



# A Hybrid Vision Transformer–Driven Feature Extraction and Machine Learning Framework for Automated Skin Burn Detection



Ahmet Yasir Duman<sup>1</sup>, Mucahit Karaduman<sup>2</sup>, Muhammed Yildirim<sup>1\*</sup>

<sup>1</sup> Department of Computer Engineering, Malatya Turgut Ozal University, 44200 Malatya, Turkey

<sup>2</sup> Department of Software Engineering, Malatya Turgut Ozal University, 44200 Malatya, Turkey

\* Correspondence: Muhammed Yildirim (muhammed.yildirim@ozal.edu.tr)

Received: 11-03-2025

Revised: 11-28-2025

Accepted: 12-07-2025

**Citation:** A. Y. Duman, M. Karaduman, and M. Yildirim, “A hybrid Vision Transformer–driven feature extraction and machine learning framework for automated skin burn detection,” *Inf. Dyn. Appl.*, vol. 4, no. 4, pp. 201–211, 2025. <https://doi.org/10.56578/ida040402>.



© 2025 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

**Abstract:** Skin burns represent a major clinical concern due to their association with pain, functional impairment, sensory damage, and even life-threatening complications. Early and accurate assessment is critical for first aid, clinical intervention, and the prevention of secondary complications. However, conventional burn diagnosis remains highly dependent on visual inspection and clinical expertise, which can introduce subjectivity and delay timely decision-making. To address these limitations, a hybrid automated skin burn detection framework was proposed, integrating transformer-based feature extraction with classical machine learning classification. In this framework, discriminative visual features were extracted using multiple Vision Transformer (ViT) architectures, including ViT-B/16, ViT-L/16, ViT-B/32, and DINOv2 (a self-supervised Vision Transformer model). The extracted features were subsequently fused. Given the resulting high-dimensional feature space, dimensionality reduction was performed using the Chi-square ( $\text{Chi}^2$ ) algorithm, through which 500 features were retained, reducing computational complexity and mitigating the risk of model overfitting. The reduced feature set was then employed for burn classification using six classifiers. Model effectiveness was assessed using accuracy, precision, sensitivity, and F1-score metrics. Experimental results demonstrated that the Support Vector Machine (SVM) classifier achieved the highest classification performance, yielding an accuracy of 82.29%. Comparable yet slightly lower accuracies were observed for the Light Gradient Boosting Machine (LGBM) (80.51%) and Extreme Gradient Boosting (XGBoost) (80.17%) classifiers. Overall, the proposed hybrid model consistently outperformed baseline models, highlighting its superior discriminative capability. These findings indicate that the proposed framework holds strong potential for integration into clinical decision support systems, offering a reliable and objective tool for automated skin burn detection.

**Keywords:** Artificial intelligence; Chi-square; Classifiers; Skin burn; Vision Transformer

## 1 Introduction

Burn injuries are one of the most common types of trauma. A burn is an injury caused by various factors, including hot objects, flames, chemicals, electric currents, radiation, and friction, damaging tissue. Burns affect people both physically and psychologically. If not treated appropriately, they can have fatal consequences [1]. Skin burns not only cause tissue and organ damage but also profoundly impact an individual’s mental health [2]. Because burns can occur for a variety of reasons, the severity of the injury can vary. To plan the burn treatment process appropriately, it is essential to quickly identify the cause and extent.

Early and correct intervention in burn treatment alleviates patient suffering and significantly reduces the risk of death. Inappropriate or delayed intervention can endanger human life. The integration of developing computer-aided systems into medical diagnostic processes has increased the possibilities for accurate and timely intervention in the healthcare field. In addition to other health studies, there are studies that have used artificial intelligence (AI) to detect burns and demonstrated successful results in treatment processes [3–5]. Skin burns consist of superficial first-degree, partial-thickness second-degree, and full-thickness third-degree burns. Automatic recognition of these graded burns with computer systems is important for clinical treatment processes. Pabitha and Vanathi [6] graded skin burns using the Region-Based Convolutional Neural Network (R-CNN) method and applied masking to these burns. Şevik et al. [7] achieved automatic classification of skin burn images using tissue-based feature extraction. It was found that

the best combination was a multilayer feedforward Artificial Neural Network (ANN) trained with a backpropagation algorithm for classification. With this method, an F-score of 74.28% was obtained.

In addition, Abazari et al. [8] conducted a classification study to select the appropriate treatment for burn wounds based on diagnosis and wound type. Similarly, Elsarta et al. [9] used deep learning methods to classify skin burns. Multi-source data were integrated to develop an AI model. Acha et al. [10] developed a segmentation and classification model, utilizing color and texture information from burn images. An 82.26% success rate was achieved with this model. Kuan et al. [11] conducted a comparative study of 20 classification algorithms for classifying burn depths using an image mining approach. The dataset was divided into both test and training sections and evaluated using 10-fold cross-validation methods. The results showed that the best classification algorithm achieved an average accuracy of 68.9%. In the 10-fold cross-validation evaluation, the best result was determined to be an average of 73.2%. Rahman et al. [12] focused on image-based diagnosis of burn type and depth. Various data augmentation processes were applied to a dataset of 29 patients. The proposed AI model was evaluated using a 5-fold cross-validation method. It was found that the average success rate of the three classes was 79%. Khan et al. [13] developed a model using a Deep Convolutional Neural Network (DCNN) for burn detection. During model development, 65% of the dataset, comprising a total of 450 images, was used for training, while the remaining 35% was used for testing. The results showed that the proposed model achieved a success rate of 79.4%.

In this study, a hybrid deep learning model was developed to classify skin burns with high accuracy. The model combines the powerful representation capabilities of different transformer-based architectures. In the first stage, features were extracted from four different architectures and then combined, aiming to use different features of the same image together. At this stage, these models were not trained and were only used for feature extraction. Feature selection was then performed on the resulting feature map using  $\text{Chi}^2$ . The aim of this step is to make the model run faster and produce more successful results. Finally, the optimized feature map was classified using six different classifiers. In the remainder of this study, the dataset used, the deep learning models, and the proposed model are detailed in Section 2. Experimental results are presented in Section 3. The conclusion is presented in Section 4.

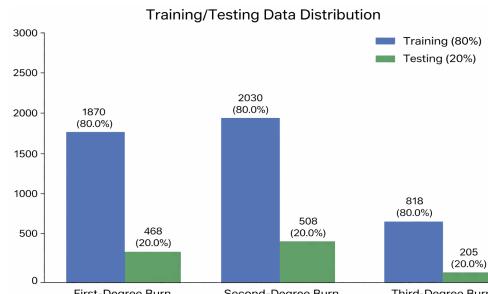
## 2 Methodology

In this section, the burn dataset used in the study is examined. The applied methods are also explained in detail.

### 2.1 Dataset

The dataset used was obtained from the publicly available Kaggle website [14]. It contains a total of 6,099 images. The 200 unlabeled images within the dataset were included solely for testing purposes; therefore, the accuracy of the proposed model's predicted values cannot be determined. Because they are unlabeled, they were excluded from the training model. Therefore, the total dataset consists of 5,899 records, of which 4,719 were used as training data and 1,180 as new test data to determine the model's accuracy. The classes and sample images of the dataset are shown in Figure 1.

Burn severities in the dataset are labeled as first-degree burns, second-degree burns, and third-degree burns. First-degree burns are ranked as the mildest, and third-degree burns are ranked as the most severe. Examples of each degree in the dataset are shown in Figure 2.



**Figure 1.** Dataset distribution



**Figure 2.** Images taken from the dataset

In Figure 2, first-degree burns are shown as the lightest, while third-degree burns are listed as the most severe.

## 2.2 Transformers and Classifiers

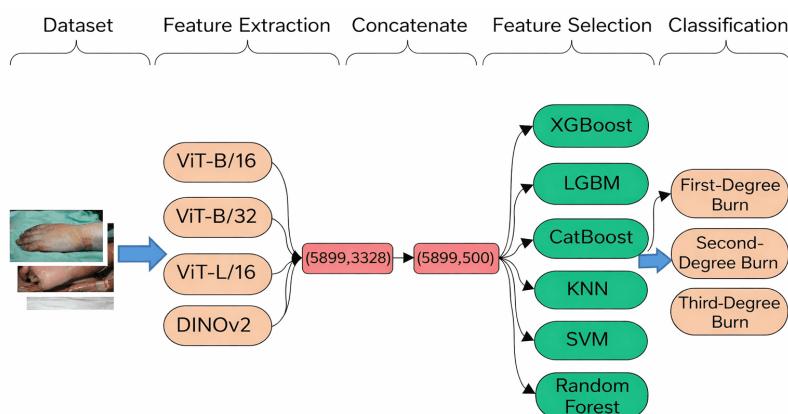
Transformer-based deep learning models, which have achieved significant success in Natural Language Processing (NLP), have demonstrated remarkable performance due to their feature extraction capabilities [15]. Among the key advantages of transformers are their ability to model long-term dependencies between input sequence elements and their capacity to support parallel processing of sequences, which is superior to recurrent networks. These strengths have led to exciting advances in several vision tasks using transformer networks [16]. The Vision Transformer (ViT) models used in this study—ViT-B/16, ViT-L/16, ViT-B/32, and DINOv2 (a self-supervised Vision Transformer model) networks—were employed for feature extraction.

ViT-B/16 is a transformer encoder model pre-trained in a supervised manner on an extensive collection of images called ImageNet-21k [17]. A standard classifier can be trained by placing a linear layer on top of the pre-trained encoder. However, in this study, in addition to being trained as a classifier, it was also used for feature extraction. The ViT-B/32 model differs from ViT-B/16 in that the images are presented to the model as a series of linearly embedded, fixed-size patches, with a resolution of  $32 \times 32$  [18]. This model was utilized to enhance efficiency in the feature extraction portion of this study. In the ViT-L/16 model, although both models belong to the ViT family, the model's capabilities are more comprehensive than the other base models [19]. The DINOv2 model was also used for feature extraction in the ViT models used in this study. The quality of the DINOv2 model was superior at both the image and pixel levels compared to various computer vision models [20]. The DINOv2 model has also been used particularly in medical imaging studies [21]. These features make it a valuable model for feature extraction in this study.

For each model, features obtained using ViTs were compared with those from six different classifiers. Three of these classifiers are XGBoost, Categorical Boosting (CatBoost), and Light Gradient Boosting Machine (LGBM), all of which are gradient boosting methods. The learning procedure in gradient boosting machines sequentially applies new models to ensure a more accurate prediction of the response variable. The basic idea is to create a new learner with maximum correlation with the negative gradient of the loss function [22]. However, the high computational cost of traditional Gradient Boosting Decision Tree (GBDT) algorithms has limited their use in large datasets. In this context, classifiers were created using the LGBM method, the CatBoost method, and the XGBoost method, which offer the advantages of gradient boosting-based algorithms in a faster and more scalable form [23–25]. In this study, in addition to gradient boosting-based methods, k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF) classifiers were also used. Despite its simple structure, the KNN classifier is one of the methods that can produce effective results in classification problems [26]. Furthermore, SVM is a powerful classification method that aims to maximize accuracy while reducing the risk of overfitting [27]. RF creates multiple random decision trees and combines their results to obtain more balanced predictions [28].

## 2.3 Proposed Model for Skin Burn Detection

In the proposed hybrid method for skin burn detection, the ViT-B/16, ViT-L/16, ViT-B/32, and DINOv2 models were used for feature extraction. These obtained features were combined and classified using XGBoost, LGBM, CatBoost, KNN, SVM, and RF. In the feature selection phase, the Chi<sup>2</sup> feature selection method was used to identify the 500 most significant features. The Chi<sup>2</sup> algorithm tests the significance of relationships between features using the Chi<sup>2</sup> statistic and infers feature size [29]. This aims to save time by reducing training time and improving classification performance. The feature vectors, initially obtained with size  $5899 \times 3328$ , were reduced to  $5899 \times 500$  using the Chi<sup>2</sup> algorithm. Figure 3 visually illustrates the general architecture of the proposed hybrid model.



**Figure 3.** Proposed model

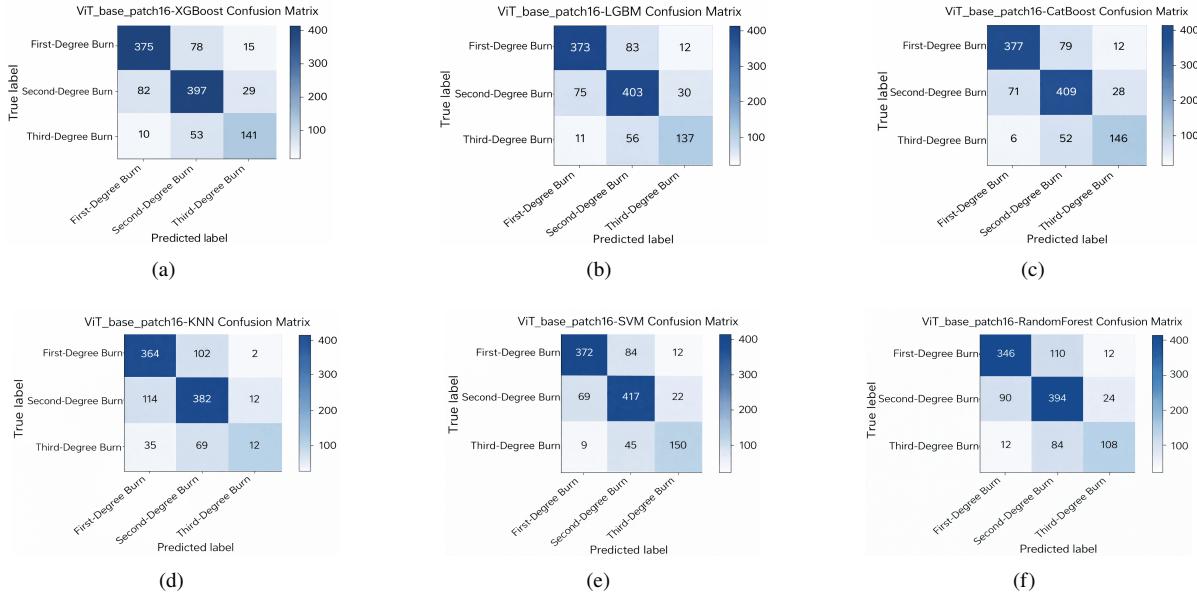
The last step of the proposed model was to evaluate the performance with different metrics after skin burns were detected.

### 3 Results

In this section, the performance of the proposed hybrid model and the other models used for comparison is examined. First, the confusion matrices of the models are presented. The performance of the classification model is measured using the values found in the confusion matrix. Accuracy, recall, precision, and F1-score metrics are used to evaluate the performance of the models [30].

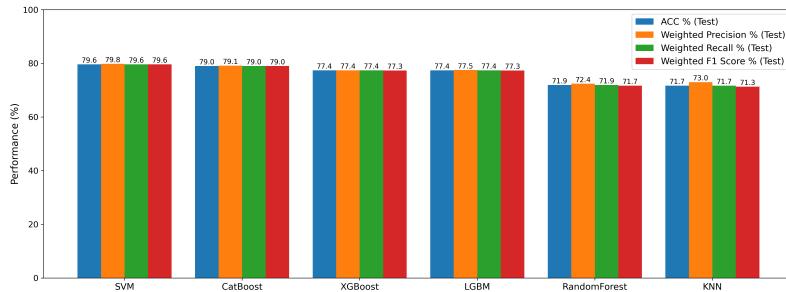
#### 3.1 Evaluation of the ViT-B/16 Model Results for Burn Detection

The features extracted by the ViT-B/16 model used in this study were fed to six different classifiers. The resulting confusion matrices for each of these classifiers are presented in Figure 4.



**Figure 4.** Confusion matrices of features extracted with ViT-B/16 in different classifiers

Figure 4 shows the most successful metrics in the SVM's confusion matrix; however, it's particularly noticeable that the model struggles to classify third-degree burns. Similarly, the results reveal that third-degree burns are generally the most challenging class to predict across all classifiers because some second-degree and third-degree burns are very similar in appearance. The distribution of success scores across classifiers is shown in Figure 5.



**Figure 5.** Comparison of classifier performance using ViT-B/16 features

The evaluation of the classifiers was conducted not only in terms of accuracy but also in terms of time and performance metrics, as shown in Table 1.

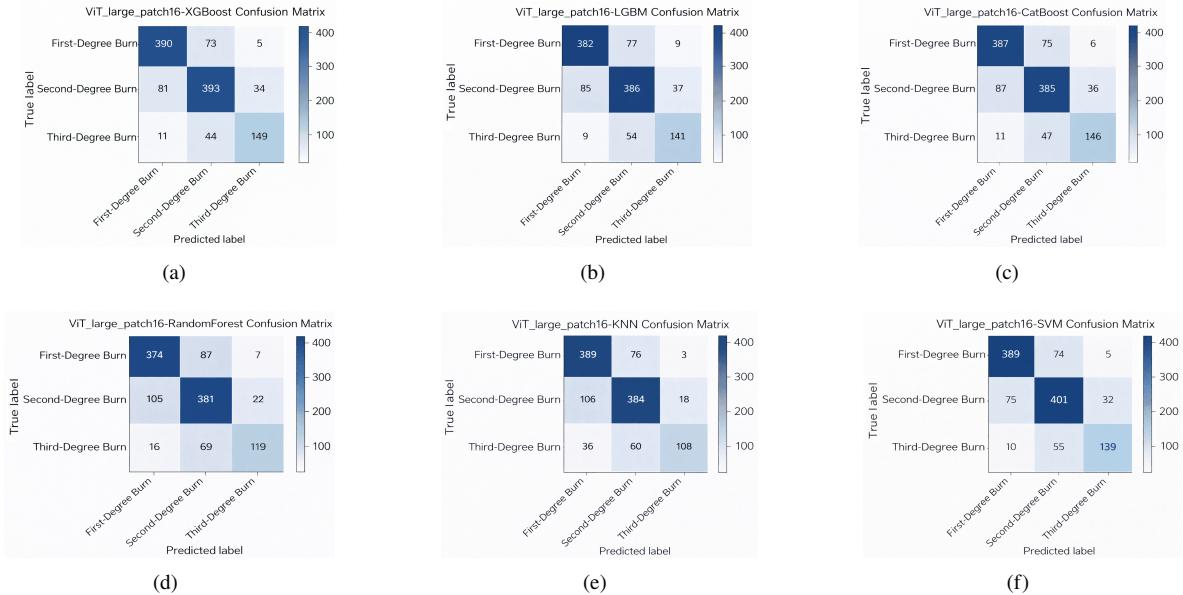
It can be seen from Figure 5 and Table 1 that the highest accuracy value was achieved with the SVM classifier using features extracted from the ViT-B/16 model. This can be attributed to the short training time and its success.

**Table 1.** Comparison of accuracy, training time, and prediction speed of classifiers using ViT-B/16 features

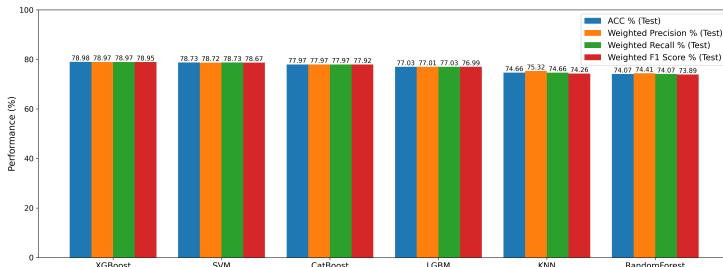
Model	Accuracy (%)	Prediction Speed (images/s)	Training Time (s)
SVM	79.58	616.9	4.41
CatBoost	78.98	32712.55	519.33
XGBoost	77.37	79480.95	72.38
LGBM	77.37	33229.57	40.49
RF	71.86	35962.58	14.87
KNN	71.69	3858.05	0.02

### 3.2 Evaluation of the ViT-L/16 Model Results for Burn Detection

Feature extraction was also performed using the ViT-L/16 model for burn detection. This model provides a more comprehensive feature representation compared to ViT-B/16, and, thanks to its broad structure, can capture detailed information. The extracted features were fed to different classifiers, and the results are presented as confusion matrices for each classifier in Figure 6.



**Figure 6.** Confusion matrices of features extracted with DINOv2 in different classifiers



**Figure 7.** Performance comparison of classifiers using ViT-L/16 features

As shown in the classification results using features obtained from the ViT-L/16 model, the highest accuracy value is achieved with the XGBoost algorithm. However, this success is quite close to that of the SVM classifier. The comparison of performance metrics of classifiers using ViT-L/16 features is presented in Figure 7.

The impact of the models on system performance and the success metrics obtained accordingly are shown in Table 2.

The classification results using features obtained from the ViT-L/16 model show that the XGBoost algorithm achieved the highest accuracy. However, this success is quite similar to that of the SVM classifier. When comparing

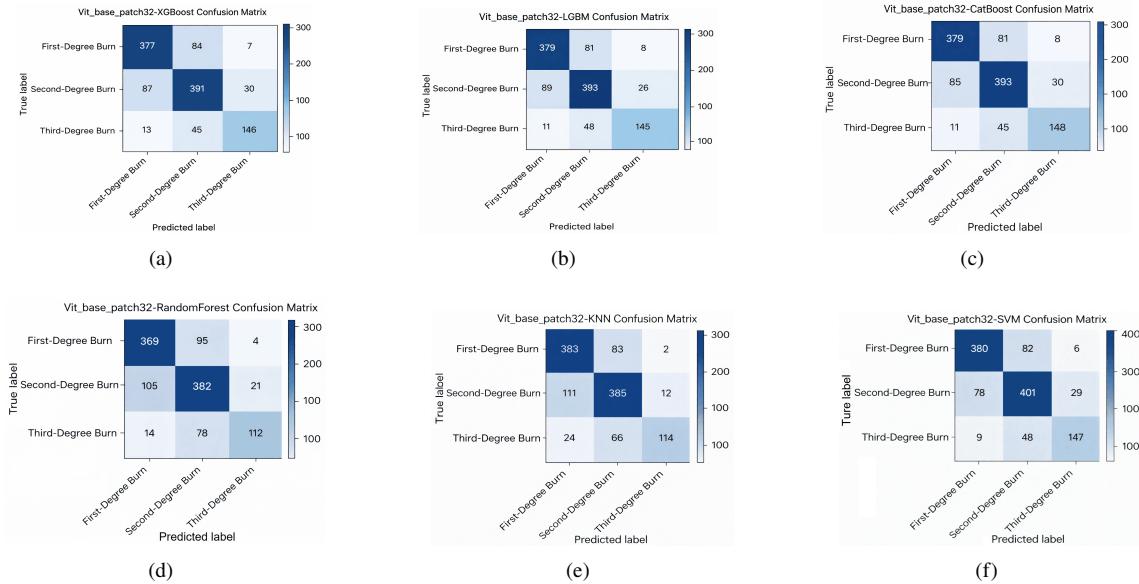
training time and prediction speed, it is evident that SVM is more efficient. However, as with the ViT-B/16 model, accurately classifying third-degree burns remains the most challenging task for all classifiers.

**Table 2.** Comparison of accuracy, training time, and prediction speed of classifiers using ViT-L/16 features

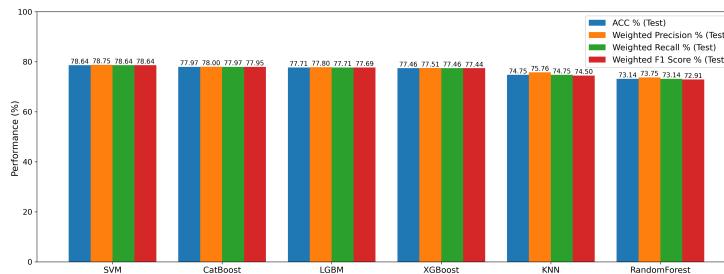
Model	Accuracy (%)	Prediction Speed (images/s)	Training Time (s)
XGBoost	78.98	65080.13	87.24
SVM	78.73	456.84	8.46
CatBoost	77.97	70115.02	639.38
LGBM	77.03	29944.45	55.83
KNN	74.66	3376.54	0.03
RF	74.07	35745.70	46.76

### 3.3 Evaluation of the ViT-B/32 Model Results for Burn Detection

The ViT-B/32 model used for burn detection works with a  $32 \times 32$  patch size, allowing for investigating how resolution-based differences in feature extraction impact model performance. The resulting features were fed to six different classifiers, and the confusion matrices of these classifiers are presented in Figure 8.



**Figure 8.** Confusion matrices of features extracted with ViT-B32 in different classifiers



**Figure 9.** Performance comparison of classifiers using ViT-B/32 features

As shown in the results obtained with the ViT-B/32 model, the SVM classifier achieved the highest accuracy. However, as with previous models (ViT-B/16 and ViT-L/16), correctly classifying third-degree burns proved to be the most challenging task for all classifiers. The comparison of performance metrics of classifiers using ViT-B/32 features is presented in Figure 9.

The system performance of the models was monitored and the resulting values are shown in Table 3.

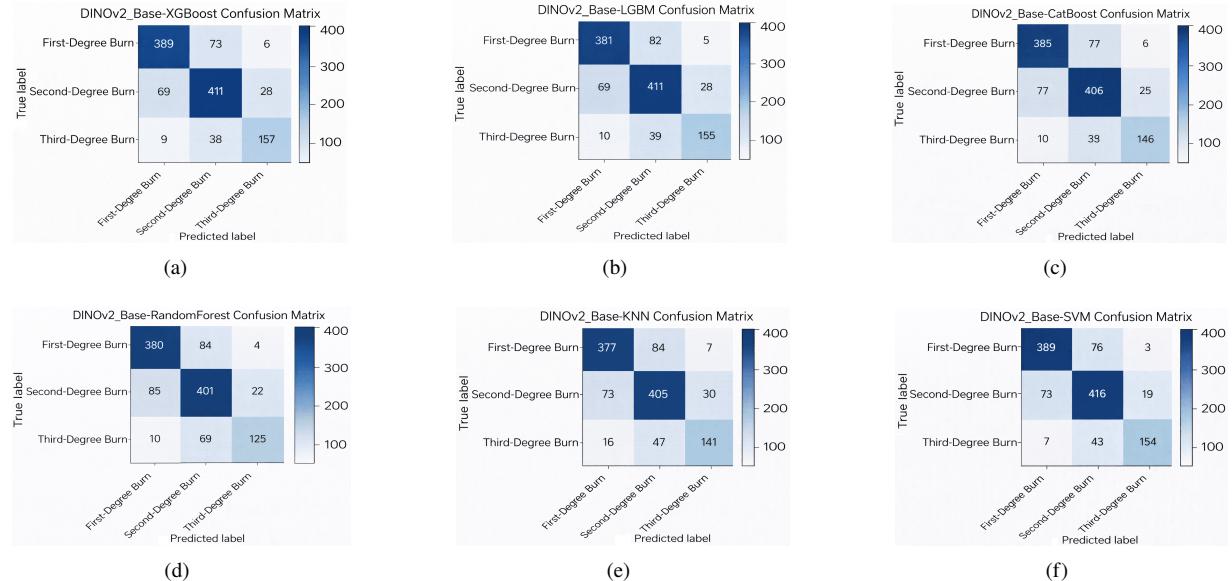
**Table 3.** Comparison of accuracy, training time, and prediction speed of classifiers using ViT-B/32 features

Model	Accuracy (%)	Prediction Speed (images/s)	Training Time (s)
SVM	78.64	473.44	5.09
CatBoost	77.97	91468.68	487.57
LGBM	77.71	28088.98	37.85
XGBoost	77.46	79301	65.22
KNN	74.75	3786.16	0.02
RF	73.14	35175.86	14.26

When Table 3 is examined and evaluated in terms of computational performance, it is determined that the KNN classifier achieves the fastest training time, and the SVM classifier achieves the highest prediction speed.

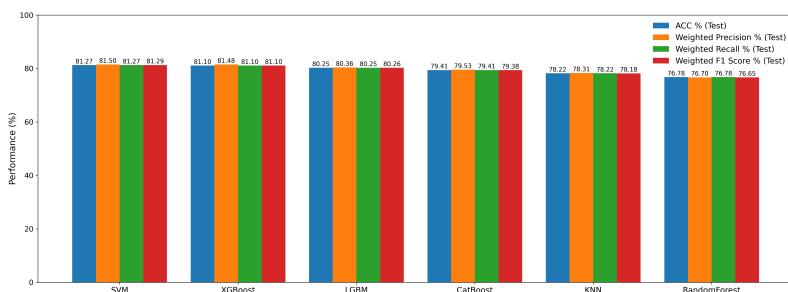
#### 3.4 Evaluation of the DINOv2 Model Results for Burn Detection

In addition to ViT-based models, the DINOv2 model was also included in this study to facilitate a comparative evaluation of model performance. The features obtained with DINOv2 were transferred to different classifiers, and the confusion matrices of these classifiers are presented in Figure 10.



**Figure 10.** Confusion matrices of features extracted with DINOv2 in different classifiers

The highest accuracy was achieved with the SVM algorithm in classifications using features extracted from the DINOv2 network. However, as with other classifiers, the model struggles with third-degree burns. The metrics calculated based on the obtained results are presented in Figure 11 for a comparative analysis of the classifiers.



**Figure 11.** Performance comparison of classifiers using DINOv2 features

The performance of the models trained with features extracted from the DINOv2 network within the dataset used

is given in Table 4.

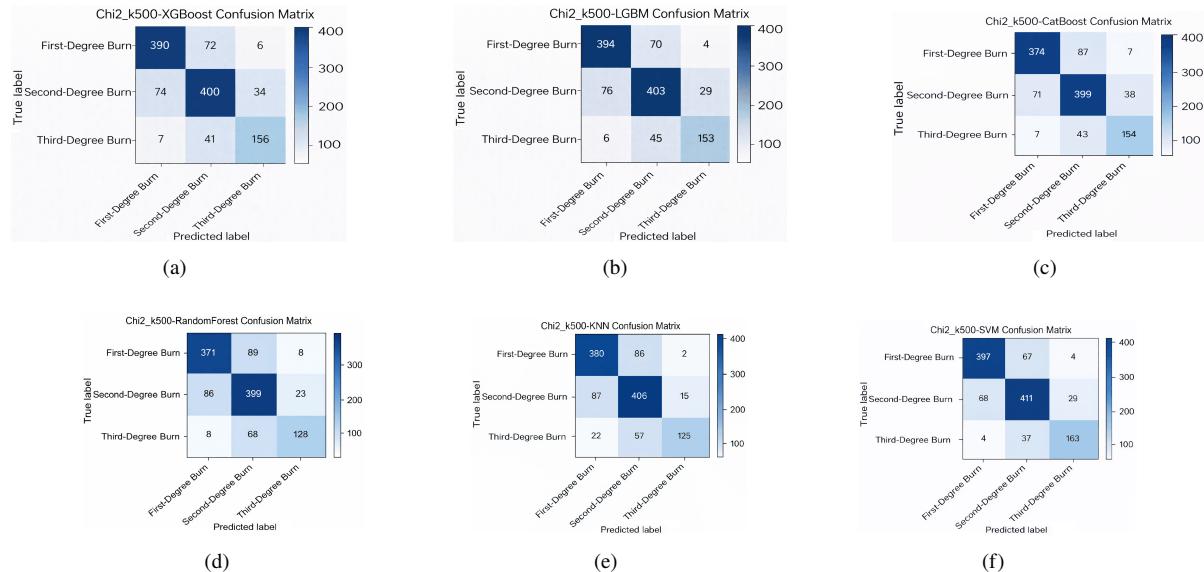
**Table 4.** Comparison of accuracy, training time, and prediction speed of classifiers using DINOv2 features

Model	Accuracy (%)	Prediction Speed (images/s)	Training Time (s)
SVM	81.27	547.42	5.06
XGBoost	81.10	79364.97	65.78
LGBM	80.25	30819.92	37.49
CatBoost	79.41	41994.64	485.02
KNN	78.22	4167.18	0.02
RF	76.78	34803.83	14.10

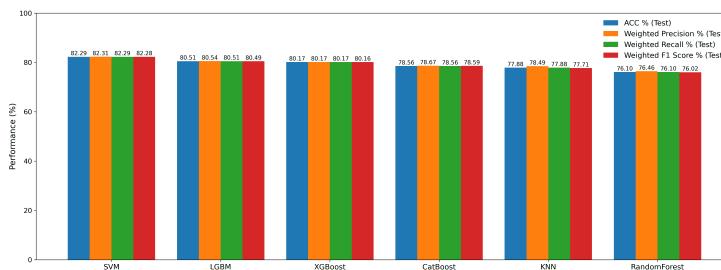
Although KNN stands out as the fastest algorithm in terms of training time, it lags behind SVM in prediction speed. When accuracy and speed metrics are evaluated together, it is concluded that SVM is the most balanced and successful classifier for this model.

### 3.5 Experimental Result of the Proposed Model

In the proposed hybrid model for determining the degree of skin burns, features obtained from four different ViT-based architectures were combined and evaluated using the  $\chi^2$  method, selecting the 500 most significant features. These features from the burn dataset were fed to six different classifiers—XGBoost, LGBM, CatBoost, KNN, SVM, and RF—and trained. The resulting confusion matrices from the training process are presented in Figure 12.



**Figure 12.** Confusion matrices of the proposed model



**Figure 13.** Performance comparison of classifiers using features of the proposed model

An examination of the confusion matrices obtained from the proposed hybrid model reveals that most classifiers can predict first- and second-degree burns with high accuracy. The SVM and LGBM models, in particular, demonstrated

quite successful performance in distinguishing these two classes. However, there was a common trend that all classifiers have difficulty in classifying third-degree burns. This is also clinically significant because third-degree burns cause severe tissue integrity disruption, and their visual similarity makes them more difficult for algorithms to distinguish.

Furthermore, while third-degree burns were more frequently confused with second-degree burns in the CatBoost and RF models, the SVM classifier performed relatively more evenly in this class. Despite the KNN model's advantage in lower training time, the misclassification rate for third-degree burns was exceptionally high. The success metrics calculated from these matrices for each classifier are detailed in Figure 13.

**Table 5.** Comparison of accuracy, training time, and prediction speed of classifiers using features of the proposed model

Model	Accuracy (%)	Prediction Speed (images/s)	Training Time (s)
SVM	82.29	835.61	3.42
LGBM	80.51	12813.90	56.91
XGBoost	80.17	41513.48	119.67
CatBoost	78.56	151636.96	92.85
KNN	77.88	5681.70	0.03
RF	76.10	7501.33	42.25

The effect of the hybrid model in terms of accuracy and system performance is shown in Table 5.

The results in Figure 12 and Table 5 show that the highest accuracy rate among the classifiers was achieved with the SVM classifier at 82.29%. LGBM and XGBoost achieved accuracies of 80.51% and 80.17%, values close to that of SVM. CatBoost and KNN achieved accuracies of 78.56% and 77.88%. The lowest performance was observed with the RF classifier with 76.10% accuracy. In terms of computational performance, the shortest training time was achieved with the KNN classifier with 0.03 seconds. Despite its slower prediction speed, the highest accuracy value was achieved with the SVM classifier. As shown in the performance metrics of the proposed model, all classifiers can distinguish first- and second-degree burns with high accuracy. However, the classifiers achieved lower performance in classifying third-degree burns.

The proposed model had particular difficulty distinguishing between second- and third-degree burns. Figure 14 shows some images of second- and third-degree burns.

In creating the dataset shown in Figure 14, some images were taken very close up, or small sections were obtained by cropping different images specifically focusing on burns. This resulted in the second-degree and third-degree burns appearing similar. These images, and similar ones, resemble each other due to their blurriness and low-pixel quality. This is considered the primary reason why the model developed in this study confuses second-degree and third-degree burns.



**Figure 14.** Similar burn samples in the dataset

#### 4 Conclusion

A burn is tissue damage caused by exposure of the skin or underlying tissues to external factors such as heat, chemicals, electricity, sunlight, or radiation. Burns are classified according to the depth and extent of the damage. The dataset used in this study categorized burns into three classes. In this study, feature extraction was performed using different transformer-based architectures to detect skin burns. These extracted feature maps were combined to bring together different features of the same image. The proposed model was then further optimized by feature selection, aiming to perform faster. The proposed model achieved an accuracy of 82.29%. However, the study has some limitations. One of the limitations is the use of a single-center public dataset. In future studies, a system that operates in real time with more data from different centers will be developed.

## Author Contributions

Conceptualization, A.Y.D. and M.K.; methodology, A.Y.D. and M.K.; software, A.Y.D. and M.K.; validation, A.Y.D., M.K., and M.Y.; formal analysis, A.Y.D.; investigation, A.Y.D. and M.K.; resources, A.Y.D. and M.K.; data curation, A.Y.D. and M.K.; writing—original draft preparation, A.Y.D., M.K., and M.Y.; writing—review and editing, A.Y.D., M.K., and M.Y.; visualization, A.Y.D., M.K., and M.Y.; supervision, M.K. and M.Y.; project administration, M.K. and M.Y. All authors have read and agreed to the published version of the manuscript.

## Data Availability

The data used to support the research findings are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] S. A. Suha and T. F. Sanam, “A deep convolutional neural network-based approach for detecting burn severity from skin burn images,” *Mach. Learn. Appl.*, vol. 9, p. 100371, 2022. <https://doi.org/10.1016/j.mlwa.2022.100371>
- [2] C. Pham, J. Greenwood, H. Cleland, P. Woodruff, and G. Maddern, “Bioengineered skin substitutes for the management of burns: A systematic review,” *Burns*, vol. 33, no. 8, pp. 946–957, 2007. <https://doi.org/10.1016/j.burns.2007.03.020>
- [3] J. J. Lee, M. Abdolahnejad, A. Morzycki, T. Freeman, H. Chan, C. Hong, R. Joshi, and J. N. Wong, “Comparing artificial intelligence guided image assessment to current methods of burn assessment,” *J. Burn Care Res.*, vol. 46, no. 1, pp. 6–13, 2025. <https://doi.org/10.1093/jbcr/irae121>
- [4] P. K. Rangaiah, B. P. Kumar, F. Huss, and R. Augustine, “Precision diagnosis of burn injuries using imaging and predictive modeling for clinical applications,” *Sci. Rep.*, vol. 15, no. 1, p. 7604, 2025. <https://doi.org/10.1038/s41598-025-92096-4>
- [5] H. Li, Q. Bu, X. Shi, X. Xu, and J. Li, “Non-invasive medical imaging technology for the diagnosis of burn depth,” *Int. Wound J.*, vol. 21, no. 1, p. e14681, 2024. <https://doi.org/10.1111/iwj.14681>
- [6] C. Pabitha and B. Vanathi, “DenseMask RCNN: A hybrid model for skin burn image classification and severity grading,” *Neural Process. Lett.*, vol. 53, no. 1, pp. 319–337, 2021. <https://doi.org/10.1007/s11063-020-10387-5>
- [7] 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 Process.*, vol. 13, no. 11, pp. 2018–2028, 2019. <https://doi.org/10.1049/iet-ipr.2018.5899>
- [8] M. Abazari, A. Ghaffari, H. Rashidzadeh, S. M. Badeleh, and Y. Maleki, “A systematic review on classification, identification, and healing process of burn wound healing,” *Int. J. Low. Extrem. Wounds*, vol. 21, no. 1, pp. 18–30, 2022. <https://doi.org/10.1177/1534734620924857>
- [9] A. Elsarta, H. Fathalla, M. Nasser, S. Elwatany, R. Fekry, M. Ebaid, Y. Mahmoud, A. Anwar, and A. Gaber, “Integrating multi-source data for skin burn classification using deep learning,” *Comput. Biol. Med.*, vol. 195, p. 110556, 2025. <https://doi.org/10.1016/j.combiomed.2025.110556>
- [10] B. A. Acha, C. Serrano, J. I. Acha, and L. M. Roa, “Segmentation and classification of burn images by color and texture information,” *J. Biomed. Opt.*, vol. 10, no. 3, p. 034014, 2005. <https://doi.org/10.1117/1.1921227>
- [11] P. N. Kuan, S. Chua, E. B. Safawi, H. H. Wang, and W. Tiong, “A comparative study of the classification of skin burn depth in human,” *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 2–10, pp. 15–23, 2017.
- [12] M. M. Rahman, M. E. L. Masry, S. C. Gnyawali, Y. Xue, G. Gordillo, and J. P. Wachs, “A framework for advancing burn assessment with artificial intelligence,” *Mil. Med.*, vol. 190, no. Supplement\_2, pp. 387–393, 2025. <https://doi.org/10.1093/milmed/usaf198>
- [13] F. A. Khan, A. U. R. Butt, M. Asif, W. Ahmad, M. Nawaz, M. Jamjoom, and E. Alabdulkreem, “Computer-aided diagnosis for burnt skin images using deep convolutional neural network,” *Multimed. Tools Appl.*, vol. 79, no. 45, pp. 34 545–34 568, 2020. <https://doi.org/10.1007/s11042-020-08768-y>
- [14] F. Abbas, “Skin Burn Dataset,” Kaggle, 2022. <https://www.kaggle.com/datasets/faresabbasai2022/skin-burn-dataset>
- [15] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, and Y. Xu, “A survey on Vision Transformer,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, 2023. <https://doi.org/10.1109/TPAMI.2022.3152247>
- [16] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in vision: A survey,” *ACM Comput. Surv.*, vol. 54, no. 10s, p. Article 200, 2022. <https://doi.org/10.1145/3505244>
- [17] Google, “ViT-Base Patch16 224 Model,” Hugging Face. <https://huggingface.co/google/vit-base-patch16-224>

- [18] Google, “ViT-Base Patch32 224 In21k Model,” Hugging Face. <https://huggingface.co/google/vit-base-patch32-224-in21k>
- [19] R. Wightman, “ViT-Large Patch16 224 AugReg In21k Model,” Hugging Face. [https://huggingface.co/timm/vit\\_large\\_patch16\\_224.augreg\\_in21k](https://huggingface.co/timm/vit_large_patch16_224.augreg_in21k)
- [20] M. Oquab, T. Dariseti, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, and P. Bojanowski, “DINOv2: Learning robust visual features without supervision,” *arXiv preprint arXiv:2304.07193*, 2023. <https://doi.org/10.48550/arXiv.2304.07193>
- [21] X. Song, X. Xu, and P. Yan, “General purpose image encoder DINOv2 for medical image registration,” *arXiv preprint arXiv:2402.15687*, 2024. <https://doi.org/10.48550/arXiv.2402.15687>
- [22] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Front. Neurorobot.*, vol. 7, p. 21, 2013. <https://doi.org/10.3389/fnbot.2013.00021>
- [23] G. Ke, M. Qi, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, “LightGBM: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, 2017.
- [24] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, and T. Zhou, “XGBoost: Extreme gradient boosting,” R package version 0.4-2, 2015.
- [25] P. L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: Unbiased boosting with categorical features,” in *Advances in Neural Information Processing Systems*, Montréal, Canada, 2018.
- [26] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “KNN model-based approach in classification,” in *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, Berlin, Heidelberg, 2003, pp. 986–996. [https://doi.org/10.1007/978-3-540-39964-3\\_62](https://doi.org/10.1007/978-3-540-39964-3_62)
- [27] V. Jakkula, “Tutorial on support vector machine (SVM),” School of Electrical Engineering and Computer Science, Washington State University, Technical Report, 2006.
- [28] A. Parmar, R. Katariya, and V. Patel, “A review on random forest: An ensemble classifier,” in *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI 2018)*, Coimbatore, India, 2018. [https://doi.org/10.1007/978-3-030-03146-6\\_86](https://doi.org/10.1007/978-3-030-03146-6_86)
- [29] H. Liu and R. Setiono, “Feature selection via discretization,” *IEEE Trans. Knowl. Data Eng.*, vol. 9, no. 4, pp. 642–645, 1997. <https://doi.org/10.1109/69.617056>
- [30] R. Yacoubi and D. Axman, “Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models,” in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, 2020, pp. 79–91.