Volume XI Issue 3 Year 2026 | Page 902-913 | ISSN: 2527-9866
Received: ***27-06-2026*** | *Revised:* ***20-07--2026*** | *Accepted:* ***13-08-2026***

Random Forest and LightGBM Comparison for Acute Pain Diagnosis Using SMOTE on an Expert-Labeled Dataset

Wayan Andre Pratama[1], I Made Gede Sunarya[2], Putu Hendra Suputra[3]
[1,2,3]Postgraduate Programs, Computer Science Study Program, Pendidikan Ganesha University, Bali, Indonesia
e-mail: andre.pratama@student.undiksha.ac.id[1], sunarya@undiksha.ac.id[2], hendra.suputra@undiksha.ac.id[3]
*Correspondence: andre.pratama@student.undiksha.ac.id

Abstract: *Limited healthcare personnel may delay early pain assessment and encourage self-medication, increasing medication-error risk. However, evidence remains limited regarding whether bagging or boosting is more suitable for multiclass acute pain classification using imbalanced, expert-system-derived symptom data, and whether SMOTE improves performance. This study compared Random Forest as a bagging approach and LightGBM as a boosting approach for classifying nine acute pain diagnostic classes without SMOTE and with SMOTE using k_neighbors=1 and 5. The dataset comprised 2,722 records and 36 discrete symptom features. Of 125 representative symptom combinations reviewed by a medical expert, 115 were considered appropriate; the remaining records were synthetically generated using the same expert-system knowledge base and inference mechanism. Data were divided using stratified 80:20 sampling, while model configuration was evaluated using five-fold cross-validation. SMOTE was applied only to training data within each fold. LightGBM without SMOTE achieved the best performance, with 83.49% accuracy, a macro F1-score of 0.81, and a weighted F1-score of 0.83, compared with 80.18%, 0.77, and 0.80 for Random Forest. With SMOTE, Random Forest achieved 78.35% and 77.61% accuracy, while LightGBM achieved 81.10% and 82.39%. Thus, LightGBM without SMOTE performed best for this dataset. Validation using real clinical data and multiple experts is required.*

Keywords: *acute pain diagnosis, expert-based dataset, LightGBM, Random Forest, SMOTE.*

1. Introduction

Pain is one of the most common medical complaints experienced by patients across various healthcare facilities[1], [2]. Findings from previous studies indicate that pain has a high prevalence and is frequently reported as an initial complaint by patients seeking healthcare services. Pain is generally classified into acute pain and chronic pain. Acute pain is temporary pain that generally lasts for less than three months. If acute pain is not properly and adequately managed at an early stage, it may negatively affect patients and potentially progress into chronic pain. Chronic pain is prolonged discomfort that requires more complex and comprehensive management than acute pain[3].

The limited availability of healthcare professionals may hinder the optimization of early acute pain management. This condition may reduce patient satisfaction with the pain management services they receive, thereby encouraging individuals to choose self-medication as an alternativ[4]e. Other contributing factors include the perception that purchasing medication independently is more practical and time-efficient, the need for privacy, lower costs, and the considerable distance to healthcare facilities[5]. However, self-medication conducted without adequate medical knowledge may lead to medication errors, which can increase the risk of adverse effects and other health complications. Therefore, a system is required to support individuals in managing acute pain while ensuring that the process remains within a medically reliable and scientifically validated framework[6].

To address this challenge, a previous study developed an innovation in the field of information technology and artificial intelligence, namely a web-based expert system that applied forward chaining and certainty factor methods to assist in diagnosing acute pain. Expert systems have demonstrated their ability to adopt and represent expert knowledge when diagnosing diseases based on symptoms entered by patients[7]. Nevertheless, recent developments in artificial intelligence have enabled the use of machine learning algorithms that are more adaptive and potentially more accurate in classifying diseases based on patterns identified in data[8].

Compared the performance of rule-based and machine learning methods in classifying patient messages within a healthcare service portal. The study found that machine learning approaches tended to classify patient messages more accurately than rule-based methods. Another study comparing machine learning and rule-based approaches reported that machine learning was more adaptive to diverse data and could provide more flexible results when additional data or system modifications were required. Machine learning-based systems are generally more flexible and capable of handling dynamic data and new conditions that may not have been anticipated previously. Machine learning is also considered more effective in situations involving continuously evolving data or variables that are difficult to predict [9].

As an effort to improve the accuracy of acute pain diagnosis, this study aims to compare the performance of the Random Forest and Light Gradient Boosting Machine (LightGBM) algorithms in classifying symptoms. These two algorithms were selected because they have different characteristics but are both considered effective for processing complex data. Random Forest applies an ensemble approach by constructing multiple decision trees and has advantages in capturing data variability and reducing the risk of overfitting[10]. Random Forest has also been evaluated in several disease classification studies and has demonstrated strong predictive performance[11].

Meanwhile, LightGBM offers greater computational efficiency, is capable of processing large-scale datasets rapidly, and can effectively handle both numerical and categorical features. Through this comparison, the study is expected to identify the most optimal machine learning method in terms of predictive accuracy and computational efficiency. The selected method can subsequently be implemented in a more adaptive and reliable acute pain diagnosis system [12]. The selection of Random Forest and LightGBM was also based on the fundamental differences between their ensemble learning strategies. Random Forest represents a bagging approach, in which multiple decision trees are constructed independently and in parallel to reduce model variance. In contrast, LightGBM represents a boosting approach, in which decision trees are constructed sequentially, with each subsequent tree attempting to correct the prediction errors produced by the previous trees, thereby reducing model bias[13].

Comparing these two ensemble strategies is relevant for identifying the most appropriate approach for the characteristics of the dataset used in this study. The dataset consists of tabular symptom data containing sparse categorical features with relatively weak correlations among features. A study by Yıldız and Kalayci (2025) demonstrated that decision tree-based ensemble models, including both bagging and gradient boosting decision tree approaches, consistently outperformed traditional machine learning and deep learning models across various tabular medical diagnosis datasets. This performance was particularly attributed to their ability to operate effectively without requiring strong correlations among features and to provide results that are more interpretable than black-box deep learning models. These characteristics are consistent with the requirements of medical research, in which the transparency of the model's decision-making process is an important consideration in addition to predictive accuracy.

The use of machine learning also creates opportunities to apply data augmentation techniques such as the Synthetic Minority Over-sampling Technique (SMOTE). This technique can

address class imbalance, which commonly occurs in healthcare datasets, where certain symptom or disease classes may be represented by only a limited number of observations. Hasanah et al. (2024) reported that the application of SMOTE to the Random Forest algorithm produced the best performance in terms of accuracy, sensitivity, and specificity. By using SMOTE, the representation of minority classes can be strengthened, enabling machine learning models to receive more balanced exposure to the various symptom patterns contained in the dataset. Consequently, the model's ability to identify acute pain conditions, including relatively rare cases, can potentially be improved.

Therefore, this study aims to implement and compare the Random Forest and LightGBM methods using an expert-derived dataset and the SMOTE approach for acute pain diagnosis. The study is expected to provide more comprehensive and accurate results for acute pain diagnosis and to identify the most effective machine learning algorithm for developing a more adaptive acute pain diagnostic system.

2. Literature Review

Related studies in similar scientific fields and research topics play an important role in further examining the feasibility and advantages of applying machine learning methods, particularly Random Forest and LightGBM. Based on the literature exploration, the researchers identified several previous studies that are relevant to the scientific domain and research topic investigated in this study.

No.	Research Title	Method and Research Object	Main Findings
1	Classification Analysis of Diabetes Mellitus Diagnosis Based on a Comparison of Supervised Learning Algorithms [14]	Compared several supervised learning algorithms for predicting the risk of diabetes mellitus. The dataset was obtained through questionnaires completed by patients at Sylhet Diabetes Hospital and was validated by medical professionals.	Random Forest achieved the highest accuracy of 98.71%, indicating very strong performance in diagnosing diabetes mellitus.
2	Comparative Analysis of Gradient Boosting Classifier and Random Forest Algorithms for Diabetes Prediction [15]	Compared Gradient Boosting Classifier and Random Forest as supervised learning algorithms for predicting diabetes.	Gradient Boosting Classifier produced better evaluation performance and higher accuracy than Random Forest.
3	Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM) [16]	Applied LightGBM to a diabetes diagnosis dataset and compared its performance with Random Forest.	LightGBM achieved 98.1% accuracy, 98.1% AUC, 99.9% sensitivity, and 96.3% specificity, outperforming Random Forest.
4	Light Gradient Boost Tree Classifier Predictions on Appendicitis with Periodontal Disease from Biochemical and Clinical Parameters [17]	Used biochemical and clinical parameters to predict appendicitis associated with periodontal disease using LightGBM and Random Forest.	LightGBM achieved an accuracy of 98%, which was higher than the 94% accuracy obtained by Random Forest.

3. Methods

A. Research Type and Design

This study employed a comparative quantitative research design with an experimental approach. It aimed to compare the performance of Random Forest and Light Gradient Boosting Machine (LightGBM) in classifying nine acute pain diagnostic classes. The experiments were

conducted using an expert-based dataset under three conditions: without class balancing, with the SMOTE using k_neighbors=1, and with SMOTE using k_neighbors=5.

Table 1. Research Design

Aspect	Description
Research type	Comparative quantitative research
Approach	Multiclass classification experiment
Algorithms	Random Forest and LightGBM
Classification object	Acute pain diagnosis
Number of classes	Nine diagnostic classes (A002–A010)
Data treatments	Without SMOTE, SMOTE with k_neighbors=1, and SMOTE with k_neighbors=5
Main evaluation metrics	Accuracy, precision, recall, F1-score, macro average, weighted average, and confusion matrix

B. Dataset Source and Generation Process

The dataset was obtained from a web-based expert system for acute pain diagnosis that applied forward chaining and certainty factor methods. Symptom combinations were constructed according to the PQRST pain assessment framework, consisting of Provocation/Palliation, Quality, Region/Radiation, Severity, and Time. Each symptom combination was processed by the expert system's inference engine to generate a diagnostic label.

To evaluate the agreement between the system-generated diagnoses and clinical judgments, 125 representative symptom combinations were validated by a certified medical expert in anesthesia care. The validation results showed that 115 combinations, or 92%, were considered appropriate, while 10 combinations, or 8%, were considered inappropriate and excluded from the dataset.

Following the initial validation, the same expert system was used to synthetically generate additional data while maintaining the same knowledge base, inference mechanism, and symptom-combination structure. The final dataset consisted of 2,722 records, including 115 expert-validated records and additional synthetic records. Dataset size adequacy was subsequently evaluated using a learning curve.

Table 2. Summary of Dataset Sources and Characteristics

Component	Description
Primary source	Web-based acute pain diagnosis expert system
Basis for symptom generation	PQRST pain assessment framework
Initial inference methods	Forward chaining and certainty factor
Validator	Certified medical expert/anesthesia technologist
Validation sample	125 symptom combinations
Expert-confirmed data	115 records (92%)
Inappropriate data	10 records (8%); excluded
Final dataset size	2,722 records
Number of features	36 symptom features
Target type	Multiclass classification with nine diagnostic classes

C. Input Feature Structure

The input features consisted of medical symptoms encoded as discrete numerical values. Feature values ranged from 0 to 1, specifically 0, 0.2, 0.4, 0.6, 0.8, and 1.0. A value of 0 indicated that a symptom was not selected or was absent, whereas higher values represented a greater degree of confidence or symptom weight.

Table 3. Symptom Feature Groups Based on Pain Assessment

Dimension	Feature Codes	Number	Examples of Symptoms
Provocation/Palliation	M, N, O, P, Q, R	16	Mechanical stimuli, activities, comorbidities, and pain patterns during activity
Quality	S	6	Sharp, burning, dull, cramping, tingling, allodynia, and hyperalgesia
Region/Radiation	T	2	Radiating pain or pain concentrated at one point
Time	U, V, W	6	Duration of less than three months, sudden or gradual onset, and continuous or intermittent patterns
Severity	Y	6	No pain to unbearable pain
Total	-	36	All symptom features used in the dataset

D. Diagnostic Target Classes

The classification target consisted of nine acute pain diagnostic classes representing combinations of three pain types-somatic, visceral, and neuropathic-and three severity levels-mild, moderate, and severe. The class distribution was imbalanced. The largest class was A008 with 660 records, while the smallest class was A004 with 104 records.

Table 4. Dataset Class Distribution

Code	Diagnosis	Number	Percentage
A002	Mild Acute Somatic Pain	277	10.18%
A003	Mild Acute Visceral Pain	250	9.18%
A004	Mild Acute Neuropathic Pain	104	3.82%
A005	Moderate Acute Somatic Pain	588	21.60%
A006	Moderate Acute Visceral Pain	180	6.61%
A007	Moderate Acute Neuropathic Pain	193	7.09%
A008	Severe Acute Somatic Pain	660	24.25%
A009	Severe Acute Visceral Pain	210	7.71%
A010	Severe Acute Neuropathic Pain	260	9.55%
Total		2,722	100.00%

E. Data Collection Techniques

Data were collected through a literature review, data generation using the expert system, expert validation using a questionnaire, and dataset documentation. The literature review was used to establish the theoretical basis for pain assessment, symptom features, diagnostic classes, and machine learning methods. The questionnaire was used to assess the appropriateness of the symptom combinations and diagnoses generated by the expert system.

Table 5. Data Collection Techniques

No.	Data Type	Method	Instrument/Source	Output
1	Acute pain symptoms and diagnostic concepts	Literature review	Journals, books, and pain assessment guidelines	List of features and diagnostic classes
2	Symptom and diagnosis combinations	Data generation	Web-based expert system	Labeled synthetic dataset
3	Diagnostic appropriateness	Expert validation	Questionnaire and anesthesia technologist	Valid and invalid data
4	Experimental data	Documentation	Dataset files and processing results	Final training and testing dataset

F. Data Preprocessing

Preprocessing was conducted to ensure that the data were consistent, within the defined scope of the study, and suitable for processing by both algorithms. This stage was performed before data splitting and SMOTE application.

Table 6. Data Preprocessing Stages

Stage	Procedure
Data filtering	Removing 10 records that did not conform to expert validation and data outside the research scope, including class A001 and chronic pain categories
Missing-value inspection	Examining all features and labels
Duplicate inspection	Examining identical symptom and label combinations
Label consistency inspection	Matching labels with classes A002–A010
Feature transformation	Applying discrete numerical representations from 0 to 1
Label encoding	Converting class codes into numerical representations

G. Data Splitting and Validation Scheme

The final dataset was divided into training and testing sets using an 80:20 ratio and stratified sampling. Stratification was applied to preserve the proportion of each class in both subsets. Of the 2,722 records, 2,177 were used as training data and 545 were used as testing data. Model configuration exploration was performed using five-fold cross-validation. In the SMOTE scenarios, oversampling was applied exclusively to the training data within each fold. Neither the validation data nor the testing data were processed using SMOTE, thereby preventing data leakage and preserving the original data distribution.

Table 7. Dataset Splitting and Utilization

Subset	Number	Proportion
Training data	2,177	80%
Testing data	545	20%
Total	2,722	100%

H. Handling Class Imbalance Using SMOTE

SMOTE was used to generate synthetic samples for minority classes through interpolation between a minority-class observation and its nearest neighbor[18]. In general, a synthetic observation was generated using the following equation:

$$x_{\text{synthetic}} = x_i + \lambda\left(x_{\text{neighbor}} - x_i\right)$$

where λ is a random number between 0 and 1.

This study evaluated two k_neighbors values: 1 and 5. A value of k_neighbors=1 was used to accommodate classes with very limited initial samples, whereas k_neighbors=5 was used as a comparative standard configuration[19]. SMOTE was applied only to the training data after data splitting or within each cross-validation training fold [20].

Table 8. SMOTE Treatment Configurations

Condition	k_neighbors	Processed Data
Without SMOTE	-	Original training data
SMOTE-1	1	Training data/training fold
SMOTE-5	5	Training data/training fold

H. Random Forest

Random Forest constructs multiple decision trees using bootstrap sampling and random feature subset selection [21]. Each tree produces a class prediction, and the final prediction is determined through majority voting. The final configuration used 100 trees and

max_features='sqrt'. The explored max_features values included sqrt, log2, 0.5, and all features through five-fold cross-validation [22].

I. LightGBM

LightGBM constructs decision trees sequentially using a gradient boosting approach. Each new tree is designed to correct the errors produced by the preceding model. LightGBM employs histogram-based splitting and leaf-wise growth to select the leaf that provides the greatest reduction in loss[23]. The model used the default LightGBM parameter configuration consistently across all experimental scenarios, including num_leaves=31, max_depth=-1, learning_rate=0.1, and n_estimators=100[24].

Table 9. Model Configurations and Parameters

Model	Parameter/Procedure	Configuration or Function
Random Forest	n_estimators	100 trees
Random Forest	max_features	sqrt as the final configuration; compared with log2, 0.5, and all features
Random Forest	max_depth	Default implementation configuration
Random Forest	Final prediction	Majority voting
LightGBM	num_leaves	31
LightGBM	max_depth	−1, indicating no explicit depth limit
LightGBM	learning_rate	0.1
LightGBM	n_estimators	100 trees or boosting iterations
LightGBM	Growth strategy	Leaf-wise growth and histogram-based splitting

J. Model Performance Evaluation

Evaluation was conducted on the testing data using a one-vs-rest approach because the target consisted of nine classes. Under this approach, each class was independently treated as the positive class, while all remaining classes were combined as the negative class[25]. The class-specific values were subsequently aggregated using macro and weighted averages. The macro average assigned equal weight to each class and was therefore suitable for assessing model performance on minority classes. The weighted average incorporated the number of samples in each class and represented model performance under the actual dataset distribution.

4. Results and Discussion

A. Comparison of Random Forest and LightGBM Without SMOTE

The initial experiment was conducted using the original dataset without SMOTE. Both algorithms were trained and evaluated using the same data split.

Table 10. Model Performance Without SMOTE

Model	Accuracy	Macro F1	Weighted F1
Random Forest	80.18%	0.77	0.80
LightGBM	83.49%	0.81	0.83

LightGBM achieved an accuracy of 83.49%, whereas Random Forest achieved an accuracy of 80.18%. Therefore, LightGBM outperformed Random Forest by 3.31 percentage points. LightGBM also produced higher Macro F1 and Weighted F1 scores. These results indicate that LightGBM was not only more effective in predicting the overall testing data but also more balanced in recognizing the individual diagnostic classes.

B. Model Performance for Each Diagnostic Class

Class-specific evaluation was performed using precision, recall, and F1-score under the one-vs-rest approach.

Table 11. Class-Specific Precision, Recall, and F1-Score

Class	RF Precision	RF Recall	RF F1	LightGBM Precision	LightGBM Recall	LightGBM F1
A002	0.80	0.82	0.81	0.85	0.80	0.82
A003	0.83	0.68	0.75	0.80	0.78	0.79
A004	0.92	0.57	0.71	0.93	0.62	0.74
A005	0.77	0.93	0.84	0.82	0.92	0.87
A006	0.73	0.67	0.70	0.84	0.72	0.78
A007	0.83	0.64	0.72	0.73	0.69	0.71
A008	0.82	0.92	0.87	0.88	0.92	0.90
A009	0.88	0.69	0.77	0.86	0.86	0.86
A010	0.77	0.69	0.73	0.80	0.79	0.80

LightGBM achieved a higher F1-score in eight of the nine classes. Random Forest was only slightly superior for class A007, with an F1-score of 0.72 compared with 0.71 for LightGBM. Class A004 produced the lowest recall for both models, with values of 0.57 for Random Forest and 0.62 for LightGBM. The low recall was associated with A004 having the smallest number of samples, consisting of only 104 records.

C. Confusion Matrix Analysis of the Best-Performing Model

Because LightGBM without SMOTE produced the best overall performance, confusion matrix analysis was conducted on this model.

Table 12. Confusion Matrix of LightGBM Without SMOTE

Actual \ Predicted	A002	A003	A004	A005	A006	A007	A008	A009	A010
A002	**44**	4	1	3	0	0	3	0	0
A003	3	**39**	0	3	3	2	0	0	0
A004	4	4	**13**	0	0	0	0	0	0
A005	1	0	0	**108**	1	8	0	0	0
A006	0	2	0	7	**26**	0	0	0	1
A007	0	0	0	10	1	**27**	0	0	1
A008	0	0	0	0	0	0	**121**	4	7
A009	0	0	0	0	0	0	5	**36**	1
A010	0	0	0	0	0	0	9	2	**41**

Of the 545 testing records, 455 were correctly classified and 90 were incorrectly classified.

Table 13. Classification Error Patterns of LightGBM

Error Pattern	Number	Percentage
Different pain types at the same severity level	71	78.9%
Different severity levels within the same pain type	13	14.4%
Different pain types and severity levels	6	6.7%
Total errors	**90**	**100%**

Most errors occurred when the model attempted to differentiate pain types at the same severity level. The highest cross-classification error occurred between A005 and A007. Eight A005 records were predicted as A007, while ten A007 records were predicted as A005. Therefore, 18 cross-classification errors occurred between Moderate Acute Somatic Pain and Moderate Acute Neuropathic Pain.

D. Random Forest Configuration Exploration

The class_weight and max_features parameters were explored. However, the evaluated configurations did not produce consistent performance improvements over the baseline configuration.

Table 14. Summary of Random Forest Configuration Exploration

Component	Evaluated Configuration
class_weight	None and balanced
max_features	sqrt
max_features	log2
max_features	0.5
max_features	All features
Validation procedure	Five-fold cross-validation
SMOTE application	Applied only to the training data in each fold

Based on the exploration results, the final Random Forest model retained its default configuration of n_estimators=100 and max_features='sqrt'. The main adjustment was made to the evaluation pipeline by ensuring that SMOTE was applied only to the training data.

E. Effect of SMOTE Application

The effect of SMOTE was evaluated using k_neighbors values of 1 and 5.

Table 15. Performance Before and After SMOTE Application

Model	Condition	Accuracy	Macro F1
Random Forest	Without SMOTE	80.18%	0.77
Random Forest	SMOTE, k=1	78.35%	0.75
Random Forest	SMOTE, k=5	77.61%	0.75
LightGBM	Without SMOTE	83.49%	0.81
LightGBM	SMOTE, k=1	81.10%	0.78
LightGBM	SMOTE, k=5	82.39%	0.80

SMOTE reduced the performance of both algorithms. For Random Forest, the greatest reduction in accuracy occurred with k_neighbors=5, corresponding to a decrease of 2.57 percentage points. For LightGBM, the greatest reduction occurred with k_neighbors=1, corresponding to a decrease of 2.39 percentage points. These results indicate that the best configuration for both algorithms was obtained without SMOTE.

F. Feature Importance Analysis

Feature importance analysis was conducted on LightGBM without SMOTE because it was the best-performing model.

Table 16. Ten Most Influential Features

Rank	Code	Symptom	Dimension
1	T001	Radiating pain	Region/Radiation
2	V002	Sudden onset	Time
3	W001	Continuous	Time
4	V001	Gradual onset	Time
5	T002	Localized at one point	Region/Radiation
6	W002	Intermittent	Time
7	O001	Lupus/Guillain–Barré syndrome/Sjögren's syndrome/vasculitis/rheumatoid disease	Provocation/Palliation
8	P008	Diphtheria	Provocation/Palliation
9	S007	Cramping/tingling	Quality
10	P004	Lyme disease/herpes/Epstein–Barr virus/leprosy	Provocation/Palliation

Six of the ten most important features belonged to the Region/Radiation and Time dimensions. The Quality dimension was represented only by S007. These findings indicate that the model

relied primarily on pain distribution, onset, and temporal patterns when determining the diagnostic class.

G. Model Performance on the Minority Class

Class A004 had the smallest number of observations and produced the lowest recall for both algorithms. Random Forest achieved a recall of 0.57, while LightGBM achieved a recall of 0.62. The precision values for A004 were relatively high, reaching 0.92 for Random Forest and 0.93 for LightGBM. The combination of high precision and low recall indicates that predictions assigned to A004 were generally correct. However, many actual A004 observations were not successfully recognized and were instead classified as other classes, particularly A002 and A003. These findings suggest that the limited number of observations reduced the models' ability to learn the variability of class A004. Nevertheless, LightGBM produced better recall and F1-score values than Random Forest for this class.

H. Classification Error Pattern Analysis

A total of 78.9% of classification errors occurred between classes representing different pain types at the same severity level. This finding indicates that the model was more effective in distinguishing mild, moderate, and severe pain than in differentiating somatic, visceral, and neuropathic pain at the same severity level. The highest error frequency occurred between A005 and A007. Both classes represented moderate pain but differed in pain type. This error pattern indicates that several symptom combinations associated with Moderate Acute Somatic Pain and Moderate Acute Neuropathic Pain had overlapping feature representations. In contrast, errors between different severity levels within the same pain type accounted for only 13 observations, or 14.4% of all errors. This result further demonstrates that the model was more capable of recognizing severity levels than pain types.

4. Conclusions

This study compared Random Forest and Light Gradient Boosting Machine (LightGBM) for classifying acute pain diagnoses using an expert-based dataset consisting of 2,722 records, 36 symptom features, and nine diagnostic classes. The models were evaluated under three data conditions: without SMOTE, with SMOTE using k_neighbors=1, and with SMOTE using k_neighbors=5. The results showed that LightGBM without SMOTE achieved the best performance, with an accuracy of 83.49%, a macro F1-score of 0.81, and a weighted F1-score of 0.83. Random Forest without SMOTE achieved an accuracy of 80.18%, a macro F1-score of 0.77, and a weighted F1-score of 0.80. Most classification errors occurred between different pain types at the same severity level, particularly between moderate somatic and moderate neuropathic pain. The application of SMOTE did not improve model performance because its interpolation mechanism generated values that were less compatible with the discrete and ordinal characteristics of the symptom features. Therefore, LightGBM without SMOTE was identified as the most appropriate model for acute pain classification in this study. Future research should validate the model using real clinical patient data and involve multiple medical experts to strengthen label reliability. Further studies should also examine alternative imbalance-handling techniques, such as SMOTENC, Borderline-SMOTE, random oversampling, class weighting, and cost-sensitive learning. In addition, broader hyperparameter optimization, feature engineering, and explainability methods should be explored to improve performance and support the development of a more transparent and clinically reliable decision-support system.

References

[1] I. Kadek, Y. Prayoga, I. Made, G. Sunarya, and P. H. Suputra, "Klasifikasi Penyakit Demam Berdarah Menggunakan Algoritma Stacking Ensemble Learning," *SINTECH (Science and Information Technology) Journal*, vol. 8, 2025, [Online]. Available: https://doi.org/10.31598

[2] I. G. A. P. Mahendra, I. M. A. Wirawan, and I. G. A. Gunadi, "Enhancement performance of the Naïve Bayes method using AdaBoost for classification of diabetes mellitus dataset type II," *International Journal of Advances in Applied Sciences*, vol. 13, no. 3, p. 733, Sep. 2024, doi: 10.11591/ijaas.v13.i3.pp733-742.

[3] I. G. A. S. Nugraha, I. G. A. Gunadi, and L. J. E. Dewi, "Application of Bagging Ensemble Learning on Naïve Bayes Algorithm to Predict Coronary Heart Disease," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 18, no. 2, pp. 943–954, Aug. 2025, doi: 10.24036/jtip.v18i2.981.

[4] D. G. Nigatu, M. G. Teferi, G. W. Hassen, T. Zewdu, and A. Azazh, "Pain treatment practices and their impact on patient satisfaction in an Ethiopian Emergency Department: a prospective observational study," *African Journal of Emergency Medicine*, vol. 16, no. 2, p. 100964, Jun. 2026, doi: 10.1016/j.afjem.2026.100964.

[5] A. Bhattacharyya, H. Laycock, S. J. Brett, F. Beatty, and H. I. Kemp, "Health care professionals' experiences of pain management in the intensive care unit: a qualitative study," *Anaesthesia*, vol. 79, no. 6, pp. 611–626, Jun. 2024, doi: 10.1111/anae.16209.

[6] I. M. A. A. D. Putra, I. M. G. Sunarya, and I. G. A. Gunadi, "Perbandingan Algoritma Naive Bayes Berbasis Feature Selection Gain Ratio dengan Naive Bayes Kovensional dalam Prediksi Komplikasi Hipertensi," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 6, no. 1, pp. 37–49, Apr. 2024, doi: 10.35746/jtim.v6i1.488.

[7] D. G. Nigatu, M. G. Teferi, G. W. Hassen, T. Zewdu, and A. Azazh, "Pain treatment practices and their impact on patient satisfaction in an Ethiopian Emergency Department: a prospective observational study," *African Journal of Emergency Medicine*, vol. 16, no. 2, p. 100964, Jun. 2026, doi: 10.1016/j.afjem.2026.100964.

[8] I. M. W. G. Negara, I. M. A. Wirawan, and I. M. G. Sunarya, "Enhancing EEG-Based Stress Detection Using ICA, Relative Difference, and Convolutional Neural Networks," *sinkron*, vol. 9, no. 3, pp. 1073–1083, Jul. 2025, doi: 10.33395/sinkron.v9i3.14777.

[9] E. Rijcken, K. Zervanou, P. Mosteiro, F. Scheepers, M. Spruit, and U. Kaymak, "Machine learning vs. rule-based methods for document classification of electronic health records within mental health care-A systematic literature review," *Natural Language Processing Journal*, vol. 10, p. 100129, Mar. 2025, doi: 10.1016/j.nlp.2025.100129.

[10] N. E. I. Karabadji, A. Amara Korba, A. Assi, H. Seridi, S. Aridhi, and W. Dhifli, "Accuracy and diversity-aware multi-objective approach for random forest construction," *Expert Syst. Appl.*, vol. 225, p. 120138, Sep. 2023, doi: 10.1016/j.eswa.2023.120138.

[11] S. Datta, G. Datta, T. Gedeon, and M. Z. Hossain, "Painthenticate: Feature Engineering on Multimodal Physiological Signals," in *Companion Proceedings of the 27th International Conference on Multimodal Interaction*, New York, NY, USA: ACM, Oct. 2025, pp. 177–184. doi: 10.1145/3747327.3764894.

[12] F. Ratnawati, A. Tedyyana, and J. . Custer, "LightGBM for Liver Disease Detection with Hybrid Hyperparameter Optimization", J. RESTI (Rekayasa Sist. Teknol. Inf.) , vol. 10, no. 4, pp. 1025–1033, Aug. 2026, doi: 10.29207/resti.v10i4.7357.

[13] I. Gede Angga Candrawibawa, I. Gede Aris Gunadi, and L. Joni Erawati Dewi, "Multimodal ECG-PPG Clinical Fusion for Myocardial Infarction Classification Using Ensemble Learning," *INOVTEK Polbeng-Seri Informatika*, vol. XI, no. 2, 2026.

[14] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Comput. Methods Programs Biomed.*, vol. 220, p. 106773, Jun. 2022, doi: 10.1016/j.cmpb.2022.106773.

[15] S. P. Nainggolan and A. Sinaga, "COMPARATIVE ANALYSIS OF ACCURACY OF RANDOM FOREST AND GRADIENT BOOSTING CLASSIFIER ALGORITHM FOR DIABETES CLASSIFICATION," *Sebatik*, vol. 27, no. 1, pp. 97–102, Jun. 2023, doi: 10.46984/sebatik.v27i1.2157.

[16] F. Rahman, S. Hossain, J.-J. Tiang, and A.-A. Nahid, "Diabetes Prediction Using Feature Selection Algorithms and Boosting-Based Machine Learning Classifiers," *Diagnostics*, vol. 15, no. 20, p. 2622, Oct. 2025, doi: 10.3390/diagnostics15202622.

[17] P. K. Yadalam, P. V. Thirukkumaran, P. M. Natarajan, and C. M. Ardila, "Light gradient boost tree classifier predictions on appendicitis with periodontal disease from biochemical and clinical parameters," *Frontiers in Oral Health*, vol. 5, Sep. 2024, doi: 10.3389/froh.2024.1462873.

[18] A. A. G. W. S. Erlangga, I. G. A. Gunadi, and I. M. G. Sunarya, "Kombinasi Oversampling dan Undersampling dalam Menangani Class Imbalanced dan Overlapping pada Klasifikasi Data Bank Marketing," *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, vol. 7, no. 1, pp. 32–42, Apr. 2024, doi: 10.31598/jurnalresistor.v7i1.1515.

[19] I. G. Ayu Nandia Lestari, D. G. Hendra Divayana, and K. Y. Ernada Aryanto, "A Concentration Selection In Study Programs Using SMOTE Techniques With Ensemble Learning Algorithms," in *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)*, IEEE, Oct. 2023, pp. 1–6. doi: 10.1109/ICORIS60118.2023.10352192.

[20] I. Putu *et al.*, "IMPROVING BUTTERFLY FISH IMAGE CLASSIFICATION ACCURACY USING HSV FEATURE EXTRACTION AND SMOTE-BASED DATA BALANCING," vol. 10, no. 3, 2025.

[21] S. Bose and S. Bose, "Random Forests: The Wisdom of Crowds in Action," *Journal of Emerging Trends in Computer Science and Applications*, vol. 1, no. 1, pp. 67–91, Nov. 2025, doi: 10.65525/jetcsa.v1i1.5.

[22] P. Sidik, I. Made, G. Sunarya, I. Gede, and A. Gunadi, "Comparison of Random Forest and Support Vector Machine Methods in Sentiment Analysis of Student Satisfaction Questionnaire Comments at ITB STIKOM Bali," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[23] H. (Richard) Zhuang, F. Lehner, and A. T. DeGaetano, "Improved Diagnosis of Precipitation Type with LightGBM Machine Learning," *J. Appl. Meteorol. Climatol.*, vol. 63, no. 3, pp. 437–453, Mar. 2024, doi: 10.1175/JAMC-D-23-0117.1.

[24] H. Wu, Y. Mao, J. Weng, Y. Yu, and J. Wang, "Fractional light gradient boosting machine ensemble learning model: A non-causal fractional difference descent approach," *Information Fusion*, vol. 118, p. 102947, Jun. 2025, doi: 10.1016/j.inffus.2025.102947.

[25] D. Krstinić, A. K. Skelin, I. Slapničar, and M. Braović, "Multi-Label Confusion Tensor," *IEEE Access*, vol. 12, pp. 9860–9870, 2024, doi: 10.1109/ACCESS.2024.3353050.