutorial on logistic regression," 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:10351003

[18] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, and N. Villa-Vialaneix, "Random forests for big data," *Big Data Research*, vol. 9, 11 2015.

[19] 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.

[20] B. N. Patro, V. K. Kurmi, S. Kumar, and V. P. Namboodiri, "Learning semantic sentence embeddings using sequential pair-wise discriminator," *arXiv preprint arXiv:1806.00807*, 2018.