The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 23, Pages 209-219

ICRETS 2023: International Conference on Research in Engineering, Technology and Science

Bi-Directional LSTM-Based COVID-19 Detection Using Clinical Reports

Salah Bouktif
United Arab Emirates University

Akib Khanday
United Arab Emirates University

Ali Ouni
University of Quebec

Abstract: COVID-19 has affected the entire globe with its rapid spreading, causing a high transmission rate. A huge amount of people come in contact with this deadly virus, and early diagnosis of such kind of viruses may save many lives. This paper proposes an improved approach for detecting COVID-19 based on Long Short Term Memory (LSTM) and taking advantage of early clinical reports. To train the LSTM-based classifier for COVID-19 detection, various preprocessing techniques and word embeddings are employed. These techniques ensure the data is in a suitable format for the LSTM model. The proposed LSTM model is then compared against state-of-the-art ensemble models like Bagging and Random Forest, demonstrating its superior performance. The evaluation results showcase a testing accuracy of 87.15%, with a precision of 91% and a recall of 88%. These metrics indicate the effectiveness of the proposed LSTM model in accurately detecting COVID-19-positive cases. By leveraging early clinical reports and utilizing advanced deep learning techniques, our approach achieves significant improvements in COVID-19 detection compared to existing ensemble models.

Keywords: LSTM, COVID-19, Diagnosis, Detecting, Embeddings, Classifier

Introduction

The COVID-19 pandemic has emerged as a severe global health crisis, with far-reaching consequences for societies, economies, and the overall well-being of individuals on a worldwide scale (Mangal et al., 2020). Since its emergence in late 2019, this highly infectious disease has rapidly spread, leading to a significant number of infections and fatalities across the globe. The virus primarily targets the respiratory system, giving rise to a wide range of symptoms that span from mild flu-like manifestations to severe respiratory distress and organ failure (Ucar & Korkmaz, 2020). It is important to note that certain populations, particularly older adults and individuals with pre-existing health conditions, are at a heightened risk of experiencing severe outcomes. The impact of this pandemic has been profound, causing widespread disruptions and necessitating urgent measures to address its implications.

In response to the spread of COVID-19, governments and health organizations worldwide have implemented a range of measures aimed at curbing the transmission of the virus. These strategies include widespread testing, contact tracing, quarantine protocols, travel restrictions, and the development and distribution of vaccines (Horry et al., 2020). The pandemic has emphasized the critical importance of global cooperation and information sharing in effectively addressing and managing public health emergencies.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

Beyond its impact on health, COVID-19 has triggered significant disruptions to economies across the world. Lockdowns, restrictions, and social distancing measures have resulted in the closure of businesses, substantial job losses, and reduced economic activity. Sectors such as tourism, hospitality, and retail have been particularly hard-hit by these measures (Rabani et al., 2023). Governments have responded by implementing fiscal stimulus packages to alleviate the economic burden and provide support to affected individuals and businesses (Mathieu et al., 2021; Ndwandwe & Wiysonge, 2021).

The pandemic has also taken a toll on mental health, as the necessary measures to contain the virus, such as isolation, fear, and uncertainty, have contributed to heightened levels of stress, anxiety, and depression among individuals. Access to mental health services and support has become crucial in addressing the psychological consequences of the pandemic (Khanday et al., 2020).

As the long-term implications of COVID-19 continue to unfold, it becomes increasingly essential to promptly and accurately diagnose cases using various Artificial Intelligence (AI) techniques. AI has shown promising results across diverse fields, including healthcare, education, and defense. Early detection facilitated by AI-assisted diagnosis holds the potential to mitigate human losses (Moore & Offit, 2021; Yan et al., 2020). Ongoing research aims to comprehensively understand the long-term health effects, evaluate vaccine efficacy against emerging variants, and develop robust strategies to prevent future pandemics (Mathieu et al., 2021; Soleimanpour & Yaghoubi, 2021).

Machine learning, has been used in different domains ranging from engineering (Almarimi et al., 2019, Daagi et al., 2017, Daaji et al., 2021) to education (Bouktif and Manzoor, 2021) and healthcare (Gudigar et al., 2021). With its capability to analyze vast and intricate datasets, has played a pivotal role in enhancing COVID-19 detection, monitoring, and response efforts (Lazarus et al., 2021). By leveraging the power of machine learning, healthcare professionals, researchers, and policymakers can make data-driven decisions that significantly improve public health outcomes. This study proposes the development of a diagnostic system specifically designed to enhance the detection of COVID-19 cases by mining clinical reports. In particular, the research focuses on investigating the role of Natural Language Processing (NLP) in identifying COVID-19 using Long Short-Term Memory (LSTM). Given the extensive utilization of LSTM in various prediction tasks across different domains (Bouktif et al., 2020), we capitalize on its text mining capabilities (Bouktif & Awad, 2013) to effectively classify texts using word embeddings, thereby leveraging its advanced classification capabilities. a type of deep learning model known for its text mining capabilities and ability to leverage embeddings. Notably, this work deviates from previous approaches that primarily rely on image processing for patient case classification. By customizing LSTM to train on COVID-19-related word embeddings, the accuracy and methodology for detecting COVID-19 cases can be enhanced, offering a promising avenue for more effective diagnosis.

The rest of the paper is organized as follows. In Section 2, we present the background as well as the relevent literature of our work and contribution. Section 3 discusses the proposed our LSTM based Detecting appoach. An empirical evaluation of our model and results of its comparison with benchmarks are discussed in Section 4. In Section 5, we drawconclusion and give insights into our future works.

Background and Related Work

Artificial intelligence (AI) has ushered in a transformative era across various domains, encompassing healthcare, defense, economics, and education. Within the realm of disease detection and diagnosis, researchers have been actively harnessing the power of machine learning and deep learning techniques to address the complex challenges posed by life-threatening conditions like cancer and tumors (Gudigar et al., 2021). Amid the ongoing efforts to diagnose COVID-19, a wealth of literature has been meticulously compiled from reputable databases, including SCOPUS and Web of Science. This comprehensive collection of research has been facilitated through a keyword-based mechanism, primarily centered around retrieving pertinent journal articles published by esteemed entities such as Elsevier, IEEE, Taylor Francis, Wiley, MDPI, and ACM.

In the context of COVID-19 diagnosis, machine learning and deep learning have emerged as indispensable tools. For instance, Alazab et al. (2020) achieved successful detection of the Coronavirus from CT scans through the utilization of Convolutional Brain organizations, attaining an impressive accuracy rate of 94.80% in Australia and 88.43% in Jordan. Roberts et al. (2021) delved into the realm of AI-based COVID-19 detection and prediction by employing chest radiographs and CT scans. However, their study revealed strategic flaws and inherent biases within existing models, indicating that none of the models were deemed suitable for clinical implementation. Mangal et al. (2020) introduced an innovative deep neural network model named CovidAID: Coronavirus Artificial Intelligence Locator, specifically designed to prioritize patients for appropriate testing. Trained on a publicly available COVID-19 chest X-ray dataset, this model showcased an accuracy of 90.5% with 100% recall in detecting Coronavirus infections. Furthermore, Islam et al. (2020) proposed a deep learning approach that combined a Convolutional Neural Network (CNN) with Long Short-Term Memory (LSTM) to automate the diagnosis of Coronavirus based on X-ray images. Their system utilized an extensive dataset of 4,575 X-ray images, including 1,525 cases of COVID-19, and demonstrated exceptional performance metrics, boasting an accuracy of 99.4%, AUC of 99.9%, specificity of 99.2%, sensitivity of 99.3%, and F1-score of 98.9%.

Using audio analysis, Hassan et al. (2020) employed a Recurrent Neural Network (RNN), particularly LSTM, to scrutinize acoustic features such as coughing, breathing, and voice patterns for early screening and diagnosis of COVID-19. The results revealed relatively lower accuracy rates for voice samples in comparison to coughing and breathing sound samples. Furthermore, Kaya et al. (2022) proposed an innovative approach that utilized the Affine Transformation (AT) technique in conjunction with a hybrid deep learning model merging GoogleNet and LSTM for COVID-19 detection through X-ray images. Remarkably, their proposed method achieved a high classification accuracy rate of 98.97% employing a dataset from Mendeley. These examples serve as a glimpse into the recent advancements made in the field of COVID-19 detection using machine learning and deep learning methodologies. Table 1 provides a concise summary of some noteworthy works within this area, highlighting the progress achieved in combating the ongoing pandemic.

Table 1. Related work

Technique	**Contribution**	**Results**	**Research Gap**
Deep Learning	Proposed SqueezeNet based diagnostic of the coronavirus disease and classified it into 3 classes (COVID19/ Pneumonia /Normal)	The accuracy of the model achieved is 87%	Only Images are used.
Machine Learning	Used Traditional Approaches of Machine Learning for classifying the text into binary class.	The classification Accuracy achieved from this approach is 80%	Only 212 data records are used.
Deep Learning	Proposed a tailored deep neural network for classifying an x-ray image into three classes (COVID-19/ PNEUMONIA/ NORMAL)	An accuracy of 91.3% is achieved.	X-Ray images are used and there is scope for Hyperparameter tuning.
Deep Learning	Proposed nCOVnet for detecting COVID from X-ray images(Binary Classification).	The model achieved 87.3% accuracy.	Multi class classification can be performed.
Deep Learning	Deep learning approach for detecting COVID-19 into three classes (COVID-19/ PNEUMONIA /NORMAL)	The proposed approach achieves an accuracy of 97%.	Only worked on X-Ray images. CT Scan images may be used.
Deep Learning	Covid detection based on Generative Adversal Learning	Accuracy of 99% is achieved by the proposed approach.	Work can be explored to other multimodalities.

After studying the relevant work following conclusions are drawn:

- Most of the work is being performed on image datasets like X-Rays and CT Scan images using image processing techniques, with less focus on text mining.

- Also the data published by the World Health Organization (WHO) contains various metadata which need to be investigated to improve the model's efficiency.
- Word Embeddings play a major role in text mining, and using word embeddings for detecting COVID-19 may be helpful.

Methodology

The field of Natural Language Processing (NLP), an emerging discipline within Artificial Intelligence, has demonstrated promising outcomes in tackling machine translation and classification challenges. This study focuses on utilizing text mining techniques to detect COVID-19 from early clinical textual reports, aiming to expedite the diagnosis process and enable early intervention. Additionally, humans possess the ability to observe various irregularities in individuals, and healthcare professionals often document their findings in written reports when examining patients. Consequently, this research endeavors to address a similar problem by developing a classification system to identify COVID-19 or Normal cases from clinical reports. The proposed approach leverages NLP and text mining methodologies to extract relevant information and classify the reports effectively. Figure 1 provides an overview of the comprehensive framework employed in this study, illustrating the interconnected components and their contribution to the overall approach. By harnessing the power of NLP and text mining, this research aims to contribute to the early detection and diagnosis of COVID-19, potentially saving time and facilitating timely medical intervention.

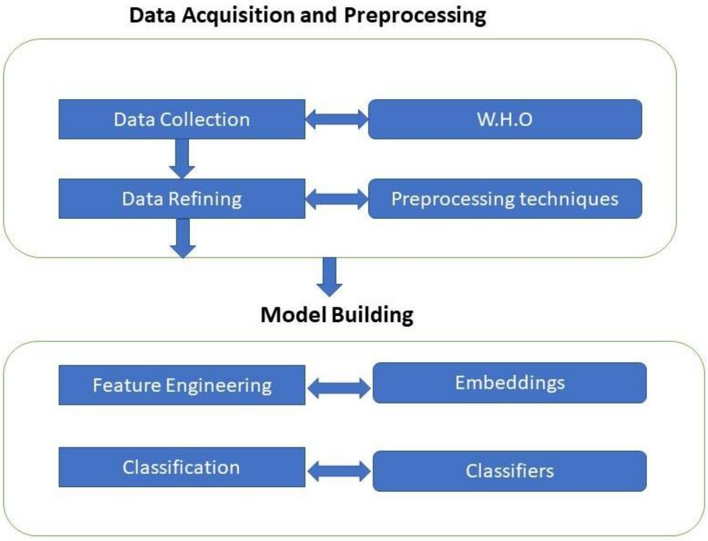

Figure 1. Overall framework for classification

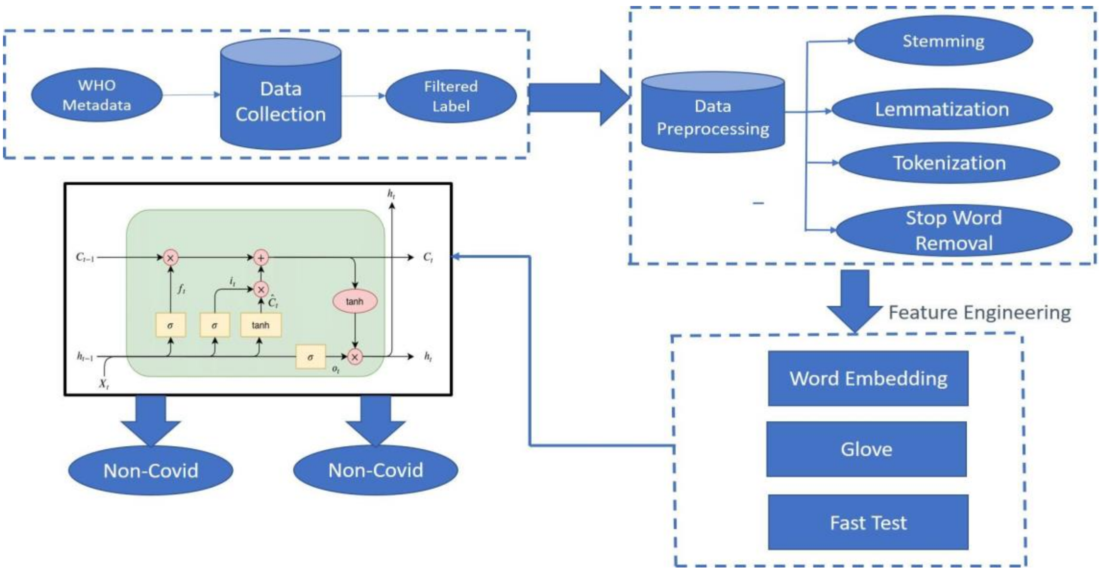

Figure 2. LSTM based COVID-19 detection.

Data Collection

A dataset on COVID-19 is sourced from a publicly available database published by the WHO. This dataset contains textual reports that encompass essential information pertaining to COVID-19, including symptoms, medical history, and laboratory test results. Each record within the collected dataset is carefully labeled to indicate whether the individual is COVID-19 positive or negative. Notably, the dataset consists of a substantial total of 915 records, offering a significant sample size that is valuable for the purposes of analysis and modeling.

Data Preprocessing

Data preprocessing is a crucial phase in the proposed methodology as it prepares the data to meet the input criteria of LSTM. Several techniques are applied during this phase, including converting the text to lowercase, removing punctuation marks, tokenizing the text into individual words, and excluding stop words. Moreover, it is recommended to consider additional methods like stemming or lemmatization to further standardize the words. In order to visualize the distribution of text lengths, Figure 3 presents a histogram that illustrates the frequency distribution of text lengths.

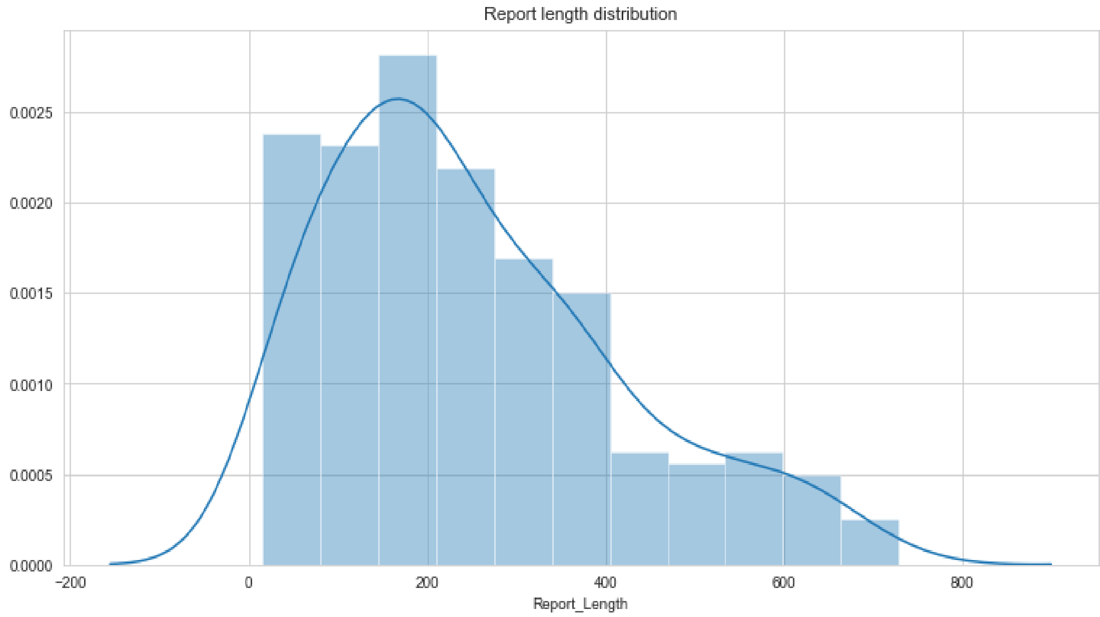

Figure 3. Clinical reports with their length

Feature Selection

In order to conduct the classification, the preprocessed text undergoes a transformation into word embeddings, which represent words as dense vectors. Pre-trained word embeddings, such as Word2Vec, GloVe, or FastText, can serve as suitable algorithms for this purpose. These embeddings capture the semantic relationships between words and enable the LSTM to comprehend the meaning of the text effectively. Furthermore, it is essential to ensure that all text sequences possess the same length by padding or truncating them as required. This step is crucial for creating fixed-length input sequences that can be processed by the LSTM. It is recommended to select an appropriate sequence length that captures sufficient contextual information from the text..

Classification

The LSTM architecture typically consists of an embedding layer to handle the input, one or more LSTM layers to capture sequential information, and one or more fully connected layers for classification. To enhance the performance of the model, one can explore different configurations of LSTM layers, such as stacked LSTMs or bidirectional LSTMs. Figure 2 provides a visual representation of the LSTM model's architecture employed in this study. To facilitate model training and evaluation, the dataset is divided into training, validation, and testing sets. The training set is used to train the LSTM model, while the validation set aids in fine-tuning hyperparameters and monitoring the model's performance. Finally, the testing set is used to assess the model's performance on unseen data. During the training process, the LSTM model is optimized by minimizing a loss function, such as binary cross-entropy. This optimization is achieved using an optimizer such as Adam or RMSprop. Hyperparameters, such as learning rate and batch size, are adjusted to optimize the model's performance. In addition, Figure 4 illustrates the architecture of the bidirectional LSTM model, which can be an alternative configuration explored in this research work.

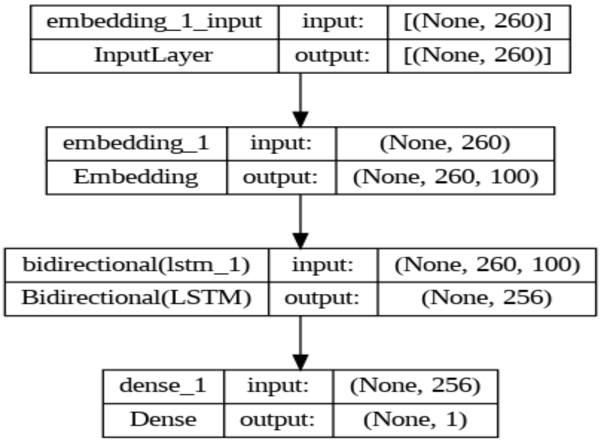

Figure 4. Bi-directional LSTM architecture

The LSTM model's effectiveness depends on the training data's quality and representativeness, the selection of word embeddings, and the design of the LSTM architecture. Regular model evaluation and validation on diverse datasets are crucial to ensure its robustness and generalizability. The Pseudo Code of the proposed framework is shown in Pseudo-Code 1.

Pseudo Code 1: Bi-Directional LSTM based COVID-19 Detection

Require: Reports, Label
Ensure: Covid or Normal
Start
For *i from 1 to n* **do**
P[i]= Report[i]+Label
Q[i]=Length(P[i])
End *For*
For *i from 1 to n* **do**
R[i]=Tokenise(P[i])
R[i]=Stopword Removal (Lemmatization(R[i]))
End *For*
For *i from 1 to n* **do**
F[i]=Embeddings(F[i])+Q[i]
End *For*
Classifier(Bi-Directional LSTM, Report[i], F[i])
End

Emperical Evaluation

In this section, we present the evaluation of the proposed LSTM model for COVID-19 case detection, building upon our previous work. The evaluation process consists of two main steps. Firstly, we evaluate the model using cross-validation to assess its performance across multiple folds of the dataset. Secondly, we compare our model with ensemble-based solutions, aiming to demonstrate the effectiveness of leveraging machine learning techniques in improving the performance of COVID-19 detection models. Indeed, One of the key contributions of this work is the utilization of machine learning techniques to enhance the accuracy of COVID-19 diagnosis based on early clinical symptoms. We achieve this by fine-tuning the hyperparameters of the LSTM model to optimize its performance. To validate the effectiveness of our customized LSTM model, we establish a set of benchmark models that serve as a basis for comparison. Specifically, we employ Bagging and Random Forest algorithms as benchmarks in our empirical evaluation.

By conducting this comprehensive evaluation, we aim to demonstrate the superiority of our proposed LSTM model in detecting COVID-19 cases compared to the benchmark models. This evaluation serves to validate the effectiveness of our approach and highlight its potential contribution to the field of COVID-19 diagnosis.

Evaluation Measures

Using the four metrics of the confusion Matrix, i.e.True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), we calculated the following evaluation measures of the our model performance.

Accuracy: It is one of the standard metrics used to evaluate the classifiers. It measures the ratio of the total number of corrected predicted instances over the total number of predictions.

$$Accuracy = T\,P + TN/T\,P + T\,N + F\,P + F\,N \qquad (1)$$

Precision: It is a metric that determines the ratio of true positives over the total predicted positives.

$$Precision = T\,P\,T\,P + F\,P \qquad (2)$$

Recall: It is a metric that determines the ratio of true positives over total actual positives.

$$Recall = T\,P/\,T\,P + F\,N \qquad (3)$$

F1-score: It is a metric used to measure the performance of the classifier/model when that model needs a balance between Precision and Recall.

$$F1 - score = 2 * P\,recision * Recall/\,P\,recision + Recall \qquad (4)$$

Experimental Settings

The sequential model is used with embeddings from the vocabulary. The sigmoid activation function is used with one dense layer. Binary Cross Entropy is utilized as a loss function and Adam as an optimization function. Ten epochs are run with a batch size of 32. The various metrics used to evaluate classification results are confusion matrix, Accuracy, Precision, Recall and F-Measure

The dataset used in this study is obtained from the public repository of ieee80231. To ensure the data is suitable for classification, various preprocessing techniques such as tokenization, lemmatization, stemming, punctuation removal, stopwords removal, and normalization are applied. The refined dataset is then utilized for training a fine-tuned LSTM classifier. The dataset is split into a training set and a testing set in an 80:20 ratio, with 80% of the data used for training the classifier and 20% for testing.

Model Performance

To evaluate the proposed model, several performance measures including Precision, Recall, and F1-score are computed. The model demonstrates an overall accuracy of 87.15%, with precision, recall, and F1-score of 89%, 87%, and 86% respectively. The results of the proposed approach are presented in Table II, providing a comprehensive summary of the model's performance. Figure 5 depicts the Confusion Matrix, providing visual insights into the classification outcomes.

The model is validated using k-fold cross-validation, in which the value of k equals 5. Five iterations are performed while validating the model; the results are shown in Figure 5. After performing the validation, the average validation score came near to the accuracy achieved on the testing dataset. Also, the results show no overfitting or underfitting of the model.

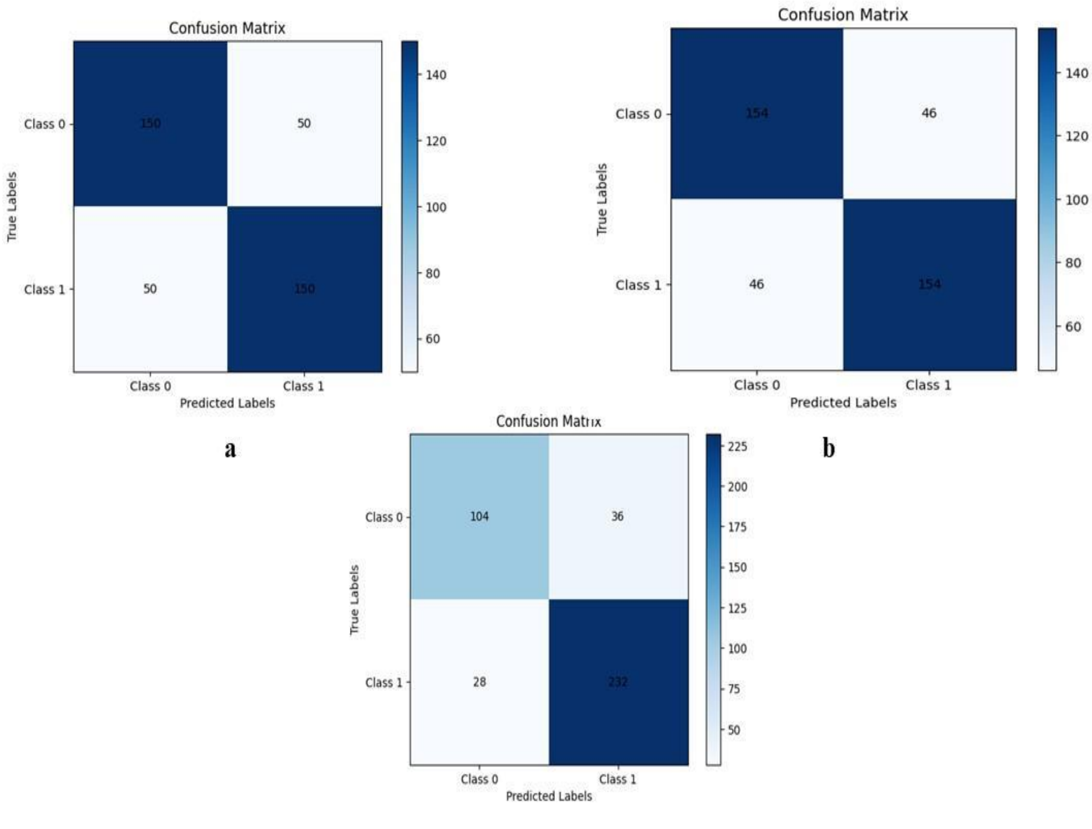

Figure 5. Confusion metrics a: Bagging, b: Random forest, c: LSTM

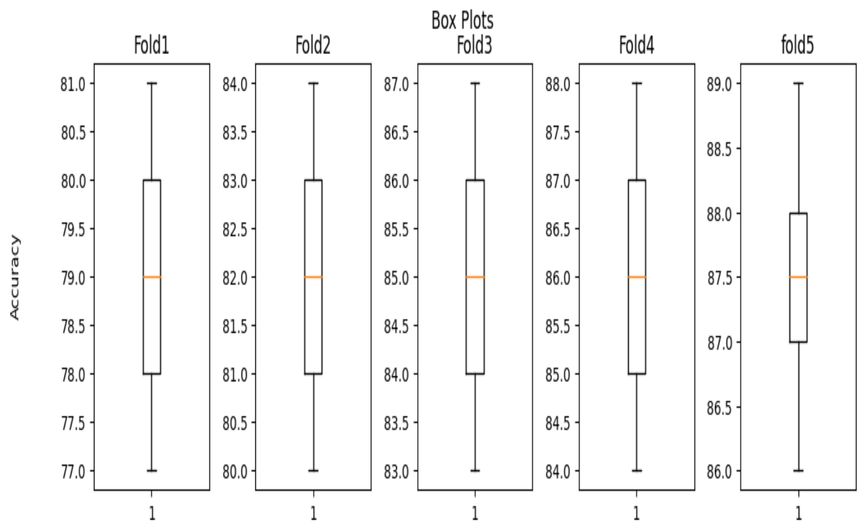

Figure 6. Cross validation

Comparison with Ensemble Model Benchmarks

The validity of the proposed approach is confirmed by conducting a comparative study with other benchmarks commonly used for COVID-19 detection. In this study, we employed 5-fold cross-validation as a model validation method. The results of the proposed approach using the 5-fold cross-validation showed impressive performance measures. The precision was determined to be 89%, indicating the accuracy of correctly identifying COVID-19 cases.

Table 2. Comparative analysis

Classifier	Class	Precision	Recall	F1-Score	Accuracy
Bagging	Normal	0.70	0.75	0.72	75%
	COVID	0.80	0.75	0.77	
Random Forest	Normal	0.77	0.76	0.76	77.5%
	COVID	0.78	0.79	0.79	
Bi-Directional LSTM	Normal	0.89	0.87	0.86	87.15%
	COVID	0.91	0.88	0.84	

The recall, representing the proportion of true COVID-19 cases detected, achieved a value of 87%. The F-measure, which combines precision and recall, was calculated at 86%. Furthermore, the overall accuracy of the proposed approach was measured at 87.15%. Random Forest, one of the benchmark algorithms used in the comparative study, also exhibited relatively good performance. It achieved a precision of 78%, recall of 77%, F-measure of 77.5%, and an accuracy of 77.5%. Table 2 presents the mean values of these different performance measures, providing a comprehensive overview of the comparative study between our proposed LSTM model and other ensemble machine learning algorithms, namely Bagging and Random Forest. These results showcase the superior performance of our LSTM model in COVID-19 detection compared to the benchmark algorithms.

Conclusion

This study presents the utilization of a deep learning algorithm, specifically the Bi-directional LSTM, for the classification of clinical textual reports into COVID-19 and Non-COVID-19 cases. Our approach diverges from previous methods that primarily rely on image processing for patient case classification. Instead, we leverage the potential of LSTM by customizing the Bi-directional LSTM variant to train on COVID-19-related word embeddings. Furthermore, we compare the performance of our model with state-of-the-art benchmarks in the field. The dataset used in this research is obtained from the World Health Organization (WHO) website, where we filter and extract the necessary textual attributes along with their corresponding labels. To facilitate the binary classification task, Word Embeddings are employed in conjunction with the LSTM model. The experimental results, obtained through 5-fold cross-validation, showcase a testing accuracy of 87%, demonstrating the model's effectiveness in accurately classifying clinical reports. As part of future directions, we propose the fusion of two models: one trained on Clinical Symptoms and the other on Clinical Reports. This fusion approach aims to enhance the efficiency of the classification process by leveraging information from both sources. A second avenue for future research is the development of COVID-19 explainable models, drawing inspiration from our previous work in promoting interpretability (Bouktif and Awad, 2013). These explainable models will provide insights into the decision-making process of the classification system, enhancing transparency and understanding in the context of COVID-19 diagnosis.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Research in Engineering, Technology and Science (www.icrets.net) held in Budapest/Hungary on July 06-09, 2023.

*This work was supported by ASPIRE Award for Research Excellence (AARE-2020). (ASPIRE award number AARE20-100 grant #21T55) .

References

Alazab, M., Awajan, A., Mesleh, A., Abraham, A., Jatana, V., & Alhyari, S. (2020). COVID-19 prediction and detection using deep learning. *International Journal of Computer Information Systems and Industrial Management Applications*, *12*(June), 168–181.

Almarimi, N., Ouni, A., Bouktif, S., Mkaouer, M. W., Kula, R. G., & Saied, M. A. (2019). Web service api recommendation for automated mashup creation using multi-objective evolutionary search. *Applied Soft Computing*, *85*, 105830.

Bouktif, S. & Awad, M. A. (2013). Ant colony based approach to predict stock market movement from mood collected on twitter. In *Proceedings of the 2013 IEEE/ACM İnternational Conference on Advances in Social Networks Analysis and Mining*, 837–845.

Bouktif, S., Fiaz, A., & Awad, M. (2020). Augmented textual features-based stock market prediction. *IEEE Access, 8,* 40269–40282.

Bouktif, S. & Manzoor, A. (2021). Artificial intelligence as a gear to preserve effectiveness of learning and educational systems in pandemic time. In 2021 *IEEE Global Engineering Education Conference (EDUCON),* 1703–1711.

Daagi, M., Ouniy, A., Kessentini, M., Gammoudi, M. M., & Bouktif, S. (2017). Web service interface decomposition using formal concept analysis. In *2017 IEEE International Conference on Web Services (ICWS)*, 172–179.

Daaji, M., Ouni, A., Gammoudi, M. M., Bouktif, S., & Mkaouer, M. W. (2021). Multi-criteria web services selection: Balancing the quality of design and quality of service. *ACM Transactions on Internet Technology (TOIT), 22*(1):1–31.13

Ghoshal, B., & Tucker, A. (2020). Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. *ArXiv Preprint ArXiv:2003.10769*.

Gudigar, A., Raghavendra, U., Nayak, S., Ooi, C. P., Chan, W. Y., Gangavarapu, M. R., Dharmik, C., Samanth, J., Kadri, N. A., Hasikin, K., & others. (2021). Role of artificial intelligence in COVID-19 detection. *Sensors*, *21*(23), 8045.

Hassan, A., Shahin, I., & Alsabek, M. B. (2020). COVID-19 detection system using recurrent neural networks. *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*, 1–5.

Horry, M. J., Chakraborty, S., Paul, M., Ulhaq, A., Pradhan, B., Saha, M., & Shukla, N. (2020). COVID-19 detection through transfer learning using multimodal imaging data. *IEEE Access*, *8*, 149808–149824.

Islam, M. Z., Islam, M. M., & Asraf, A. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Informatics in Medicine Unlocked*, *20*, 100412.

Kaya, Y., Yiner, Z., Kaya, M., & Kuncan, F. (2022). A new approach to COVID-19 detection from X-ray images using angle transformation with GoogleNet and LSTM. *Measurement Science and Technology*, *33*(12), 124011.

Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., & Mohi Ud Din, M. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*, *12*, 731–739.

Lazarus, J. V, Ratzan, S. C., Palayew, A., Gostin, L. O., Larson, H. J., Rabin, K., Kimball, S., & El-Mohandes, A. (2021). A global survey of potential acceptance of a COVID-19 vaccine. *Nature Medicine*, *27*(2), 225–228.

Mangal, A., Kalia, S., Rajgopal, H., Rangarajan, K., Namboodiri, V., Banerjee, S., & Arora, C. (2020). CovidAID: COVID-19 detection using chest X-ray. *ArXiv Preprint ArXiv:2004.09803*.

Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., Appel, C., Giattino, C., & Rodés-Guirao, L. (2021). A global database of COVID-19 vaccinations. *Nature Human Behaviour*, *5*(7), 947–953.

Moore, J. P., & Offit, P. A. (2021). SARS-CoV-2 Vaccines and the growing threat of viral variants. *JAMA*, *325*(9), 821. https://doi.org/10.1001/jama.2021.1114

Ndwandwe, D., & Wiysonge, C. S. (2021). COVID-19 vaccines. *Current Opinion in Immunology*, *71*, 111–116.

Rabani, S. T., Khanday, A. M. U. D., Khan, Q. R., Hajam, U. A., Imran, A. S., & Kastrati, Z. (2023). Detecting suicidality on social media: Machine learning at rescue. *Egyptian Informatics Journal*, *24*(2), 291–302.

Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A. I., Etmann, C., McCague, C., Beer, L., & others. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*, *3*(3), 199–217.

Soleimanpour, S., & Yaghoubi, A. (2021). COVID-19 vaccine: where are we now and where should we go? *Expert Review of Vaccines*, *20*(1), 23–44.

Ucar, F., & Korkmaz, D. (2020). COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Medical Hypotheses*, *140*, 109761.

Yan, L., Zhang, H.-T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., & others. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence*, *2*(5), 283-288.

Author Information

Salah Bouktif
College of Information Technology
United Arab Emirates University
Al Ain, UAE
Contact email : *salahb@uaeu.ac.ae*

Akib Mohi Ud Din Khanday
College of Information Technology
United Arab Emirates University
Al Ain, UAE

Ali Ouni
École de technologie supérieure,
ÉTS Montréal , University of Quebec,
Montreal, Canada.

To cite this article:

Bouktif, S., Khanday, A. & Ouni A. (2023). Bi-directional LSTM-based Covıd-19 detection using clinical reports. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)*, *23*, 209-219.