Text Mining Strategies: RoBERTa Optimization for Efficient Pain Assessment in Hospice Care

YU-JU LIN[1], PI-SHAN HSU[2], BING-JUN CAI[3], TZUNG-PEI HONG[4], RUNG-CHING CHEN[5]
[1]Taichung Veterans General Hospital Taichung, city, TAIWAN
[2]Taichung Veterans General Hospital Taichung, city, TAIWAN
[3]Department of Information Management, Chaoyang University of Technology Taichung, TAIWAN
[4]Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung, TAIWAN
[5]Department of Information Management, Chaoyang University of Technology Taichung, TAIWAN

Abstract: —The hospice unit in medical care offers comprehensive, personalized care to patients, yet the recent epidemic and associated illnesses have strained medical resources, leading to a shortage in capacity. The necessity for frequent physiological documentation and patient assessments places a considerable burden on the nursing staff, particularly in the context of limited personnel. This study addresses this challenge by leveraging natural language processing to aid in the evaluation of pain indices, aiming to enhance implementation quality and reduce associated costs. Three BERT models—BERT, MacBERT, and RoBERTa were employed for training purposes. Among these models, RoBERTa demonstrated exceptional performance, achieving an impressive accuracy rate of 99%. This research highlights the potential of natural language processing tools, specifically the RoBERTa model, in alleviating the workload of nursing staff and improving the efficiency of pain assessment in hospice care during times of heightened demand and limited resources.

Key-words: —BERT, Machine Learning, Natural Language Processing, Transfer learning, Medical Language Processing.

Received: December 19, 2023. Revised: August 13, 2024. Accepted: September 17, 2024. Published: October 8, 2024.

1. Introduction

Since the onset of the COVID-19 pandemic in 2019, numerous industries across the globe have undergone digital transformations, and healthcare is no exception. The world is now placing increased emphasis on healthcare capacity, exploring ways to alleviate the workload of healthcare professionals through existing technologies, and seeking methods to provide more accurate insights into patients' conditions. One specific area of medical care that demands attention is "hospice care," characterized by proactive and personalized support for terminally ill patients who are unresponsive to curative treatments. Through meticulous recording of physiological data and routine measurements, healthcare practitioners gain a comprehensive understanding of the patient's condition, facilitating effective pain management and relief from other physical discomforts [1].

In times of epidemics, the healthcare workforce often faces staff shortages, amplifying the already significant burden on healthcare professionals dealing with a multitude of tasks arising from a surge in patients. This study addresses these challenges by employing a preprocessing method proven effective in recent years within the field of natural language processing. The pain index, assessed using a Visual Analog Scale (VAS) ranging from 0 to 10 (Figure 1), provides a nuanced measurement where higher values indicate more severe pain. The subsequent categorization simplifies the classification into four levels, aiding in the assessment of care plans.

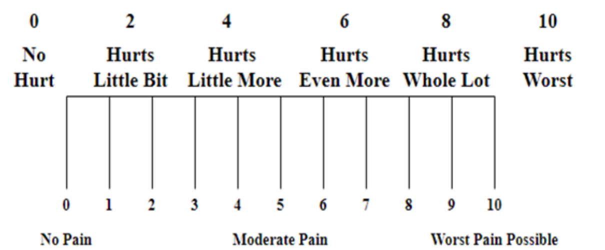

Fig.1. VAS Pain Classification

BERT (Bidirectional Encoder Representations from Transformers) plays a pivotal role in processing extensive physiological data and condition descriptions. Leveraging BERT for preprocessing, a well-trained model can predict relative assessment results by assimilating substantial physiological data and text [2, 3]. This application not only alleviates the burden on physicians in determining pain indices but also facilitates a more comprehensive understanding of a patient's condition, aiding in tailoring appropriate treatments [4]. The dataset undergoes preprocessing, with an 8:2 ratio division into training and test sets. Three models, namely BERT, Mac-BERT, and RoBERTa, will be employed for training, and their respective accuracies will be compared.

The significance of natural language processing in healthcare lies in its ability to extract pertinent information from medical texts and convert it into data rich in cru-cial medical insights. This transformative process enhances the quality of medical execution while concurrently reducing costs and addressing the challenge of staff shortages [5]. Once the model is constructed, it undergoes iterative training, testing, and validation with substantial datasets. The continual refinement throughout this process enhances the model's accuracy. Furthermore, the generated data facilitates diverse medical scientific studies, contributing to ongoing advancements in the field.

This research consists of five parts. Section 2 presents some related works. Section 3 explains our research methodology. Section 4 shows the experiment and results, and section 5 shows the conclusion and future works.

2. Related Works

The term "hospice care," as defined by the National Hospice Organization (NHO), provides support and care for patients in the final stages of a terminal illness so that they can live as comfortable a life as possible [1]. Hospice care recognizes the value of life and accepts death as a natural part of life. A major focus is on neither prolonging nor hastening the death of patients. Transfer learning is a technique for transferring knowledge from a pre-trained model to a new model, usually when the new task has limited annotated training data. Although insufficient to train an accurate model, supplementing a small dataset with a model pre-trained on richer data from similar tasks can surpass training solely on the limited data [6]. The trained weights reflect priors about related inputs learned from larger datasets, priming the model to better extract features even for limited data [7].

Masked language modeling to randomly mask some percentage of input tokens and then train the model to fill these masks correctly. For example, an original sentence "I like to eat meat" might become "I like to eat [Mask]" where [Mask] replaces "meat." Training on these corrupted inputs allows models to learn relationships between words and their contexts [8]. The percentage of masked tokens, such as 10%, controls the prediction difficulty. Combined with uncorrupted data, this teaches models to impute missing words and overcome noisy inputs.

Various clinical trial records in the medical field are usually unstructured text, so natural language processing (NLP) is used to extract information from a large number of documents and provide them to other experiments. Many methods have been developed in the medical field for parsing clinical trial texts, such as CT-BERT, which uses the ClinicalTrals.gov dataset for fine-tuning [9].

Name Entity Recognition (NER) is a research focus in natural language processing to identify and categorize the components of named entities in text. BERT pre-trained language model performs well on the task of NER. Contextual information [10].

RoBERTa is an improved version of the BERT model, which has many differences and improvements in the overall model size and performance compared with the original BERT. In addition, RoBERTa also has some differences in the training method. In the training task, it removes the "subsequence prediction" and uses dynamic masks, i.e., each time a sequence is an input to the model, a new mask pattern is generated.

MacBERT is a simple and effective model proposed by other scholars to check the effectiveness of BERT in non-English languages. In some aspects, it is an improvement of RoBERTa, especially in the way of MLM as a corrected text mask. Some research results show that MacBERT has a relatively excellent performance in most of the NLP tasks in Chinese.

Next Sentence Prediction. Fig.2. This prediction will specify two sentences in the text, and determine whether the second sentence in the text after the first sentence, can be seen as a reordering of the text paragraphs, the paragraphs of an article randomly disrupted, and this prediction, trying to restore the original text, need to have a certain understanding of the general meaning of the text.

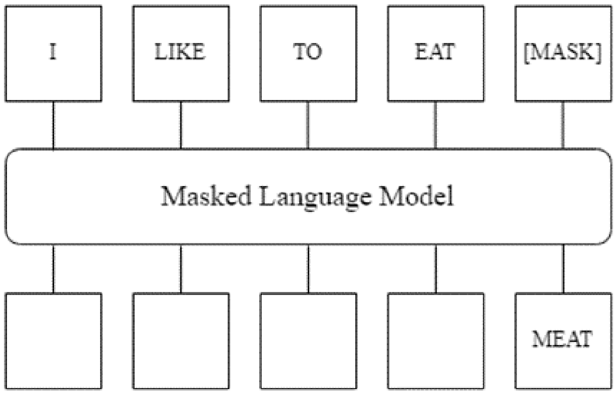

Fig.2. Masked Language Model

3. Methodology

In this study, we leverage data provided by Taichung Veterans General Hospital to employ the BERT model for analyzing medical orders. Specifically, we explore three distinct models: MacBERT, RoBERTa, and BERT.

Step 1: Data Import

The dataset from Taichung Veterans General Hospital is imported for analysis.

Step 2: Pre-processing

This involves filtering required fields, eliminating data with null values, and addressing the uneven distribution of pain indices. Pain indices 2 and 3 are merged into pain index 2, resulting in 8,261 data points split into training and test sets in an 8:2 ratio.

Step 3: Model Selection

We utilize three BERT models: BERT, RoBERTa, and MacBERT. RoBERTa, a modified BERT version, incorporates dynamic masks for adaptable recognition strategies during training. MacBERT, an improved RoBERTa, employs MLM masks, offering diverse results for this study.

Step 4: Model Training

This phase focuses on adjusting training parameters, equivalent to fine-tuning for BERT models, and training with labeled data.

Step 5: Model Evaluation

Evaluation is performed using the Confusion Matrix, comparing Accuracy, Precision, Recall, and F1 scores for the three models.

The dataset, detailed in Table 1, consists of 13,118 items with 8 columns. The experiment concentrates on classifying pain index levels) based on patient conditions, aiding healthcare providers in understanding and responding to patient needs.

This experiment focused on the classification of the pain index in the text, using the classification of the original data set, according to the patient's condition, to provide healthcare providers with information about the patient's status and to respond to it. The levels of care are Level 0 continuous care, Level 1 continuous monitoring, Level 2 review of the care plan, and Level 3 emergency response .

Table.1.Pain Levels Description

PainLevel	Description
0	Continuing Care
1	Continuous Monitoring
2	View Care Plans
3	Emergency Response

The text is a written description of the patient by a healthcare provider at a fixed time, which is used to rate the current condition in Table.2. The text of the patient's status related and physiological data records. The field contents were modified by the physician from the "str_others" field.

The study will focus on the "Pain" and "str_others_rm_default_phrase" columns."str_others_rm_default_phrase" is a description of the patient's current condition by the health care provider, while "pain" is an index given by the physician of the patient's status and the description of "str_others_rm_default_phrase," and is used to determine what the patient's condition is and whether adjustments to the current care plan are needed. The index determines the patient's condition and the need to adjust the current care plan.

Table.2. Data set text

Pain	str_others_rm_default_phrase
Continuous Monitoring(1)	The left chest Pig tail was connected to a sterile chest drainage bottle, and the pressure regulator on the wall was used to maintain -15cmH2O suction flow, and the drainage fluid was smooth and yellow.
Continuing Care(0)	The case left at 19:00 and did not return to the room
Continuing Care(0)	Assist SPA Nursing, patient comfort
View Plan (2)	The patient was drowsy and could still wake up after calling, but it was difficult to inhale and swallow, so Morphine sulfate tab 15mg at 0:00 am 1TAB PO Q6H was not taken.

The data set used in this study was provided by Taichung Veterans General Hospital, and the original data consisted of 13118 items. The number of valid data was 2256 for 0, 4785 for 1, 1072 for 2, and 148 for 3.

Before the model training, the data is processed first. This pre-processing method::

Check if there are missing values and delete them.

Delete the columns other than "str_others_rm_default_phrase" and "Pain" and change both of them to "Label" and "Review", Table 3.

When examining the data divisions, an imbalance was found, so the imbalanced divisions were merged. The pain index 2 is increased from 1072 to 1220, Table 4.

After the above steps, 80% of the data set will be randomly selected as the training set and 20% as the test set.

Table 3. Description of data columns (after pre-processing)

Column	Description
Label	Pain Level（0,1,2）
Review	Remarks after removal of invalid words

Table.4. Number of Label fields by the level of data (after merging)

PainLevel	Number of data
0	2256
1	4785
2	1220

The split between the training set and the test set is divided by 8261 data processed in a ratio of 8:2, with 6608 data in the test set and 1653 data in the test set.

Confusion Matrix is a visualization tool specifically designed for supervised learning. Each column of the matrix represents a class of instance predictions, and each row represents an actual class of instances.

In Table 5, TP (True Positive) is True Positive and predicts correctly; TN (True Negative) is True Negative and predicts correctly; FP (False Positive) is False Positive and predicts incorrectly; FN (False Negative) is False Negative and predicts incorrectly.

Table 5 Confusion Matrix

	prediction is true	predictions are false
The real situation is true	True Positive(TP)	False Negative(FN)
The real situation is false	False Positive(FP)	True Negative(TN)

After the confusion matrix is generated, various evaluation metrics can be calculated based on the TP, FP, FN, and TN in the matrix.

Accuracy is the most common metric used in classification questions to determine the merit of a model and indicates how many of the samples with positive predictions are correct.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \text{.............(1)}$$

Precision indicates the probability that the prediction is correct in the case of a positive result.

$$Precision = \frac{TP}{TP+FP} \text{........................(2)}$$

Recall indicates the probability of a positive prediction from all positive true data.

$$Recall = \frac{TP}{TP+FN} \text{...........................(3)}$$

F1-score is the summed average of the accuracy and recall rates.

$$F1 = \frac{2*(Precision*Recall)}{Precision+Recall} \quad \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

4. Experiments

4.1 Experimental Environment

The experiment was conducted by the Google Colab environment, which has the following features compared to other editing environments: Various machine learning and deep learning frameworks can be used, and different suites can be imported according to requirements. Free to use, and do not have to install other software, you can directly in the browser to edit. Google Colab is stored on Google Drive by default and can be concatenated when editing, allowing users to manage files in the same environment.

In Table 6, the parameters of the three models were adjusted to be the same, so the differences could be compared.

Table 6. Parameter Setting

train epoch	2
batch size	32
max length	128
datasets	8261
learning rate	0.000003

4.2 Bert Model Training and Results

This study was conducted on Colab Notebooks, using the text pre-training technique, BERT, proposed by Google, and Sklearn in the python suite to help us train the models. The total number of datasets used in this experiment is 8261, and the models used are all from the Hugging Face website, using the models "Bert-base-Chinese", "Roberta-base" and "Chinese-macbert-base".

Table.7. According to the model training results, the accuracy of BERT is 97%, the accuracy of MacBERT is 98%, and finally, the accuracy of RoBERTa is 99%.

Table.7. Model Training Results (Accuracy)

Model	Accuracy
RoBERTa	99%
MacBERT	98%
BERT	97%

The model evaluation is assisted by sklearn's matrics function which we calculate, in Table.8. The evaluation scores of the three models are very different; RoBERTa stands out among them, probably because the model can be adjusted to accommodate larger batch sizes, and the dynamic masking in the model is also effective in this type of text, thus producing superior results.

Table.8. Model Testing

	BERT	RoBERTa	MacBERT
Accuracy	97%	99%	98%
Precision	97%	98%	98%
Recall	95%	98%	97%
F1	95%	98%	97%

5. Conclusion

With a relative imbalance in the ratio of the three data categories, the future availability of more data is expected to yield more balanced results. The study concluded that the model achieved 99% accuracy in predicting pain level 2, with specific challenges in distinguishing other levels due to smaller sample sizes and contextual similarities. Future experiments should explore different textual content and recognition strategies to address these challenges. The successful prediction of pain information in medical text suggests the potential to reduce manual assessments, alleviating the burden on healthcare workers. Subsequent studies will expand variables and data volume, including blood pressure, heart rate, respiratory rate, or body temperature. Integrating data prediction and textual modeling aims to offer a more comprehensive understanding of the patient's condition, aiding healthcare providers in making informed decisions and providing better medical care.

Acknowledgment. This paper is supported by the Ministry of Science and Technology, Taiwan. The Nos are MOST-111-2221-E-324-020 and MOST-111-2622-E-324 -002.

References

[1]. Lee, Moen. "An exploration of the current status of regional hospice and palliative care service model in Taiwan - A case study of Kaohsiung area." Master's thesis, Department of Social Welfare, National Chung Cheng University, Chiayi County, Taiwan, (2003).

[2]. J. Devlin, M.W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, pages 4171–4186, 2019.

[3]. B. Cai, P. Hsu, Y. Lin, R. Chen, K. You, Predicting pain severity category by vital signs in hospice ward: An application of Palliative Care Outcomes Collaboration (PCOC) Pain scale measurement, ICMHI '23: Proceedings of the 2023 7th International Conference on Medical and Health InformaticsMay 2023, pp. 72–76, https://doi.org/10.1145/3608298.3608313

[4]. R. Tang et al., "Embedding Electronic Health Records to Learn BERT-based Models for Diagnostic Decision Support," 2021 IEEE 9th International Conference on Healthcare Informatics (ICHI), pp. 311-319, 2021, doi: 10.1109/ICHI52183.2021.00055.

[5]. N. Liu, Q. Hu, H. Xu, X. Xu, and M. Chen, "Med-BERT: A Pretraining Framework for Medical Records Named Entity Recognition," in IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5600-5608, 2022, doi: 10.1109/TII.2021.3131180.

[6]. S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2010, doi: 10.1109/TKDE.2009.191.

[7]. H. Liang, W. Fu, and F. Yi, "A Survey of Recent Advances in Transfer Learning," 2019 IEEE 19th International Conference on Communication Technology (ICCT), pp. 1516-1523, 2019, doi: 10.1109/ICCT46805.2019.8947072.

[8]. 4. Experiments 4.1 Experimental Environment 4.2 Bert Model Training and Results 5. Conclusion References

[9]. Y. Zhao, R. Cao, J. Bai, W. Ma, and H. Shinnou, "Determining the Logical Relation between Two Sentences by Using the Masked Language Model of BERT," 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), pp. 228-231, 2020, doi: 10.1109/TAAI51410.2020.00049.

[10]. X. Liu, G. L. Hersch, I. Khalil, and M. Devarakonda, "Clinical Trial Information Extraction with BERT," 2021 IEEE 9th

[11]. International Conference on Healthcare Informatics (ICHI), pp. 505-506, 2021, doi: 10.1109/ICHI52183.2021.00092.

[12]. W. Zhang, S. Jiang, S. Zhao, K. Hou, Y. Liu, and L. Zhang, "A BERT-BiLSTM-CRF Model for Chinese Electronic Medical Records Named Entity Recognition," 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA), pp. 166-169, 2019, doi: 10.1109/ICICTA49267.2019.00043.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.