Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 6, December 2026, pp. 5606~5616
ISSN: 2302-9285, DOI: 10.11591/eei.v15i6.11751

Accurate burn severity classification via texture analysis and supervised learning

Zaure Kabdrakhmanova[1], Aizhan Tlebaldinova[1], Zbigniew Omiotek[2], Markhaba Karmenova[3], Symbat Nurgaliyeva[4]

[1]School of Digital Technology and Artificial Intelligence, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan
[2]Department of Electronics and Information Technology, Faculty of Electrical Engineering and Computer Science, Lublin University of Technology, Lublin, Poland
[3]Department of Artificial Intelligence and Information Technologies, School of Information Technology and Natural Sciences, Sarsen Amanzholov East Kazakhstan University, Ust-Kamenogorsk, Kazakhstan
[4]School of Software Engineering, Astana IT University, Astana, Kazakhstan

Article Info

Article history:

Received Oct 30, 2025
Revised Sep 22, 2026
Accepted Oct 4, 2026

Keywords:

Burn injuries
Burn severity classification
Feature selection
Machine learning
Texture analysis

ABSTRACT

This paper presents a methodology for adapting machine learning (ML) algorithms to determine an optimal approach to burn severity classification and improving classification accuracy. The dataset for multiclass classification includes 1,500 images evenly distributed into three burn severity classes, and the dataset for binary classification includes 1,000 images evenly distributed into two classes. All images were processed with the MaZda 4.6 software package, which computed 297 texture features for each region of interest (ROI) based on different statistical descriptors. In a preprocessing step, the images were converted to 8-bit grayscale format and normalized for luminance. ROIs were manually selected with expert consultation to capture representative texture regions of burn tissue. Six feature selection methods were applied to identify the most informative descriptors. Subsequently, more than ten ML models were trained based on these features. The combination of a multilayer perceptron (MLP) classifier and recursive feature elimination (RFE) yielded the highest multiclass classification accuracy of 79% with 11 selected features. For binary classification, the MLP classifier combined with RELIEF achieved an accuracy of 96% with eight selected features. These results demonstrate the potential of feature selection combined with MLP for burn severity classification.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Markhaba Karmenova
Department of Artificial Intelligence and Information Technologies
School of Information Technology and Natural Sciences
Sarsen Amanzholov East Kazakhstan University
30th Guards Division St., 36, Ust-Kamenogorsk, Kazakhstan
Email: mkarmenova@vku.edu.kz

1. INTRODUCTION

Skin lesions are pathological changes in the structure or appearance of the skin caused by various factors, ranging from trauma and burns to infectious diseases and cancer. These lesions may affect all layers of the skin, from the epidermis to the subcutaneous tissue, and can have a profound impact on an individual's physical and psycho-emotional well-being. This issue is particularly acute in the case of burn injuries, which are often associated with intense pain, a high risk of infection, prolonged recovery periods, and permanent

cosmetic defects. According to a global study by Huang *et al.* [1], approximately 9 million burn cases and 111,000 burn-related deaths were recorded worldwide in 2019. Non-fatal burns frequently lead to long-term disability (LTD), stigmatization, and social isolation, particularly in regions with limited access to quality healthcare. The rapid and accurate determination of burn depth is therefore critical for selecting appropriate treatment strategies and predicting recovery outcomes, particularly in emergency medical settings. Rostami *et al.* [2] and Anisuzzaman *et al.* [3] noted that conventional burn diagnosis largely relies on visual examination of the affected area and the clinician's subjective judgement, which may result in diagnostic errors, particularly when clinical signs are inconclusive. Rangaiah *et al.* [4] and Asif *et al.* [5] reported that more objective diagnostic techniques, including histological examination and infrared thermography, may not always be feasible in emergency settings because they require specialised equipment, appropriate facilities, and trained personnel. Consequently, there is growing interest in the development of non-invasive, automated methods for assessing burn severity through image analysis. Jaén *et al.* [6] developed a web-based system for the automated diagnosis of skin lesions based on deep learning (DL) techniques. Six pretrained convolutional neural network (CNN) models—VGG16, VGG19, DenseNet201, InceptionV3, MobileNetV2, and Xception—were evaluated on the ISIC dataset, which comprises six types of skin lesions. Hyperparameter tuning and performance validation were conducted for each model, with VGG16 achieving the highest accuracy of 86%. A promising research direction involves the application of machine learning (ML) algorithms to texture analysis of medical images [6]–[8]. These advanced methods enable the extraction of latent features from damaged skin regions and facilitate classification with high accuracy. In particular, texture analysis provides a quantitative assessment of structural patterns in skin lesions, thereby significantly enhancing diagnostic precision. However, the high dimensionality of the extracted feature space necessitates effective feature selection techniques and the careful choice of appropriate classification models. In recent years, the classification of skin burns has gained increasing importance in the field of medicine due to the urgent need for rapid diagnostic decisions, assessment of tissue damage severity, and planning of appropriate therapeutic interventions. A review of current research demonstrates a wide array of approaches applied to this problem, ranging from traditional image classification methods to advanced DL techniques. Although deep neural networks have demonstrated promising performance in medical image analysis, their clinical applicability remains constrained by their black-box nature, limited transparency in feature attribution, and dependence on large-scale labelled datasets, which are often difficult to obtain in medical domains. Furthermore, the existing literature rarely provides a comprehensive evaluation of the interaction between different feature selection algorithms and traditional ML classifiers, particularly under both binary and multiclass classification scenarios. This highlights a significant gap, especially in studies involving relatively small and specialized datasets, where the careful selection of relevant features becomes critical for improving classification performance and ensuring model interpretability.

Elsarta *et al.* [9], Lee *et al.* [10], and Boissin *et al.* [11] investigated DL methods incorporating multimodal data, ultrasound images, and clinical parameters. These studies combined heterogeneous data sources with CNN-based models for clinically oriented classification tasks and reported accuracies ranging from 88% to 94%. Suha and Sanam [12], Lee *et al.* [13], Chauhan and Goyal [14], and Cirillo *et al.* [15] investigated CNNs for burn-depth classification and assessment. In particular, Chauhan and Goyal [14] proposed the body part-specific burn severity assessment model (BPBSAM), which incorporates the anatomical location of the affected area. These approaches employ end-to-end training and automated feature extraction, while some also account for the spatial and anatomical characteristics of burn injuries. The reported classification accuracies ranged from 85% to 92%. Liu *et al.* [16], Dittakan and Nookong [17], Yıldız *et al.* [18], and Kuan *et al.* [19] investigated frameworks integrating burn-area segmentation with subsequent classification. These studies employed architectures such as DeepLabV3+, U-Net, and YOLOv8 to localise affected regions before classification, thereby reducing the influence of background information. The reported accuracies ranged from 87% to 92%. In contrast, Yadav *et al.* [20], Şevik *et al.* [21], and Zhang *et al.* [22] investigated conventional ML approaches based on texture descriptors, including features derived from the grey-level co-occurrence matrix (GLCM). Classification was performed with support vector machines (SVMs), k-nearest neighbours (KNN), random forests (RFs), and artificial neural networks (ANNs). These approaches offer greater model interpretability and facilitate the analysis of feature importance; however, their reported accuracies ranged from approximately 65% to 80%. Suha and Sanam [23] proposed a hybrid approach combining KNN and ANN, thereby integrating local instance-based classification with the modelling of nonlinear relationships. Their method achieved an accuracy of approximately 94.6%. The Kaggle dataset described in [24] provides annotated burn images for model training and validation. Finally, Wilson *et al.* [25] reviewed ML and optical imaging methods for burn severity assessment and discussed their principal limitations.

An analysis of the reviewed studies indicates that DL methods achieve reported accuracies of approximately 85–94% in burn classification, whereas conventional ML approaches based on texture features achieve accuracies of approximately 65–80%. However, the influence of different feature selection methods

on classifier performance remains insufficiently investigated, particularly in comparisons between binary and multiclass classification. Therefore, the principal research gap lies in the limited comparative analysis of feature selection methods and ML algorithms across different burn severity classification tasks.

This study proposes a systematic approach that combines texture descriptors, six feature selection methods, and multiple ML classifiers for binary and multiclass burn severity classification. Existing DL methods, in particular CNNs, require large amounts of labelled data, whereas the proposed approach offers greater interpretability, computational efficiency, and suitability for working with small and medium-sized medical datasets. Thus, this approach is valuable for practical application in clinical settings, particularly in resource-constrained environments.

Rapid automated classification of burn severity could provide valuable support for clinical decision-making, particularly during emergency triage. Such systems may accelerate clinical decision-making, reduce diagnostic errors, and improve the accuracy of tissue-damage assessment. This is particularly important in resource-constrained settings, where access to specialised diagnostic methods and experienced specialists may be limited.

2. METHOD

The dataset was compiled from publicly available images retrieved from the Kaggle and Roboflow repositories. Owing to the characteristics of these open-access datasets, detailed demographic metadata, including patient age, ethnicity, and skin type, were unavailable for some images. To mitigate this limitation, the dataset was reviewed by a burn-care specialist and curated to ensure a balanced distribution across burn severity classes. Visual inspection confirmed variability in skin tones and imaging conditions, indicating a degree of data diversity. However, the absence of explicit demographic annotations remains a limitation of this study. A burn-care specialist assisted in verifying the class labels and manually delineating the region of interests (ROIs) to capture clinically relevant areas of burned tissue. Although manual ROI delineation may introduce a degree of subjectivity, consistent selection criteria were applied across all samples. The incorporation of automated segmentation methods could further reduce operator dependence and enhance the reproducibility of the classification pipeline. For multiclass classification, the dataset comprised 1,500 images evenly distributed across three severity classes: mild, moderate, and severe. The binary classification dataset comprised 1,000 images evenly distributed between the mild and severe classes. Each dataset was split into a 70% training set and a 30% test set through stratified sampling to preserve the class distribution across both subsets. To ensure reproducibility, a fixed random seed (random_state=100) was applied during the data splitting process. Model performance was subsequently evaluated through 10-fold cross-validation on the training set to obtain more robust performance estimates and reduce evaluation variability. All images were converted from RGB to 8-bit grayscale and normalized to minimize brightness variation while preserving natural texture characteristics. Each image was cropped to 200×200 pixels before analysis. Figure 1 illustrates representative images of burn injuries from the dataset used in this study, the three burn severity classes considered in the classification task. Specifically, Figures 1(a)–(c) present examples of first-degree (mild), second-degree (moderate), and third-degree (severe) burns, respectively.



(a) (b) (c)

Figure 1. Sample images of the three degrees of skin burns selected in this research; (a) mild (first-degree) burn, (b) moderate (second-degree) burn, and (c) severe (third-degree) burn

Texture features were extracted with MaZda 4.6, which computed 297 descriptors across six feature categories: histogram, GLCM, run-length matrix, gradient map, autoregressive model, and wavelet transform. To reduce feature dimensionality, six selection methods were applied: Fisher, ANOVA, RELIEF,

sequential forward selection (SFS), sequential backward selection (SBS), and recursive feature elimination (RFE). These methods yielded compact subsets containing the most informative features. Figure 2 presents the general workflow of the proposed classification model.

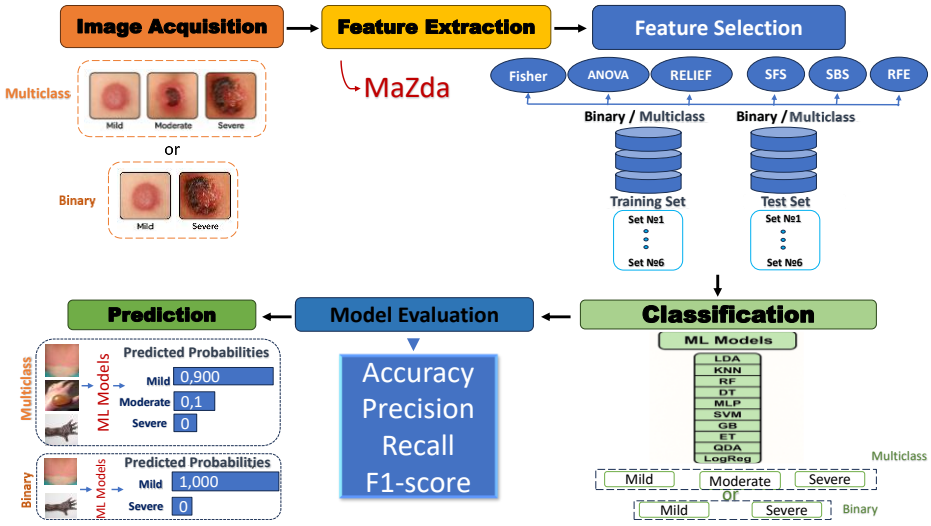


Figure 2. General workflow diagram of the proposed burn lesion classification model

A range of supervised-learning algorithms was evaluated for burn severity classification, including linear discriminant analysis (LDA), SVM, KNN, decision tree (DT), RF, extra trees (ET), multilayer perceptron (MLP), logistic regression (LR), and gradient boosting (GB). Hyperparameters were optimized through grid search, and model performance was assessed in both multiclass and binary settings based on standard evaluation metrics, including accuracy, precision, recall, and F1-score.

All experiments were conducted on a computer equipped with a 12th Gen Intel®, Core™ i5-12500H processor operating at 2.50 GHz, 8 GB of RAM, and Intel® Iris® Xe Graphics, running Windows 11 64-bit. The experiments were implemented in Python 3.12.4 with the scikit-learn library (version 1.8.0). graphics processing unit (GPU) acceleration was not employed.

The implementation of the proposed pipeline, including preprocessing, feature extraction, feature selection, and classification stages, is available in a public GitHub repository: https://github.com/Zaure1982376/Burn severity-classification-via-texture-analysis-and-supervised-learning.

3. RESULTS AND DISCUSSION

This section presents the outcomes of model training and evaluation for both multiclass and binary classification of burn severity levels. Figure 3 presents the training and validation accuracy curves for multiclass classification. Among the tested models, the RFE-MLP achieved the highest validation accuracy of 79%, combining high precision with low model complexity. These results demonstrate the effectiveness of RFE for identifying a compact feature subset for multiclass burn severity classification.

Figure 4 presents the binary classification results, in which burn images were classified as either mild or severe. The RELIEF–MLP configuration achieved the highest validation accuracy of 96%, demonstrating its effectiveness in distinguishing between mild and severe burns.

The study evaluated numerous combinations of feature selection methods, classifiers, and feature-subset sizes. Table 1 summarizes the representative results for multiclass and binary classification, including the performance metrics, number of selected features, and optimal model hyperparameters. Models with a reduced number of features were preferred, as they mitigate overfitting, eliminate redundant information, and improve generalization while maintaining comparable classification performance. The performance across all evaluated model and feature selection combinations ranged from 60% to 79% for multiclass classification. For binary classification, accuracy ranged from 76% to 96%. These results indicate greater separability in the binary setting compared to the multiclass task. The lower accuracy observed in multiclass classification can be attributed to the greater complexity of distinguishing among three burn severity levels and to the partial overlap of texture characteristics between adjacent severity classes. Table 2 summarizes the main hyperparameters considered during classifier optimization.

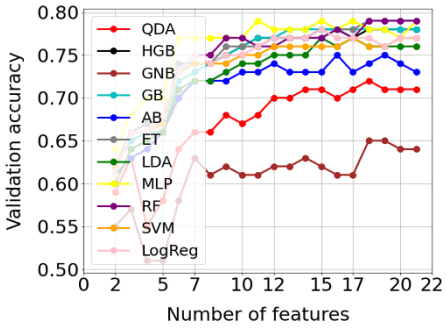


Figure 3. Training and validation accuracy curves obtained with RFE for multiclass classification

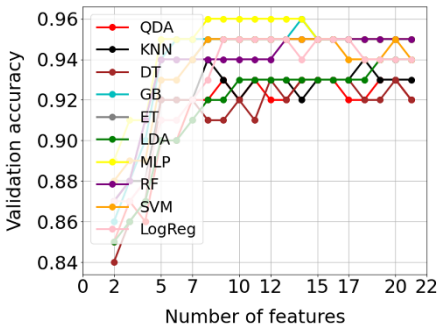


Figure 4. Training and validation accuracy curves obtained with RELIEF for binary classification

Table 1. Optimal model and feature selection configurations for multiclass and binary classification

Feature selection	Classifier	Task type	Number of features	Accuracy	Precision	Recall	F1-score	Optimal model parameters
RFE	MLP	Multiclass	11	0.79	0.78	0.78	0.78	activation: 'relu', alpha: 0.0001, hidden_layer_sizes: (5,), max_iter: 20000, solver: 'lbfgs'
RELIEF	MLP	Binary	8	0.96	0.95	0.95	0.95	{'clf__activation': 'relu', 'clf__alpha': 1.0, 'clf__hidden_layer_sizes': (6,), 'clf__max_iter': 20000, 'clf__solver': 'lbfgs'}
FISHER	ET	Multiclass	16	0.79	0.78	0.78	0.78	max_depth: 20, min_samples_split: 5, n_estimators: 200
SBS	MLP	Binary	19	0.96	0.95	0.94	0.95	{'clf__activation': 'relu', 'clf__alpha': 1.0, 'clf__hidden_layer_sizes': (6,), 'clf__max_iter': 20000, 'clf__solver': 'lbfgs'}

Table 2. Hyperparameter search spaces for classifier optimization

Classifier	Hyperparameter	Search values
LDA	solver	{svd, lsqr, eigen}
SVM	C	{0.001, 0.01, 0.1, 1.0}
	kernel	{linear, poly, rbf, sigmoid}
	gamma	{scale, auto}
	probability	{True}
KNN	n_neighbors	{1,2,3,4,5,6,7,8,9,10}
DT	criterion	{gini, entropy}
	max_depth	{1,2,3,4,5,6,7,8,9,10}
MLP	hidden_layer_sizes	{(1,), (2,), (3,), (4,), (5,), (6,), (7,), (8,), (9,), (10,)}
	activation	{relu}
	solver	{lbfgs}
	alpha	{0.0001, 0.001, 0.01, 0.1, 1.0}
	max_iter	{20000}
RF	n_estimators	{50, 100, 150, 300, 500}
	max_depth	{2, 3, 4, 5, 6, 7, 8, 9, 10}
GB	n_estimators	{50, 100, 200}
	learning_rate	{0.01, 0.1, 0.2}
	max_depth	{3, 5, 7}
AdaBoost	n_estimators	{50, 100, 200}
	learning_rate	{0.01, 0.1, 1}
HistGradientBoosting	max_iter	{100, 200, 300}
	learning_rate	{0.01, 0.1, 0.2}
	max_depth	{None, 3, 5}
ET	n_estimators	{50, 100, 200}
	max_depth	{None, 10, 20}
	min_samples_split	{2, 5, 10}
Qualitative data analysis (QDA)	reg_param	{0.0, 0.1, 0.5, 1.0}
GNB	var_smoothing	{1e-9, 1e-8, 1e-7, 1e-6}
LogReg	C	{0.01, 0.1, 1, 10}
	solver	{liblinear, lbfgs, saga}

The optimal hyperparameter values for the best-performing models are reported in the last column of Table 1. Table 3 summarizes the multiclass and binary classification results in terms of precision, recall, and F1-score. The results show that the largest classification errors in the multiclass task occur for Class 1, whereas the binary classification models demonstrate higher precision, recall, and F1-score values.

Table 3. Evaluation results for the multiclass and binary classification models

Model	Class	Precision	Recall	F1-score
MODEL_11_RFE_MLP	Class0	0.77	0.84	0.80
	Class1	0.71	0.67	0.69
	Class2	0.86	0.84	0.85
MODEL_16_FISHER_ET	Class0	0.74	0.77	0.75
	Class1	0.69	0.64	0.66
	Class2	0.85	0.87	0.86
MODEL_8_RELIEF_MLP	Class0	0.93	0.97	0.95
	Class1	0.97	0.91	0.94
MODEL_19_SBS_MLP	Class0	0.93	0.97	0.95
	Class1	0.97	0.91	0.94

Figure 5 compares representative ROIs from the three burn severity classes-Figure 5(a) mild, Figure 5(b) moderate, and Figure 5(c) severe - with their corresponding GLCMs, calculated at a pixel distance of d=1 and orientations θ=0°, 45°, 90°, 135°. Here, *i* represents the grey level of the reference pixel (Y-axis), while *j* represents the grey level of its neighbouring pixel (X-axis) at the specified distance and orientation.

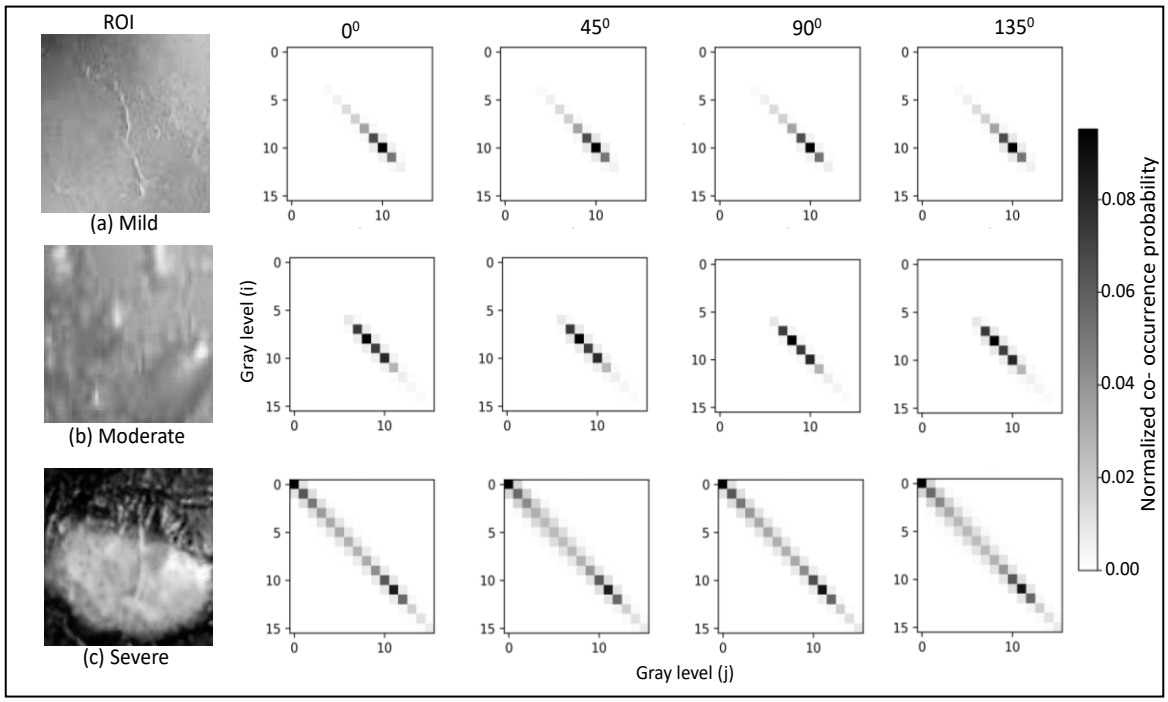


Figure 5. Normalized GLCM for three burn severity classes; (a) mild, (b) moderate, and (c) severe burn ROIs, computed at four orientations (θ=0°, 45°, 90°, 135°)

Figure 6 presents the confusion matrix of the best multiclass classification model (RFE-MLP). The model demonstrates effective recognition across all three burn severity classes, with the highest classification performance observed for the severe burn category.

Figure 7 displays the confusion matrix of the best binary classification model (RELIEF-MLP). The model demonstrates high classification performance in distinguishing between mild and severe burns, consistent with the precision, recall, and F1-score values reported in Table 3.

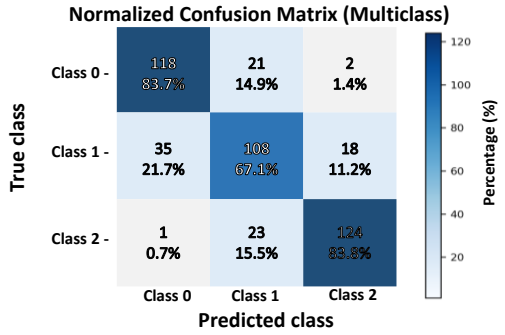


Figure 6. Confusion matrix of the best multiclass classification model (RFE-MLP)

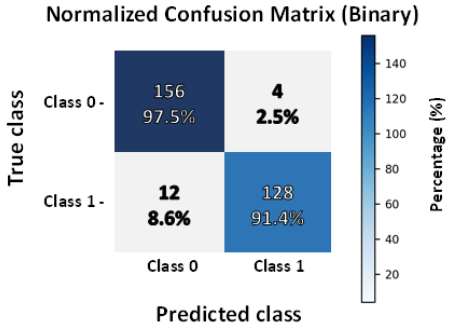


Figure 7. Confusion matrix of the best binary classification model (RELIEF-MLP)

Overall, the findings indicate that texture-based ML models, particularly MLP classifiers combined with RFE or RELIEF, offer accurate and computationally efficient solutions for burn severity classification. The binary classifier achieved higher accuracy than the multiclass classifier (96% vs. 79%), which is expected given the higher complexity of differentiating between three severity levels. The results also demonstrate that a reduced feature set (8–11 features) is sufficient to achieve high performance, confirming that effective feature selection can substantially enhance accuracy and interpretability while minimizing computational costs. The superior performance of MLP combined with RFE and RELIEF can be attributed to the complementary roles of feature selection and nonlinear classification. Starting from the original set of 297 texture descriptors, RFE and RELIEF reduce redundant and weakly discriminative information and retain compact subsets of features that are more relevant to burn severity differentiation. The MLP classifier can then model nonlinear relationships among these selected texture characteristics. This combination reduces the dimensionality of the input space while preserving discriminative information, which may explain the high accuracy achieved with only 11 features for multiclass classification and 8 features for binary classification. Although the present study did not report the identities and relative contributions of the selected features, the reduction from 297 descriptors to 11 and 8 features suggests that compact feature subsets retained substantial discriminative information. A detailed analysis of the contribution and ranking of individual texture features could provide additional insight into their diagnostic relevance and enhance the interpretability of the classification models. To validate the effectiveness of the proposed methodology, the obtained results were compared with other recent studies on automated burn injury classification. As shown in Table 4, the proposed approach demonstrates competitive performance relative to previously reported DL and hybrid methods. However, direct comparison should be interpreted with caution because the studies differ in dataset composition, sample size, burn severity definitions, preprocessing procedures, and evaluation protocols. In contrast to end-to-end DL approaches, the proposed method combines the explicit extraction of texture descriptors with subsequent feature selection and conventional ML classification. This framework offers a computationally efficient alternative that can be trained and evaluated without GPU acceleration; however, its performance remains dependent on the quality of the manually delineated ROIs and predefined texture features. In contrast, DL methods learn hierarchical image representations directly from input data. Their computational cost and generalization performance depend on the model architecture, dataset, and training procedure.

The proposed approach has potential for future application in clinical decision-support systems. Following external validation on independently collected clinical data, the developed models could potentially be incorporated into diagnostic software or lightweight applications for rapid burn severity assessment, particularly in settings with limited access to specialist care. Potential extensions of the proposed system include clinical validation involving medical experts, expansion of the model's functional capabilities, and adaptation to real-world healthcare environments, including telemedicine applications. Furthermore, deployment in real-world clinical environments requires consideration of variations in operator practices and imaging devices, together with compatibility with existing hospital information systems. Future work should therefore assess the robustness of the proposed pipeline under varying acquisition conditions (e.g., different cameras or lighting) and explore lightweight integration strategies compatible with legacy hospital diagnostic infrastructure, minimizing the need for specialized hardware.

Error analysis: most misclassifications occurred between neighbouring classes, particularly between mild and moderate burns, which exhibited similar texture patterns. In contrast, severe burns are more accurately classified, as they exhibit more distinctive features. These findings indicate that intermediate classes are more difficult to separate, which affects the overall performance in the multiclass setting. This

behavior is also reflected in Table 3 and the multiclass confusion matrix Figure 6, where the largest classification errors are observed for the intermediate severity class.

Limitations: despite the promising results, this study has several limitations, including the limited dataset size, heterogeneous image-acquisition conditions, reliance on publicly available images, and subjectivity inherent in manual ROI delineation. These factors may limit the generalizability of the developed models. Furthermore, because the images were obtained from publicly available repositories and complete demographic and image acquisition metadata were unavailable, the generalizability of the proposed models across different patient populations, skin types, imaging devices, and clinical settings cannot yet be fully established. External validation on independently collected clinical datasets with standardized metadata is therefore required before the proposed approach can be considered for clinical deployment.

Table 4. Comparison with previous studies

No	Study objective	Method type	Methods	Accuracy (%)	Reference
1	Burn images (1st, 2nd, and 3rd degree)	ML (texture-based)	MLP+RFE	79	This study
2	Burn images (1st and 3rd degree)	ML (texture-based)	MLP+RELIEF	96	This study
3	Diagnosis of skin diseases	DL	CNN+transfer learning	-	[6]–[8]
4	Acute burn assessment and surgery prediction	DL	CNN+SVM	64.7	[11]
5	Multiclass classification	DL	VGG16+transfer learning	95.63	[12]
6	Burn severity assessment	Hybrid/DL	CNN+SVM, body-part-specific model	84.85	[14]
7	Burn segmentation and classification	DL	YOLOv7	75.12	[18]
8	Burn depth classification	ML (feature based)	Feature extraction+SVM	82.43, 80.74	[20], [22]
9	Skins burn color-image classification	ML/DL	Texture features+ANN/SegNet, Color+texture+depth features+RF	-	[21]
10	Burn severity prediction	Hybrid	CNN+ML/ANN	-	[23]
11	Burn diagnosis (feature-based)	ML (texture-based)	Feature extraction+SVM/ML	-	[24]

Note: references [7]–[16], [17]–[25] correspond to articles listed in the bibliography of this paper and reflect recent research in the field of computer-aided burn diagnosis.

4. CONCLUSION

This study assessed the feasibility of automated burn severity classification based on texture features extracted from medical images. The analysis encompassed a complete image-processing pipeline comprising image acquisition, preprocessing, manual ROI delineation, extraction of 297 texture features with the specialised MaZda software, and subsequent feature selection through six filter and wrapper methods: Fisher score, ANOVA, RELIEF, SFS, SBS, and RFE. Multiple ML classifiers were subsequently trained and evaluated. The configurations that achieved the highest classification accuracy combined the MLP classifier with different feature selection methods. For multiclass classification, MLP combined with RFE achieved an accuracy of 79%, whereas for binary classification, MLP combined with RELIEF achieved an accuracy of 96%. The optimal configurations comprised 11 features for multiclass classification and eight features for binary classification. These results indicate that compact feature subsets can retain substantial discriminative information.

Despite these promising results, a fully integrated and deployable diagnostic system has not yet been implemented. Therefore, future research and development directions include: i) development of a complete software solution with an integrated user interface and an embedded classification module; ii) implementation of an automatic roi segmentation algorithm to reduce dependence on manual roi selection; iii) integration and adaptation of DL models, particularly cnns, to enhance classification accuracy and facilitate automated feature extraction; iv) expansion of the dataset by incorporating additional medical images acquired under real-world clinical conditions; and v) clinical validation of the proposed methodology involving medical experts.

Overall, the experimental results demonstrate the potential of texture-based ML for automated burn severity classification. Compact feature subsets combined with conventional ML models provide a computationally efficient framework that does not require GPU acceleration. However, the present findings should be considered preliminary because of the limitations of the publicly available dataset, manual ROI selection, and the absence of external clinical validation. Further evaluation on independently collected clinical data is required to establish the generalizability and practical applicability of the proposed approach before its potential integration into clinical decision-support systems.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the expert assistance of the combustiologists and surgeons at the Academician N. D. Batpenov National Scientific Center for Traumatology and Orthopedics (Astana, Kazakhstan) in identifying and delineating regions of interest in the medical images.

FUNDING INFORMATION

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan, grant number AP23486396.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Zaure Kabdrakhmanova	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓
Aizhan Tlebaldinova		✓				✓		✓	✓	✓	✓	✓		✓
Zbigniew Omiotek	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓
Markhaba Karmenova		✓				✓			✓	✓				✓
Symbat Nurgaliyeva	✓				✓		✓		✓	✓		✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

INFORMED CONSENT

This study involved the secondary analysis of publicly available, de-identified images obtained from the Kaggle and Roboflow repositories. The authors did not directly recruit participants or collect data from patients. Therefore, additional informed consent was not required for this study.

ETHICAL APPROVAL

This study was based exclusively on publicly available, de-identified image datasets obtained from Kaggle and Roboflow. No participants were directly recruited, and the authors did not collect or process any personally identifiable information. Therefore, institutional ethical approval was not required for this secondary analysis of publicly available data.

DATA AVAILABILITY

The dataset supporting the findings of this study is publicly available on Kaggle (https://www.kaggle.com/datasets/shubhambaid/skin-burn-dataset). The code used for image preprocessing, feature extraction, and classification is available in a public GitHub repository: https://github.com/Zaure1982376/Burn severity-classification-via-texture-analysis-and-supervised-learning.

REFERENCES

[1] S. Huang, H.-Z. Lin, and X. Wei, "Global, regional and national burden of injuries caused by fire, heat, and hot substances from 1990 to 2021," *PLOS One*, vol. 20, no. 5, May 2025, doi: 10.1371/journal.pone.0324481.

[2] B. Rostami, D. M. Anisuzzaman, C. Wang, S. Gopalakrishnan, J. Niezgoda, and Z. Yu, "Multiclass wound image classification using an ensemble deep CNN-based classifier," *Computers in Biology and Medicine*, vol. 134, p. 104536, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104536.

[3] D. M. Anisuzzaman, Y. Patel, B. Rostami, J. Niezgoda, S. Gopalakrishnan, and Z. Yu, "Multi-modal wound classification using wound image and location by deep neural network," *Scientific Reports*, vol. 12, no. 1, Nov. 2022, doi: 10.1038/s41598-022-21813-0.
[4] P. K. B. Rangaiah, B. P. P. Kumar, F. Huss, and R. Augustine, "Precision diagnosis of burn injuries using imaging and predictive modeling for clinical applications," *Scientific Reports*, vol. 15, no. 1, Mar. 2025, doi: 10.1038/s41598-025-92096-4.
[5] A. Asif, C. Poyiatzis, F. J. Raheman, and D. M. Rojoa, "The use of infrared thermography (IRT) in burns depth assessment: A diagnostic accuracy meta-analysis," *European Burn Journal*, vol. 3, no. 3, pp. 432–446, Sep. 2022, doi: 10.3390/ebj3030038.
[6] E. S. Jaén, Z. Omiotek, and C. Pinzón, "Web-based system for the diagnosis of skin lesions using deep convolutional neural networks and transfer learning techniques," in *Proceedings of the 22nd LACCEI International Multi-Conference for Engineering, Education and Technology (LACCEI 2024): "Sustainable Engineering for a Diverse, Equitable, and Inclusive Future at the Service of Education, Research, and Industry for a Societ*, 2024, doi: 10.18687/LACCEI2024.1.1.940.
[7] S. Aksoy, P. Demircioglu, and I. Bogrekci, "Deep learning-based web application for automated skin lesion classification and analysis," *Dermato*, vol. 5, no. 2, Apr. 2025, doi: 10.3390/dermato5020007.
[8] C. W. Chang *et al.*, "Application of multiple deep learning models for automatic burn wound assessment," *Burns*, vol. 49, no. 5, pp. 1039–1051, Aug. 2023, doi: 10.1016/j.burns.2022.07.006.
[9] A. Elsarta *et al.*, "Integrating multi-source data for skin burn classification using deep learning," *Computers in Biology and Medicine*, vol. 195, Sep. 2025, doi: 10.1016/j.compbiomed.2025.110556.
[10] S. Lee *et al.*, "A deep learning model for burn depth classification using ultrasound imaging," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 125, Jan. 2022, doi: 10.1016/j.jmbbm.2021.104930.
[11] C. Boissin *et al.*, "Development and evaluation of deep learning algorithms for assessment of acute burns and the need for surgery," *Scientific Reports*, vol. 13, no. 1, Jan. 2023, doi: 10.1038/s41598-023-28164-4.
[12] S. A. Suha and T. F. Sanam, "A deep convolutional neural network-based approach for detecting burn severity from skin burn images," *Machine Learning with Applications*, vol. 9, Sep. 2022, doi: 10.1016/j.mlwa.2022.100371.
[13] J. J. Lee *et al.*, "Comparing artificial intelligence guided image assessment to current methods of burn assessment," *Journal of Burn Care & Research*, vol. 46, no. 1, pp. 6–13, Jan. 2025, doi: 10.1093/jbcr/irae121.
[14] J. Chauhan and P. Goyal, "BPBSAM: Body part-specific burn severity assessment model," *Burns*, vol. 46, no. 6, pp. 1407–1423, Sep. 2020, doi: 10.1016/j.burns.2020.03.007.
[15] M. D. Cirillo, R. Mirdell, F. Sjöberg, and T. D. Pham, "Time-independent prediction of burn depth using deep convolutional neural networks," *Journal of Burn Care & Research*, vol. 40, no. 6, pp. 857–863, Oct. 2019, doi: 10.1093/jbcr/irz103.
[16] H. Liu, K. Yue, S. Cheng, W. Li, and Z. Fu, "A framework for automatic burn image segmentation and burn depth diagnosis using deep learning," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–12, Apr. 2021, doi: 10.1155/2021/5514224.
[17] K. Dittakan and S. Nookong, "Burn wound severity detection using segmentation model: A comparative study between U-NET and YOLOv8," *Journal of Applied Informatics and Technology*, vol. 7, no. 1, pp. 153–173, 2024.
[18] M. Yıldız *et al.*, "Segmentation and classification of skin burn images with artificial intelligence: Development of a mobile application," *Burns*, vol. 50, no. 4, pp. 966–979, May 2024, doi: 10.1016/j.burns.2024.01.007.
[19] P. N. Kuan, S. Chua, E. B. Safawi, and H. H. Wang, "A comparative study of segmentation algorithms in the classification of human skin burn depth," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 10, no. 1, pp. 145–150, Feb. 2020, doi: 10.18517/ijaseit.10.1.10227.
[20] D. P. Yadav, A. Sharma, M. Singh, and A. Goyal, "Feature extraction based machine learning for human burn diagnosis from burn images," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–7, 2019, doi: 10.1109/JTEHM.2019.2923628.
[21] U. Şevik, E. Karakullukçu, T. Berber, Y. Akbaş, and S. Türkyılmaz, "Automatic classification of skin burn colour images using texture-based feature extraction," *IET Image Processing*, vol. 13, no. 11, pp. 2018–2028, Sep. 2019, doi: 10.1049/iet-ipr.2018.5899.
[22] X. Zhang, Q. Zhang, P. Li, J. You, J. Sun, and J. Zhou, "Multi-feature extraction and selection method to diagnose burn depth from burn images," *Electronics*, vol. 13, no. 18, Sep. 2024, doi: 10.3390/electronics13183665.
[23] S. A. Suha and T. F. Sanam, "Predicting burn injury intensity from skin burn images with hybrid machine learning technique," in *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*, Dec. 2022, pp. 5–8, doi: 10.1109/ICECE57408.2022.10088521.
[24] "Skin burn severity detection dataset," *Kaggle*. https://www.kaggle.com/datasets/shubhambaid/skin-burn-dataset, (accessed Nov. 28, 2024).
[25] R. H. Wilson *et al.*, "Review of machine learning for optical imaging of burn wound severity assessment," *Journal of Biomedical Optics*, vol. 29, no. 02, pp. 1–18, Feb. 2024, doi: 10.1117/1.JBO.29.2.020901.

BIOGRAPHIES OF AUTHORS



Zaure Kabdrakhmanova is currently a doctoral student at the School of Digital Technology and Artificial Intelligence (AI), D. Serikbayev East Kazakhstan Technical University, Kazakhstan. She was born in August 1982. In 2006, she obtained her bachelor's degree from the Faculty of Physics, Mathematics and Technology at Sarsen Amanzholov East Kazakhstan University, Kazakhstan. From 2006 to 2012, she worked as a lecturer at the Department of Mathematical Modelling and Computer Technologies of the same university. In 2014, she received her master's degree in Pedagogical Sciences, specializing in Computer Science. In 2023, she enrolled in a doctoral programme at D. Serikbayev East Kazakhstan Technical University. She can be contacted at email: zaure_k_82@mail.ru.



Aizhan Tlebaldinova is an associate professor at the School of Digital Technology and AI, D. Serikbayev East Kazakhstan Technical University, Kazakhstan. She received her B.Sc. and M.Sc. degrees in Applied Mathematics from East Kazakhstan State University, Kazakhstan. In 2016, she was awarded her Ph.D. in Information Systems (by industry) from D. Serikbayev East Kazakhstan Technical University. Her primary research interests include big data, data mining, ML, image processing, and pattern recognition. She can be contacted at email: tlebaldinovazhan@gmail.com.



Zbigniew Omiotek is a professor at the Faculty of Electrical Engineering and Computer Science, Department of Electronics and Information Technology, Lublin University of Technology, Poland (since 2022). He received his Ph.D. in Electrical Engineering from Lublin University of Technology in 2014. In 2022, he earned the doctor of science (D.Sc., habilitation) in Automatics, Electronics, and Electrical Engineering for work on ML algorithms in combustion-process diagnostics and medical diagnosis. His research focuses on traditional and deep learning models for technical and medical diagnostics, including detection and segmentation in computer vision, analysis of physiological and biomedical signals (EMG, MMG, ECG, respiratory sounds), medical imaging (CT, X-ray, MRI, ultrasound), combustion flame-intensity signals, and microscopic imaging of hydrophobized building materials. He can be contacted at email: z.omiotek@pollub.pl.



Markhaba Karmenova is an associate professor of Information Technology at the School of Natural Sciences, Sarsen Amanzholov East Kazakhstan University, Kazakhstan. She holds an M.Sc. in Computer Science from East Kazakhstan State University, Kazakhstan, and a Ph.D. in Information Systems (by industry) from D. Serikbayev East Kazakhstan Technical University, Kazakhstan. She has authored and co-authored more than 15 research papers. Her research interests include the applications of AI, spatial data analysis, ML, and IT for education and training. She can be contacted at email: mkarmenova@vku.edu.kz.



Symbat Nurgaliyeva is an assistant professor at Astana IT University, Kazakhstan. She obtained her bachelor's degree in Computer Science from S. Amanzholov East Kazakhstan State University, Ust-Kamenogorsk, Kazakhstan, in 2010, and her master's degree in Computer Science from the same university in 2014. She holds a Ph.D. in Computer Science, with her doctoral dissertation focused on enhancing the quality of education for future computer science teachers through the theoretical and practical integration of robotics technologies, specifically the use of mobile robots in the educational process. Her research interests include robotics technologies in education, the creation of mobile robots, ML, and neural networks. She can be contacted at email: symbat.nurgaliyeva@astanait.edu.kz.