Enhancing Skin Cancer Diagnosis through the Integration of Deep Learning and Machine Learning Approaches

Araştırma Makalesi/Research Article

Yahya DOĞAN[1], Cüneyt ÖZDEMİR[1]

[1]Bilgisayar Mühendisliği, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Siirt Üniversitesi, Siirt, Türkiye

yahyadogan@siirt.edu.tr, cozdemir@siirt.edu.tr

(Geliş/Received:14.05.2024; Kabul/Accepted:21.10.2024)
DOI: 10.17671/gazibtd.1484037

Abstract— Skin cancer is a disease characterized by the uncontrolled proliferation of skin cells, typically manifesting as lesions or abnormal growths. Early diagnosis is critical for improving treatment outcomes. This study proposes an innovative approach to skin cancer diagnosis by integrating modern deep learning models with traditional machine learning algorithms. A three-phase methodology was developed. In the first phase, meaningful features were extracted from skin lesion images using various transfer learning models, including Xception, VGG16, ResNet152V2, InceptionV3, InceptionResNetV2, MobileNetV2, EfficientNetB2, and DenseNet201. In the second phase, dimensionality reduction was performed using Principal Component Analysis (PCA). In the final phase, the reduced feature sets were classified using K-Nearest Neighbors (KNN) and Random Forest (RF) algorithms. Experimental results demonstrated that the highest accuracy of 91.28% was achieved through the combination of DenseNet201 for feature extraction, PCA for dimensionality reduction, and Random Forest for classification. These findings highlight the effectiveness of integrating transfer learning models, dimensionality reduction techniques, and machine learning algorithms in enhancing the accuracy of skin cancer diagnosis.

Keywords— Skin cancer diagnosis, feature selection, dimensionality reduction, transfer learning, machine learning

Cilt Kanseri Teşhisi için Hibrit Derin Öğrenme ve Makine Öğrenmesi Yaklaşımı

Özet— Cilt kanseri, cilt hücrelerinin kontrolsüz çoğalması sonucu ortaya çıkan ve genellikle lezyonlar veya yeni büyümeler olarak kendini gösteren bir hastalıktır. Erken teşhis, tedavi sonuçlarını iyileştirmek için kritik bir rol oynamaktadır. Bu çalışmada, modern derin öğrenme modelleri ile geleneksel makine öğrenimi algoritmalarını birleştirerek cilt kanseri teşhisinde yenilikçi bir yöntem sunulmaktadır. Üç aşamalı bir metodoloji geliştirilmiştir. İlk aşamada, cilt lezyonlarının görüntülerinden anlamlı özellikler çıkarılmış ve bu amaçla Xception, VGG16, ResNet152V2, InceptionV3, InceptionResNetV2, MobileNetV2, EfficientNetB2 ve DenseNet201 gibi çeşitli transfer öğrenme modelleri değerlendirilmiştir. İkinci aşamada, Temel Bileşen Analizi (PCA) ile özellik boyutlarının azaltılması sağlanmış ve üçüncü aşamada ise bu indirgenmiş özellikler, K-En Yakın Komşular (KNN) ve Rastgele Orman (RF) algoritmaları ile sınıflandırılmıştır. Yapılan deneyler sonucunda en yüksek doğruluk %91.28 ile DenseNet201 modelinden elde edilen özelliklerin PCA ile boyutlarının azaltılarak RF algoritması ile sınıflandırılmasıyla elde edilmiştir. Bu bulgular, transfer öğrenme modelleri ile yapılan özellik çıkarma işlemlerinin, PCA ile boyut azaltmanın ve makine öğrenmesi algoritmalarının cilt kanseri teşhisinde yüksek performans sağladığını göstermektedir.

Anahtar Kelimeler— cilt kanseri teşhisi, öznitelik seçimi, boyut azaltma, transfer öğrenme, makine öğrenmesi

1. INTRODUCTION

The number of cancer patients continues to increase due to factors such as rising radiation exposure, environmental changes, viruses, and shifts in dietary habits [1]. Skin cancer emerges as a neoplastic growth resulting from the uncontrolled proliferation of cells within the skin, which is the body's largest organ. Acting as a protective barrier, the skin covers the body's outer layers and contains structures like hair follicles and sweat glands [2]. Each year, thousands of lives are lost to skin cancer [3],[4]. Early diagnosis can enhance therapeutic effectiveness and improve survival outcomes. Patients diagnosed at an early stage are more likely to respond to less invasive treatments and achieve favorable outcomes [5]. When detected early, skin cancer is highly treatable, with five-year survival rates reaching up to 99% [6].

Skin cancer commonly develops in sun-exposed areas, particularly on the face, neck, hands, and arms. The most prevalent types are basal, squamous, and melanoma, each requiring distinct treatment approaches [7],[8]. Traditionally, skin cancer diagnosis relies on physical examination and visual assessment of lesions. However, these methods have limitations, including diagnostic variability among physicians and the inability to detect deeper lesions [9],[10]. Furthermore, limited access to specialist dermatologists, especially in developing countries and rural regions, can hinder early diagnosis [11].

Automated diagnostic systems hold great promise for the early detection of skin cancer. Artificial intelligence-based methods, such as deep learning models, can identify and classify skin lesions, operating independently of dermatologists and potentially improving diagnostic accuracy [12]. However, challenges remain regarding the integration and reliability of these systems in clinical practice. Their performance compared to human dermatologists has not yet been fully validated, and large datasets are required for effective training and long-term sustainability [13]. Several studies have demonstrated the success of deep learning models in detecting and classifying skin lesions, accelerating the diagnostic process, and enhancing accuracy [14]. These models are increasingly being used as alternatives to traditional methods for skin cancer diagnosis [15]. By analyzing skin lesions, they can identify various cancer types, contributing to early detection efforts [16]. Furthermore, their ability to function independently of dermatologists increases access to diagnosis, particularly for patients in remote areas.

Research demonstrates that employing multiple deep learning models can significantly enhance accuracy in skin cancer classification, as ensemble approaches combine the strengths of different architectures to improve predictive performance [17]. In particular, transfer learning has emerged as a powerful technique, enabling models pre-trained on large datasets, such as ImageNet, to be fine-tuned for skin cancer diagnosis. This method not only accelerates the training process but also mitigates the issue of limited domain-specific data by leveraging knowledge from other visual recognition tasks [18]. However, the deployment of deep learning models in this field faces several challenges. The need for large, high-quality labeled datasets is a primary constraint, as collecting and annotating dermatological images requires significant time, financial resources, and expertise. Furthermore, variability in image quality and data sources—including differences in resolution, lighting, and skin tones—can affect model performance and generalizability across populations [13]. The scarcity of annotated datasets also introduces the risk of overfitting, where the model performs well on the training data but struggles with unseen data in real-world scenarios.

The literature features numerous studies utilizing machine learning and deep learning techniques for skin cancer classification [19–21]. Garg et al. [5] proposed a methodology employing image processing techniques to detect melanoma skin cancer, integrating the ABCD rule, which evaluates asymmetry, border irregularity, color, and diameter. After performing illumination correction and lesion segmentation, the model achieved an accuracy of 91.6%. Pham et al. [22] utilized deep CNNs with data augmentation to classify skin lesions, compiling a dataset from multiple sources including the ISBI Challenge, the ISIC Archive, and the PH2 dataset. They evaluated the results using classifiers such as support vector machines, random forests, and artificial neural networks, and achieved an accuracy of 89%. Remya et al. [23] developed a framework combining deep learning and multi-modality data analysis to classify skin lesions. This approach demonstrated significant potential for early detection and accurate diagnosis, showing improvements in key parameters like sensitivity, specificity, and accuracy compared to traditional methods. The model's performance on real-world datasets suggests strong clinical application potential. Fahad et al. [24] proposed a shallow CNN architecture that effectively classifies skin lesions with reduced computational resources, achieving a test accuracy of 98.81% on the heavily imbalanced ISIC 2020 dermoscopic dataset. They used preprocessing and data augmentation techniques to remove image noise and balance the dataset, and compared the model's performance against different transfer learning models, showing superior speed and accuracy. Sivakumar et al. [19] introduced a ResNet50-based model that enhances accuracy through a comprehensive process including data acquisition, preprocessing, segmentation, feature extraction, and classification. The model achieved a peak accuracy of 94% and an F1-score of 93.9% for malignant melanoma detection on the ISIC dataset, demonstrating significant improvement over traditional methods. Furthermore, an application developed by the researchers facilitates faster and more accurate diagnosis of malignant melanoma. Manimurugan et al. [21] presented a novel deep learning architecture combining context-aware convolutional and recurrent neural networks (CA-CNN-RNN) for analyzing skin cancer images. Using H&E stained images from the Cancer Genome Atlas database, the model encoded local features into high-dimensional representations, achieving superior performance across

various classification architectures such as DarkNet-53, VGG-19, ResNet50, and Inception. The CA-CNN-RNN model demonstrated superior metrics, achieving 97.14% accuracy, 96.49% precision, 98.21% recall, and 96.50% F1 score.

This study presents a three-stage methodology for classifying different types of skin cancer, focusing on the synergy between deep learning and traditional machine learning techniques. In the first stage, several transfer learning models—including Xception, VGG16, ResNet152V2, InceptionV3, InceptionResNetV2, MobileNetV2, EfficientNetB2, and DenseNet201—were assessed for their ability to extract meaningful features from skin lesion images. These models, pre-trained on large datasets, were leveraged to transfer learned features, thereby addressing the challenge of limited domain-specific data and accelerating the training process. In the second stage, the features extracted from the best-performing transfer learning model were processed through PCA for dimensionality reduction. PCA helps retain essential information while reducing the feature space, which mitigates the risk of overfitting and decreases the computational burden during classification. Finally, in the third stage, the reduced feature set was classified using two machine learning algorithms: KNN and RF. KNN, known for its simplicity and effectiveness in classification tasks, offers robust performance when the dimensionality is reduced. RF, on the other hand, provides high flexibility and resilience against overfitting through its ensemble-based structure.

3. MATERIAL AND METHODS

This section provides a detailed description of the datasets, proposed methodology, training procedures, and evaluation metrics employed in this study. The goal is to ensure transparency and reproducibility of the research process. First, the dataset used for model training and testing is introduced, including its structure and class distribution. Following this, the proposed hybrid approach, which integrates transfer learning models with traditional machine learning algorithms, is explained in detail. The training details, such as hyperparameters, batch sizes, learning rates, and training epochs, are also discussed. Finally, the performance metrics used to evaluate the effectiveness of the models—such as accuracy, precision, recall, and F1-score—are outlined, providing a comprehensive view of the evaluation criteria used to assess the classification performance.

3.1. Dataset

The dataset utilized in this study was obtained from the Kaggle online platform, provided by the International Skin Imaging Collaboration (ISIC), a global initiative dedicated to advancing melanoma diagnosis. The images used in this research are available at Kaggle: https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign . The ISIC Archive is recognized as the most extensive public repository of dermoscopic images of skin lesions, with all images undergoing rigorous quality control to ensure consistency and reliability. The dataset comprises 1440 images of benign skin lesions and 1197 images of malignant lesions, all standardized to a resolution of 224x224 pixels to ensure compatibility with deep learning models. For this study, the dataset was divided into separate training and test sets. The training set contains 1440 benign and 1197 malignant lesion images, yielding a total of 2637 images. The test set consists of 360 benign and 300 malignant images, amounting to 660 images. Additionally, 10% of the training data was randomly selected and reserved as a validation set to monitor the model's performance during training and prevent overfitting. Figure 1 presents the distribution of benign and malignant lesions in the training set, while Figure 2 illustrates the class proportions in the dataset. This carefully curated dataset ensures that the models are trained and evaluated on high-quality, balanced data, contributing to reliable and meaningful classification outcomes.

Figure 1. Distribution of benign and malignant lesion images in the training set. The bar chart shows that the training set consists of 1440 benign images and 1197 malignant images, highlighting the slight class imbalance present in the dataset.

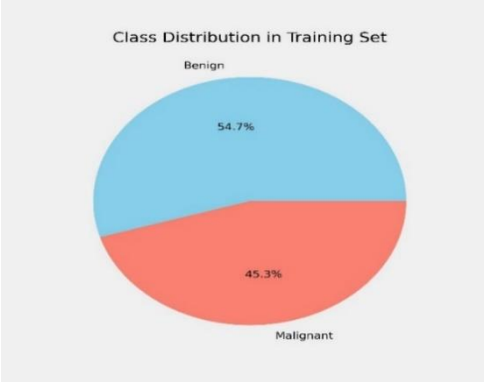

Figure 2. Proportional distribution of benign and malignant lesion images in the training set. The pie chart illustrates that 54.7% of the images represent benign lesions, while 45.3% correspond to malignant lesions.

Figure 3 provides representative sample images from the dataset, illustrating both benign and malignant skin lesions.

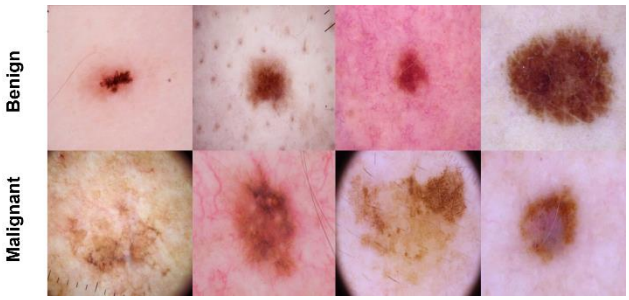

Figure 3: Randomly Selected Benign and Malignant Images from the Dataset

3.1. Proposed Approach

This section presents a detailed explanation of the proposed methodology, which integrates transfer learning models with traditional machine learning algorithms to classify skin cancer. The approach combines the strengths of both techniques, leveraging deep learning for feature extraction and machine learning for effective classification. The proposed framework utilizes a combination of feature extraction through deep learning models, dimensionality reduction techniques, and machine learning classifiers to enhance classification performance. The following subsections provide a comprehensive overview of each component and the workflow involved, highlighting the steps taken to ensure optimal model performance and generalization.

Initially, the DenseNet201 model was employed to extract meaningful features from the skin lesion images. This deep-learning architecture was selected due to its proven ability to capture complex patterns and fine-grained details, which are essential for differentiating between benign and malignant skin lesions. In this context, the DenseNet201 model performs a series of convolutional operations on the input image, denoted as x, to generate a comprehensive feature representation. The output of the DenseNet201 model consists of a feature vector that encapsulates the relevant information extracted from each image, providing a lower-dimensional representation suitable for further processing. Each element within the feature vector corresponds to a distinct characteristic or pattern identified in the image during the feature extraction process. This process is represented mathematically as the function $F(x)$, where x refers to the input image.

$$F(x) = [f_1, f_2, \dots, f_n] \quad (1)$$

Where $F(x)$ denotes the extracted feature vector, with $[f_1, f_2, \dots, f_n]$ representing the individual features. Each f_i (where i=0, 1, 2, …, n) is a numerical value that captures a specific attribute of the input image. This feature vector serves as a crucial input for subsequent stages, such as dimensionality reduction and classification, facilitating accurate and efficient diagnosis of skin cancer.

Secondly, the PCA method was applied to reduce the dimensionality of the obtained feature vectors. PCA performs dimensionality reduction by performing principal component analysis on the feature vectors. Accordingly, the covariance matrix C of the feature vector is computed, and transformation is carried out based on the eigenvalues and eigenvectors of this matrix. These operations are provided mathematically in Equations 2, 3, and 4.

$$C = \frac{1}{n}(X - \bar{x})^T(X - \bar{x}) \quad (2)$$

Where C represents the covariance matrix of the feature vector, $F(x)$ denotes the data matrix formed by feature vectors, and $\bar{x}$ represents the mean of the feature vectors.

$$C_{vi} = \lambda_i v_i \quad (3)$$

Where v_i represents the eigenvectors, forming the principal components of the PCA transformation, λ_i denotes the eigenvalues determining the significance of the eigenvectors, and Z represents the matrix of newly obtained feature vectors after dimensionality reduction.

$$Z = X_{v_{selected}} \quad (4)$$

After dimensionality reduction, feature vectors are classified using the RF classifier. RF operates by constructing multiple decision trees during training and outputting the mode of the classes for classification. For each new example, every decision tree within the forest independently provides a classification. The final class is then determined by taking the majority vote across all trees, meaning the class predicted by the most trees becomes the overall prediction. Mathematically, this process is expressed as:

$$y = mode(y_{trees}) \quad (5)$$

Where y represents the predicted class of the new example, y_{trees} denotes the set of class predictions from all decision trees in the forest, and the mode operator indicates the most frequently occurring value among these predictions.

3.1. Metrics

The performance of the proposed skin cancer classification model was assessed using several key metrics to evaluate its accuracy, reliability, and ability to distinguish between benign and malignant lesions. The following metrics were employed to ensure a comprehensive performance evaluation:

Accuracy represents the overall effectiveness of the model by calculating the percentage of correctly classified cases among all predictions. For skin cancer classification, it reflects how well the model identifies both benign and malignant lesions accurately.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (6)$$

Precision measures the proportion of true malignant cases among all instances predicted as malignant. In skin cancer diagnosis, a high precision rate ensures that the model minimizes false positives, reducing unnecessary patient anxiety and avoiding unnecessary biopsies or treatments.

$$Precision = TP / (TP + FP) \quad (7)$$

Recall, also known as sensitivity, reflects the model's ability to correctly identify all malignant lesions. In the context of skin cancer, recall is crucial as missing a malignant case (false negative) could delay treatment, potentially worsening patient outcomes.

$$Recall = TP / (TP + FN) \quad (8)$$

The F1 Score provides a balance between precision and recall, making it an important metric in scenarios where class imbalance exists, such as in the skin cancer dataset. A high F1 Score ensures that the model performs well in both identifying true positives and minimizing false positives.

$$\text{F1 Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (9)$$

The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve evaluates the model's capability to distinguish between benign and malignant lesions across different thresholds. A high AUC value indicates that the model is effective in differentiating between the two classes, which is essential for reliable skin cancer diagnosis.

3.1. Training details

The training process for the skin cancer classification model was configured with several key hyperparameters to ensure optimal performance. All input images were resized to 224x224 pixels and normalized by scaling pixel values to the range [0, 1] (1/255.0) to facilitate faster convergence. The Adam optimizer was employed for weight updates due to its adaptive learning rate and ability to handle sparse gradients effectively. As the task involved binary classification, binary cross-entropy was used as the loss function to measure the discrepancy between predicted probabilities and true labels. The model was trained for a maximum of 100 epochs with a batch size of 32, balancing training speed and memory efficiency. To prevent overfitting, early stopping with a patience value of 6 was applied, meaning the training process stopped if the validation performance failed to improve for six consecutive epochs. Table 1 shows the hyperparameters used in the training phase of Transfer learning models used as feature extractors.

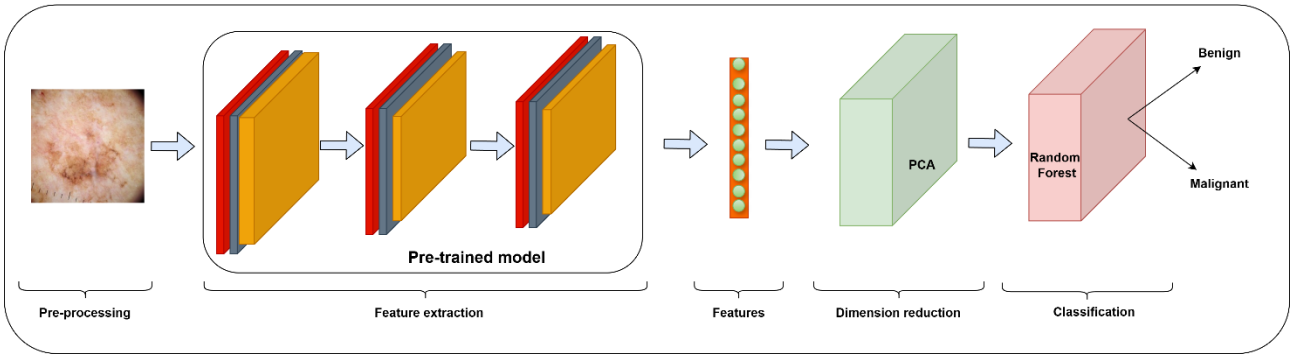

Figure 4: The proposed architecture involves several stages. Initially, preprocessing was applied to the images before training. The DenseNet201 model was then used to extract relevant features. Following this, the most significant feature vectors were selected using the PCA dimensionality reduction algorithm. Finally, the RF algorithm was employed to classify the reduced feature set

Table 1: Hyperparameters

Hyperparameters	**Value**
Image size	224x224
Normalization	Yes (1/255.0)
Optimization Algorithm	Adam
Early Stopping	Yes (Patience: 6)
Epoch	100
Batch Size	32
Loss Function	Binary cross entropy

4. EXPERİMENT AND RESULTS

This section presents the experimental results obtained using the proposed approach. The experiments were conducted with several popular transfer learning models, including Xception, VGG16, ResNet152V2, InceptionV3, InceptionResNetV2, MobileNetV2, EfficientNetB2, and DenseNet201, to evaluate their performance in classifying benign and malignant skin lesions. During preprocessing, only normalization (scaling pixel values to the range [0,1]) was applied, without the use of any additional preprocessing or data augmentation techniques. Feature extraction was performed using the selected transfer

learning models, where only the output layer with a sigmoid activation function was utilized, with no additional hidden layers added. Binary cross-entropy was employed as the loss function for optimization. A 10% subset of the training data was reserved for validation, while the images designated for testing, as outlined in the dataset section, were used for the final evaluation. The summarized performance metrics of each model are presented in Table 2.

Among all the models, DenseNet201 delivered the best performance, achieving a test accuracy of 90.61%, an F1 Score of 90.60%, and an AUC of 90.50%. This model's superior performance can be attributed to its architecture, which effectively captures complex features through densely connected layers, making it highly suitable for differentiating between benign and malignant skin lesions. Additionally, DenseNet201 maintained a well-balanced performance between precision (90.60%) and recall (90.61%), indicating that it effectively identified malignant cases without generating an excessive number of false positives.

The Xception model also demonstrated strong performance, with a test accuracy of 86.97% and an F1 Score of 86.98%, making it the second-best model in the study. The high accuracy of Xception reflects its ability to capture spatial and abstract features efficiently, contributing to reliable classification results. Similarly, ResNet152V2 performed well, achieving 86.52% accuracy, closely trailing Xception. Both models' high F1 Scores (86.98% and 86.49%) confirm their suitability for skin cancer classification tasks.

Other models, such as InceptionV3, VGG16, and InceptionResNetV2, produced moderate performance with test accuracies ranging between 81.52% and 85.00%. While these models demonstrated acceptable precision and recall values, their overall performance was slightly inferior to that of DenseNet201 and Xception. These results suggest that while these models are effective in identifying patterns in skin lesions, they may not generalize as well to unseen data compared to more advanced architectures like DenseNet201.

MobileNetV2 and EfficientNetB2 performed poorly, failing to achieve satisfactory results. MobileNetV2, with a test accuracy of 57.42%, exhibited poor generalization, as reflected in its low F1 Score of 44.65%. Although its precision (76.09%) was relatively high, this model failed to maintain adequate recall, leading to an imbalanced classification performance. EfficientNetB2 demonstrated even weaker performance, with an accuracy of 54.55% and a precision of only 29.75%, despite achieving a recall of 54.55%. The low F1 Score (38.50%) indicates that EfficientNetB2 was not well-suited for this classification task. Figure 5 presents the loss and accuracy graphs for DenseNet201, the model that achieved the highest performance in this study. In contrast, Figure 6 displays the loss and accuracy graphs for EfficientNetB2, which demonstrated the lowest performance. These graphs provide insights into the training process of both models, highlighting DenseNet201's superior ability to converge effectively and achieve high accuracy while illustrating the challenges faced by EfficientNetB2 in learning from the dataset. Figure 7 illustrates the ROC curve for the DenseNet201 model, which provides a visual representation of the model's ability to distinguish between benign and malignant skin lesions at various threshold settings. The area under the curve (AUC) indicates the overall effectiveness of the model in terms of sensitivity and specificity. A higher AUC suggests better performance in correctly classifying positive and negative instances. Figure 8 shows the confusion matrix, which details the model's classification outcomes by comparing the predicted labels with the actual labels. The matrix provides insight into the number of true positives, true negatives, false positives, and false negatives, allowing for a more detailed evaluation of the model's accuracy and error distribution.

In summary, the experimental results highlight that DenseNet201 outperformed all other models, providing the most accurate and balanced classification performance. In contrast, MobileNetV2 and EfficientNetB2 were less effective, suggesting that lightweight architectures may struggle with the complexities of skin cancer detection. These findings underscore the importance of selecting robust deep learning models, such as DenseNet201, for tasks requiring high precision and reliability in medical diagnostic.

Table 2: Performance Results of Transfer Learning Models

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**	**AUC (%)**
Xception	86.97	87.02	86.97	86.98	86.97
VGG16	81.52	84.41	81.52	81.40	82.61
ResNet152V2	86.52	86.51	86.52	86.49	86.25
InceptionV3	84.55	84.78	84.55	84.43	83.94
InceptionResNetV2	85.00	86.02	85.00	84.74	84.03
MobileNetV2	57.42	76.09	57.42	44.65	53.17
DenseNet201	**90.61**	**90.60**	**90.61**	**90.60**	**90.50**
EfficientNetB2	54.55	29.75	54.55	38.50	50.00

The performance of deep learning models is largely influenced by their ability to capture and learn complex patterns within the data. While these models excel at extracting deep features, the resulting high-dimensional feature space often leads to increased computational demands, making model training both time-consuming and resource-intensive. To mitigate these challenges, dimensionality reduction techniques like PCA are employed. PCA reduces the number of features by identifying the most significant components that preserve the essential variability in the data, thereby enabling more efficient and faster processing.

In the second phase of the experimental studies, rather than directly training transfer learning models—which can be computationally expensive and time-consuming—features were first extracted using the DenseNet201 model. DenseNet201 was selected based on its superior performance in earlier experiments, as it demonstrated the best classification accuracy and balance across key metrics. The high-dimensional features obtained from this model were then subjected to dimensionality reduction using PCA to minimize complexity without losing critical information.

After reducing the feature set with PCA, two machine learning algorithms, KNN and RF, were used to classify the reduced features. This approach combines the deep feature extraction power of DenseNet201 with the efficiency of traditional machine learning classifiers. The RF and KNN algorithms were chosen for their complementary strengths—RF for its robustness and ability to handle large datasets, and KNN for its simplicity and effectiveness in smaller, well-structured datasets.

The classification results obtained using this approach were evaluated on the test set, and the performance metrics, including accuracy, precision, recall, F1 score, and AUC, are presented in Table 3. The results demonstrate the effectiveness of integrating deep learning with dimensionality reduction and machine learning classifiers, offering a more computationally efficient solution while maintaining high classification performance.

This combined methodology allows for a scalable, efficient approach to skin cancer diagnosis, enabling the model to perform well even when computational resources are limited. The success of this approach highlights the potential of using feature extraction followed by dimensionality reduction and machine learning classification as a viable alternative to fully training deep learning models in resource-constrained environments.

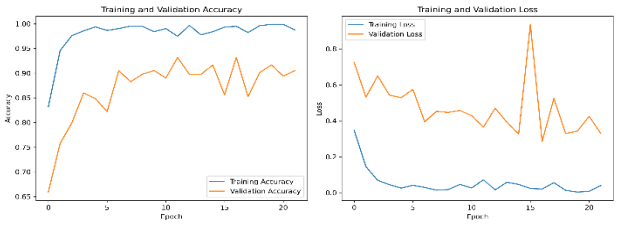

Figure 5: Loss and accuracy graph for DenseNet201

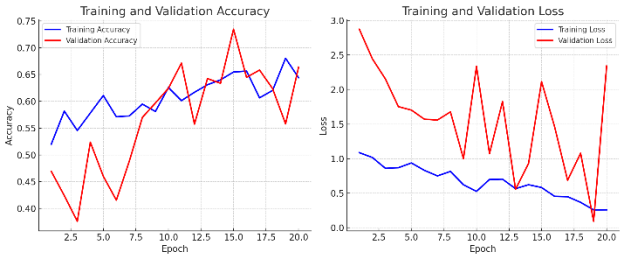

Figure 6: Loss and accuracy graph for EfficientNetB2

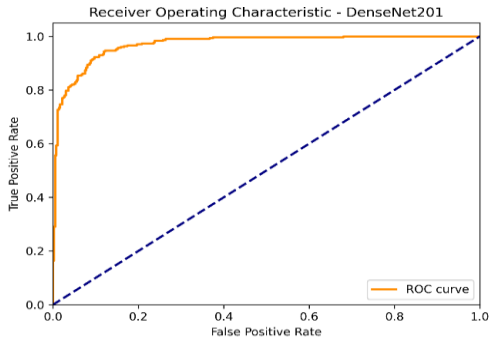

Figure 7: Roc curve for DenseNet201

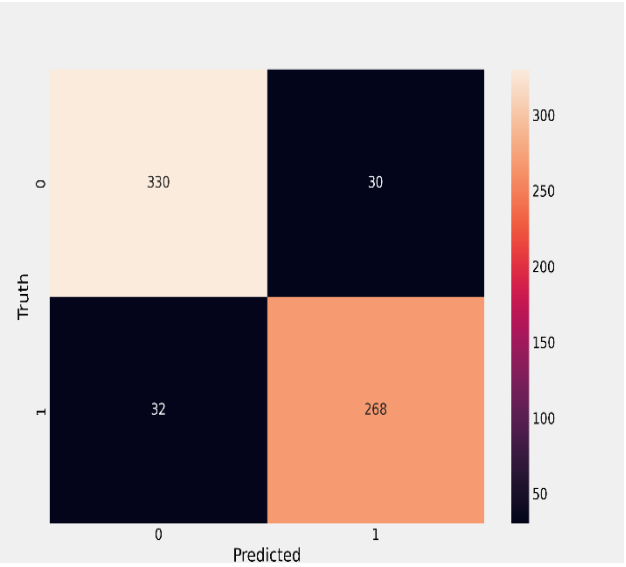

Figure 8: Confusion matrix

Table 3: Results of the Hybrid Model Approach

Model	Test Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC (%)
KNN	91.1	91.07	91.1	91.09	90.85
Random Forest	91.28	91.58	91.28	91.32	91.69

Table 3 presents the performance metrics of hybrid models utilizing different machine learning algorithms after feature extraction by a deep learning model. The KNN classifier yielded notable results with a test accuracy of 91.1%, precision of 91.07%, recall of 91.1%, F1 score of 91.09%, and an AUC of 90.85%. In comparison, the Random Forest classifier demonstrated marginally superior performance, achieving a test accuracy of 91.28%, precision of 91.58%, recall of 91.28%, F1 score of 91.32%, and an AUC of 91.69%.

The results indicate that both classifiers when applied to the features extracted from the DenseNet201 model and processed through PCA, exhibit high performance metrics. However, the RF classifier consistently performs better than KNN across all evaluated metrics. This indicates that while KNN delivers solid performance, the Random Forest model demonstrates higher classification accuracy and AUC, proving to be more effective in handling dimensionality-reduced features. These findings underline the efficacy of using a deep learning model for feature extraction combined with PCA for dimensionality reduction, followed by classification with machine learning algorithms. Figure 9 illustrates the qualitative performance of the proposed approach by comparing the predicted outputs with ground truth on randomly selected test images. This figure further validates the effectiveness and reliability of the hybrid model approach.

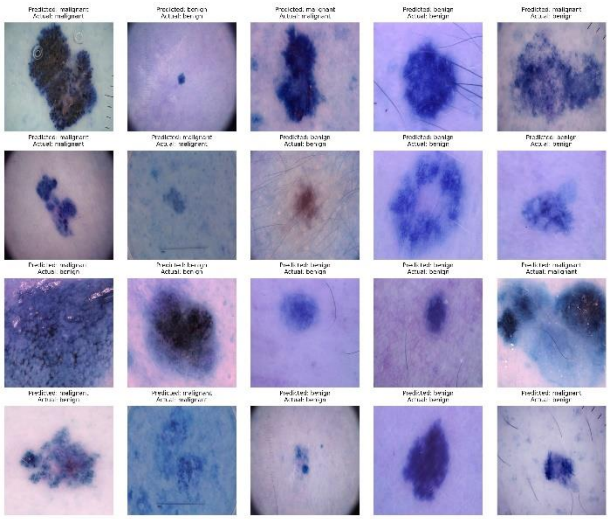

Figure 9: Results of the qualitative assessment of the proposed approach's performance on test images

5. CONCLUSION

This study demonstrates the effectiveness of a hybrid approach that integrates deep learning models with machine learning algorithms for the early detection of skin cancer. In the first phase, several transfer learning models were evaluated, with DenseNet201 emerging as the top performer, achieving high accuracy, precision, recall, F1 score, and AUC. Given its superior performance, DenseNet201 was selected for feature extraction in the subsequent phase, and PCA was employed to reduce the dimensionality of the extracted features.

The reduced features were classified using machine learning algorithms, with RF delivering the best overall performance, surpassing the results of KNN. findings highlight that integrating dimensionality reduction with machine learning classifiers can enhance performance beyond what is achievable by deep learning models alone. Specifically, the RF classifier demonstrated robustness in handling the dimensionality-reduced feature space, providing superior accuracy and balanced performance.

This approach underscores that machine learning techniques, when applied to features extracted from transfer learning models, can offer meaningful improvements in classification tasks. Rather than aiming for state-of-the-art results, the study emphasizes the practicality and efficiency of combining these methodologies to achieve reliable outcomes.

In conclusion, the integration of deep learning with machine learning techniques presents a practical approach for enhancing skin cancer diagnosis, with considerable potential benefits for clinical practice. Future studies will focus on expanding the dataset and incorporating more diverse skin lesion types to improve the model's generalization and robustness.

REFERENCES

[1] Alam, T. M., Khan, M. M. A., Iqbal, M. A., Abdul, W., & Mushtaq, M. (2019). Cervical cancer prediction through different screening methods using data mining. IJACSA) *International Journal of Advanced Computer Science and Applications*, 10(2).

[2] Rognoni, E., & Watt, F. M. (2018). Skin cell heterogeneity in development, wound healing, and cancer. Trends in cell biology, 28(9), 709-722.

[3] Fitzmaurice C, Akinyemiju TF, Al Lami FH, Alam T, Alizadeh-Navaei R, Allen C, et al. (2018) Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the global burden of disease study. JAMA Oncol.;4(11):1553–68

[4] Dinehart, S. M. (2000). The treatment of actinic keratoses. *Journal of the American Academy of Dermatology*, 42(1), S25-S28

[5] Flohr, C., and Hay, R. 2021. Putting the burden of skin diseases on the global map. **British Journal of Dermatology**, 184(2): 189-190.

[6] Skin Cancer Facts & Statistics [Internet]. 2021. Available from: https://www.skincancer.org/skin-cancer information/skin-cancer-facts/

[7] Ergün, E., & Kılıç, K. (2021). Derin öğrenme ile artırılmış görüntü seti üzerinden cilt kanseri tespiti. *Black Sea Journal of Engineering and Science*, 4(4), 192-200.

[8] Andrew, T. W., Alrawi, M., & Lovat, P. (2021). Reduction in skin cancer diagnoses in the UK during the COVID-19 pandemic. Clinical and Experimental Dermatology, 46(1),145-146.

[9] Lukaviciute, L., Ganceviciene, R., Navickas, P., Navickas, A., Grigaitiene, J., and Zouboulis, C. C. 2020. Anxiety, depression, and suicidal ideation amongst patients with facial dermatoses (acne, rosacea, perioral dermatitis, and folliculitis) in Lithuania. *Dermatology*, 236(4): 314-322.

[10] Arnold, J. D., Yoon, S., and Kirkorian, A. Y. 2019. The national burden of inpatient dermatology in adults. **Journal of the American Academy of Dermatology**, 80(2): 425-432.

[11] Feng, H., Berk-Krauss, J., Feng, P. W., and Stein, J. A. 2018. Comparison of dermatologist density between urban and rural counties in the United States. **JAMA Dermatology**, 154(11): 1265-1271.

[12] Assiri, A., et al. 2013. Hospital outbreak of Middle East respiratory syndrome coronavirus. **The New England Journal of Medicine**, 369(5): 407-416.

[13] W. Yue, Z. Wang, H. Chen, A. Payne, & X. Liu, "Machine learning with applications in breast cancer diagnosis and prognosis", *Designs*, vol. 2, no. 2, p. 13, 2018.

[14] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, & S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model", *Scientific Reports*, vol. 7, no. 1, 2017.

[15] C. AKYEL and N. Arici, "A new approach to hair noise cleansing and lesion segmentation in images of skin cancer", *Politeknik Dergisi*, vol. 23, no. 3, p. 821-828, 2020. DOI: 10.2339/politeknik.645395

[16] A. Karli, "Cilt kanseri görüntüleri kullanılarak eğitilen efficientnet-b3 mimarisinde hiperparametre seçiminin sınıflandırma performansına etkisinin incelenmesi", *Fırat Üniversitesi Mühendislik Bilimleri Dergisi,* vol. 36, no. 1, p. 499-507, 2024. DOI: 10.35234/fumbd.1426044

[17] N. Kausar, A. Hameed, M. Sattar, R. Ashraf, A. Imran, M. Abidinet al., "Multiclass skin cancer classification using ensemble of fine-tuned deep learning models", Applied Sciences, vol. 11, no. 22, p. 10593, 2021. DOI: 10.3390/app112210593

[18] A. Mahbod, G. Schaefer, C. Wang, G. Dorffner, R. Ecker, & I. Ellinger, "Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification", Computer Methods and Programs in Biomedicine, vol. 193, p. 105475, 2020. DOI: 10.1016/j.cmpb.2020.105475

[19] Sivakumar, M. S., Leo, L. M., Gurumekala, T., Sindhu, V., & Priyadharshini, A. S. (2024). Deep learning in skin lesion analysis for malignant melanoma cancer identification. Multimedia Tools and Applications, 83(6), 17833-17853.

[20] Gajera, H. K., Nayak, D. R., & Zaveri, M. A. (2023). A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. Biomedical Signal Processing and Control, 79, 104186.

[21] Manimurugan, S. (2023). Hybrid high performance intelligent computing approach of CACNN and RNN for skin cancer image grading. Soft Computing, 27(1), 579-589.

[22] Pham, T.C., Luong, C.M., Visani, M. and Hoang, V.D., 2018, March. Deep CNN and data augmentation for skin lesion classification. In Asian Conference on Intelligent Information and Database Systems (pp. 573-582). Springer, Cham.

[23] Remya, S., Anjali, T., & Sugumaran, V. (2024). A Novel Transfer Learning Framework for Multimodal Skin Lesion Analysis. IEEE Access.

[24] Fahad, N. M., Sakib, S., Raiaan, M. A. K., & Mukta, M. S. H. (2023, February). Skinnet-8: An efficient cnn architecture for classifying skin cancer on an imbalanced dataset. In 2023 International conference on electrical, computer and communication engineering (ECCE) (pp. 1-6). IEEE.