Intelligent Methods in Engineering Sciences 5(1): 20-30, 2026

IMIENS

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

https://www.imiens.org

International
Open Access

April, 2026

e-ISSN 2979-9236

Research Article

https://doi.org/10.58190/imiens.2026.168

Classification of Skin Lesions with Deep Features Based on SqueezeNet Using Machine Learning Methods

Samet Akcakaya [a,*], ***Ilknur Cevik Tekin*** [a], ***Yavuz Selim TASPINAR*** [b]

[a] *Beysehir Ali Akkanat Faculty of Business Administration, Management Information Systems, Selcuk University, Konya, Türkiye*
[b] *Technology Faculty, Mechatronics Engineering, Selcuk University, Konya, Türkiye*

ARTICLE INFO

Article history:
Received 25 December 2025
Accepted 14 April 2026

Keywords:
Skin Lesion,
SqueezeNet,
Skin Image Analysis,
ISIC 2019 Dataset,
Image Processing,
Classification.

ABSTRACT

Early and accurate identification of skin lesions is one of the most important factors that improves treatment success in dermatological diseases. Deep learning and image processing algorithms can provide promising results in diagnostic processes by learning complex texture, pigment, and structural features that are difficult for the human eye to distinguish. In this study, a hybrid classification approach based on deep feature extraction and machine learning is proposed. A total of 1,763 dermoscopic images, including eight lesion classes, selected from the ISIC 2019 skin lesion dataset were used. Deep features were extracted using a pretrained SqueezeNet model and analyzed using a 10-fold cross-validation method. The extracted features were evaluated using Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Logistic Regression (LR) classifiers. Experimental results show that the proposed model achieved accuracy values ranging from 96.7% to 89.6% and F1-score values ranging from 96.6% to 89.5%, successfully classifying skin lesions. The KNN model showed relatively lower discrimination capability compared to the other models. All models achieved 100% accuracy in the BCC (Basal Cell Carcinoma) class, while ANN and LR models also classified the SCC (Squamous Cell Carcinoma) class without error. Among all classes, the lowest performance was observed in the DF (Dermatofibroma) class. This study demonstrates that high classification performance can be achieved in multi-class (8 classes) dermatological analysis using lightweight architectures without requiring computationally heavy deep learning models. However, further validation on larger and more diverse datasets is required before real-world clinical applicability can be considered.

1. INTRODUCTION

Skin lesions are among the most common health problems worldwide, and skin cancer is considered one of the most serious diseases affecting public health. The World Health Organization (WHO) indicates that the impact of skin cancer is expected to continue in the coming years, with an increase in the number of new cases worldwide [1]. Early and accurate diagnosis of skin lesions enables timely intervention in dermatological conditions. However, the analysis of skin lesions is highly complex due to morphological diversity, color variations, and irregular boundaries. The presence of different characteristic features in lesions makes accurate identification and the selection of appropriate treatment methods critical. Variations in diagnosis among dermatologists and errors arising from subjective evaluations make it difficult to reliably classify skin lesions based solely on clinical examination. This situation has increased the need for more objective, reproducible, and accurate technologies to support clinical decision-making processes.

In recent years, image processing algorithms and deep learning methods in medical image analysis have demonstrated the ability to analyze complex features of skin lesions, such as color, texture, pigmentation, and structural patterns, and to learn details that are difficult for the human eye to detect. The development of large-scale image datasets has enabled the training of these models, contributing significantly to dermatological screening processes. The integration of such systems is important for allowing clinicians to understand the visual cues used by artificial intelligence models in decision-making, which is essential for reliability and practical use. These systems can achieve high accuracy and may help reduce error rates. In addition, through interpretable and reliable decision

* **Corresponding Author:** sametakcaa@gmail.com

mechanisms that support research workflows, it becomes possible to develop new models and design systems that assist in the classification of skin lesions.

In this study, the diagnostic performance of the model was evaluated using ISIC (International Skin Imaging Collaboration) 2019 skin lesion data (1,763 images from eight classes). The use of the SqueezeNet architecture aims to provide a computationally efficient alternative to existing models in the literature [2]. A review of the literature shows that studies in this field have been shaped by different approaches and research focuses. Studies related to the classification of skin lesions are summarized below.

Gouda et al. examined benign, malignant, non-melanocytic, and melanocytic tumors using CNN-based methods on the ISIC 2018 dataset. In their study, they observed that the Inception model achieved better performance with an accuracy of 85.7%. They also reported that deeper networks (e.g., ResNet50) do not always yield better results in medical imaging, and that relatively shallower but wider architectures such as InceptionV3 may be more effective in capturing general features [3].

Nigar et al. analyzed eight different lesion classes in the ISIC 2019 dataset by establishing a preprocessing pipeline to manage class imbalance and image variability. The F1-score increased from 75% to 94.45% after preprocessing. They demonstrated that the ResNet-18 model outperformed InceptionV3 and enhanced model interpretability by integrating the LIME framework [4].

Iqbal et al. developed a CNN-based method for eight different lesion classes in the ISIC 2019 dataset and reported that their proposed Deep Convolutional Neural Network (DCNN) model achieved higher accuracy compared to models such as MobileNetV2 and ResNet50. They also stated that problem-specific architectures designed for imbalanced datasets may provide more stable performance than general-purpose pretrained networks [5].

Kassem et al. modified the GoogLeNet architecture and achieved an accuracy of 94.92% on the ISIC 2019 dataset. Instead of computationally expensive networks such as VGG, they preferred GoogLeNet and showed that their hybrid approach improved the reliability of the model against unseen lesions that may be encountered in clinical environments [6].

Goyal et al. reviewed recent studies in dermoscopy, clinical imaging, and histopathology, reporting that while dermatologists achieved an accuracy of 86.6%, artificial intelligence-based systems could reach up to 95%. They also emphasized that model performance decreases with increasing complexity, and that the balanced accuracy dropped from 88.5% in the 7-class ISIC 2018 dataset to 63.6% in the 8-class ISIC 2019 dataset. Additionally, they highlighted that the lack of ethnic diversity in training datasets negatively affects model generalization [7].

Shetty et al. proposed a customized CNN architecture on the HAM10000 dataset and applied data augmentation techniques including horizontal flipping and random geometric transformations. By resizing and normalizing the images, they reduced overfitting and achieved a generalization performance of 95.18%. Although their experimental results showed that the Random Forest (RF) algorithm achieved the highest accuracy of 87% among traditional methods, their proposed CNN model reached an accuracy of 95.18%, outperforming both classical approaches and complex transfer learning models such as InceptionV3 (95.14%). Their findings also emphasized the importance of data augmentation and pixel normalization [8].

Olayah et al. proposed a hybrid approach combining image processing techniques with deep learning architectures. They applied Contrast Limited Adaptive Histogram Equalization (CLAHE), mean filtering, and the Geometric Active Contour (GAC) algorithm to enhance images and separate lesion regions from healthy tissue. Using pretrained networks such as AlexNet, GoogLeNet, and VGG16 as classifiers, they achieved an accuracy of 96.10% and a sensitivity of 88.90%, demonstrating improved performance compared to individual architectures. Their study suggests that combining the strengths of different CNN architectures can be effective in distinguishing complex skin lesions [9].

Ahammed et al. focused on preprocessing techniques and applied a digital hair removal protocol. Instead of manual or simple thresholding methods, they used a GrabCut algorithm based on the HSV color space without requiring user intervention. Their results showed that the SVM classifier achieved accuracy rates of 95% and 97% on the ISIC 2019 and HAM10000 datasets, respectively, outperforming KNN and Random Tree (RT) methods and indicating that classical machine learning can still be a strong alternative to deep learning approaches [10].

Wang et al. developed an innovative framework called SSD-KD based on Knowledge Distillation (KD). They incorporated a weighted loss function and self-supervised learning strategies to improve sensitivity to minority classes. Their experimental results showed that the MobileNetV2 model trained with distilled knowledge achieved an accuracy of 85% in classifying eight different skin diseases. Their study demonstrates that high-performance deep learning models can be adapted to resource-constrained embedded systems [11].

Hameed et al. designed a cloud-based mobile system based on the SqueezeNet model for automatic skin disease detection. They developed a client-server architecture that transfers images from mobile devices to a server. Using this expert system, named i-Rash, they focused on classifying four classes: acne, eczema, psoriasis, and healthy skin. By applying random sampling techniques to

reduce dataset imbalance, they achieved accuracy rates of 97.21% and sensitivity of 98.14% on 1,856 images. The classification time of 0.09 seconds indicates that the proposed lightweight architecture may support real-time decision-making with high sensitivity [12].

In light of the existing literature, the contributions of this study to the analysis of dermatological images can be summarized as follows:

• In contrast to computationally intensive deep architectures such as ResNet and DenseNet, the SqueezeNet architecture was employed to extract feature representations with high discriminative capability at a lower computational cost.

• Instead of reducing the problem to a binary classification (benign/malignant), a multi-class classification approach covering eight different lesion classes was adopted. This approach aims to improve the discrimination of visually similar lesion types.

• To reduce variations in diagnosis caused by subjective interpretation among clinicians, a model was developed to analyze the structural and color characteristics of lesions using objective criteria and high classification performance.

• Despite the class imbalance, which is a common challenge in medical imaging, this study presents a methodological example of developing models that can operate reliably on relatively limited datasets. The remainder of the paper is organized as follows: Section 2 describes the dataset, methods, and performance metrics under the Materials and Methods section. Section 3 presents the experimental results and discussion. Finally, Section 4 summarizes the conclusions, contributions, limitations, and potential application areas of the study.

2. MATERIAL AND METHODS

2.1. *ISIC 2019 Skin Lesion Images for Classification (SLIFC)*

Although the ISIC 2019 dataset archive contains 25,331 images, a subset consisting of 1,763 images across eight classes was used in this study [13–16]. The primary reason for this selection is the severe class imbalance present in the original dataset. While the NV class constitutes more than half of the dataset, classes such as DF and VASC represent less than 1% of the total data. To prevent the model from developing a bias toward the majority class, a more balanced distribution was established among the selected samples. The distribution of the selected dataset is as follows: 211 images for AK (Actinic Keratosis), 229 images for BCC (Basal Cell Carcinoma), 215 images for BKL (Benign Keratosis), 184 images for DF (Dermatofibroma), 252 images for MEL, 261 images for NV (Melanocytic Nevus), 231 images for SCC (Squamous Cell Carcinoma), and 180 images for VASC (Vascular Lesion).

To improve reproducibility and enhance model performance, a comprehensive preprocessing pipeline was applied to the raw images. Since the image sizes and pixel values vary across the dataset, all dermoscopic images were first resized to 227×227×3 pixels using bilinear interpolation, which corresponds to the standard input size of the SqueezeNet architecture. Subsequently, normalization based on ImageNet statistics was applied to accelerate model convergence. In order to address class imbalance and improve the generalization capability of the model, several data augmentation techniques were applied during the training phase, including random rotation up to 180 degrees, horizontal and vertical flipping, zooming, and brightness adjustment. Sample images from the SLIFC dataset are presented in Figure 1.

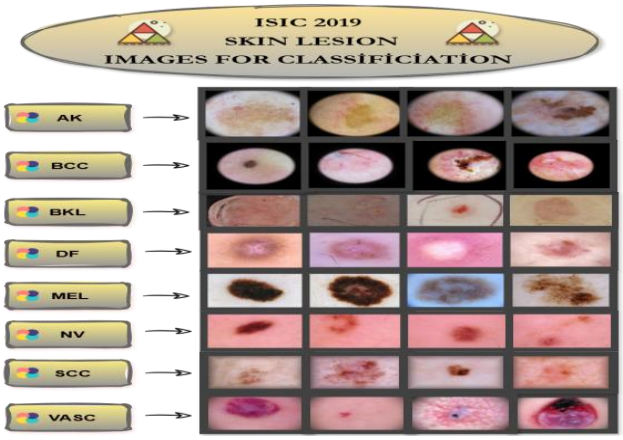


Figure 1. Sample images by class in SLIFC.

2.2. *Convolutional Neural Networks (CNN)*

Convolutional Neural Networks (CNNs) are increasingly used in medical image analysis due to their ability to learn discriminative features from raw image data without manual intervention [17]. They have demonstrated high performance, particularly in tasks such as image classification, object recognition, and lesion detection. CNNs perform a multi-layered feature extraction process by using convolutional filters to capture local patterns in images. These filters automatically learn visual cues such as color distribution, edge structures, asymmetry, and texture characteristics of lesions. CNN-based approaches for dermatological images can achieve performance comparable to that of expert clinicians when trained on large-scale datasets. Unlike traditional machine learning methods, CNN architectures do not require manual feature extraction or extensive preprocessing; instead, they optimize filter weights during training to automatically identify patterns such as lesion boundaries, color asymmetry, and dermatological textures. With modern CNN architectures optimized in terms of depth, width, and parameter efficiency, high accuracy, sensitivity, and specificity values can be achieved in multi-class tasks such as skin lesion classification. In such analyses, CNNs first detect lesion boundaries and color

characteristics, transform the image into a processable representation, and ultimately classify the lesion as malignant or benign. This structure enables improved diagnostic performance, particularly in problems involving high variability and visual similarity, such as skin lesion classification.

2.3. *SqueezeNet Architecture*

The success of deep learning models, particularly convolutional neural networks, is generally associated with increasing network depth and the number of parameters. However, for tasks such as skin lesion classification that require integration into mobile or embedded systems, computational cost and memory requirements are as critical as classification accuracy. SqueezeNet reduces the computational burden of high-dimensional convolutional layers by minimizing model size while maintaining high classification performance.

Due to its low parameter count and compact architecture, SqueezeNet efficiently utilizes convolution operations, which are the fundamental building blocks of the model, making it suitable for resource-constrained environments. While reducing computational cost, it is capable of effectively learning fine-grained details such as pigment networks, color distributions, and boundary irregularities in high-resolution dermoscopic images, thereby supporting accurate classification. This characteristic provides a significant advantage in medical imaging tasks involving small and detailed structures, such as skin lesions. In this study, the pretrained SqueezeNet model trained on the ImageNet dataset was utilized not for direct classification but for deep feature extraction. The input images were propagated through the network, and feature representations were obtained from the conv10 layer, where the final features are generated. Each dermoscopic image was provided as input to the SqueezeNet model, and a forward pass operation was performed. The 1000-dimensional output vector obtained from the final layer of the model was used as a high-level feature representation of the image. These feature vectors contain semantic information learned from large-scale data and represent the discriminative characteristics of the images. The extracted features were then used as input to machine learning algorithms, including Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Logistic Regression (LR). This approach enables the effective use of deep feature representations together with classical classification methods.

Thanks to its ability to operate on devices with limited computational resources and its high parameter efficiency, SqueezeNet provides a fast and reliable architectural solution for the automatic analysis of dermoscopic images.

2.4. *Artificial Neural Network (ANN)*

Artificial Neural Networks (ANNs), inspired by the learning and decision-making principles of the human brain, consist of an input layer, one or more hidden layers, and an output layer. Artificial neurons in each layer process incoming signals by applying weights and transmit the results to the next layer through an activation function. This structure enables the system to learn complex and non-linear relationships among inputs. ANN models have the capability to model relationships within data, recognize patterns, and solve classification problems. By mathematically modeling the working principles and learning mechanisms of biological neural systems, ANNs are particularly effective for problems such as skin lesion classification, where visual features are multi-dimensional and heterogeneous.

In dermoscopic images of skin lesions, factors such as pigment intensity, texture organization, boundary clarity, and color variations play a critical role in determining lesion type. ANN can statistically model the relationships among these multiple variables and improve its learning capacity as the number of samples increases. Especially in small- to medium-sized datasets, ANN-based classifiers can achieve high accuracy when combined with feature representations such as color histograms, texture statistics, or edge density obtained after feature extraction. An important advantage of ANN is that it can be trained relatively quickly even with simple architectures and does not require the high computational resources needed by complex CNN models. In skin lesion classification, non-linear activation functions are essential for enabling the network to learn not only linear relationships but also the complexity of lesion boundaries. The training process is based on calculating the error between the predicted lesion class and the true pathological label and propagating this error back through the network using the backpropagation algorithm. The weights between neurons are updated using optimization algorithms to reduce the model error. ANN can take as input either handcrafted features such as color, texture, and shape or feature vectors extracted from CNN layers, and use them for classification. In this way, the model learns from training data and gains the ability to generalize to previously unseen lesion images.

2.5. *K-Nearest Neighbor (KNN)*

The ISIC 2019 dataset is one of the most widely used open-access data sources in artificial intelligence-based classification studies aimed at supporting experts in skin cancer diagnosis. Despite its simplicity, the K-Nearest Neighbor (KNN) algorithm serves as a strong baseline method for evaluating performance in image-based classification tasks. KNN establishes a decision mechanism by comparing each test image with samples in the training set and determining similarity based on the nearest neighbors. In this study, feature vectors obtained

from pretrained deep neural networks were used as input to the KNN classifier. This approach enables classification to be performed on more informative feature representations rather than raw pixel values, which can improve classification performance. However, due to the high dimensionality of the data, the computational cost of KNN may increase, which can become a limitation in large-scale test scenarios. To address this issue, various approaches such as dimensionality reduction techniques and weighted KNN methods have been proposed in the literature. When supported by deep feature extraction on the ISIC 2019 dataset, the KNN algorithm provides a relatively low-complexity yet reliable classification method and serves as an important reference baseline [18].

2.6. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) is a powerful supervised learning method frequently used in the classification of skin lesions. The SVM algorithm aims to find the optimal hyperplane that best separates classes by mapping the data into a high-dimensional feature space. This approach can achieve high classification performance, particularly in problems where clear boundaries exist between classes [19]. In studies conducted on the ISIC 2019 dataset, SVM is commonly applied to feature vectors extracted from deep learning models, and complex visual patterns are classified using linear or non-linear kernel functions. Among these, the Radial Basis Function (RBF) kernel is widely preferred for modeling complex boundaries between lesion types. Due to class imbalance in the dataset, techniques such as class weighting or resampling are often used to improve SVM performance. In the literature, SVM models supported by deep feature representations have been reported to achieve accuracy levels exceeding 85% and, in some cases, performance comparable to baseline CNN models. Therefore, SVM serves as a reliable alternative in medical imaging research and decision support systems.

2.7. *Logistic Regression (LR)*

Logistic Regression (LR) is one of the fundamental statistical learning methods used for classifying samples in a dataset as malignant or benign. The algorithm constructs a linear decision boundary and estimates the probability of a sample belonging to a particular class using the logit function [20]. For high-dimensional dermoscopic images in the ISIC 2019 dataset, LR is typically applied using pre-extracted feature vectors, such as deep features obtained from CNN-based models. This approach allows the algorithm to better adapt to complex patterns. In particular, the use of L2 regularization (Ridge regularization) helps reduce the risk of overfitting and improves generalization performance. LR models supported by deep feature representations have been reported to achieve accuracy levels above 80%, indicating that the model can provide satisfactory performance with relatively low complexity. In addition, the probabilistic output structure of LR enhances model interpretability in decision support systems and can contribute to early-stage analysis. Therefore, LR serves as an important baseline classification method in terms of both statistical interpretability and generalization capability.

2.8. *Confusion Matrix (CM)*

The confusion matrix is one of the fundamental evaluation tools used to assess the performance of classification algorithms. It provides a visual representation of the relationship between predicted and actual class labels, enabling a detailed analysis not only of overall accuracy but also of performance metrics such as sensitivity, specificity, precision, and F1-score [21]. Due to the multi-class structure of skin lesion images, the confusion matrix presents correct and incorrect predictions for each class separately, allowing the identification of lesion types where the model performs well or poorly. This is particularly useful when comparing traditional classifiers such as KNN, SVM, and LR with deep learning-based approaches, as it helps to interpret model behavior more effectively. In the literature, models supported by CNN-based feature extraction have been observed to achieve higher accuracy in benign classes, while error rates tend to increase in relatively rare classes such as melanoma. Therefore, the confusion matrix plays an important role in model optimization, analysis of class imbalance, and identification of sources of misclassification in skin lesion analysis.

2.9. *Performance Metrics*

Based on the results obtained from the models, metric analysis was performed using Accuracy, F1-score, Precision, and Recall in order to evaluate the performance of the proposed approach and determine the most suitable classification model. The fundamental formulas used for calculating these metrics are presented in Table 1.

Table 1. Calculation formula.

Metric	Equation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall (Sensitivity)	$\frac{TP}{TP+FN}$
F1-Score	$\frac{2\times(Precision\times Recall)}{Precision+Recall}$

2.10. *Cross Validation*

In this study, a 10-fold cross-validation (k = 10) method was employed to evaluate the reliability and generalization capability of the proposed SqueezeNet-based skin lesion classification model. This approach enables repeated training and testing of the model on different data subsets, allowing performance to be assessed in a stable manner

without being affected by random data partitioning. In this way, the risks of overfitting and underfitting are reduced. The ISIC 2019 skin lesion dataset used in this study was randomly divided into 10 equal subsets. In each iteration, 9 subsets were used for training and 1 subset was used for validation, and this process was repeated 10 times. Accuracy, recall, precision, and F1-score values were calculated in each iteration, and the overall performance of the model was determined by averaging the results across all iterations. As reported in the literature, 10-fold cross-validation is one of the standard approaches that provides an effective balance between bias and variance during model evaluation [22]. While lower fold numbers may reduce the stability of validation results, higher fold numbers significantly increase computational cost and processing time. Therefore, a 10-fold structure was selected as a balanced solution in terms of both computational efficiency and generalization performance. The presence of multiple images belonging to the same lesion in both training and test sets may lead to memorization by the model and artificially inflated performance. To prevent this issue, the dataset was partitioned at the lesion level using the metadata parameter (lesion_id). Accordingly, all images belonging to a specific lesion were grouped such that they were included entirely in either the training set or the test set in each iteration. This strategy ensures that the model learns to generalize lesion characteristics rather than memorizing specific images. This multi-stage validation process implemented on the SqueezeNet architecture demonstrates the consistency of the model across different data subsets and provides evidence supporting its reliable performance.

3. EXPERIMENTAL RESULTS

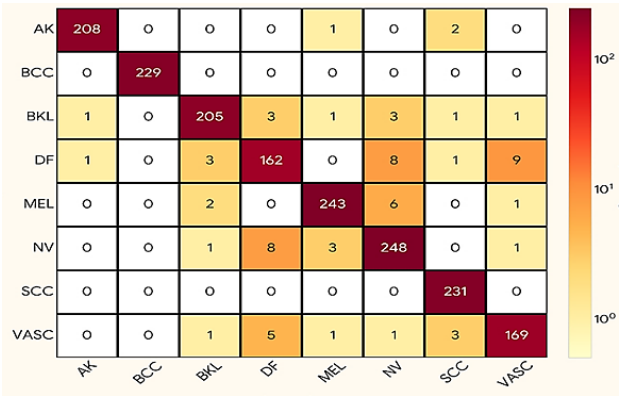
In this dermatological study, classification analysis was performed on the SLIFC dataset using image processing and machine learning methods to identify skin lesions from dermoscopic images. KNN, ANN, SVM, and LR algorithms were employed. The experiments were conducted on a computer equipped with an Intel Core i5-11400H 2.70 GHz processor, an NVIDIA GeForce RTX 3050 Ti graphics card, and 32 GB of RAM. The Python programming language was used.

For the ANN model, the following parameters were used: 100 hidden layers, ReLU activation function, regularization parameter of 0.0001, and 200 iterations. For the KNN method, the number of neighbors was set to 5, with Euclidean distance as the metric and uniform weighting. For the SVM model, the parameters included cost = 1, regression loss epsilon = 0.10, RBF kernel, and 100 iterations. For the LR method, Ridge (L2) regularization was applied. Training and testing processes were performed using the cross-validation method. To ensure reliable model evaluation and prevent data leakage, a 10-fold cross-validation strategy with lesion-level grouping was implemented. The dataset was divided into 10 equal parts; in each iteration, 9 parts were used for training and 1 part for testing. Data augmentation procedures were applied only to the training folds to prevent test data from influencing the training process. Additionally, all images belonging to the same lesion were kept within the same fold to avoid data leakage.

The subset prepared from the ISIC 2019 dataset was selected to establish a controlled experimental setup while preserving class diversity. The dataset was organized to represent different lesion types, ensuring that each class was included in the analysis process. Using SqueezeNet, a 1000-dimensional feature vector was extracted from each image, and these feature representations were used as input for the classification models (ANN, KNN, SVM, and LR). The use of a pretrained network allows the transfer of complex texture representations learned from natural images to a new task, such as skin lesion classification, which can contribute to faster convergence and improved generalization performance. The final layer of the model was selected for feature extraction. This layer produces a 1000-dimensional feature vector for each image through a global average pooling (GAP) operation, representing high-level semantic information.

The confusion matrices obtained from the machine learning models are presented below. The confusion matrix corresponding to the ANN method is shown in Table 2.

Table 2. Confusion Matrix of ANN Model

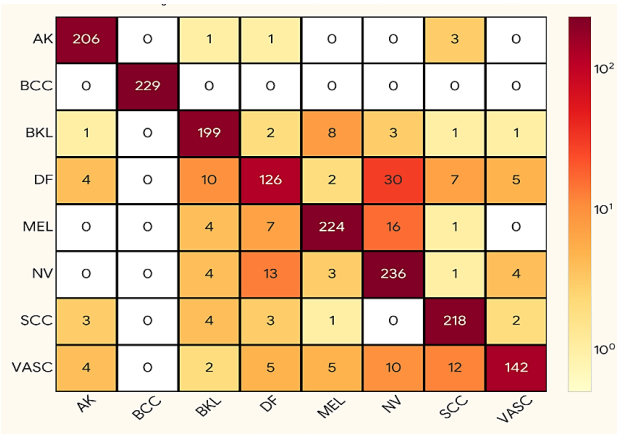


According to the table 2, When the classification results of the ANN model are examined, it is observed that the model generally achieves high accuracy. In the AK class, 208 samples were correctly classified, with only 3 misclassifications (1 MEL, 2 SCC). All 229 samples in the BCC class were correctly classified, and similarly, all 231 samples in the SCC class were predicted without error. This indicates that these classes have more distinguishable characteristics. In the BKL class, 205 samples were correctly classified, with 10 misclassifications observed. In the DF class, relatively lower performance was observed, with 162 correct classifications and 22 errors. Most of these misclassifications were associated with the

NV and VASC classes. In the MEL class, 243 samples were correctly classified with 9 errors, while in the NV class, 248 samples were correctly classified with 13 errors. In the VASC class, 169 samples were correctly classified with 11 misclassifications.

Overall, the model operates with a low error rate in most classes; however, a higher level of confusion is observed in the DF class compared to the others. A more detailed analysis of the ANN model outputs shows that all samples in the BCC and SCC classes were correctly identified, corresponding to a 100% accuracy rate for these classes, which is important in terms of reliability. This suggests that the model is effective in learning the discriminative features of these lesions. The deep features obtained from the SqueezeNet layer appear to encode the vascular structures of BCC and the textural characteristics of SCC with high sensitivity. When combined with the non-linear activation functions of ANN, this may result in the formation of effective decision boundaries. On the other hand, the DF class represents the most challenging case for the model. The 22 misclassifications in this class were mainly associated with the VASC and NV classes. However, considering the overall results, the model exhibits minimal errors in 6 out of 8 classes, indicating that the hybrid structure combining SqueezeNet features with ANN provides a consistent classification performance. The confusion matrix of the KNN method is presented in Table 3.

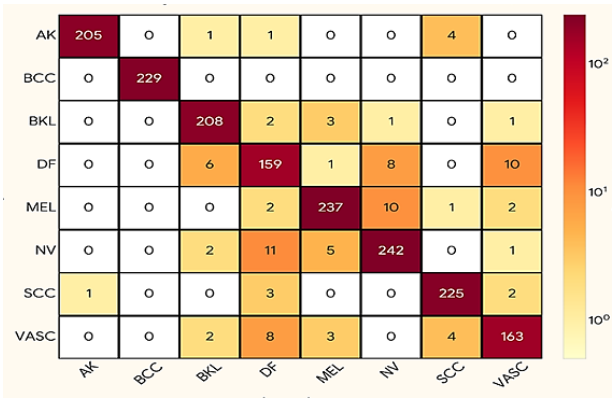
Table 3. Confusion Matrix of KNN Model



According to the table 3, When the classification results of the KNN model are examined, it is observed that the model achieves high accuracy in some classes; however, it produces more errors overall compared to the ANN model. In the AK class, 206 samples were correctly classified, with a total of 5 misclassifications (1 BKL, 1 DF, and 3 SCC). All 229 samples in the BCC class were correctly classified, with no errors observed. In the BKL class, 199 samples were correctly classified, while 16 misclassifications were recorded. In the DF class, the model performance decreased significantly, with 126 correct classifications and 58 misclassifications. Most of these errors were associated with the NV class. In the MEL class, 224 samples were correctly classified with 28 errors, while in the NV class, 236 samples were correctly classified with 25 errors. In the SCC class, 218 samples were correctly classified with 13 misclassifications. In the VASC class, 142 samples were correctly classified, with a total of 38 errors. These misclassifications were mainly associated with the NV and SCC classes.

A detailed examination of the model outputs shows that all samples in the BCC and SCC classes were correctly identified, corresponding to a 100% accuracy rate for these classes, indicating the model's effectiveness for these categories. In contrast, the DF class represents the most challenging case for the model, where misclassifications are mainly attributed to morphological similarities with the VASC and NV classes. The relatively lower performance compared to other models may be due to the high dimensionality of the feature vectors (1000 features), which can affect distance-based calculations. Nevertheless, the KNN model demonstrates a reasonable generalization capability in modeling complex lesion features. The confusion matrix of the SVM method is presented in Table 4.

Table 4. Confusion Matrix of SVM Model

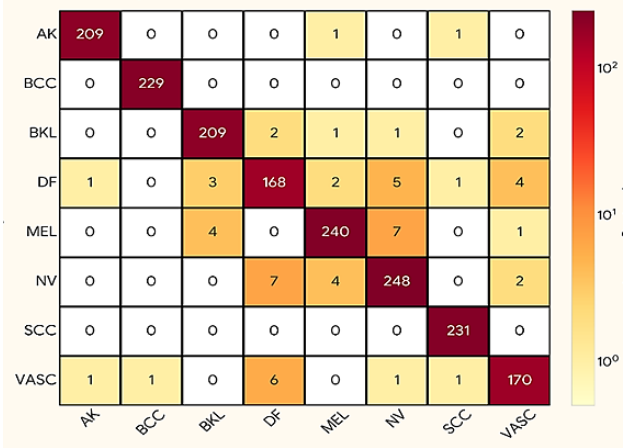


According to the table 4, When the classification results of the SVM model are examined, it is observed that the model generally performs in a balanced manner with high accuracy. In the AK class, 205 samples were correctly classified, with a total of 6 misclassifications (1 BKL, 1 DF, and 4 SCC). All 229 samples in the BCC class were correctly classified, with no errors observed. In the BKL class, 208 samples were correctly classified, with 7 misclassifications recorded. In the DF class, 159 samples were correctly classified, while 25 misclassifications were observed. Most of these errors were associated with the NV and VASC classes. In the MEL class, 237 samples were correctly classified with 15 errors, while in the NV class, 242 samples were correctly classified with 19 errors. In the SCC class, 225 samples were correctly classified with 6 misclassifications, whereas in the VASC class, 163 samples were correctly classified with 17 errors.

Compared to KNN, the SVM classifier provides a clear improvement. Although the model demonstrates generally high performance, it remains slightly behind LR and ANN

in some classes. The BCC class was once again classified without error. In the MEL class, 237 correct predictions indicate an acceptable performance level. The low number of errors in the AK class (6 errors) suggests that the model has a strong ability to distinguish between classes. The most challenging class for the model is DF. The concentration of errors in the VASC and NV classes indicates that the visual boundaries between lesion types may become less distinguishable for this model as well.The confusion matrix of the LR method is presented in Table 5.

Table 5. Confusion Matrix of LR Model



According to the table 5, When the classification results of the LR model are examined, it is observed that the model generally operates with high accuracy and a low error rate. In the AK class, 209 samples were correctly classified, with only 2 misclassifications (1 MEL, 1 SCC). All 229 samples in the BCC class were correctly classified, with no errors observed. Similarly, all 231 samples in the SCC class were correctly classified. In the BKL class, 209 samples were correctly classified, with a total of 6 misclassifications. In the DF class, 168 samples were correctly classified, while 19 misclassifications were observed. Most of these errors were associated with the NV and VASC classes. In the MEL class, 240 samples were correctly classified with 12 errors, while in the NV class, 248 samples were correctly classified with 13 errors. In the VASC class, 170 samples were correctly classified, with 10 misclassifications observed.

Despite its relatively simple structure, the LR model produced one of the most notable results in this study. While achieving zero error in the BCC and SCC classes, it also demonstrated near-perfect performance in the AK class with only 2 misclassifications. This high performance suggests that the deep features extracted from the SqueezeNet architecture may be linearly separable. When the error distribution is examined, it is observed that, similar to other algorithms, the model is affected by morphological similarities in the DF and MEL classes; however, the overall error rate remains relatively low. These findings indicate that, in addition to complex classifiers, simpler models supported by well-extracted features can also provide reliable performance.

3.1. Performance Metrics

Based on the results obtained from the models, metric analysis was performed using Accuracy, F1-score, Precision, and Recall in order to evaluate the performance of the method and to determine the most suitable classification model. The comparative results for each metric are presented in Table 6.

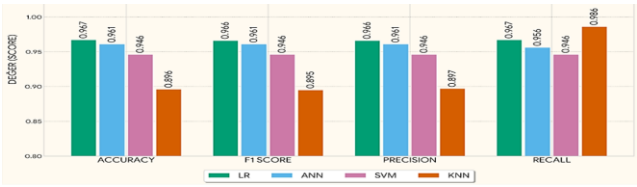
Table 6. Performance Metrics of All Models

	Accuracy	F1 Score	Precision	Recall
ANN	0.961	0.961	0.961	0.956
KNN	0.896	0.895	0.897	0.896
SVM	0.946	0.946	0.946	0.946
LR	0.967	0.966	0.966	0.967

According to the results presented in the table, the LR model appears to be the most successful, achieving an accuracy of 0.967. The lowest classification performance was obtained by the KNN model with an accuracy of 0.896, indicating that it has the lowest discriminative capability among the evaluated models. The ANN model demonstrated a stable performance with an accuracy of 0.961. Finally, the SVM model achieved an accuracy of 0.946, indicating a reliable level of performance.

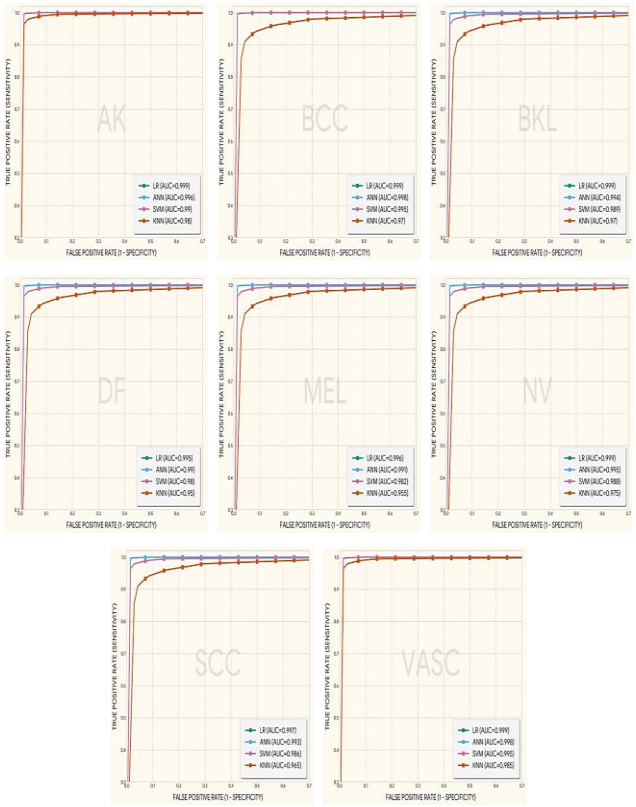
The high accuracy values may be attributed to several methodological factors. First, the SqueezeNet architecture, pretrained on ImageNet, enables the effective transfer of complex visual representations to the skin lesion classification task through transfer learning. Second, the 1000-dimensional feature vector obtained from the conv10 and Global Average Pooling (GAP) layers provides a rich representation of discriminative features among lesion classes, which may facilitate linear separability, particularly for models such as LR. Finally, the application of comprehensive data augmentation techniques (rotation, zooming, and brightness adjustment) and balanced sampling strategies may contribute to improved learning of minority classes (such as AK and VASC), thereby enhancing overall performance. The results are presented in Table 7.

Table 7. Overall Performance



When the other performance metrics are examined, it is observed that the values of F1-score, Precision, and Recall are generally consistent with the classification performances of the models. Overall, all models demonstrate classification accuracies of 89.6% and above. The ROC curves of all models for eight different skin lesion classes are presented in Table 8.

Table 8. ROC curves for all models for eight different classes of skin lesions.



In order to evaluate the classification performance of the developed model in terms of sensitivity–specificity balance, Receiver Operating Characteristic (ROC) curves were generated and the Area Under the Curve (AUC) values were calculated. According to the ROC analysis presented in Table 8, the proposed SqueezeNet-based feature extraction method provides a high level of discriminative capability. The strong convergence of the ROC curves toward the upper-left corner for all classes indicates that the models achieve a favorable balance between sensitivity and specificity. (The graph axes are presented within the ranges of 0–0.7 and 0.3–1.0.) When the performance hierarchy among the models is examined, the highest performance was achieved by the LR and ANN models, which obtained the highest AUC values across all classes. For example, in the AK and VASC classes, the LR model reached an AUC value of 0.999, indicating an almost perfect level of separation. This suggests that the features extracted from SqueezeNet may enable even relatively simple models, such as LR, to achieve performance comparable to deep learning models. The ANN model closely followed LR, with AUC scores ranging between 0.991 and 0.998. The SVM model demonstrated stable performance with AUC values in the range of 0.982–0.999, although it remained slightly behind LR and ANN. Compared to the other three models, KNN exhibited a relatively lower performance profile. Overall, when considering the average of results above 0.95, the findings suggest that the proposed SqueezeNet-based hybrid architecture achieves a high level of recognition and generalization in class discrimination, despite the presence of imbalanced data distribution.

The perfect classification performance observed in the BCC and SCC classes may be attributed to the distinctive characteristics of these lesions, which are effectively represented by the SqueezeNet architecture. However, further validation on larger and more diverse datasets is necessary. To support the validity of the observed 100% accuracy values, a methodological self-validation process was conducted. First, to prevent data leakage, all images were grouped based on the lesion_id parameter and subjected to lesion-level cross-validation. This approach ensured that the model learned pathological features rather than memorizing individual images. Second, the perfect scores in the BCC and SCC classes may be explained by the presence of distinctive dermoscopic features, such as telangiectasia and ulceration, which can be effectively captured by deep learning filters. Nevertheless, it should be noted that these results are limited to the subset derived from the ISIC 2019 dataset. In real-world scenarios, variations in factors such as skin type, imaging devices, and lighting conditions may introduce performance deviations. In conclusion, based on the accuracy values obtained throughout the training process, it can be stated that the classification performance of the models is satisfactory.

4. CONCLUSIONS

Due to the global increase in the number of skin lesions, the need for automated diagnostic systems that can accelerate accurate diagnosis processes and reduce the workload of professionals in this field is increasing. The selected subset used in this study is important in terms of representing different lesion types, including all classes in the analysis process, and enabling comparative evaluation of performance across different classes. In this study, automatic classification of skin lesions was achieved by utilizing the deep feature extraction capability of the SqueezeNet architecture in combination with classical machine learning algorithms (ANN, KNN, SVM, and LR). The ISIC 2019 dataset used in this study includes eight different skin lesion classes (AK, BCC, BKL, DF, MEL, NV, SCC, VASC), and model performances were evaluated using 10-fold cross-validation (k=10). This approach enabled the model not only to recognize images but also to generalize lesion characteristics.

The results obtained from comprehensive experimental analyses indicate that the models achieved generally high classification performance. In particular, when integrated with the LR classifier, the SqueezeNet-based hybrid architecture achieved the highest overall accuracy of 96.7% and an AUC value of approximately 0.99. Only a limited number of errors were observed in the MEL and DF classes. The ANN model closely followed LR with an

accuracy of 96.1%, demonstrating a stable and reliable performance. The 100% accuracy observed in the BCC and SCC classes may be attributed to the more distinctive visual characteristics of these lesion types. In multi-class dermatological image analysis, such results are often associated with class-specific feature distinctiveness or dataset characteristics. It should also be considered that these results may be influenced by the limited dataset size and the selected subset structure. On the other hand, the performance decrease observed in the DF class may be attributed to its visual similarity with other lesion types.

The superior performance of the LR and ANN models compared to other classifiers suggests that the deep features extracted by the SqueezeNet model are effectively utilized and contribute to the sensitivity of the system. In line with this, recent studies in the literature (ISIC 2019 / HAM10000) that employ SqueezeNet and hybrid architectures report comparable performance levels, supporting the validity of the high accuracy obtained in this study. Misclassifications in the KNN model were largely concentrated in the DF class (22 errors). The SVM method ranked third with an accuracy of 94.5%; although it showed strong discriminative performance in the NV and MEL classes, it exhibited relatively higher error rates in the DF and VASC classes. This indicates that performance varies across classes and that error rates tend to increase in visually similar categories.

The KNN method achieved the lowest accuracy (91.8%) and demonstrated weaker class separation capability. In particular, misclassifications increased in the VASC (38 errors) and DF (58 errors) classes. The relatively lower performance in the DF class indicates that the model does not have equal discriminative power across all classes. A comparison of the model results is presented in Table 9.

Table 9. Comparison of Model Results

	Accuracy (%)	F1 Score (%)	Number of Correct Classifications	Number of Incorrect Classifications	Most Successful Classes	Classes that are difficult
ANN	96.1	96.1	1695	68	BCC, SCC	DF
LR	96.7	96.6	1704	59	BCC, SCC	DF,MEL
SVM	94.6	94.6	1668	95	NV, MEL	DF, VASC
KNN	89.6	89.5	1580	183	BCC	DF, VASC

The deep features obtained from SqueezeNet, together with the relatively balanced class distribution of the selected subset consisting of 1,763 images and the applied data augmentation techniques, may have contributed to improving the generalization performance of the model. Despite its low parameter count, SqueezeNet provides high feature extraction capability, offering an effective solution through the combination of deep learning-based feature extraction and classical classifiers. When used in combination with ANN and LR methods, a high level of reliability may be achieved in the computer-aided diagnosis of skin lesions. The high classification performance obtained in this study suggests that the deep features extracted by the SqueezeNet model possess strong discriminative power. In particular, the high-level semantic representations provided by the pretrained model may enhance the performance of classical machine learning algorithms. However, the observed misclassifications between certain classes indicate that model performance may be influenced by dataset characteristics and that additional validation studies on different datasets may be required. The ability of SqueezeNet to produce competitive results with significantly lower computational cost compared to high-parameter models such as ResNet and VGG further highlights the effectiveness of combining deep features with classical machine learning methods. This may provide an advantage, especially in applications with limited computational resources.

In future work, the generalization capability of the model can be improved by extending the training process to include datasets with diverse demographic characteristics. Applying techniques to reduce data imbalance and incorporating explainable methods may further enhance the effectiveness of the diagnostic process. Testing the system in real-world scenarios through mobile applications may facilitate its practical implementation. The small model size and fast processing capability of SqueezeNet make it suitable for mobile platforms. Patients may be able to perform preliminary self-screening at home, and smart dermatoscopic devices integrated with computing systems could be developed to display lesion classifications in real time. Future developments may enable evaluations based on live dermoscopic imaging. The system may also be potentially useful in computer-aided dermatological assessment processes. However, in order to generalize the findings directly to real clinical environments, further studies on larger, more diverse, and clinically validated datasets are required. Additionally, the system could be designed to assist in hospital workflow management, patient prioritization, and as a training tool for medical students and dermatology residents, thereby supporting learning processes and clinical decision-making.

Declaration of Ethical Standards

The article does not contain any studies with human or animal subjects.

Credit Authorship Contribution Statement

Authors individually were responsible for the ideation, modeling, analysis, and writing of this article.

Declaration of Competing Interest

Author claims that there are no conflicts of interest.

Funding / Acknowledgements

No funding was received from any organization for this research.

Data Availability

The dataset is open source that can Access through: https://www.kaggle.com/datasets/salviohexia/isic-2019-skin-lesion-images-for-classification

References

[1] H. Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.

[2] Iandola, F.N., et al., SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360, 2016.

[3] W. Gouda et al., "Detection of skin cancer based on skin lesion images using deep learning," *Healthcare*, vol. 10, no. 6, p. 1057, 2022, doi: 10.3390/healthcare10061057.

[4] N. Nigar et al., "A deep learning approach based on explainable artificial intelligence for skin lesion classification," *IEEE Access*, vol. 10, pp. 113715–113725, 2022, doi: 10.1109/ACCESS.2022.3212277.

[5] I. Iqbal et al., "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," *Computerized Medical Imaging and Graphics*, vol. 88, p. 101843, 2021, doi: 10.1016/j.compmedimag.2020.101843.

[6] M. A. Kassem et al., "Machine learning and deep learning methods for skin lesion classification and diagnosis: A systematic review," *Diagnostics*, vol. 11, no. 8, p. 1390, 2021, doi: 10.3390/diagnostics11081390.

[7] M. Goyal et al., "Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities," *Computers in Biology and Medicine*, vol. 127, p. 104065, 2020, doi: 10.1016/j.compbiomed.2020.104065.

[8] B. Shetty et al., "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," *Scientific Reports*, vol. 12, no. 1, p. 18134, 2022, doi: 10.1038/s41598-022-22644-9.

[9] F. Olayah et al., "AI techniques of dermoscopy image analysis for the early detection of skin lesions based on combined CNN features," *Diagnostics*, vol. 13, no. 7, p. 1314, 2023, doi: 10.3390/diagnostics13071314.

[10] M. Ahammed, M. Al Mamun, and M. S. Uddin, "A machine learning approach for skin disease detection and classification using image segmentation," *Healthcare Analytics*, vol. 2, p. 100122, 2022, doi: 10.1016/j.health.2022.100122.

[11] Y. Wang et al., "SSD-KD: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images," *Medical Image Analysis*, vol. 84, p. 102693, 2023, doi: 10.1016/j.media.2023.102693.

[12] N. Hameed et al., "Mobile-based skin lesions classification using convolution neural network," 2023.

[13] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, p. 180161, 2018, doi: 10.1038/sdata.2018.161.

[14] N. C. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging," in Proc. IEEE ISBI, 2018, pp. 168–172, doi: 10.1109/ISBI.2018.8363547.

[15] M. Combalia et al., "BCN20000: Dermoscopic lesions in the wild," *arXiv preprint arXiv:1908.02288*, 2019.

[16] ISIC, "ISIC 2019 Skin Lesion Images for Classification," [Online]. Available: https://kaggle.com/datasets/shajinrp/skin-lesion-image.

[17] Y. LeCun et al., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[18] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.

[19] V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[20] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. New York: Wiley, 2013, doi: 10.1002/9781118548387.

[21] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.

[22] J. Allgaier and R. Pryss, "Cross-validation visualized: A narrative guide to advanced methods," *Machine Learning and Knowledge Extraction*, vol. 6, no. 2, pp. 1378–1388, 2024, doi: 10.3390/make6020071.