Artificial Intelligence Powered Medical Applications For The Performance Evaluation Of Clinical Models On Sequential Clinical Texts

Vanshika Goel

Vishwakarma University, Pune, Maharashtra, India

[1]*Received: 07 November 2023; Accepted: 23 January 2024; Published: 04 February 2024*

ABSTRACT

Experts use clinical terminology to assess or draw conclusions about particular outcomes based on patients' symptoms. The specialists' clinical terms might be referred to as aspects, and the main work in any sentimental analysis is aspect extraction. The results produced by a pretrained model are the primary focus of the current natural language processing techniques for aspect extraction from a given text. These pre-trained models perform admirably for everyday English, but they struggle to produce precise results when it comes to clinical terminologies. to discuss the problems with clinical terminology. There are several transformer-based learning clinical models that outperform these clinical data, including Clinical Big Bird, Clinician Bert, and Clinical Long former. Large datasets yield better results from these models. This research compares the performance of these models and assesses coherence and context in addition to evaluating the Bio models' performance with sequential clinical content. The most effective outcomes of these models will be shared with an artificial bee colony for additional optimization and integration with AI-based applications.

INTRODUCTION

People utilize standard English language when conversing with one other in daily life, but medical professionals, such as doctors, have a very different communication style and employ specific medical terminology. Our sentiment analysis system performs exceptionally well when used to communicative English; however, it is not as effective when applied to clinical terminology entered into an individual's Electronic Health Record or when patients communicate with doctors directly through third-party services such as chatbots. The standard coarse-level (document or sentence) sentiment analysis approaches are unable to meet the expectations of users because more precise information about the exact terminology that the doctor uses to describe a symptom is needed. Aspects are the precise medical phrases that the doctor uses to explain a symptom or illness. For instance, when a patient communicates his symptoms to a physician, those descriptions may not match up with what the physician enters in the patient's electronic health record since in a medical record, a physician always enters the symptom's technical equivalent.[1, 2, 3] Consequently, it becomes imperative to assess the data entered at the aspect level.

The goal of this research is to analyze each bio model's performance in real-time interactions with apps that feed sequential text into the model. We project some aspects onto bio models, specifically Clinical Bert, Clinical Long former, and Clinical Big Bird, and develop a set of language attributes linked with those aspects.[6] Subsequently, the output produced by the models specified in the first stage will be compared, and the model that produces the best results will be selected.

[1] *How to cite the article:* Goel V., February 2024; Artificial Intelligence Powered Medical Applications For The Performance Evaluation Of Clinical Models On Sequential Clinical Texts; *International Journal of Innovations in Scientific Engineering*, Jan-Jun 2024, Vol 19, 19-25

Third, in order to maximize outcomes and choose the ideal working framework for medical terminology, the second stage's results will be fed into the Artificial Bee Colony in the third step. To the best of our knowledge, this is the first study assessing these bio models' performance when sequential clinical language is entered, and the models' coherence and context are assessed. In addition to examining these models' performance in terms of efficiency, bias, and f1 score, we also need to examine if they can consistently preserve the current context when interacting with any external applications.

In summary, the following is a list of this paper's contributions:

Clinical Bert, Clinical Long former, and Clinical Big Bird are the three clinical models that will be used in this novel feature selection and comparison of linguistic feature extraction. The best linguistic model will be chosen based on the output that each model produces when tested against the same dataset. In-depth tests are carried out on real-world datasets to confirm the viability of the suggested framework and to clearly illustrate how the aspect term extraction performance can be enhanced by the chosen implicit characteristics. Artificial Bee Colony will receive the greatest output from the aforementioned models in order to optimize the outcomes.

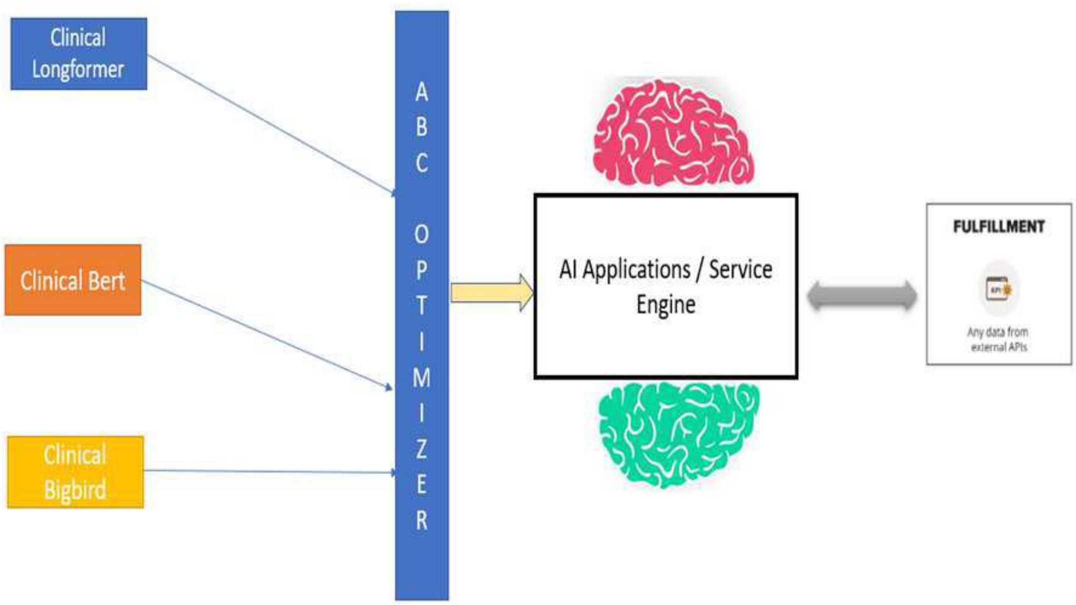

Fig 1 shows an overview of the clinical model's ABC performance analysis.

RELATED WORK

A. Extracting Aspects

NLP can be used for tasks particular to different domains thanks to the development of multiple language models and NLP. However, this use is limited if a domain has its own jargon that has meanings distinct from those of standard English. One general-purpose language model that has been trained on a sizable corpus of material from the internet is called BERT. It may even extract some domain-specific data that is helpful, but it won't be effective for subtle domain-specific information, like medical data. Additionally, as deep learning techniques can integrate text features to extract a new representation via several hidden layers, they can be applied to NLP applications as well. Even while these deep learning models beat machine learning-based models, they are not as good at managing linguistic quirks and aspects like word ordering and spelling. In this case, we could move away from relying just on the pre-trained embeddings and instead use an altered architecture model, like Clinical Big Bird, which has a unique sparse attention

(IJISE) 2024, Vol. No. 19, Jan-Jun **e-ISSN: 2454-6402, p-ISSN: 2454-812X**

mechanism that makes it more scalable and memory-efficient for processing lengthy context sequences. Clinical Big Bird captures dependencies across distant contexts, making it possible to process massive medical documents—like electronic health records (EHRs)—efficiently. Another example of a transformer-based model is Clinical Longformer, a language model that is specifically intended for use with clinical texts. It tackles the problem of processing huge texts similarly to Clinical Big Bird, but instead of using sparse attention, it makes effective use of the sliding window approach to capture long-range dependencies. This enables Clinical Longformer to efficiently process large volumes of clinical documents. For the purpose of clinical language understanding, the original BERT (Bidirectional Encoder Representations from Transformers) model has been modified into clinical BERT. BERT is a pre-trained language model that captures bidirectional contextual information, revolutionizing NLP activities. Clinical BERT can acquire domain-specific representations and comprehend medical terminology effectively because it has been trained on a vast amount of clinical content.

Clinical models, such Clinical Big Bird, Clinical BERT, and Clinical Longformer, have advanced to the point where they can now do NLP tasks with medical datasets better than any normal language model. [4] This research aims to assess the performance of clinical models and identify the advantages and disadvantages of each model in relation to Named Entity or Relation Extraction. These models may perform better when a substantial volume of clinical text is presented at once, as they have been extensively trained on numerous publicly available medical datasets.

This research aims to examine how well these clinical models perform when given clinical material in a sequence similar to user input, and whether or not they are able to preserve the coherence and context. coherence.[5] They would evaluate how well each model identified signs and illnesses. In this study, we also suggest integrating these clinical models with the ABC algorithm to select the optimal linguistic features prior to tweaking the hyperparameters and performing performance analysis.

B. Using ABC for Feature Selection

Feature selection presents an ever-tougher issue in the fields of data mining, with enormous amounts of dimensions that are rapidly expanding. We choose the most informative features using the ABC method to ensure consistency in feature selection. We look for a possible subset of features using the artificial bees.

To lessen the batch effect, ABC applies batch correction to the correction set iteratively. The following are the steps that are involved in active batch correction:

Split the corrective set up into multiple smaller batches (mini batches, for example).

•In order to obtain predictions or embeddings, apply the trained model to every micro batch.

•Determine the batch effect using suitable metrics, such as mean or variance, inside each mini-batch or among mini-batches.

• Modify the embeddings or predictions in accordance with the estimated batch effect.

•Iterate through the aforementioned processes several times until the batch impact is either completely eradicated or greatly diminished.

We assess how well the revised predictions or embeddings perform on the intended downstream tasks or assessment measures after applying ABC. To determine how well ABC reduces batch effects, compare the performance of the corrected models to the original versions.

Iterate through the aforementioned processes several times until the batch impact is either completely eradicated or greatly diminished.

We assess how well the revised predictions or embeddings perform on the intended downstream tasks or assessment measures after applying ABC. To determine how well ABC reduces batch effects, compare the performance of the corrected models to the original versions.

To increase the efficacy of the batch correction, repeat the ABC process iteratively by modifying the hyperparameters, batch sizes, or other factors based on the evaluation results. On the correction set, repeat the active batch correction procedure and iteratively assess the outcomes [11,12].

DETAIL OF THE PROBLEM

A. Formulating problems

The clinical Big Bird, clinical long former, and clinical BERT models' accuracy can be compared using the following mathematical representation:

Let X1 represent the Clinical Big Bird model's accuracy.

Let X2 represent the Clinical Long former model's accuracy.

Let X3 represent the Clinical BERT model's accuracy, and so forth. Our goal is to evaluate these models' accuracy and identify the more accurate model.

I in Xn } X1 > X2 AND X1 > X3... X1 > Xn

OR

X2>X1 AND X2>X3..X2>Xn

OR

Xi>Xy AND Xi>Xz…Xi>Xn]

These claims suggest that a model performs better in terms of accuracy if its accuracy is higher than that of the other two models. But take note that this depiction is predicated on the idea that accuracy is the only factor to be compared.

For a thorough assessment, additional variables like model size, computational effectiveness, or particular work needs are also taken into account. The table below contains the formulas for a few measures.

C. The Optimization Algorithm for ABC

The deterministic feature of the ABC method led to its selection, ensuring accurate optimization of the outcomes without any ambiguity.

Stage I: Begin counting the likely answers for ABC.

Every solution is a representation of the model's parameters. There are three different kinds of bees: scout bees, employee bees, and observer bees. Each has a distinct role that is well defined.

Stage II: Worker bees' phase Using mutation, crossover, and random initialization, new candidate solutions are defined and assessed for their fitness throughout this phase. Then in the employee bee phase, the model will be run on an existing dataset, and the results will be checked for optimization.[19]

Stage III: The Onlooker Bees Phase: During this stage, the most appropriate solution is chosen based on its fitness value and is also updated in the solution population. The precision of the new solutions is also assessed during this step.[20]

Stage IV: the "scout bees" phase: To preserve diversity, find candidate-based solutions that are stalled in local optima and replace them with fresh, random solutions. In this stage, the solutions from stages II and III are examined for suitability in order to determine which is the best.

Step V: Evaluation: Using a fitness function, assess the population's level of fitness and choose the best candidate solution to be the go-forward option.

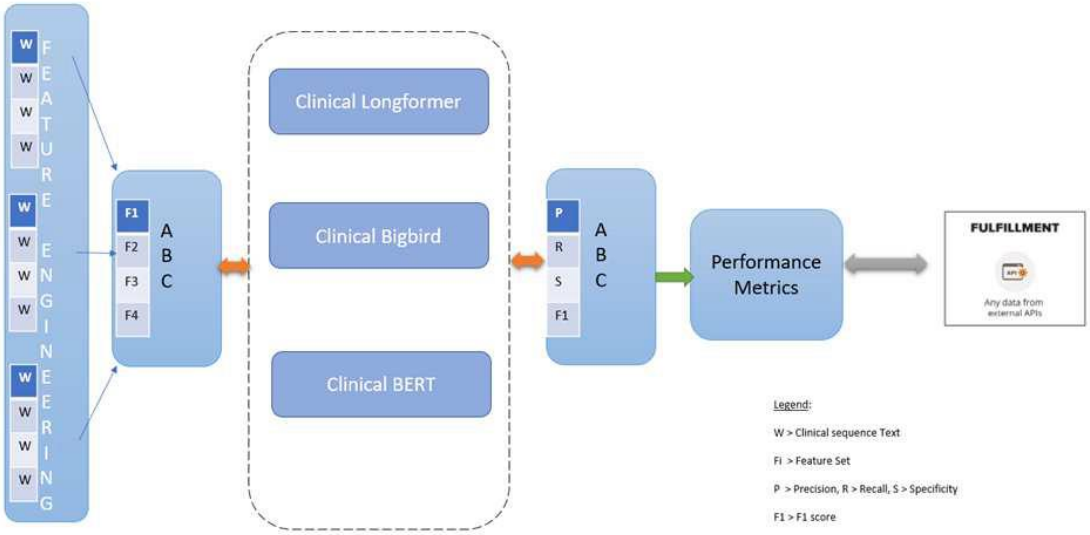

Fig. 2. Performance Framework.

RESULTS

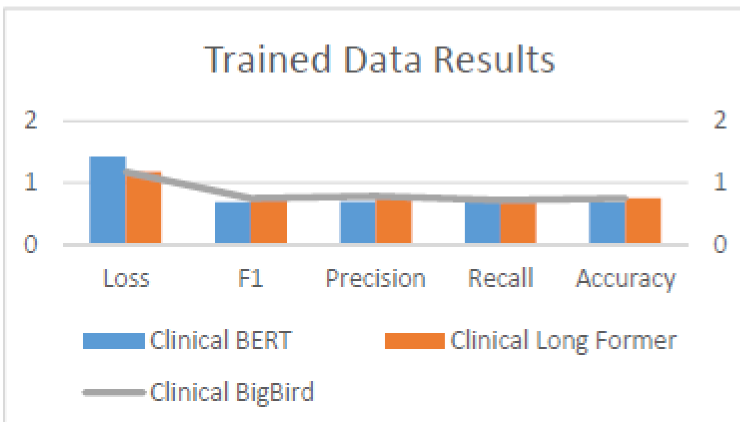

Fig. 3. Results on Trained Dataset

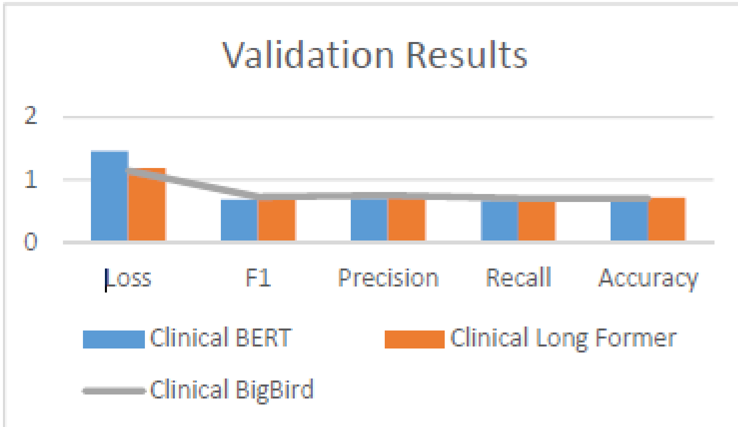

Fig 4: Test Dataset Outcomes

We can infer from Figs. 3 and 4 that while Clinical Big Bird and Clinical Longformer perform comparably, Clinical Bigbird took longer to train the model than Clinical Longformer. Before merging the results with the ABC algorithm, the models must be regularly exposed to newer datasets and undergo ongoing training and fine tuning [15,16,17].

CONCLUSION AND FUTURE WORK

In this work, we propose to use ABC to evaluate clinical models' performance when they interact with a real-time system that processes sequential clinical text. We are attempting to assess the coherence and context of these bio models. The purpose is to demonstrate that these bio models can be applied to chatbots relevant to Next Gen AI [7–10]. The best-performing model is chosen to be utilized for medical chatbot integration after the bio models are further refined and improved utilizing ABC. We also want to examine any bias or constraints present in these clinical models. Along with examining how well clinical models work in real-time with updates and medical knowledge, we will also look into how feasible it is to change these models to include the most recent medical information during chatbot conversations.

REFERENCES

1. Sonal Sharma, Sandeep Kumar, Anand Nayyar (2019) Logarithmic Spiral Based Local Search in Artificial Bee Colony Algorithm.
2. Sonal Sharma, Sandeep Kumar, Kavita Sharma (2018) Improved Gbest artificial bee colony algorithm for constraints optimization problems..
3. Harish Sharma, Sonal Sharma, Sandeep Kumar(2016) Lbest, Gbest artificial bee colony algorithm..
4. Kim SM, Hovy E (2004) Determining the sentiment of opinions. In: Proceedings of the 20th International Conference on Computational Linguistics, pp 1367–1373.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Kumar HK, Harish B (2018) A new feature selection method for sentiment analysis in short text. JIntell Syst 29(1):1122–1134.
7. Kuo RJ, Huang SL, Zulvia FE et al (2018) Artificial bee colony-based support vector machines with feature selection and parameter optimization for rule extraction. Knowl Inf Syst 55(1):253–274.
8. Lafferty JD, McCallum A, Pereira FC (2001) Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning
9. Lai CH, Liu DR, Lien KS (2021) A hybrid of XGBoost and aspect based review mining with attention neural network for user preference prediction. Int J Mach Learn Cybern 12(5):1203–1217
10. Li F, Han C, Huang M et al (2010) Structure-aware review mining and summarization. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp 653–661
11. Li H, Pun CM, Xu F et al (2021) A hybrid feature selection algorithm based on a discrete artificial bee colony for Parkinson’s diagnosis. ACM Trans Internet Technol 21(3):1–22

12. Li X, Bing L, Li P et al (2018) Aspect term extraction with history attention and selective transformation. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp 4194–4200
13. Liao M, Li J, Zhang H et al (2019) Coupling global and local context for unsupervised aspect extraction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp 4579–4589
14. Liu P, Joty S, Meng H (2015) Fine-grained opinion mining with recurrent neural networks and word embeddings. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp 1433–1443
15. Ma J, Cheng JC, Xu Z et al (2020) Identification of the most influential areas for air pollution control using XGBoost and grid importance rank. J Clean Prod 274(122):835
16. R. Kaladevi, S. Saidineesha, P. Keerthi Priya, K. M. Nithiya and S. Sai Gayatri, "Chatbot For Healthcare Using Machine Learning," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-4, doi: 10.1109/ICCCI56745.2023.10128261
17. Pavan Badempet, Prashanth Cheerala and Shiva Prasad Anagondi. A Healthcare System using Machine Learning Techniques for Disease Prediction with Chatbot Assistance. ScienceOpen Preprints. 2023. DOI: 10.14293/PR2199.000474.v1
18. M. Ahmed, H. U. Khan and E. U. Munir, "Conversational AI: An Explication of Few-Shot Learning Problem in Transformers-Based Chatbot Systems," in IEEE Transactions on Computational Social Systems, doi: 10.1109/TCSS.2023.3281492.keywords: {Chatbots; Artificial intelligence; Oral communication; Task analysis; Business; Transformers; Deep learning; Chatbot; conversational artificial intelligence (AI);dialog management; few-shot learning; question answering (QA);transformers},
19. N. P. Krishnam, A. Bora, R. S. V. R. Swathi, A. Gehlot, S. Talwar and T. Raghu, "AI-Based advanced Talk-chatbot for Implementation," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1808-1814, doi: 10.1109/ICACITE57410.2023.10182611.keywords:{Training;Schedules;Terminology;Chatbots;Prediction algorithms;Software;Pupils;bilingual English Arabic; Talkbot; conversational agent; academic counselling; NLP; deep learning},