Sentiment Prediction for Social Information Retrieval: A Comparative Study of Machine Learning and Deep Learning Approaches

Aicha BOUBEKEUR[1,*], Fouzia BENCHIKHA[2,3] and Naila MARIR[2,3]

[1] University of Ibn Khaldoun Tiaret, Algeria
[2] LIRE Laboratory of Constantine2 University, Algeria
[3] University of Constantine2 - Abdelhamid Mehri Constantine, Algeria
{aicha.boubekeur,fouzia.benchikha,naila.marir}@univ-constantine2.dz
*Corresponding author: aicha.boubekeur@univ-constantine2.dz

Abstract. Sentiment analysis plays a pivotal role in social information retrieval, enabling the extraction of valuable insights from user-generated content. In this study, we conduct a comprehensive comparative analysis of machine learning and deep learning approaches for sentiment prediction in the context of social media data, with a specific focus on the COVID-19 vaccine discourse. We investigate the performance of traditional machine learning classifiers, including Naive Bayes, Support Vector Machines, K-Nearest Neighbors, and Decision Tree, in conjunction with the TF-IDF representation model. In parallel, we assess the efficacy of deep learning models, such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and a hybrid LSTM-CNN architecture, utilizing Word Embedding representation. Notably, the CNN model with Word2Vec vectorization demonstrates the highest performance. While the accuracy of the combined model, featuring the two LSTM-CNN classifiers, is slightly lower for our specific problem.

Keywords: Sentiment analysis, social information retrieval, Comparative analysis, Machine learning, Deep learning.

1 Introduction

In recent years, the proliferation of social media has transformed the landscape of information sharing and communication. It not only redefined the way we interact but has also given rise to vast reservoirs of user generated content, opening up unprecedented opportunities for gaining insights into public opinions and attitudes.

Sentiment analysis, a subfield of natural language processing (NLP), has emerged as a crucial tool for deciphering sentiments expressed within this massive data. It enables to decode the emotions, opinions, and attitudes of users, shedding light on their reactions to a wide range of topics, from political events to consumer products.

C. A. Kerrache et al. (eds.), *Proceedings of the International Conference on Emerging Intelligent Systems for Sustainable Development (ICEIS 2024)*, Advances in Intelligent Systems Research 184,
https://doi.org/10.2991/978-94-6463-496-9_2

As social media continues to grow, the need for more effective sentiment prediction techniques becomes increasingly apparent. To address this problem, researchers have explored various sentiment prediction techniques, including machine learning and deep learning approaches [11] [12] [13] [14] [16].

In the present work, we focus on a comparative evaluation of machine learning and deep learning approaches for sentiment prediction in social information retrieval. We are particularly interested in analyzing sentiment and opinion of people throw their tweets during the COVID19 pandemic. Indeed, tweeters had polarized opinions on the effectiveness of vaccines and on the vaccination process itself. Sentiment analysis of tweets will help to understand the dynamics of vaccination and to design and develop effective preventive measures to increase vaccine uptake, by carefully monitoring conversations on social media.

In order to efficiently and correctly predict public opinion towards marketed Covid-19 vaccines, classical and deep ML-based prediction models are compared. Their effectiveness was evaluated using different pre-processing strategies, such as Stop Words Removing, Stemming, N-grams and Word Embedding weighting and representation schemes, based respectively on the term frequency and inverse frequency of TF-IDF and Word2Vec documents. These models are divided into two distinct classes: Deep Learning algorithms such as CNN, LSTM and the combination (LSTM-CNN) and classical algorithms such as Support Vector Machines (SVM), Decision Tree (DT), Multinomial Naïve Bayes (MNB) and KNearest Neighbors (KNN) as presented in **Fig. 1.**. In the end, evaluation metrics measure their performances.

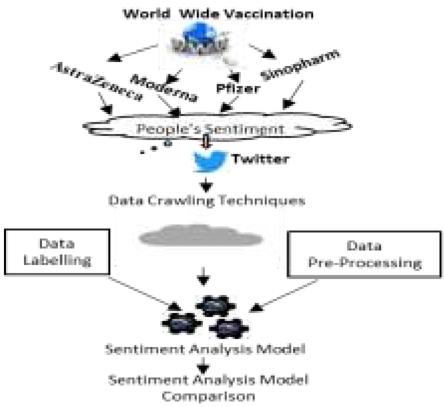

Fig. 1. Block diagram of the implementation.

The main objective of our study is to rigorously compare the performance of machine learning and deep learning techniques in the realm of sentiment analysis within the context of social media data. By doing so, we aim to advance our understanding of the strengths and weaknesses of these methods and contribute to the development of more accurate and efficient sentiment prediction models.

The remaining parts of the paper are organized, as follows: Section 2 reviews the related works. Section 3 describes the methodology. Section 4 presents the experimental results. Section 5 present the limitations and future research perspectives of this study. Finally, we conclude the paper in Section 6.

2 Related Work

In the context of social information retrieval, the proliferation of user-generated content on social media has pushed researchers to monitor public emotion or sentiment and assist in decision-making [17] [18]. Although there has been an increase in the number of studies focused on COVID'19. Several researchers have proposed solutions to identify the concerns of the confined public. A significant number of this work is based on the use of lexicon, machine learning and deep learning techniques as shown in **Table 1.**, and the choice of technique depends on the specific requirements and use case of the sentiment analysis task:

2.1 Approaches-Based Lexicon

In [2], the researchers have studied public attitude following the announcement of the first COVID-19 vaccine. They have evaluated their tweets on polarity scores which differ from one country to another. In this period, public sentiment and number of tweets are considerably increased. Subsequently, they have summer slowly diminished. A similar study in [1] went to identify and compare vaccine - related sentiments in tweets in operator other additional dimensions. Negative feelings wore mainly on a set of constraints. This study was based on the lexicon and a keyword function to categorize tweets into depending on their theme and feeling.

2.2 Approaches-Based Traditional ML

In [6], the researchers have designed a model thematic to identify citizens' concerns about the ineffectiveness of COVID-19 vaccines in exploiting a frequent NLP tool. In [3] and [8], Others researchers have released two solutions for closely monitoring popular tweets retweeted. In order to increase the number of retweets, one of them determined the popularity of tweets by exploiting information complementary [8] and the other examined the polarity of popular tweets [3]. Searching for relevant messages published depended to a large extent on the selected hashtags. The classifier having the best precision is the model that predicts the retweet-ability of tweets [8]. The authors have Thus discovered that LR classifiers outperformed others ML classifiers evaluated on two datasets, despite the difference in partition sizes and different periods observation [3]. In [4], the authors proposed an approach sentiment analysis to filter fake news about the COVID-19 epidemic in tweets related to Moroccan Corona.

2.3 Approaches-based DL Methods

In [5], a solution for detecting informative tweets was proposed. This preventive solution would limit irrelevant information and avoid the spread of negative feelings. The authors applied a set of DL models on Twitter data to filter informative tweets in depending on their content. The classifier was trained on a labeled dataset. In [7], researchers have revealed that the majority of Twitter messages had a neutral position and the number of tweets in favor of vaccines East higher than that of those against, along the first month following the announcement of the first vaccine effective, the day its marketing was authorized, a peak of unfavorable tweets was checked in as result interesting. The authors are gone so to label manually the 1% of the dataset to locate the citizens. The BERT classifier was the best performing. In [9], a study on the analysis of tweets containing at least one hashtag generated categories on topics different. Aside from the message and its source, its timing is essential. In [10], an exploration of the different characteristics of tweeters reveal that women are more likely to have hesitant opinions on vaccination more than men. Likewise, the older public tends to be pro-vaccine. The low- income community income and/or religious background is likely to have divided opinions on vaccination.

Table 1. Overview of included studies.

Works	Data source	Approach	Classification algorithm	Valuation and various
[1]	Real-time on Twitter	Lexicon	Vader and Gensim library keyword	
[2]	Real-time on Twitter	Lexicon	Vader	— polarity score — effectiveness : 90%
[3]	Self collected COVID-19 dataset	ML	- MNB, LinearSVC, AdabbostClassifier - Trigram and tfidf score - Logistic Regression	— Accuracy: 81.4 — smallest dataset+ trigram+ tfidf accuracy: 81% — largest dataset + bigrams + tfidf: 75%
[4]	Moroccan corona tweets	ML	- TextBlob - LR, DT, RF, NB, SVM, MLP, Gradient Boosting	
[5]	WNUT 2020 SharedTask2 dataset	DL	MVEDL (ROBERTa,BERTweet, CT-BERT)	— Accuracy 91.75% — F1score 91.14
[6]	Self COVID 19 vaccine dataset	ML	TextBlob	No performs metrics were reported
[7]	Self COVID-19 vaccine datast	ML, DL	MNB, RF, SVM, Bi-LSTM, CNN, BERT	— BERT: Accuracy78.94% — SVM: accuracy 76,23% — Bidirectional LSTM integrated — with GloVe:Accuracy 74,7%
[8]	Self collected COVID-19 dataset	ML, DL	-Train:EVC (RF, SGD, LR) -Test:SVC, RF, CNN-LSTM, BERT	— EVC:Accuracy 95.04% — AUC ROC
[9]	Self collected COVID-19 dataset	DL	Spacy Library	
[10]	Self collected COVID-19 dataset	DL	VADER, XLNeT, LDA	Accuracy 63%

Abbreviations included in Table 1. *Multi-Layer Perceptron (MLP), Bidirectional Encoder Representations from Transformers (BERT), Valence Dictionary for Sentiment Analysis (VADAR), Stoch Gradient Descent (SGD), Ensemble Voting Classifier (EVC), Logistic Regression (LR), Global Vectors for Word Representations (GLOVE), Ensemble Voting Classifier (EVC), the Area Under the Receiver Operating Characteristic Curve (AUC ROC), Majority Voting technique-based Ensemble Deep Learning (MVEDL).*

2.4 Discussion

The disparities between related works are closely linked to the source and nature of the data used for extraction, at the period observations, at pre-processing stages, etc. In addition, the different classifiers, methods feature extraction and fine-tuning of various parameters make intra- and inter- work comparisons difficult. A small modification of their settings fundamentals can have a significant impact on the overall performance of a classifier.

In contrast, ML classification techniques are mainly the most used and the DL methods present several advantages over more traditional approaches:

- The category lexicon is simple and does not require data labeled. It leans exclusively on lexicons of annotated words and do not take into account the information context-specific or information-specific to the field and require constant updating of their dictionary to include new words or abbreviations.
- The traditional ML category need a feature design that will take time,
- The DL category automates the feature learning process.

However, the hybrid methods that combine several DL classifiers gave the most encouraging results.

3 Methodology

The problem at the core of our research is the need to extract sentiment information from user-generated content on social media platforms. To formally define the sentiment analysis problem, let:

- D represent the training dataset comprising tweets.
- di denote an individual piece of text data within the dataset D.
- S(di) be the sentiment label assigned to data point di, where S(di) is {Positive, Negative}.

The task of sentiment analysis can be represented as a function f, mapping each data point di to its corresponding sentiment label: f: D → S(di).

Our main objective is to design and implement an accurate sentiment prediction model, f, using machine learning and deep learning techniques, such that it assigns the correct sentiment label to each piece of tweets data in the dataset D. In order to achieve this, we present the overall framework architecture. The developed framework consists of four phases: (a) the data collection phase; (b) the pre-processing phase; (c) the deep analysis phase; and (d) the classification phase. **Fig. 2.** illustrates the employed framework for predicting sentiments based on the learning models.

3.1 Data collection phase

The primary objective of this phase is to extract valuable insights from social media dialogues concerning COVID-19 vaccines. To accomplish this, we performed a systematic data gathering operation on Twitter, concentrating specifically on tweets related to four prominent vaccine types: Sinovac, AstraZeneca, Pfizer, and Moderna, all conveyed in the English language. This rigorous data collection endeavor took place during the first half of 2021, dictated by the constraints of the API, which permits access to tweets for a maximum duration of seven to nine days. Each week, we diligently retrieved and processed the data, culminating in the assembly of a substantial dataset comprising approximately 50,000 tweets. The constructed dataset provides a basis for learning and evaluating sentiment analysis models.

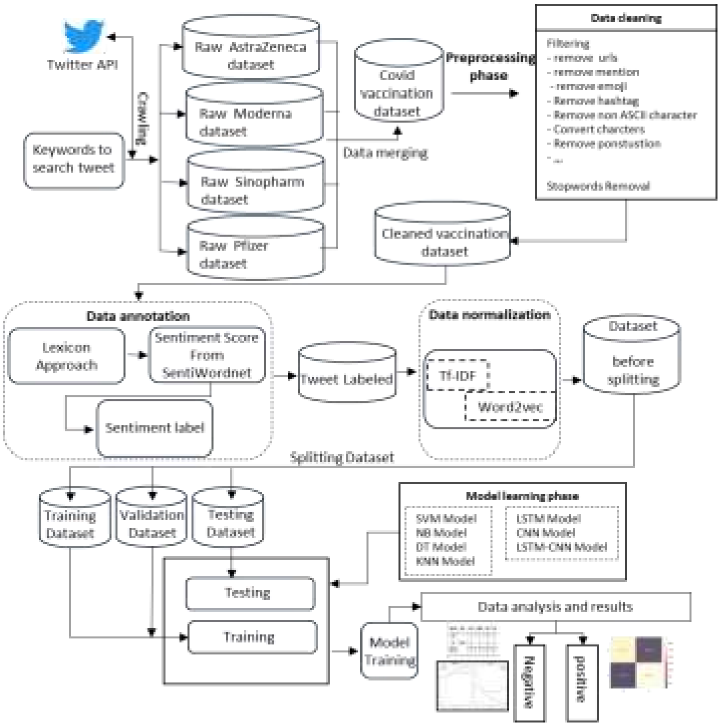

Fig. 2. Sentiment analysis process.

3.2 Data preprocessing phase

Data cleaning. The cleaning of the initial Twitter data involved a multi-stage NLP procedure: The first stage focused on deletions, including HTML tags, retweets ("RT"), URLs, mentions with usernames ("@"), hashtags (#) along with their terms, non-ASCII characters, numbers, and converting emojis to words. Additionally, text was converted to lowercase. The second stage involved the removal of punctuation and stop words, as well as stemming and lemmatization. The final stage primarily focused on eliminating empty and duplicate tweets.

Data annotation. The employed automatic annotation was conducted through heuristic methods, incorporating the SentiWordNet lexicographic sentiment dictionary [16]. This lexicon provides words along with their associated polarity scores, facilitating sentiment identification. So, it exhibits a sentiment distribution with 25.7% positive, 22.3% negative, and 52% neutral tweets.

Data normalization. To prepare data for machine and deep learning models, feature extraction from the initial data was performed. Our evaluation encompassed various techniques, including TF-IDF, bag of words (BoW), and word embedding (Word2Vec), were considered and evaluated. In addition, N-grams (unigrams, bigrams, and trigrams) were used to develop feature vectors.

3.3 Model learning phase

In our quest to predict sentiment with precision, we have devised a set of models, both in the realm of machine learning and deep learning, and each relies on a distinct set of mechanisms:

Machine learning models. These models, known for their simplicity, are built upon the foundation of domain expertise and manual feature selection. They start by standardizing text data through preprocessing and removing any irrelevant information. Then, feature extraction techniques, such as TF-IDF and N-grams, are applied to represent the text as numerical features that can be fed into classifiers.

LSTM model. In this model (see Fig. 3), the embedding layer first transforms the input into a sequence of embedding vectors, which are then passed to the LSTM layer. Each LSTM cell selects the information to be preserved and generates a new encoding vector based on the previous storage. We enhance the categorization of classes through the inclusion of two dense layers. This configuration includes an LSTM layer with 200 dimensions (units), followed by a dense layer using the 'ReLu' activation function and a final inactive layer of a neuron utilizing the 'softmax' function.

CNN model. This model accepts input data at the embedding layer. The convolution layer extracts feature and generates feature maps, while the pooling layer reduces the dimensionality of these feature maps. The first dense layer employs the "ReLu" activation function, and the second layer employs the "softmax" function. In its implementation, we incorporate two 1D convoluted layers, followed by MaxPooling1D and GlobalAveragePooling1D layers, respectively. Dropout layers with a rate of 0.2 are employed to mitigate overfitting. Additionally, the convolutional layers are equipped with L2 regularization (λ=10^{-4}) and He initialization (consistent with the 'ReLu' activation function). The output of the convolutional and pooling layers feeds into a dense layer, which in turn provides the extracted features to a hidden layer for classification. The dense layer is also endowed with L2 regularization and He initialization. (see Fig. 3)

LSTM-CNN model. In this configuration (see Fig. 3), both LSTM and CNN are employed sequentially with identical layers, neurons, and parameters. The embedding layer first transforms the inputs into embedding vectors. Subsequently, the LSTM layer processes each embedding vector, preserving information, and generating encoding vectors. The convolutional layer then processes the output, creating a series of feature maps, which are combined with the pooling layer. To enhance class categorization based on input attributes, two dense layers are employed. The first dense layer utilizes the 'ReLu' activation, while the second takes into account a set of data to predict the output. The LSTM analyzes the syntactic structure of a tweet, while the convolutional layer extracts feature such as positive and negative polarity data. In this setup, a 200-unit LSTM layer is inserted between the Embedding and the first convolutional layer to implement the LSTM-CNN model. The optimizer employs a suitable loss function for binary classification tasks, with the learning rate set to ($\alpha = 10^{-4}$).

Fig. 3. Our corresponding LSTM, CNN and LSTM-CNN implementations

For sentiment analysis of tweets, LSTM outperforms in understanding context and sentences, even with little data. CNN stands out for its speed and accuracy in identifying local features. LSTM-CNN combines the strengths of both models for comprehensive and detailed analysis. The final choice depends on the task and the available data.

4 Data Analysis and Results

The same data distribution was adopted for all classification algorithms, 80% for training and 20% for testing. The performance of sentiment classifiers is evaluated using metrics such as accuracy, precision, recall, F1 score, ROC curve, and area under the curve (AUC):

4.1 ML testing results

The results of **Table 2.** show that:

- The Multinomial NB classifier achieved very good levels of accuracy and precision with Bigram, recall with Ngram and F1 score with Trigram.
- The Decision classifier tree achieved very good levels of Accuracy, Precision and F1 score with Ngram and Recall with Trigram.
- The SVM classifier achieved very good levels of Accuracy, Precision, Recall and F1 score with Ngram.
- The KNN classifier achieved very good levels of Accuracy and Precision with Bigram and recall and F1 score with Ngram.

The SVM algorithm performs well and outperforms all three algorithms.

Table 2. Results of ML classifiers

Classical Model	Features (TF-IDF)	Accuracy	Precision	Recall	F-measure
MNB	Ngram	76.87	76.63	100	86.76
	Bigram	77.61	82.66	96.99	86.82
	Trigram	76.44	76.97	98.46	86.40
DT	Ngram	83.34	87.47	89.67	88.55
	Bigram	76.55	82.66	87.50	85.01
	Trigram	76.78	79.06	94.62	86.14
SVM	Ngram	85.35	84.51	99.02	91.19
	Bigram	79.46	80.09	97.13	87.79
	Trigram	77.45	78.18	97.55	86.80
KNN	Ngram	80.58	80.50	98.25	88.49
	Bigram	78.09	82.44	90.43	86.25
	Trigram	76.58	78.58	95.25	86.11

4.2 DL testing results

For each model (LSTM, CNN and LSTM-CNN), specific parameters are defined to optimize their performance for the analysis of the sentiments of the tweets. Their parameterizations (in **Table 3.**) illustrates different strategies adapted to their specific architectures. The LSTM, with its 64 embedding dimensions and 200 cells, is optimized to capture sequential dependencies in 5 epochs, while the CNN uses Conv1D filters and various kernels to extract local features over 8 epochs, integrating dropout and L2 regularization to improve generalization. The LSTM-CNN model combines these approaches, exploiting both LSTM sequence capture and local CNN pattern extraction, building a hybrid architecture. All models use the Adam optimizer with a learning rate of 1e-4 and the BinaryCrossentropy loss function, highlighting a consistent approach for efficient optimization and improved performance in binary classification.

Table 3. Our DL models parameters

Parameters	LSTM	CNN	LSTM-CNN
Embedding Dimension	64	64	64
Epochs	5	8	8
Batch Size	128	128	128
Filters	-	64, 32	64, 32
Kernel Size	-	3, 2	3, 2
Pool Size	-	2 (MaxPooling1D)	2 (MaxPooling1D)
Dropout	0.2	0.2, 0.2	0.2, 0.2
Word Embeddings	Pre-trained	Pre-trained	Pre-trained
Learning Rate (Adam)	1e-4	1e-4	1e-4
Regularization	L2 (LSTM), None (others)	L2 (Conv1D, Dense)	L2 (LSTM, Conv1D, Dense)
Total params	864,993	658,721	896,833
Trainable params	864,993	658,721	896,833
Non-trainable params	0	0	0

Word embeddings play an important role in the performance and effectiveness of these models. Word embeddings can either be learned during model training or pre-trained (like Word2Vec). These parameter choices are based on common practices in the field of NLP, but can be adjusted according to the specificities of the dataset and computing resources. These configurations effectively extract relevant features from tweets while improving the generalization and performance of the models. So, the choice of model and parameter configuration depends on our specific Twitter sentiment analysis needs.

In this context, the best score of 0.90, for accuracy and precision (see **Table 4.**), is given by CNN followed by LSTM-CNN with an accuracy of 0.86 and a precision of 0.87. Regarding recall, LSTM-CNN and CNN share the same score with a value equal to 0.88. For F1-Score, CNN is once again the best with a score of 0.89.

However, CNN architecture outperformed LSTM and LSTM-CNN because it is more important to capture positive and negative expressions than to identify sentence structure and focusing on syntax can sometimes reduce sentiment classification performance.

Table 4. DL testing results

Deep model	Features	Accuracy	Specificity	Recall	F1-score
LSTM	Word2Vec	0.84	0.82	0.92	0.88
CNN		0.90	0.90	0.88	0.89
LSTM-CNN		0.86	0.87	0.88	0.85

Loss and Accuracy graph. The LSTM model shows slight overfitting since the traditional mechanism, such as dropout and regularization, does not work as well for LSTMs as for CNNs. A better solution would be to reduce the number of LSTM units in the layer from 200 to 100, which makes the model simpler and improves its generalization ability. The high complexity of the hybrid LSTM-CNN model could also be a contributing factor to the observed overfit.

Moreover, in the following figures (Fig. 4, Fig. 5 and Fig. 6), the loss curves decrease monotonically for the training set, which is desirable, and they are not optimized for the test set. They are therefore bound to exhibit fluctuations. The accuracy curve on the test set is generally increasing, which validates the models for the next stage of performance analysis.

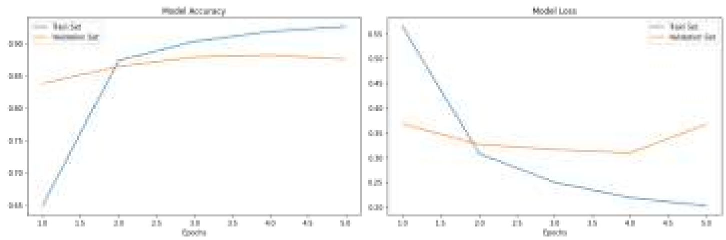

Fig. 4. Loss and Accuracy LSTM graph

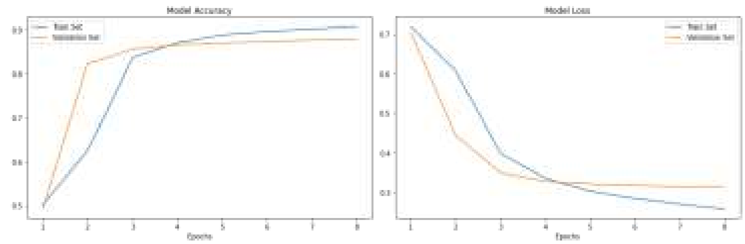

Fig. 5. Loss and Accuracy CNN graph

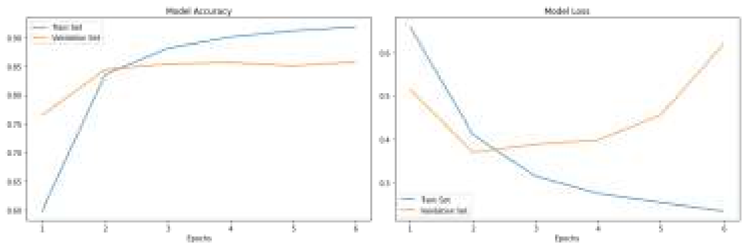

Fig. 6. Loss and Accuracy LSTM-CNN graph

Accuracy training and testing. The CNN model surpasses all other models in terms of test accuracy. This is due to the fact that for sentiment analysis, the syntax, which is usually well captured by RNN-based models, is not as important as the characteristics of sentences that are usually well captured by CNN-based models.

On the other hand, the LSTM-CNN model shows high performance on the train partition due to its higher model complexity. Indeed, its under-adjustment is minimal and such an optimization is done at the cost of a generalization error, as shown by its rather low score on the test partition. (See **Table 5.**)

The three models therefore perform better than the classic ML models NB, SVM, DT and KNN for the same task. ML models are limited to an accuracy of up to 0.85.

Table 5. The training and testing accuracies for the DL models

Model	LSTM	CNN	LSTM-CNN
Training Accuracy	0.93	0.91	0.92
Testing Accuracy	0.88	0.89	0.86

Confusion matrices, ROC curves and AUC. The AUC scores retained for these three models are 0.87 LSTM, 0.89 CNN and 0.86 LSTM-CNN. The CNN model gave the best score with a value of 0.89. The AUC scores of the three models are between 0.8 and 0.9, so this is "excellent discrimination" power (see **Fig. 7.** and **Fig. 8.**).

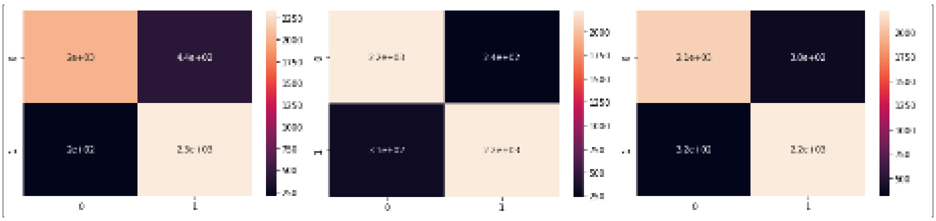

Fig. 7. The confusion matrices for the three DL models

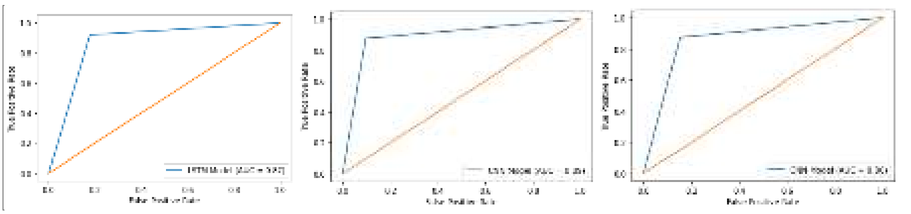

Fig. 8. ROC Curves for the DL models

To deal with overfitting, regularization and dropout mechanisms have been integrated, particularly for convolutional layers. Additionally, it was observed that the LSTM model was more efficient for validation and test partitions, with early shutdown and reduction in the number of memory cells.

In conclusion, the experimental results reveal that the CNN model outperforms the LSTM and LSTM-CNN models on the majority of evaluation metrics because it assigns relatively greater importance to features, such as words and sentences (short-text), conveying strong feelings in relation to the syntactic structure of the text. All three DL models offer a significant improvement over more traditional ML models, supporting the growing adoption of deep learning techniques in the fields of data science and machine learning.

5 Limitations and future research perspectives

Although the selected experimental results prove that the effectiveness, capacity and efficiency of DL models improved the tweet classification performance, some limitations may hinder the proper functioning or performance of our sentiment analysis process:

- The process does not use effective data initialization and preprocessing techniques. Instead of relying on pre-established NLP preprocessing techniques, a more advanced preprocessing technique, such as standard normalization that takes into account cases of disambiguation, negation, sarcasm or irony, and mixed emotions, would be extremely supportive.
- The exclusive focus on specific data expressed only in English and originating only from the social platform Twitter may not represent the sentiments of the general public around the world. Results may be biased based on the demographics of all Twitter users and may not generalize to the attitudes of the broad Big Web public, including multiple modes, such as text, images, videos, and audio, or to understand the long-term context, sentiments can change and evolve.
- The performance of the learning models is evaluated and compared on test tweets which belong to the same distribution as the training tweets. It would therefore be desirable to exploit models trained for new tweets which would obviously belong to a new similar distribution. It would therefore be advantageous to examine this generalization problem outside of the same distribution and to construct robust and more advanced neural models.
- The RNN-based model used an LSTM layer where the information flow is unidirectional. It would therefore be beneficial to test different types of RNNs outside of our LSTM model. For example, using bidirectional LSTM over the LSTM-CNN model could yield an even better result, because each token passed to the CNN layer would contain the information of all other tokens in the original input.
- Using the Word Embedding layer in the neural network rather than feature extraction using Word2Vec pre-trained embedding would be an interesting experiment to undertake.
- The sentiment classes used (positive, negative or neutral) may not capture all of the various emotional intensities that a tweeter may express. It is therefore important to extend this to a finer analysis of feelings, which includes for example various emotional intensities, such as strongly positive, positive, neutral, negative and strongly negative, and more specific categories, such as happy, sad, in anger or surprise.

- The current process takes advantage of sentiment calculation methods and ML and DL learning algorithms, but the choice of the best performing model always remains based on experimental results. An in-depth study of the three deep neural models and similar large-scale and more advanced models would be very beneficial to the NLP research community because it would allow understanding the precise role of each layer of the network and then using them as basic building blocks for better neural networks and for finer multilingual and multimodal sentiment analysis accessible to a wide public.

6 CONCLUSION

This present work aims to explore and compare the performance of machine and deep learning algorithms. In machine learning, text is preprocessed to remove stop words, normalize the text and represent it as numerical features based on TF-IDF or bag-of-words. The cleaned text is then fed into MNB, SVM, DT and KNN for classification. In deep learning, preprocessed text is encoded using pretrained word embeddings such as Word2Vec. These embeddings capture patterns in the text, which are then integrated into LSTM, CNN, and LSTM-CNN. At last, the results indicate that the CNN model outperforms all other DL models, achieving 90% accuracy at the cost of quite a long training time. It performs 6% better than LSTM and 4% better than LSTM-CNN in terms of accuracy. Because syntax is not as important as positive or negative in sentiment classification. CNN also outperforms other ML architectures on the constructed dataset. These DL models, individually or in combination, go a long way in achieving high accuracy and insight in sentiment analysis.

At last, the decision-making process towards an effective and successful vaccination campaign can be guided by the involvement of the general target public by listening to them and responding to their expectations, concerns and difficulties linked to vaccination. It will therefore be recommended to evaluate all content generated across fully social platforms in order to broaden the scope of the results and better understand their applicability in various long-term contexts. For this, it is imperative to take an in-depth look at current advances in finer-grained sentiment analysis for a comprehensive understanding of recent advances by leveraging large-scale models, such as Large Language Models (LLM) and pre-trained models (PTM).

References

1. Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., Sharif,S: An analysis of COVID-19 vaccine sentiments and opinions on Twitter, Int. J. Infectious Diseases, vol. 108, pp. 256–262, Jul. (2021)
2. Liu, S., Liu, J.:Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis, Vaccine, vol. 39, no. 39, pp. 5499–5505, Sep. (2021)

3. Chakraborty,K., Bhatia, S., Bhattacharyya, S., Platos, J., Bag, R., ,Hassanien, A. E.:Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media, Appl. Soft Comput., vol. 97, Dec. (2020)
4. Madani, Y., Erritali, M., Bouikhalene, B.:Using artificial intelligence techniques for detecting COVID-19 epidemic fake news in Moroccan tweets, Results Phys., vol. 25, Jun. 2021, Art. no. 104266
5. Malla, S., Alphonse, P. J. A.: COVID-19 outbreak: An ensemble pretrained deep learning model for detecting informative tweets, Appl. Soft Comput., vol. 107, Aug. (2021), Art. no. 107495
6. Sv, P., Tandon, J., Vikas, Hinduja, H.:Indian citizen's perspective about side effects of COVID-19 vaccine—A machine learning study, Diabetes Metabolic Syndrome, Clin. Res. Rev., vol. 15, no. 4, Jul. (2021)
7. Cotfas,L.-A. , Delcea, C., Roxin, I., Ioanas, C., Gherai, D. S., Tajariol F.: The longest month: Analyzing COVID-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement,: IEEE Access, vol. 9, pp. 33203–33223, (2021)
8. Mahdikhani, M.:Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of COVID-19 pandemic, Int. J. Inf. Manage. Data Insights, vol. 2, no. 1, Apr. (2022)
9. K. Petersen, J. M. Gerken, "#COVID-19: An exploratory investigation of hashtag usage on Twitter," Health Policy, vol. 125, no. 4, pp. 541–547, Apr. (2021)
10. Lyu, H., Wang, J., Wu,W., Duong, V., Zhang, X., Dye, T. D., Luo, J.:Social media study of public opinions on potential COVID-19 vaccines: Informing dissent, disparities, and dissemination, Intell. Med., vol. 2, no. 1, pp. 1–12, Feb. (2022)
11. N. Braig, A. Benz, S. Voth, J. Breitenbach, and R. Buettner, "Machine learning techniques for sentiment analysis of COVID-19-related Twitter data," IEEE Access, vol. 11, pp. 14778–14803, Nov. 2023.
12. McCallum A., Nigam, K.: A comparison of event models for naive Bayes text classification, in Proc. AAAI Workshop Learn. Text Categorization, vol. 752, no. 1, (1998), pp. 41–48 14
13. Bengio,Y. LeCun,Y. , Hinton,G. E.:Deep learning for AI, Commun. ACM, vol. 64, pp. 58–65, Jun. (2021)
14. Gandhi,U. D. , Malarvizhi Kumar, P., Chandra Babu G., Karthick, G.:Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM), Wireless Personal Communications, (2021).
15. Muhammad,P. F., Kusumaningrum, R. , Wibowo, A.:Sentiment Analysis Using Word2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews, in Procedia Computer Science, Semarang, (2021)
16. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta, (2010).
17. Najar F., Bouguila, N.: On smoothing and scaling language model for sentiment-based information retrieval, Adv Data Anal Classif, vol. 17, pp. 725–744, (2023)
18. Jmal S., Ahmed A. A. I., Jalota, C.: Natural Language Processing in Improving Information Retrieval and Knowledge Discovery in Healthcare Conversational Agents, Journal of Artificial Intelligence and Machine Learning in Management, vol. 7, no. 1, pp. 34–47, (2023)