

USING VITAL SIGNS AND NURSING ASSESSMENT DOCUMENT FOR CLINICAL PAIN SCALE APPLICATION WITH ENSEMBLE NATURAL LANGUAGE PROCESSING AND MACHINE LEARNING

YU-JU LIN¹, PI-SHAN HSU¹, BING-JUN CAI², RUNG-CHING CHEN^{2,*}
AND WILLIAM ERIC MANONGGA²

¹Department of Family Medicine
Taichung Veterans General Hospital
No. 1650, Taiwan Boulevard Sect. 4, Taichung 40705, Taiwan
{ nps798; pshsu3 }@gmail.com

²Department of Information Management
Chaoyang University of Technology
No. 168, Jifeng East Road, Wufeng District, Taichung 413310, Taiwan
s11014617@gm.cyut.edu.tw; s11014907@cyut.edu.tw

*Corresponding author: crching@cyut.edu.tw

Received October 2023; revised February 2024

ABSTRACT. *Managing hospital resources is complex due to the imbalance between the number of patients, medical doctors, and nurses. Patients' care programs are assigned by medical doctors, who base their decisions on nursing assessments and vital signs data. However, there is a significant time gap between data collection and decision-making. This paper proposes a method to measure a patient's pain scale using nursing assessment data (text) and vital signs information (numerical). Using our proposed method, the pain scale of each patient can be measured immediately. We propose three models to predict pain scale depending on available data. The first model uses pre-trained language models such as BERT (Bidirectional Encoder Representations from Transformers), roBERTa (Robustly Optimized BERT-Pretraining Approach), and MacBERT (Masked Language Modeling Correction BERT) to encode nursing assessment data into numerical data and predict the pain scale. We use vital signs information for the second model and perform a machine learning model search with different algorithms such as decision tree, random forest, logistic regression, and support vector machine. Finally, we employ ensemble learning to combine the predictions from the previous two models. The results from the ensemble learning model achieve pain scale estimation with an accuracy of 99.45%.*

Keywords: Machine learning, Natural language processing, BERT, Classification model, Vital signs

1. Introduction. Insufficient resources for hospital staff is a common issue in many countries. This highlights the importance of management for inpatient care programs. Targeted care programs are essential for critically ill inpatients to avoid suffering and an increased mortality risk. However, when a patient recovers and becomes stable, the care level should be reduced to ease the workload for hospital staff. Medical doctors use patient information and standard data collection to make informed decisions.

The Palliative Care Outcomes Collaborative [1], PCOC, is a clinical assessment tool used to measure palliative care outcomes and quality. It was developed and established by healthcare professionals and researchers in Australia. The tool includes the Palliative Care Problem Severity Scale (PCPSS) and the Assessment of Patient Functional Status

Scale (AKPS), which help assess pain, other symptoms, and psychological/psychiatric problems.

Many hospitals and clinics in Taiwan have adopted the PCOC protocol, with nurses primarily responsible for recording patient data. Patient information is gathered, including Body Temperature (BT), Pulse Rate (PR), Respiratory Rate (RR), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Mean Arterial Pressure (MAP), and nursing assessments in text format. Nurses record numerical data (BT, PR, RR, SBP, DBP, and MAP) three times a day. Skilled nurses also perform nursing assessments in text format once a day. A medical doctor will then review all the collected information and make decisions based on that.

According to a study in Taiwan [2], it was observed that the ratio of care-to-patient is unbalanced. To alleviate the burden on the medical staff and medical doctors, we propose the implementation of technology. The main motivation of our research is to reduce the workload of medical doctors and expedite the process of assigning care levels to each patient. Additionally, it can help ward staff improve their precision by providing them with real-time updates on patient pain scales and care levels.

To validate our proposed model, we will conduct a three-step experiment. The details of each step are outlined below.

- 1) We use pre-trained language models: BERT (Bidirectional Encoder Representations from Transformers) [3], roBERTa (Robustly Optimized BERT-Pretraining Approach) [4], and MacBERT (Masked Language Modeling Correction BERT) [5], to encode nursing assessment text and make pain scale predictions.

- 2) We comprehensively search for the most suitable machine learning model to predict the pain scale using patient information such as BT, PR, RR, SBP, DBP, and MAP.

- 3) We utilize ensemble learning with the random forests model to aggregate the prediction outputs and generate the final pain scale prediction.

The contribution of our research is as follows: 1) We assessed the effectiveness of different pretrained BERT models in processing medical texts; 2) We employed different ML models to predict the patient's pain scale by using the vital signs data; 3) We proposed a novel approach to classify the patient's pain scale by combining NLP, ML, and ensemble techniques using text and also numeric data.

The rest of the paper is organized as follows. First, we provide an overview of the related literature in Section 2. Then, in Section 3, we explain the methodology used, including the data collection process. Section 4 presents the experiment and discusses the results. Finally, we summarize the contributions of our study and suggest potential directions for future research in Section 5.

2. Related Work. Patient information is critical in healthcare as it plays a fundamental role in treatment decisions and improving patient outcomes. Traditional methods of estimating pain scales rely on human observation, review, and reporting. However, a growing interest is in using machine learning techniques to enhance pain scale estimation. With the help of machine learning algorithms, it is possible to automate the process of pain scale estimation. This section comprehensively reviews the current literature on using machine learning algorithms and artificial intelligence applications in pain scale estimation.

Studies, such as in [6,7], have utilized electronic health information to identify patients suffering from chronic pain. By using available data elements such as pain diagnostic codes (International Classification of Disease, revision 9; ICD-9), patient-reported pain scores, and opioid prescription medications, these studies estimated pain scales. The accuracy of the algorithm's results was validated through medical chart reviews.

Various machine learning algorithms, including decision tree, random forest, Logistic Regression (LR), and Support Vector Machine (SVM), are utilized to classify pain scales based on physiological signals. In [7], a combination of tree-based models and other machine learning techniques, such as BiLSTM [8], was used to extract features and generate output for XGBoost [9], resulting in high accuracy in predicting pain scales.

Using an SVM and K-Nearest Neighbor (KNN) to classify pain scales is proved in research such as [10-12]. Machine learning techniques, such as SVM, can classify the pain scale with using two separate feature selection strategies (univariate feature selection and sequential forward selection) during the feature extraction [13]. Similarly, using an SVM model, [14] identified low back pain based on features extracted from motion sensor data.

Patient chart data is not the only information that can be used to estimate pain scale. A facial image and facial expression are valid indicators of pain [15,16], and they appear in the observational scales for pain assessment. Deep learning approach using convolutional neural networks and facial images was used to estimate the pain scale [17]. The facial expression approach is able to obtain high accuracy in predicting pain scales. However, it is computationally expensive and unsuitable for hospitals with many patients. Moreover, using the facial picture of the patient can raise some privacy concern issues.

Nursing assessment text or clinical note requires an expert to review and label the pain scale for patients. However, text or documents cannot be processed directly using machine learning. NLP is required to process the text or documents before being processed by ML [18].

Transformers [19] are neural network architectures designed to work with text data. It is used as a base architecture for Large Language Models (LLM) and Generative Pre-trained Transformers (GPT) [20]. Text classification uses a transformer encoder to encode text data into context represent vectors. The encoded vector is used as an input to the classification model.

Bidirectional Encoder Representations from Transformers (BERT) [3] is a powerful natural language processing model developed by Google. It uses a transformer architecture to process and understand the context of words in the text. BERT's pre-training and fine-tuning approach helps it to learn language representations and achieve state-of-the-art performance on various NLP tasks. By considering both preceding and following words, BERT captures comprehensive contextual information. Google provides different pre-trained versions of BERT, such as BERT-base and BERT-large. However, fine-tuning with a specific dataset is required for particular tasks or languages. roBERTa [4] applies a new strategy to optimizing BERT during training, including longer training time, larger batch size, and dynamic masks. BERT is also able to process text in languages other than English, which is done in many researches. Richardson and Wicaksana [21] compared IndoBERT-lite, built based on the ALBERT (A Lite BERT) architecture, and roBERTA in text mining for Indonesian language question answering application. Nicolae et al. [22] proposed ALR-BERT, a lite Romanian BERT trained on a large corpus solely in Romanian language that scales better than the original Romanian BERT. Alawadh et al. [23] used an Attention-Enriched Mini-BERT to analyze fake news in the Arabic language.

The difficulty to process Chinese text in NLP is significantly higher compared to English text because there are many types of Chinese morphemes, meanings, and complex combinations of words. MacBERT [5] is a transformer-based for Chinese NLP, which proposed the Whole Word Masking (WWM) method to process Chinese vocabulary more accurately. Compared with the original BERT masking method, WWM masks the whole Chinese words to capture the relationship between words better, preserve the complete contextual information, and improve the model's prediction accuracy. In some research, BERT has been used to classify news in Chinese, such as text classification [21,22], tagging

in Chinese judicial documents [26], and government policy plan texts [27]. The accuracy of text classification is further increased by combining it with other neural networks.

In practice, estimating the pain scale requires the patient information from a medical device (e.g., Vital Signs) and a description from observation (e.g., nursing assessment text). Existing research used only measured values or clinical notes to predict pain scales. Lin et al. [28] used machine learning techniques to predict a patient’s pain level from the recorded vital signs. In our research, we consider using “Vital Signs” and “Nursing assessment text” as the input for our model. We proposed to create a model to manage each data type, then use ensemble learning algorithm [29] to combine and make a final prediction.

3. Research Methodology. Traditional pain scale is decided by a doctor to understand patient’s pain and make a better treatment. In this research, our research team proposed a novel method to predict the patient’s pain scale automatically. The method used the text processing by BERT system for natural language processing to label pain scale automatically firstly. Next, we put vital signs to machine learning model to predict the pain scale. Finally, we ensemble the output of BERT system and vital pain scale. The detail of operation includes 1) the dataset setup; 2) data preparation and preprocessing for the collected data; 3) nursing assessment text classification for BERT system; 4) vital signs classification model generation; 5) Pain Scale Classifier Model (PSC-Model) uses ensemble learning which combines BERT system output and vital signs information. We describe the operations as the following.

3.1. Dataset. We collected data from patients admitted to the Taichung Veterans General Hospital hospice ward between July 2021 and June 2022. All identifiable patient information is removed to maintain privacy and adhere to ethical standards. The dataset is presented in two parts: nursing assessment text data and vital signs data, both in a tabular format.

The first dataset, the nursing assessment text dataset, contains the record date, nursing assessment text and the pain scale. The pain scale value is left blank until a medical doctor reviews and evaluates it. The nurse uses a mix of English and Chinese to fill out the nursing assessment text. English is mostly used for technical terms or medicine names. The dataset includes a pain scale that a licensed medical professional has carefully reviewed and completed. The scale ranges from 0 to 3, where 0 indicates the need for “Continuing Care”, 1 suggests the need for “Continuous Monitoring”, 2 suggests the need to “View Care Plan”, and 3 indicates the need for “Emergency Handling”.

The vital signs dataset contains standard information collected using medical equipment such as thermometers and heart rate monitors. It encompasses all the necessary information a nurse must collect during a routine checkup such as Body Temperature (BT), Pulse Rate (PR), Respiratory Rate (RR), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), and Mean Arterial Pressure (MAP).

3.2. Data preparation and preprocessing. With the nursing assessment text dataset, we can use NLP and Text Classification models to classify pain scales. However, the vital signs dataset, which has important information, needs to have the pain scale information. To address the requirement, the nursing assessment text and vital signs data are merged to become two new datasets. The new datasets have all the necessary information to train our models.

The merged dataset 1 is the combination of vital signs data and the pain scale information obtained from the nursing assessment text for the same patient and date. The pain scale added to the merged dataset 1 is rescaled to range between 0 and 2. Pain scales 2

and 3 are merged together to become pain scale 2. The reason is because the action taken for pain scales 2 and 3 are found to be the same.

The merged dataset 2 merged the nursing assessment text dataset with the vital signs dataset. We map the value of BT, PR, RR, SBP, DBP, and MAP for the same patient and day. The pain scale in merged dataset 2 is also rescaled to the range from 0 to 2 to keep the label consistent among datasets. The new merged dataset will have the same number of data with nursing assessment text dataset.

Table 1 shows the number of records and the distribution of data in the datasets. We split the dataset into the training and testing set with the ratio of 80 : 20. The splitting was randomly and discussed the dataset with doctor. After splitting the dataset into the training and testing set, the nursing assessment text dataset and the merged dataset 1 have 6,528 data in the training set and 1,633 data in the test set. Merged dataset 2 has 10,494 data for training and 2,624 data for testing.

TABLE 1. Pain scales label distribution in each dataset

Pain scale	Nursing assessment text dataset	Merged dataset 1	Merged dataset 2
0	2,256	3,879	2,256
1	4,785	7,584	4,785
2	1,120	1,655	1,120
Total	8,161	13,118	8,161

3.3. Nursing assessment text classification (NTC-BERT). BERT [3] (Bidirectional Encoder Representations from Transformers) is a cutting-edge machine learning model used for Natural Language Processing (NLP) tasks. It has been pre-trained in 104 languages, including Chinese, using Masked Language Modeling (MLM) on the Wikipedia website. The mentioned training method has enabled BERT to understand the context of documents and encode text documents into a vector.

Our method used mixed language from the Chinese and English languages. The difference between Chinese and English is the word tokenizers, or how they break down the text into individual units or tokens. Although BERT can be fine-tuned for specific tasks, our study did not involve such fine-tuning. Instead, we experimented with various pre-trained versions of BERT, including roBERTa [4] and MacBERT [5], which differ from BERT in their pre-training techniques. As shown in Figure 1, NTC-BERT is a combination of a BERT followed by a linear layer which acts as a classifier. NTC-BERT is trained using the nursing assessment text data. BERT will encode the text data, and the output will go through a fully connected layer to predict the pain scales.

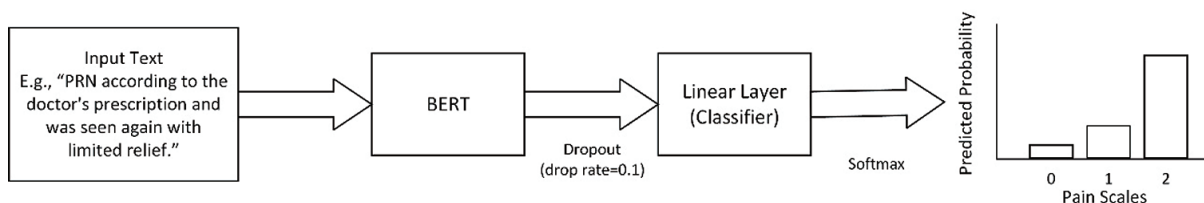


FIGURE 1. The overview of NTC-BERT text classification

3.4. Vital Signs Classification Model (VSC-Model). To select an appropriate machine learning algorithm that can classify pain levels based on data on vital signs. To achieve the objective, we have explored various algorithms, conducting experiments and do hyperparameter tuning to optimize their performance. After testing each model, we analyzed their implementations using the Model Search (MS) framework. Our method allowed us to optimize each model’s capabilities and ultimately reach the best performance. The six indices of features are Body Temperature (BT), Pulse Rate (PR), Respiratory Rate (RR), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), and Mean Arterial Pressure (MAP) which are selected by two doctors based on their domain knowledge.

Figure 2 presents an overview of the process involved in training the VSC-Model. We conducted a model search to classify vital signs data by experimenting with four classification models: decision tree, random forest, logistic regression, and Support Vector Machine (SVM). Each model was trained with specific hyperparameters using the training dataset and evaluated based on the accuracy of the testing dataset. The VSC-Model was trained using the merged dataset 1 which is the combination of the vital sign data and the pain scale labels from the nursing assessment dataset.

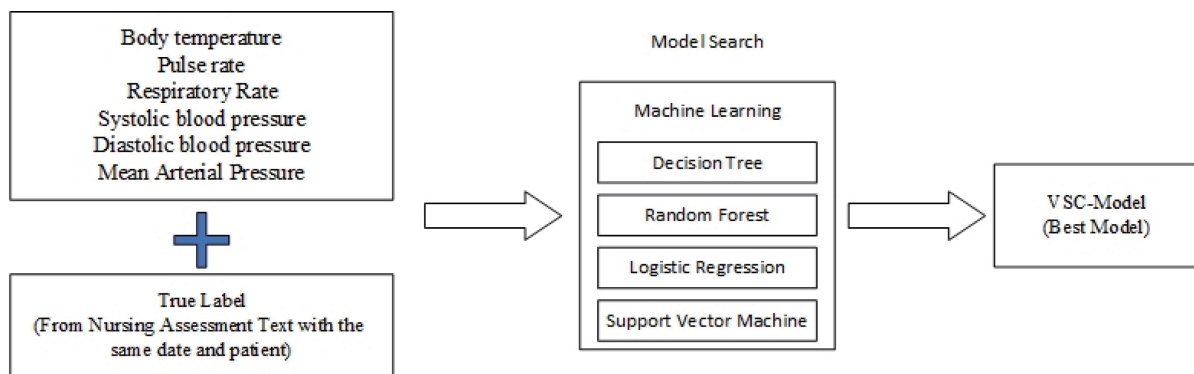


FIGURE 2. The overview of the VSC-Model training process

3.5. Pain Scale Classifier Model (PSC-Model) using ensemble learning. The PSC-Model is the improvement to the two previous models. We trained the two previous models (NTC-BERT VSC-Model) using the merged dataset 2. As already explained in previous Sections 3.3 and 3.4, the NTC-BERT will work using the text data while VSC-Model will predict the pain scale using the numerical data. The data used on both models are from the same patient.

The prediction results from each model are combined using ensemble methods. The ensemble method combines multiple learning models to make better predictions compared to a single model. The proposed combined prediction model is called the Pain Scale Classifier Model (PSC-Model). The PSC-Model model only works when nursing assessment text and vital signs data are available since the outputs from NTC-BERT and VSC-Model are needed in the ensemble method. Random forests algorithm is selected as an ensemble algorithm to combine NTC-BERT and VSC-Model. The outcome from the random forest is used as the PSC-Model’s prediction. The PSC-Model is then evaluated using the test set to verify its performance.

3.6. Evaluation metrics. Evaluation metrics play a crucial role in machine learning by providing a quantitative measure of the models performance. We use a range of evaluation metrics, such as accuracy, precision, recall, and F1 score, to validate the prediction results

of our models. These multiple evaluation metrics allow us to assess the performance of the models from different perspectives, ensuring that we have a comprehensive understanding of how well they perform.

Our models are designed to predict three levels of pain scales. The evaluation metrics begin by calculating the values for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP represents the number of instances correctly predicted to a specific class. In contrast, TN represents the number of instances correctly predicted as not belonging to a specific class. FP represents the number of instances incorrectly predicted to a particular class. Finally, FN represents the number of instances incorrectly predicted as not belonging to a specific class.

Accuracy (Equation (1)) represents the proportion of correctly classified instances out of the total number. It is commonly used to evaluate the performance of classification problems. However, our dataset has a slight class imbalance (the sample from the pain scale equal is $> 50\%$ from the dataset). The accuracy might only reflect in some classes. We use Precision (Equation (2)) and Recall (Equation (3)) to validate the balance of the trade-off between FP and FN, and F1-score (Equation (4)) averages Precision and Recall.

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

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

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

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

To summarize our methodology, we proposed three classification models to predict pain scale, which can be used in different environments for newly admitted patients with vital signs data from basic checking. VSC-Model can be used to predict the pain scale. After a skilled nurse checks and makes a note for the patient, we proposed NTC-BERT to predict pain scale from the nursing assessment text. Finally, PSC-Model is the improvement to the first two models. It incorporates ensemble learning to make the final pain scale prediction when both nursing assessment text and vital signs data are available.

4. Experimental Result and Analysis. Our experiment was divided into three parts. The first part involved the NTC-BERT model, where different versions of BERT are tested to encode nursing assessment text and classify it into pain scales. The second part involved the VSC-Model, where we performed machine learning model search to find the best model for classifying pain scale based on a patient's vital signs. In the last experiment, NTC-BERT and VSC-Model are combined using random forests as an ensemble learning algorithm to improve the results from the two models and make the final prediction.

4.1. Nursing assessment text classifier BERT (NTC-BERT). In our experiment, we employed various versions of the pre-trained BERT model provided by Hugging Face. Table 2 presents the NTC-BERT results using different BERT versions for text encoding (the best results are shown in bold). After combining with a classification layer, the model using BERT-base demonstrates superior performance in terms of accuracy, recall, and F1-score. roBERTa performs slightly worse compared to BERT-base, while MacBERT yields the least favorable results.

Taking BERT-base as the best model for NTC-BERT, Table 3 shows the precision, recall and F1-score for each category of pain scale. From the results, it is shown that

TABLE 2. The performance of NTC-BERT with the test dataset using the different encoding

BERT version	Accuracy	Precision	Recall	F1-score
BERT-base [3]	82.74%	80.23%	79.20%	79.68%
roBERTa [4]	81.15%	78.43%	78.39%	78.36%
MacBERT [5]	75.83%	75.9%	75.83%	75.76%

TABLE 3. Performance of NTC-BERT using BERT-base as encoding in each class

Pain scale	Precision	Recall	F1-score
0	81.06%	76.30%	78.61%
1	85.41%	88.07%	86.72%
2	74.22%	73.24%	73.73%

the highest results are found in pain scale 1 with 85.41% precision, 88.07% recall, and 86.72% F1-score. The next best result is found in pain scale 0, while pain scale category two being the worst. As patients in pain scale category two require immediate attention and emergency handling as their condition is critical, a precision of 74.22% and recall rate of only 73.24% is not good enough, which shows that many patients have the potential to be misclassified and will not receive the necessary treatment.

4.2. Vital Signs Classification Model (VSC-Model). In our experiment, a model search using the training set was done on various machine learning models. The results of the experiments for VSC-Model are presented in Table 4 (the best results are shown in bold). The results from the experiment indicate that random forest outperforms other models in three evaluation metrics: accuracy, precision, and F1-score, while being the second best in terms of recall score. The result from the decision tree algorithm is slightly worse than the random forest, while logistic regression and SVM both give low results and are not good enough to classify the pain scale from vital sign data.

TABLE 4. The performance of the machine learning model with the test dataset

Model	Accuracy	Precision	Recall	F1-score
Decision Tree	95.84%	95.85%	95.84%	95.83%
Random Forest	96.72%	96.97%	94.89%	95.88%
Logistic Regression	72.23%	66.01%	72.03%	65.11%
SVM	72.00%	61.71%	72.00%	61.52%

Table 5 presents the performance of the VSC-Model using random forest on predicting each pain scale. The model shows high overall performance for all three pain scales. However, the model shows lower recall score for a pain scale 2. The result indicates that the model tends to misclassify patients with pain scale 2 and can lead to wrong treatment by giving regular care programs to severely ill patients.

4.3. Pain Scale Classifier Model (PSC-Model). From the previous two experiments, we got the best configuration for NTC-BERT and VSC-Model. In the third part of the experiment, we combine the two models using random forest for the ensemble method. The output from the NTC-BERT and VSC-Model will be processed by the random forest to get the final pain scale prediction for the patient.

TABLE 5. The performance VSC-Model using random forest in each class

Pain scale	Precision	Recall	F1-score
0	96.67%	96.12%	96.43%
1	96.53%	98.46%	97.49%
2	97.65%	90.09%	93.71%

Table 6 shows the performance comparison of the three models (the best result is shown in bold). The PSC-Model gives a very good performance with 99.45% accuracy, 99.54% precision, 99.16% recall, and 99.35% F1-score. The result from the experiment shows the superior performance of the PSC-Model, which uses ensemble method, compared to the two separate models.

TABLE 6. Performance comparison between NTC-BERT, VSC-Model, and PSC-Model

Model	Accuracy	Precision	Recall	F1-score
NTC-BERT	82.74%	80.23%	79.20%	79.68%
VSC-Model	96.72%	96.97%	94.89%	95.88%
PSC-Model	99.45%	99.54%	99.16%	99.35%

The prediction performance of PSC-Model for each separate class is presented in Table 7. The proposed method shows high performance for each class, which is close to 100% for pain scales 0 and 1, and only slightly lower for pain scale 2.

TABLE 7. The performance PSC-Model in each class

Pain scale	Precision	Recall	F1-score
0	99.78%	99.34%	99.56%
1	99.26%	99.78%	99.52%
2	99.58%	98.36%	98.97%

4.4. Result discussion. After knowing the results of the experiment, in this section we will discuss about two important findings. The first is about the lower performance when predicting a pain scale 2 patient. The second is on how our proposed PSC-Model can improve the prediction performance compared to the separated model.

From all three experiments using NTC-BERT, VSC-Model, and PSC-Model, the results always show lower performance when detecting patient with pain scale 2. The result could be attributed to the lower data available for pain scale 2. In the merged dataset 1 which is used in VSC-Model and PSC-Model, only 1,655 data of pain scale 2 are available from a total of 13,188 records. While in the merged dataset 2 which is used to train the NTC-BERT, only 1,120 records are of pain scale 2 from total 8,161 records. The low number is also already a merged result from the pain scales 2 and 3 in the original dataset. The low performance shows that while pain scales 2 and 3 from the original dataset might correspond to the same treatment action from the medical staffs, the features are not the same, which makes it difficult for the model to learn the pattern to make correct predictions. To improve the result, more data need to be collected for pain scale 2. It is even better if the dataset has enough data to train the model to predict the original pain scales 2 and 3, and let the medical professionals decide what actions to do for the patient.

The results of the experiments display how the ensemble method can improve the pain scale prediction performance. Table 8 presents the prediction performance comparison on

TABLE 8. The comparison of each class prediction performance for NTC-BERT, VSC-Model, and PSC-Model

Pain scale	Model	Precision	Recall	F1-score
0	NTC-BERT	81.06%	76.30%	78.61%
	VSC-Model	96.67%	96.12%	96.43%
	PSC-Model	99.78%	99.34%	99.56%
1	NTC-BERT	85.41%	88.07%	86.72%
	VSC-Model	96.53%	98.46%	97.49%
	PSC-Model	99.26%	99.78%	99.52%
2	NTC-BERT	74.22%	73.24%	73.73%
	VSC-Model	97.65%	90.09%	93.71%
	PSC-Model	99.58%	98.36%	98.97%

each pain scale for each model. By using ensemble methods, our proposed PSC-Model gives better prediction accuracy than the separated NTC-BERT and VSC-Model, which is more significant in the prediction of pain scale 2, especially for the NTC-BERT, which might misclassify the pain scale of the patients, leading to wrong treatment. The only downside of the PSC-model is the need of both nursing assessment text and vital sign data before being able to be used.

5. Conclusions. Our study uses patient data from Taichung Veterans General Hospital in Taiwan to create the PSC-Model model for predicting pain scale. From the experiment, PSC-Model can achieve a very high score, near 100%, in predicting the pain scale from patient information. The dataset used for this study included mixed patient information, which comprised nursing assessment text and vital signs data.

To handle each data type efficiently, we devised three classification models: NTC-BERT, a text classification model, was employed to predict pain scale from nursing assessment text; VSC-Model was used to predict pain scale from vital signs information; and the PSC-Model was utilized as an ensemble learning algorithm to combine the predictions from the previous two models and generate the final prediction. Even with lower performance, both NTC-BERT and VSC-Model can accelerate the process of reviewing and categorizing pain scales for a patient. Skilled nurses can gather patient information and the system can assign a pain scale accordingly based on the available data. When a more comprehensive data is already collected, the PSC-Model can assign a more accurate pain scale for the patient.

Our high-performance models enable us to predict pain scales immediately after data collection, which can be used to assign different care programs to patients. The proposed method will reduce the workload of hospital staff by speeding up the patient information review process and prioritizing patient care programs effectively.

Moreover, our study demonstrates the potential of using mixed data types and ensemble methods to create a better classification model in medical applications, especially with text or documents requiring specialists to review and analyze. In our future work, we plan to increase the number of data to ensure better prediction accuracy. We also plan to experiment with newer technology, such as the GPT Model, and create a language model that can understand the context of nursing assessment or clinical notes, and recommend the response action in human language.

Acknowledgment. This paper is supported by Taichung Veterans General Hospital, Taiwan and NTSC Project No. NTSC-112-2221-E-324-003-MY3, NTSC-112-2221-E-324-011-MY2, Taiwan.

Institutional Review Board Statement. The research was approved by the Institutional Review Board (IRB) of Taichung Veterans General Hospital (code TCVGH-IRB CE22441B) (Date of approval: November 2, 2022) and was conducted according to the Declaration of Helsinki.

REFERENCES

- [1] Palliative Care Outcomes Collaborative (PCOC), *Clinical Protocol – University of Wollongong*, <https://www.uow.edu.au/ahsri/pcoc/palliative-care/clinical-protocol/>, Accessed on Sep. 27, 2023.
- [2] C.-M. Chang and L. Hsien-Feng, A study on the turnover behavior factors and workload of nursing staff, *Chang Gung Nurs.*, vol.32, no.1, pp.18-31, DOI: 10.6386/CGN.202103.32(1).0002, 2021.
- [3] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, MN, pp.4171-4186, DOI: 10.18653/v1/N19-1423, 2019.
- [4] Z. Liu, W. Lin, Y. Shi and J. Zhao, A robustly optimized bert pre-training approach with post-training, *Proc. of the 20th Chinese National Conference on Computational Linguistics (CCL2021)*, 2021.
- [5] Y. Cui, W. Che, T. Liu, B. Qin and Z. Yang, Pre-training with whole word masking for Chinese BERT, *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol.29, DOI: 10.1109/TASLP.2021.3124365, 2021.
- [6] T. Y. Tian, I. Zlateva and D. R. Anderson, Using electronic health records data to identify patients with chronic pain in a primary care setting, *J. Am. Med. Informatics Assoc.*, vol.20, no.E2, DOI: 10.1136/amiajnl-2013-001856, 2013.
- [7] F. Pouroumran, Y. Lin and S. Kamarthi, Personalized deep Bi-LSTM RNN based model for pain intensity classification using EDA signal, *Sensors*, vol.22, no.21, DOI: 10.3390/s22218087, 2022.
- [8] Z. Huang, W. Xu and K. Yu, Bidirectional LSTM-CRF models for sequence tagging, *arXiv Preprint*, arXiv: 1508.01991, 2015.
- [9] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, pp.785-794, DOI: 10.1145/2939672.2939785, 2016.
- [10] R. Cao, S. A. H. Aqajari, E. K. Naeini and A. M. Rahmani, Objective pain assessment using wrist-based PPG signals: A respiratory rate based method, *Proc. of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, DOI: 10.1109/EMB C46164.2021.9630002, 2021.
- [11] E. K. Naeini et al., Pain recognition with electrocardiographic features in postoperative patients: Method validation study, *J. Med. Internet Res.*, vol.23, no.5, DOI: 10.2196/25079, 2021.
- [12] F. Pouroumran, S. Radhakrishnan and S. Kamarthi, Exploration of physiological sensors, features, and machine learning models for pain intensity estimation, *PLoS One*, vol.16, no.7, DOI: 10.1371/journal.pone.0254108, 2021.
- [13] E. Campbell, A. Phinyomark and E. Scheme, Feature extraction and selection for pain recognition using peripheral physiological signals, *Front. Neurosci.*, vol.13, no.5, DOI: 10.3389/fnins.2019.00437, 2019.
- [14] M. Abdollahi et al., Using a motion sensor to categorize nonspecific low back pain patients: A machine learning approach, *Sensors (Switzerland)*, vol.20, no.12, DOI: 10.3390/s20123600, 2020.
- [15] E. Sheu, J. Versloot, R. Nader, D. Kerr and K. D. Craig, Pain in the elderly: Validity of facial expression components of observational measures, *Clin. J. Pain*, vol.27, no.7, DOI: 10.1097/AJP.0b013e31820f52e1, 2011.
- [16] M. Kunz, V. Mylius, S. Scharmann, K. Schepelman and S. Lautenbacher, Influence of dementia on multiple components of pain, *Eur. J. Pain*, vol.13, no.3, DOI: 10.1016/j.ejpain.2008.05.001, 2009.
- [17] H. Baumgartl, D. Flathau, S. Bayerlein, D. Sauter, I. J. Timm and R. Buettner, Pain level assessment for infants using facial expression scores, *2021 IEEE 45th Annual Computers, Software,*

- and Applications Conference (COMPSAC)*, pp.590-598, DOI: 10.1109/COMPSAC51774.2021.00087, 2021.
- [18] J. F. Allen, Natural language processing, in *Encyclopedia of Cognitive Science*, John Wiley & Sons, Ltd., DOI: 10.1002/0470018860.s00078, 2006.
- [19] A. Vaswani et al., Attention is all you need, *Neural Information Processing Systems*, 2017.
- [20] R. Alec, W. Jeffrey, C. Rewon, L. David, A. Dario and S. Ilya, Language models are unsupervised multitask learners | Enhanced reader, *OpenAI Blog*, vol.1, no.8, 2019.
- [21] B. Richardson and A. Wicaksana, Comparison of IndoBERT-lite and RoBERTa in text mining for Indonesian language question answering application, *International Journal of Innovative Computing, Information and Control*, vol.18, no.6, pp.1719-1734, DOI: 10.24507/ijicic.18.06.1719, 2022.
- [22] D. C. Nicolae, R. K. Yadav and D. Tufiş, A Lite Romanian BERT: ALR-BERT, *Computers*, vol.11, no.4, DOI: 10.3390/computers11040057, 2022.
- [23] H. M. Alawadh, A. Alabrah, T. Meraj and H. T. Rauf, Attention-enriched Mini-BERT fake news analyzer using the Arabic language, *Futur. Internet*, vol.15, no.2, DOI: 10.3390/fi15020044, 2023.
- [24] Y. Liu, H. Yang, Z. Li and M. Zhang, A lightweight annotation guideline of Chinese semantic role labeling, *J. Chinese Inf. Process.*, vol.34, no.4, pp.10-20, 2020.
- [25] X. Chen, P. Cong and S. Lv, A long-text classification method of Chinese news based on BERT and CNN, *IEEE Access*, vol.10, pp.34046-34057, DOI: 10.1109/ACCESS.2022.3162614, 2022.
- [26] M. Dai and C. L. Liu, Multi-label classification of Chinese judicial documents based on BERT, *Proc. of 2020 IEEE International Conference on Big Data (Big Data 2020)*, DOI: 10.1109/BigData50022.2020.9377969, 2020.
- [27] B. Yu, C. Deng and L. Bu, Policy text classification algorithm based on BERT, *Proc. of 2022 11th International Conference of Information and Communication Technology (ICTech 2022)*, DOI: 10.1109/ICTech55460.2022.00103, 2022.
- [28] Y.-J. Lin, B.-J. Cai, P.-S. Hsu, R.-C. Chen and K.-X. You, Predicting pain severity category by vital signs in hospice ward: An application of Palliative Care Outcomes Collaboration (PCOC) pain scale measurement, *Proc. of the 2023 7th International Conference on Medical and Health Informatics (ICMHI'23)*, New York, NY, USA, pp.72-76, DOI: 10.1145/3608298.3608313, 2023.
- [29] Z.-H. Zhou, Ensemble learning, in *Encyclopedia of Biometrics*, S. Z. Li and A. Jain (eds.), Boston, MA, Springer US, DOI: 10.1007/978-0-387-73003-5_293, 2009.

Author Biography



Yu-Ju Lin received his bachelor degree from National Cheng Kung University. Dr. Lin currently serves as the Chief Resident in the Department of Family Medicine at Taichung Veterans General Hospital. His research focuses on natural language processing (NLP) and its application in medicine.



Pi-Shan Hsu graduated from College of Medicine of Chung-Shan Medical University, Taichung, Taiwan and got Master degree from Business Administration of Tunghai University, Taiwan. Dr. Hsu is now the Director, Department of Family Medicine, Taichung Veterans General Hospital. Her major interests include family medicine, preventive medication, geriatric medicine, community medicine, long term care, and hospice care.



Bing-Jun Cai received the B.S. degree from the Department of Information Management, Chaoyang University of Technology, Taiwan. He received his Master's Degree from Chaoyang University of Technology, Taiwan, in 2023. His research focuses on deep learning, machine learning, and natural language processing (NLP).



Rung-Ching Chen received the B.S. degree from the Department of Electrical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, in 1987, the M.S. degree from the Institute of Computer Engineering, National Taiwan University of Science and Technology, in 1990, and the Ph.D. degree in Computer Science from the Department of Applied Mathematics, National Chung Hsing University, in 1998.

Prof. Chen is currently a Distinguished Professor with the Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan. He is listed in the Top 2% Scientists Worldwide in A.I. by Stanford University. His research concerns include network technology, pattern recognition, knowledge engineering, the Internet of Things, data analysis, and artificial intelligence.



William Eric Manongga received a B.Sc. degree in Information Technology from President University, Indonesia, and an MBA degree in Business Engineering from Swiss-German University, Indonesia. He had a career spanning 10 years in the software industry, specializing in developing enterprise applications.

He is currently pursuing a Ph.D. degree in Information Management at Chaoyang University of Technology, Taiwan. His research interests include computer vision, autonomous driving, and AI-related applications.