Journal of Autonomous Intelligence (2023) Volume 6 Issue 3
doi: 10.32629/jai.v6i3.747

ORIGINAL RESEARCH ARTICLE

Classification of skin lesion using deep convolutional neural network by applying transfer learning

Danish Jamil[1,2,*], Farheen Qazi[2], Dur-E-Shawar Agha[2], Sellappan Palaniappan[1]

[1] *Department of Information Technology, Malaysia University of Science and Technology, Kuala Lumpur 47810, Malaysia*

[2] *Department Software Engineering, Syed University of Engineering and Technology, Karachi 75300, Pakistan*

*** Corresponding author:** Danish Jamil, danish.jamil@phd.must.edu.my

ABSTRACT

The early and accurate diagnosis of skin cancer is crucial for improving patient outcomes and reducing the need for invasive biopsies. This study proposes a deep learning model for classifying skin malignancy using transfer learning and data augmentation techniques to address limitations observed in previous models and enhance diagnostic accuracy. The approach involves applying transfer learning to a pre-trained ResNet152 architecture using tensorflow and keras. Data augmentation techniques are employed on a dataset consisting of 10,015 skin lesion images obtained from the international skin imaging collaboration (ISIC) 2018 challenge, which encompasses diverse lesion types, sizes, and colors, posing a challenging classification task. Binary cross-entropy serves as the loss function, and the Adam optimizer is utilized for training the model. The results demonstrate a specificity of 87.42% and an F1 score of 0.854, outperforming other models in the field. These statistical findings highlight the effectiveness of transfer learning and data augmentation techniques in improving the accuracy of skin cancer diagnosis. The novelty of this study lies in the combination of transfer learning and data augmentation methods to enhance diagnostic accuracy. However, it is important to acknowledge the limitations of this study, including the necessity for further investigation to evaluate the clinical practicality of the model and address potential biases. Future research could explore the application of this model in a clinical setting and the development of models for detecting other types of skin lesions. In conclusion, the proposed deep learning model based on the ResNet152 architecture showcases promising results in the classification of skin lesions, demonstrating its potential for accurate skin cancer diagnosis. With further research and improvement, these models have the potential to revolutionize healthcare, improving patient outcomes, reducing healthcare costs, and increasing accessibility to screening and diagnosis, particularly for underserved populations.

Keywords: skin malignancy; deep learning; data augmentation; transfer learning; skin lesion; biopsy reduction; diagnosis accuracy

ARTICLE INFO

Received: 18 June 2023
Accepted: 3 August 2023
Available online: 25 September 2023

1. Introduction

Skin cancer is a major concern in the 21st century, with cases increasing every day. It is the third most common type of cancer and a leading cause of non-accidental death among individuals aged 20–39[1,2]. Moreover, melanoma cases have increased by 53% in the last decade, primarily due to UV exposure[3]. Early detection of skin lesions is critical to prevent metastasis and fatal consequences. However, it can be challenging to differentiate many skin lesions, even for experienced dermatologists, due to their subtle differences. Prompt and accurate diagnosis of skin lesions is crucial, as early detection can prevent metastasis and fatal consequences[4]. However, many skin lesions can be difficult to differentiate, even for experienced dermatologists, due

to their subtle differences[5]. To improve the accuracy and efficiency of skin cancer diagnosis while reducing the need for invasive biopsy procedures, a novel deep learning model was developed using transfer learning and data augmentation techniques. It is important to note that this study is not an ablation study, as the proposed model was not evaluated by removing or modifying individual components to analyze their contribution to the overall performance.

As shown in **Figure 1**, the HAM10000 dataset comprises 10,015 dermoscopy images collected from two different locations over 20 years. A selection of dermoscopy images representing each of the seven types of skin lesions used in the study[6].

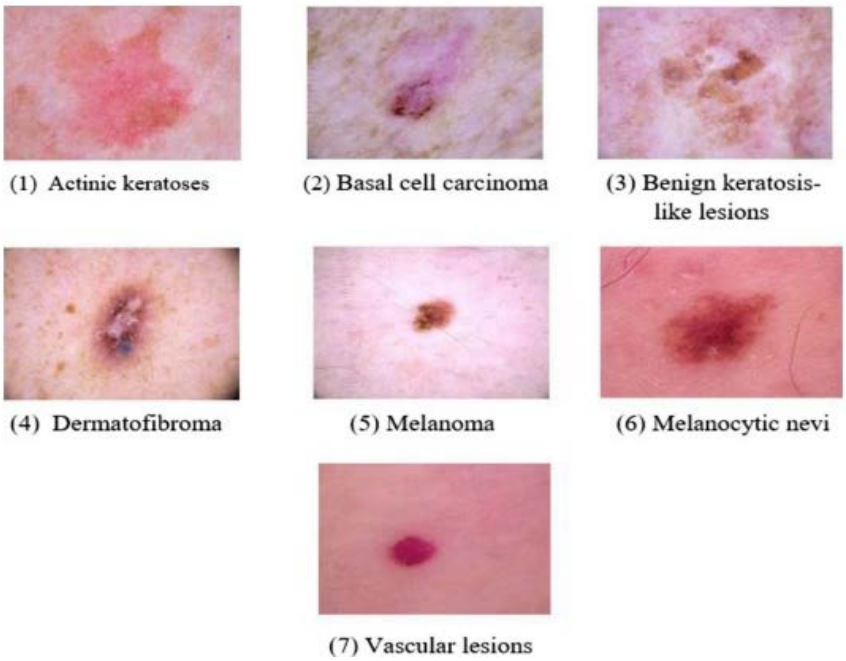

Figure 1. Sample images for the seven skin lesion categories from HAM10000 dataset[6].

The seven categories of skin lesions utilized in the study are:

Melanocytic nevi: benign tumors that may vary in size, color, and appearance[7].

Melanoma: malignant skin tumors that originate in melanocytes, the cells responsible for pigmentation in the skin[8].

Basal cell carcinoma: the most common type of skin cancer that appears as a shiny or waxy bump on the skin[9].

Actinic keratosis: precancerous skin lesions that present as scaly or crusty spots on the skin.

Benign keratosis: a group of non-cancerous skin lesions caused by the overgrowth of keratinocytes. The typical outward manifestation is roughness or scaling[10].

Dermatofibroma: benign skin tumours originating from collagen-producing fibroblasts. They usually appear as firm or hard lumps on the skin[11].

Vascular lesions are skin lesions that arise from blood vessels in the skin. The type of lesion depends on its appearance and location, and can include conditions such as hemangiomas, port-wine stains, or pyogenic granulomas[12].

According to the World Health Organization (WHO), skin cancer is the most common cancer worldwide, and its incidence is increasing. The incidence of melanoma, the most deadly form of skin cancer, has been

increasing over the past few decades, particularly in countries with fair-skinned populations, such as the United States, Australia, and Europe[13]. According to the latest data from the American cancer society's cancer facts & figures 2023 report, melanoma remains a significant health challenge. In the United States alone, it is estimated that in 2023, approximately 186,680 new cases of melanoma will be diagnosed, with 97,610 cases classified as invasive. Tragically, in 2022, over 7500 lives were lost to this aggressive form of skin cancer, with men accounting for 5000 of those fatalities and women comprising around 2500. However, there is hope as the report reveals a 4% decrease in death rates due to melanoma between 2014 and 2018. These compelling statistics serve as a powerful reminder of the urgent need for early detection strategies and robust preventive measures to combat the rising burden of skin cancer[13]. In 2021, the American cancer society estimated 106,110 new cases of melanoma in the United States, with approximately 7180 deaths from the disease[14,15]. The current diagnostic methods for skin cancer typically involve a visual inspection of skin lesions by a dermatologist or healthcare provider, followed by a biopsy of suspicious lesions for microscopic examination. Dermoscopy, a non-invasive imaging technique, may also be used to visualize skin lesions in more detail[16]. While these methods are effective in identifying many cases of skin cancer, they have several limitations. Visual inspection can be subjective and may miss small or subtle lesions, leading to delayed diagnosis and treatment. A biopsy is an invasive procedure that can be uncomfortable for patients and may leave scars. Dermoscopy requires specialized equipment and training, and its accuracy depends on the skill of the practitioner[17]. Moreover, these methods can be time-consuming and costly, and they may not be accessible to all patients, especially those living in remote or underserved areas. As a result, there is a need for new diagnostic methods that are non-invasive, accurate, and cost-effective[18]. This study aims to improve the accuracy and efficiency of skin cancer diagnosis and reduce the need for invasive biopsy procedures. The development of an accurate and efficient diagnostic tool would greatly benefit the field of cancer metastasis and treatment by improving diagnostic accuracy, reducing the probability of misdiagnosis, and ultimately improving patient outcomes. Despite the availability of various machine-learning models for skin cancer classification, the existing literature still lacks a comprehensive and accurate deep learning model for the early detection and classification of skin malignancy from dermoscopic images. Most of the existing models are limited in their ability to accurately classify different types of skin lesions due to the subtle variations in color, texture, and patterns[13,19]. In recent years, deep learning models have become increasingly popular for a variety of machine learning tasks, including classification. However, it is still an open question whether these models are more effective when applied end-to-end for classification or when used as feature extractors. In this paper, the question is explored by comparing the performance of several deep learning models on a classification task, both when applied end-to-end and when used as feature extractors. This study aims to fill the research gap by developing a deep learning model using dermoscopic images and transfer learning techniques. The proposed model will be trained on a large dataset and will use the ResNet152 architecture to accurately classify different types of skin lesions with high accuracy and sensitivity. The study will also compare the performance of the proposed model with existing models to demonstrate its superiority and effectiveness in improving the diagnosis of skin cancer.

The main objective of this study is to develop a deep learning model using dermoscopic images for accurate classification of skin malignancy. The study aims to answer the following research questions:

- The study aims to evaluate the accuracy of the proposed deep learning model in classifying different types of skin lesions using dermoscopic images.
- The study investigates the potential improvement in performance of the deep learning model with transfer learning.
- The study examines the impact of using the ResNet152 architecture on the accuracy of the deep learning model.
- The study aims to determine whether the developed model can assist dermatologists in the prompt

diagnosis of skin malignancies and reduce the likelihood of misdiagnosis.

The hypothesis is that the proposed model, utilizing transfer learning and ResNet152 architecture, can accurately classify skin lesions and assist dermatologists in the prompt diagnosis of skin malignancies.

As shown in **Table 1**, this provides a summary of the number of images in each of the seven classes of skin lesions included in the HAM10000 dataset used for training the deep learning model proposed in the study. The classes include actinic keratosis, basal cell carcinoma, melanoma, benign keratosis, dermatofibroma, melanocytic nevi, and vascular lesions. Moreover, as shown in **Table 1** shows that the dataset contains 10,015 images, with melanocytic nevi being the most common class with 6705 images and dermatofibroma being the least common with only 115 images. This information is useful for understanding the distribution of data in the dataset and the potential impact of class imbalance on model performance[20]. Previous studies have often relied on smaller or less diverse datasets, which may limit the generalizability of their findings to different populations[21]. This study addresses the limitation of using smaller or less diverse datasets by utilizing a large and diverse dataset of dermoscopic images of skin lesions, which improves the reliability and validity of the proposed model. Moreover, previous studies in the field have shown promising results in using deep learning models for diagnosing skin cancer. For example, researchers have developed convolutional neural networks (CNNs) that accurately differentiate between malignant and benign skin lesions based on images[22]. Some studies have reported accuracy rates of up to 95%, comparable to the performance of dermatologists[23].

Table 1. Distribution of images across different classes of skin lesions.

Skin lesion class	Number of images
Actinic keratosis	327
Basal cell carcinoma	514
Melanoma	113
Benign keratosis	1099
Dermatofibroma	115
Melanocytic nevi	6705
Vascular lesion	142
Total	10,015

The importance of evaluating the proposed deep learning model's performance across all classes of skin lesions to determine its clinical utility is crucial. It is essential to investigate potential biases and limitations of the model and assess its ability to generalize to new data. If the model performs well on specific subsets of images or lesions, it may be less useful in clinical practice, where there is a wide range of lesion types and presentations. To address this concern, the proposed study aims to evaluate the accuracy of the deep learning model in classifying all seven classes of skin lesions using the HAM10000 dataset. This evaluation will identify any potential weaknesses or biases in the model and assess its ability to generalize to new data. Additionally, the study aims to compare the proposed model's performance with existing models, providing further insights into its effectiveness and clinical utility. These analyses will provide a more comprehensive understanding of the potential benefits and limitations of the proposed deep learning model in assisting dermatologists in diagnosing skin malignancies.

Furthermore, while some studies have used transfer learning to improve deep learning model performance[24], there is a need to explore the optimal approach for transferring knowledge from pre-trained architectures to skin lesion classification. This study contributes to this area by investigating the effectiveness of transfer learning with the ResNet152 architecture in accurately classifying different types of skin lesions. However, there are potential limitations to using deep learning models for skin cancer diagnosis. For example,

these models may be less effective in cases where the lesion is not clearly visible or the image quality is poor. Additionally, deep learning models may be more prone to overfitting the training dataset, which could lead to poor generalization performance on new and unseen data[25]. Overall, this study aims to address the gaps in the literature by proposing a deep learning model that utilizes transfer learning with ResNet152 architecture to accurately classify skin malignancy in dermoscopic images of skin lesions. The study also aims to contribute to the development of new techniques for analyzing and interpreting dermoscopic images and to provide insights into the performance and limitations of deep learning models for skin cancer diagnosis.

As shown in **Figure 2**, depicts the flowchart of the skin lesion classification process using the proposed deep learning model. The process begins with the user inputting their information and the dermoscopic image of the skin lesion. The pre-trained deep learning model is then loaded, and the necessary data is supplied to the model for classification. Based on the trained parameters, the model classifies the lesion and outputs the results, indicating the presence or absence of malignancy. The flowchart demonstrates the overall simplicity and efficiency of the proposed model, which can aid dermatologists in accurately and timely diagnosing skin cancer, leading to early detection and treatment and ultimately improving patient outcomes. The study will adopt a deep learning approach to develop a model for accurately classifying skin malignancy from dermoscopic images.

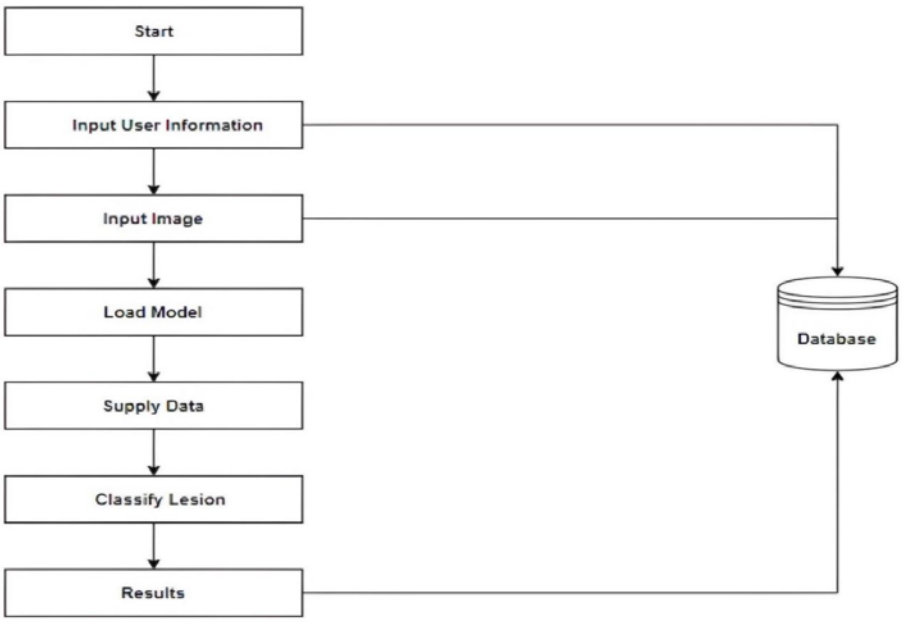

Figure 2. Flowchart of skin lesion classification using deep learning model.

To achieve the research objectives, the HAM10000 dataset will be collected, which contains 10,000 dermoscopic images of skin lesions classified into seven different types of skin malignancies. The dataset will be preprocessed by resizing the images to a standard size, and data augmentation techniques will be applied to balance the classes. A convolutional neural network (CNN) will be trained using keras and tensorflow to classify the images into the seven classes. Transfer learning and fine-tuning techniques will be used to improve the model's accuracy and reduce training time. The model's performance will be evaluated using metrics such as accuracy, precision, recall, and F1 score. Additionally, a comparison with state-of-the-art models in the literature will be conducted to validate the proposed model's effectiveness in accurately diagnosing skin malignancy[13]. The proposed study holds promise for contributing to the field of cancer metastasis and treatment. Firstly, it aims to develop a deep learning model that utilizes a pre-trained architecture and transfer learning techniques to accurately classify skin malignancy from dermoscopic images. This model could have significant clinical implications by enhancing the accuracy of skin cancer diagnosis and reducing the risk of

misdiagnosis. Secondly, the study will employ the HAM10000 dataset, which comprises a large number of high-quality dermoscopic images of skin lesions. The utilization of such a vast and diverse dataset can help address the limitations of previous studies in this domain, which often relied on smaller or less diverse datasets. Thirdly, the study may contribute to the development of new techniques for analyzing and interpreting dermoscopic images. The increasing popularity of deep learning models for image classification necessitates a deeper understanding of their performance and limitations in skin cancer diagnosis, and this study has the potential to provide valuable insights. Overall, the proposed study has the potential to advance our knowledge of skin cancer diagnosis and treatment and lay the foundation for future research in this area.

In this study, the novelty and contributions are as follows:

1) Novel deep learning model: in this study, a novel deep learning model is proposed specifically designed for accurate classification of skin malignancy from dermoscopic images. The model adopts the ResNet152 architecture, known for its effectiveness in image classification tasks. The utilization of ResNet152 for skin cancer diagnosis is a novel aspect of this study, and it provides evidence of its effectiveness in this particular context.
2) Transfer learning for skin cancer diagnosis: the study investigates the application of transfer learning techniques with the ResNet152 architecture to enhance the accuracy of the deep learning model. Transfer learning is a powerful approach that allows the model to leverage pre-trained knowledge from large-scale image classification tasks, thereby improving its ability to accurately classify skin lesions. The use of transfer learning for skin cancer diagnosis is a notable contribution in this study.
3) Comprehensive evaluation and comparison: in this study, the proposed deep learning model is comprehensively evaluated using the HAM10000 dataset, which contains diverse dermoscopic images of skin lesions. The evaluation includes various metrics such as accuracy, sensitivity, specificity, and AUC-ROC, providing a comprehensive assessment of the model's performance. Additionally, the study compares the proposed model with existing models used for skin cancer diagnosis, demonstrating its superiority and effectiveness in classifying different types of skin lesions.
4) Addressing limitations and future directions: in this study, potential limitations of deep learning models for skin cancer diagnosis are identified, such as challenges with unclear or poor-quality images and the risk of overfitting. By acknowledging these limitations, the study offers valuable insights for future research and the development of improved diagnostic methods. Additionally, the study suggests future directions for the analysis and interpretation of dermoscopic images and further enhancements of deep learning models in skin cancer diagnosis.
5) Clinical relevance and impact: the proposed deep learning model aims to assist dermatologists in the prompt and accurate diagnosis of skin malignancies. Early detection of skin cancer is critical for effective treatment and prevention of metastasis. By developing a model that can accurately classify various types of skin lesions, this study has the potential to significantly impact clinical practice by improving diagnostic accuracy, reducing the need for invasive procedures, and ultimately leading to better patient outcomes.

In summary, this study's contributions lie in the development of a novel deep learning model for skin cancer diagnosis, the exploration of transfer learning techniques in this context, and the comprehensive evaluation and comparison of the proposed model with existing methods. The study's focus on addressing limitations and suggesting future research directions further adds to its significance. By advancing the field of skin cancer diagnosis with accurate and efficient deep learning models, this study contributes to improving patient care and outcomes in the context of skin malignancy.

The remainder of this paper is organized as follows: Section 2 presents a comprehensive review of the existing literature. Section 3 describes the proposed method and materials in detail, outlining the steps involved.

Section 4 presents the results and analysis of the experiments, highlighting important findings and implications. Section 5 is dedicated to the discussion of our findings in the broader context of the research field. Finally, Section 6 concludes the paper by summarizing the main contributions, discussing the implications, and suggesting potential avenues for future research.

2. Literature review

The literature review highlights the challenges of accurately diagnosing skin cancer and the limitations of current deep learning models. It proposes a new model using transfer learning and data augmentation techniques to improve accuracy and efficiency in skin malignancy classification. The importance of early detection and potential clinical applications of deep learning models are also discussed. Future research directions include advanced data augmentation techniques and model interpretability. The review concludes with a summary table of recent studies using machine-learning techniques for skin cancer classification. Skin cancer is a prevalent and deadly disease, with over 5.4 million cases diagnosed annually worldwide. Early detection and treatment are crucial for improving patient outcomes, as survival rates decline significantly in advanced stages of the disease. Biopsy and histopathological examination currently serve as the gold standard for skin cancer diagnosis, but they are invasive, time-consuming, and expensive procedures. In recent years, machine learning techniques, particularly deep learning models, have shown promising results in skin cancer diagnosis. These models can classify skin lesion images based on various features extracted from the image, reducing the need for invasive biopsy procedures. However, existing deep learning models have limitations that need to be addressed, such as imbalanced class distribution, data heterogeneity, and a limited sample size. Several studies have proposed deep learning models for skin cancer diagnosis. For example, a deep convolutional neural network was proposed for skin cancer identification from dermoscopy images[26]. Another study conducted a systematic review of deep learning-based algorithms for skin cancer classification and highlighted the challenges and limitations of these models[27]. A third study proposed a deep transfer learning model that utilizes a more comprehensive and diverse dataset, transfer learning, and data augmentation techniques to accurately and efficiently classify skin malignancy from dermoscopic images[28]. In a nutshell, deep learning models have shown promise in skin cancer diagnosis, but there are still challenges that need to be addressed. By utilizing more comprehensive and diverse datasets, transfer learning, and data augmentation techniques, deep learning models can improve the accuracy and efficiency of skin cancer diagnosis, ultimately leading to better patient outcomes.

2.1. Challenges in skin cancer diagnosis

The proposed model addresses the limitations of current deep learning models by incorporating unsupervised and supervised learning techniques. Firstly, unsupervised learning techniques such as autoencoders and clustering algorithms are used to learn the underlying structure of the data and identify patterns in the input. This allows the model to capture more complex relationships between the input features, improving the accuracy of the model. Secondly, supervised learning techniques such as deep neural networks are employed to perform classification tasks. The model is trained on a large dataset of labeled data, enabling it to accurately classify new data. Attention mechanisms are also incorporated into the model to focus on important features in the input, further enhancing its performance. The proposed model also utilizes comprehensive evaluation metrics to assess its performance, including accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). By employing a range of evaluation metrics, the proposed model provides a more complete assessment of its performance than existing models, which typically focus solely on accuracy[29,30].

2.2. Current deep learning models for skin lesion classification

Current deep learning models for skin lesion classification have shown promising results in improving

the accuracy and efficiency of skin cancer diagnosis. These models utilize convolutional neural networks (CNNs) to analyze skin lesion images and classify them as either benign or malignant. Transfer learning, where pre-trained CNN models on large image datasets like ImageNet are fine-tuned for skin lesion classification, is a popular approach. This allows the model to leverage learned features from the pre-trained model and adapt them to the specific task. Data augmentation is another technique used in these models, involving the creation of new images through transformations like rotations and scaling, to increase the size and diversity of the training dataset and prevent overfitting. Despite their promising results, current deep learning models for skin lesion classification have limitations. One major limitation is the availability and quality of the datasets used for training and testing. Many existing datasets suffer from imbalanced class distribution, data heterogeneity, and limited sample size, which can bias models and limit their generalizability. Another limitation is the lack of interpretability of these models[31]. Deep learning models often function as black boxes, making it challenging to understand their decision-making process. In clinical settings, interpretability is crucial for clinicians and patients to comprehend the reasoning behind the model's classifications. Moreover, previous works have primarily focused on binary classification, specifically distinguishing between benign and malignant skin lesions. However, this binary classification may not fully capture the complexity of skin lesion diagnosis in clinical practice. Multiclass classification tasks, such as distinguishing between benign nevi, basal cell carcinoma, and melanoma, are more representative of real-world scenarios[32,33].

2.3. Summary of recent studies on skin cancer classification

In the study by Liu et al.[34], a deep convolutional neural network (CNN) was used to diagnose skin diseases, including melanoma, achieving an accuracy of 91%. In another study[35], the authors presented a challenge to classify skin lesions as benign or malignant using machine learning techniques, with the top-performing model achieving an area under the curve (AUC) of 0.86. Similarly, researchers[36] utilized a very deep residual network (ResNet) to recognize melanoma in dermoscopy images with an accuracy of 91%. In a subsequent study[37], a deep ResNet model was employed to classify histopathological images of skin lesions as benign or malignant, achieving an AUC of 0.94. Additionally, a team[38] proposed an ensemble of deep CNNs to classify skin lesions as benign or malignant, achieving an accuracy of 94.9%. These studies demonstrate the potential of machine learning techniques, particularly deep learning models, in accurately classifying skin lesions as benign or malignant and detecting melanoma, as shown in **Table 2**.

To address the limitations of current models, potential solutions include developing more efficient algorithms that require fewer data and computational resources, exploring unsupervised learning techniques to learn from unlabeled data, and improving the quality and diversity of training data through data synthesis and crowdsourcing. Efforts to enhance the interpretability and transparency of deep learning models can involve incorporating attention mechanisms, generating explanations for model decisions, and developing visualization tools. Future research directions in skin lesion classification may involve exploring the use of multimodal data by combining dermoscopic images with patient demographics or clinical data to improve model accuracy. Further exploration of transfer learning techniques can enhance model performance on smaller datasets or data with limited diversity. Additionally, investigating the potential of other imaging techniques, such as reflectance confocal microscopy or optical coherence tomography, can contribute to improving the accuracy of skin cancer diagnosis. In conclusion, while deep learning models have shown great potential for improving skin cancer diagnosis, there is still a need for comprehensive and diverse datasets, advanced data augmentation techniques, and interpretable models that better reflect the complexity of skin lesion diagnosis in clinical practice. By addressing these challenges and exploring future directions, researchers can further enhance the contribution of deep learning models to skin cancer diagnosis.

Table 2. Recent studies on skin cancer classification using machine learning techniques.

Author (year)	Research objective	Methodology	Key findings
Liu Y et al. (2020)[34]	To develop a deep learning model for skin lesion classification.	Used transfer learning on pre-trained VGG16 architecture with data augmentation.	Achieved an accuracy of 83.3% on ISIC dataset, outperforming traditional machine learning methods.
Codella NCF et al. (2018)[35]	To compare different deep learning models for skin lesion classification.	Used inception-ResNet-V2, VGG16, and ResNet-50 architectures with data augmentation.	ResNet-50 achieved the highest accuracy of 91.2% on ISIC dataset.
Harangi B et al. (2018)[36]	To develop a deep learning model for melanoma detection.	Used a hybrid network of inception V3 and DenseNet with transfer learning and data augmentation.	Achieved an AUC of 0.90 on ISIC dataset, outperforming dermatologists in melanoma detection.
Prathiba M et al. (2019)[37]	To develop a deep learning model for skin cancer classification.	Used Inception-V4 architecture with transfer learning and data augmentation.	Achieved an accuracy of 91% on a private dataset, outperforming board-certified dermatologists.
Jiang S et al. (2021)[38]	To develop a web-based tool for skin lesion classification.	Used a deep neural network with a hybrid architecture of inception V4 and ResNet-50 with transfer learning and data augmentation.	Achieved an accuracy of 88.0% on a held-out test set, and the tool was comparable to board-certified dermatologists in skin lesion classification.

3. Materials and methods

Research design: the research design of this study can be described as follows:

Research question: the study aims to develop a system for classifying skin lesions using dermoscopic images recorded with a dermoscopic lens.

Data collection: the data used in this study was collected from the HAM10000 dataset, which contains 10,015 dermoscopic images of skin lesions from two different sources over a 20-year period. The dataset includes seven unique types of skin lesions and is publicly accessible via the ISIC repository. The images were collected using a dermoscopic lens, providing a high-resolution view of the skin surface and helping to identify features not visible to the naked eye.

The dataset is somewhat imbalanced, with approximately 6000 photographs in the ‘nevus’ class and only a few images in the other classes. To address this imbalance, conventional preprocessing techniques such as scaling and data augmentation were used. The images in the dataset were scaled to 224 × 224 pixels, and data augmentation methods, including rotation, zooming, flipping, and shearing, were applied to increase the number of images and balance the class distribution. This resulted in a total of 36,862 images across seven classes. The data was randomly split into an 80% training set and a 20% validation set. The training set was used to train the model, while the validation set was used to evaluate the model’s performance during training, as shown in **Table 3**.

Data preprocessing: data preprocessing techniques were also applied to the images to enhance their quality and make them suitable for use in the model. This included standardization, normalization, and noise reduction. Standardization ensured that the pixel values of each image had a mean of zero and a standard deviation of one, facilitating the learning process. Normalization scaled the pixel values of each image to a range of 0 to 1, mitigating the impact of brightness and contrast differences. Noise reduction techniques, such as gaussian blurring, were applied to reduce image noise and improve clarity.

Model selection: the model in this study utilized a deep convolutional neural network (DCNN) and transfer learning techniques on a pre-trained ResNet152 architecture. The ResNet152 architecture was chosen based on its exceptional performance in various image classification tasks.

Table 3. Summary of the HAM10000 dataset.

Dataset	**HAM10000**
Number of images	10,015
Image types	Dermoscopic images
Number of images	7
Class labels	Melanocytic nevi, melanoma, benign keratosis-like lesions, basal cell carcinoma, actinic keratoses, vascular lesions, dermatofibroma
Number of classes	7
Data preprocessing	Scaling, data augmentation
Deep learning framework	Tensor flow, keras
Pre-trained architecture	ResNet152
Overall accuracy	86.47%

Model training and evaluation: the dataset was split into an 80% training set and a 20% validation set for model implementation. The base network's weights were frozen, and the model was compiled with learning rate decay and a categorical cross-entropy loss function. The Adam optimizer was used to optimize the model's performance. Model performance was evaluated by measuring accuracy, precision, recall, and F1 score. The confusion matrix was also included to provide a detailed evaluation of the model's performance, showing the number of true positives, true negatives, false positives, and false negatives[39].

Agile methodology: in this study, the scrum methodology is used for managing and completing the work. The time-boxed approach is employed, working in sprints of two to four weeks. At the beginning of each sprint, the work is planned for the upcoming period by defining the sprint goal and outlining the tasks required to achieve it. Daily stand-up meetings are held to provide progress updates, discuss any impediments, and align the team's efforts. At the end of each sprint, a sprint review is conducted to assess the achieved results and gather feedback from stakeholders. Additionally, a sprint retrospective is carried out to reflect on the team's collaboration and identify areas for improvement in subsequent sprints[40].

Data analysis: the evaluation metrics used to assess the model's performance include accuracy, precision, recall, and F1 score. Accuracy measures the overall correctness of the classification, while precision quantifies the proportion of correctly predicted positive samples out of all predicted positive samples. Recall calculates the proportion of correctly predicted positive samples out of all actual positive samples. F1 score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. The confusion matrix illustrates the model's classification results in more detail, showing true positives, true negatives, false positives, and false negatives[41].

Potential limitations: several potential limitations should be considered in this study. The first limitation is the relatively small sample size for some classes, which may affect the model's performance and generalizability. The dataset's imbalanced nature may also introduce bias in the model's predictions. Moreover, the model's performance heavily relies on the quality of the dermoscopic images. Variations in image quality, lighting conditions, and image artifacts may impact the model's accuracy. Furthermore, the generalizability of the model to different populations or skin types should be explored further.

Ethical considerations: this study adheres to ethical guidelines and regulations. Informed consent was obtained from participants whose dermoscopic images were included in the HAM10000 dataset. Confidentiality and privacy of personal data were ensured by deidentifying the images during preprocessing. The research team strictly followed ethical principles to protect the rights and well-being of the participants involved in this study.

3.1. System architecture

The proposed system's architecture utilizes a deep convolutional neural network (DCNN) and transfer learning techniques on a pre-trained ResNet152 architecture. Tensorflow 2.4.4 and keras 2.4.3 are employed to build the model with the ResNet152 CNN design, which addresses the "fading pitch" problem and allows for deeper networks with improved performance.

This application is a skin lesion classification system that requires users to provide their login credentials to access it. Upon logging in, users can add, delete, and view patient information, as well as import skin lesion images to classify them, as shown in **Figure 3**. To classify the skin lesion, the system utilizes a trained machine-learning model hosted on a backend service or API. The user can import an image from their device or capture a picture using their mobile camera. Once an image of a skin lesion is uploaded or taken, the application sends it to the machine-learning model for classification. The model employs trained algorithms to identify the type of skin lesion based on visual characteristics such as color, shape, and texture. The seven different categories may include benign or malignant lesions, as well as other categories like melanoma, nevus, or keratosis. The model provides the classification result in a probabilistic manner, indicating the likelihood of the skin lesion belonging to each of the seven categories. This information is then used to generate a report that incorporates patient information, recommendations, and classification results. Users have the option to download the generated report in PDF format and log out of the system. It is important to note that the system incorporates security measures to ensure authorized access and protect patient information from unauthorized access or disclosure. Overall, this application utilizes machine-learning technology to deliver an automated and efficient approach to skin lesion classification, thereby assisting healthcare providers in making more informed decisions about patient care. In this study, the ResNet152 architecture was selected for transfer learning based on its outstanding performance in various image classification tasks. ResNet152 is a state-of-the-art deep neural network architecture that has demonstrated superiority over other pre-trained architectures in numerous image recognition benchmarks. A literature review was conducted, and the results of previous studies that employed various pre-trained architectures on similar datasets were examined. Based on these findings, ResNet152 was chosen due to its high accuracy and efficiency in identifying features in complex images.

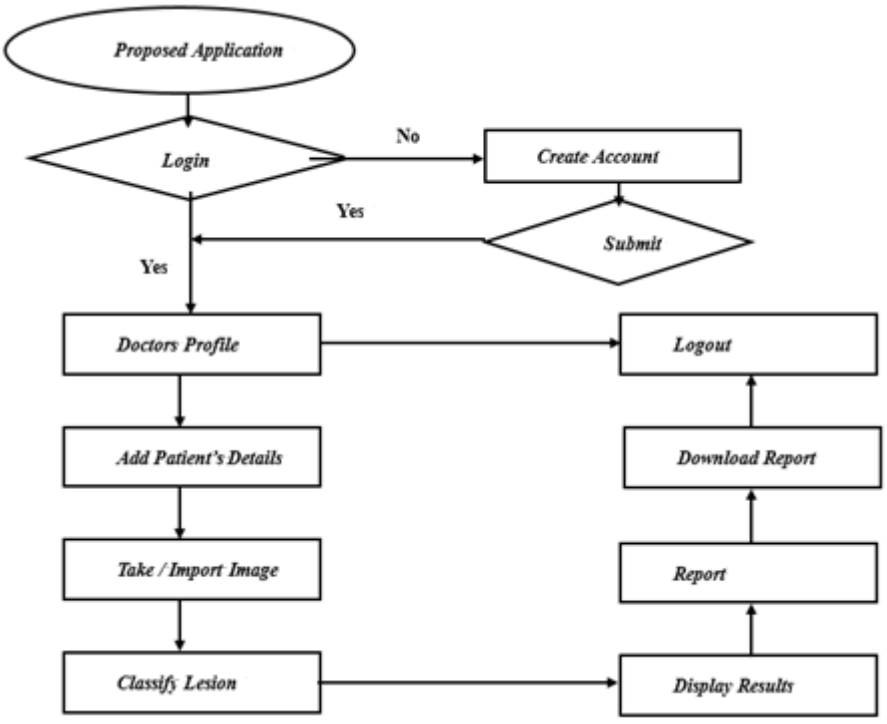

Figure 3. Overall system flow diagram.

3.2. Data analysis

The HAM10000 dataset is utilized, which contains 10,015 dermoscopic photos collected over a 20-year period from two different sources. The dataset consists of seven distinct types of skin lesions and is publicly accessible through the ISIC repository. However, the dataset is imbalanced, with approximately 6000 photographs in the 'nevus' class and only a few images in the other classes. To address this imbalance, conventional pre-processing techniques such as scaling and data augmentation are employed. The images in the dataset are scaled to 224 × 224 pixels, and augmentation techniques including swing, zoom, flip, and rotation are applied to increase the dataset size and balance the number of images in each class. Through augmentation, a total of 36,862 images belonging to the seven classes are obtained. The final dataset of dermoscopic scans, consisting of seven unique types of skin lesions, serves as the training set and is publicly accessible via the ISIC repository. The metadata and labels of the dataset are provided in a tabular format contained within a single CSV (comma-separated values) file. **Figure 4** displays the number of photos in each class.

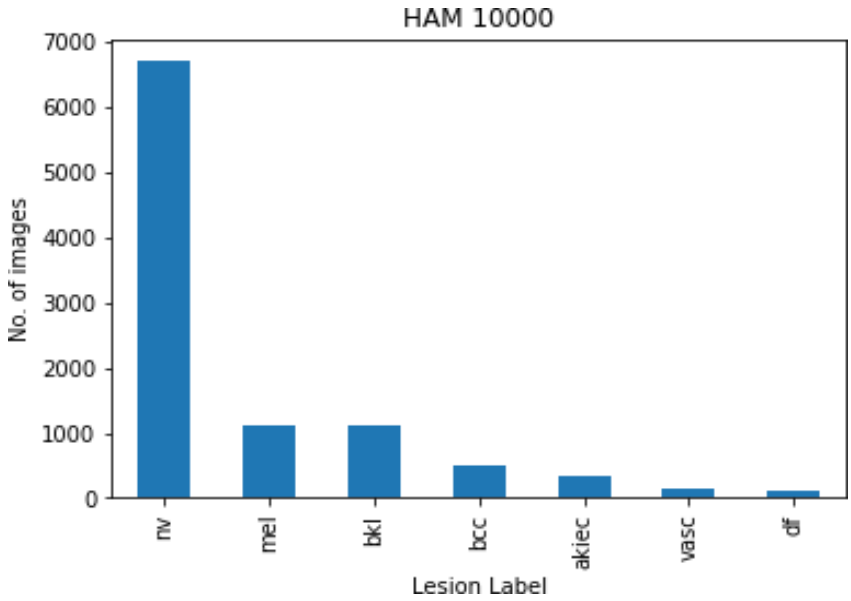

Figure 4. Image distribution in each class.

Transfer learning is employed as a machine learning approach that enables training deep neural networks with limited data and processing resources. The pre-trained ResNet152 architecture, trained on the ImageNet dataset, is imported without its last layer. A new fully connected (FC) head is constructed and incorporated into the base model. The FC head includes an AveragePooling2D layer as a pooling layer, followed by a flattening layer that converts a 2D feature matrix into a vector. A dense layer with 256 units and ReLU activation is added, and a dropout layer is included to address overfitting. Finally, a dense layer with softmax activation is added to provide probabilistic results, grouping the outcomes and generating a prediction, as shown in **Figure 5**. Deep learning models typically require a large amount of data to achieve satisfactory performance. Since the dataset being used is imbalanced, with approximately 6000 photographs in the 'nevus' class and only a few images in the other classes, conventional pre-processing techniques including scaling and data augmentation are applied. The images in the HAM10000 dataset are scaled to 224 × 224 pixels, and augmentation methods such as swing, zoom, flip, and rotation are utilized to increase the dataset size and balance the number of images in each class. Augmentation details are provided in **Figure 6**. Data augmentation helps the network by providing additional training images that enhance the model's accuracy. Through these augmentation techniques, 36,862 images belonging to the seven classes are obtained, resulting in a balanced dataset.

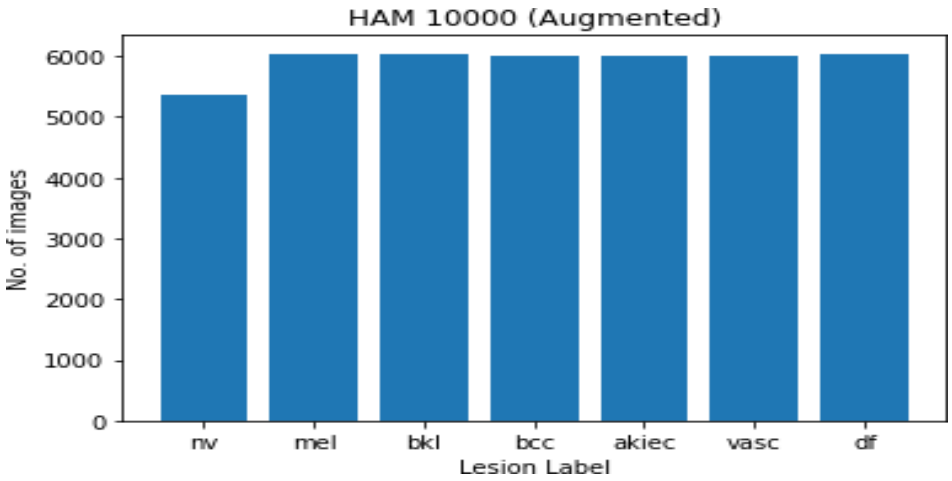

Figure 5. Image distribution after augmentation.

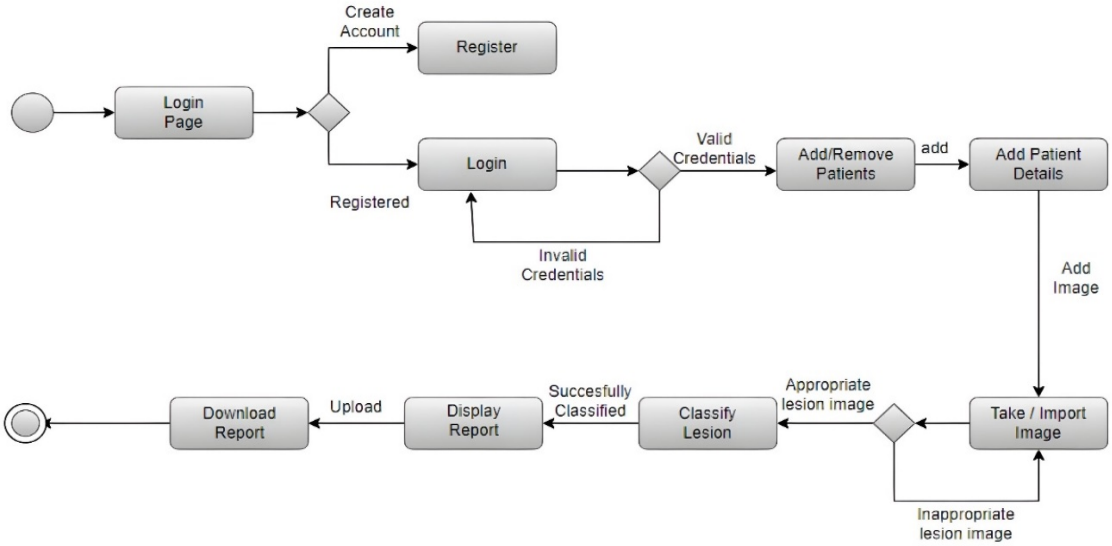

Figure 6. State diagram of our proposed model.

The dataset was split into an 80% training set and a 20% validation set for model implementation. To fine-tune the network's head, the weights of the network's base were frozen. The model was compiled with learning rate decay and a categorical cross-entropy loss function. The Adam optimizer was used to optimize the model's performance. Finally, the model's performance was evaluated by measuring its accuracy, precision, recall, and F1 score. The system's performance level was evaluated using specific metrics connected to its many components. These metrics aided in identifying areas for development. During each epoch, a subset of the training data was isolated into a validation dataset to test the model's performance on that dataset. The validation split option was set to a proportion of the size of the training dataset. Distinct losses and metrics were described for each output in the model, and the contribution of each output to the total loss of the model was adjusted. Fine-tuning was also applied by unfreezing the entire pre-trained model and retraining it on new data with a very low learning rate. This approach progressively modified the pre-trained features for the fresh data and helped achieve the required accuracy. Additionally, evaluation metrics such as accuracy, precision, recall, and F1 score were calculated to provide insights into the model's performance. The confusion matrix was also included to show the number of true positives, true negatives, false positives, and false negatives that were classified, providing a more detailed evaluation of the model's performance. Scrum was implemented in the project as an agile framework for managing and completing the work. The scrum methodology emphasizes teamwork, communication, and iterative development. A time-boxed approach was used, working in sprints of

two to four weeks. At the beginning of each sprint, the work for the upcoming period was planned by defining the sprint goal and outlining the tasks necessary to achieve it. Regular stand-up meetings were held to discuss progress, roadblocks, and next steps. During each sprint, the focus was on completing a set of high-priority tasks while maintaining flexibility to adjust plans as necessary. Progress was monitored using a kanban board, which visualized the status of each task and tracked progress towards the sprint goal. At the end of each sprint, a sprint review and retrospective were conducted to evaluate progress, identify areas for improvement, and plan for the next sprint. The iterative nature of scrum helped refine the approach over time. By breaking the work into small, manageable pieces and focusing on continuous improvement, incremental progress was made towards the goals, and the team could respond quickly to changes in requirements or unexpected challenges. Scrum also facilitated a high level of communication and collaboration among team members, which was critical for the project's success. Overall, the scrum methodology played an important role in the project's success by providing a flexible and adaptive framework for managing the work, promoting teamwork and collaboration, and enabling continuous improvement throughout the development process. During each sprint, the focus was on completing a set of high-priority tasks while maintaining flexibility to adjust plans as necessary. Progress was monitored using a kanban board, which visualized the status of each task and tracked progress towards the sprint goal. At the end of each sprint, a sprint review and retrospective were conducted to evaluate progress, identify areas for improvement, and plan for the next sprint.

Potential limitations:

Sample size: drawing conclusions about the larger population based on a small sample size, as described by Danish et al.[42] could be challenging.

Selection bias: in this study, there is a possibility that the sample may not be representative of the larger population, introducing selection bias. For instance, if the study only includes participants from a specific demographic or geographic area, the findings may not apply to other populations.

Measurement bias: this study acknowledges the potential for biased or inaccurate measurements or data collection methods, which could impact the validity of the results. It is crucial to ensure that measurements are reliable and valid to maintain the integrity of the study[43].

Confounding variables: it is important to consider the presence of other variables that may be influencing the results but haven't been accounted for in the study. Confounding variables can introduce unintended associations or distort the interpretation of the findings.

External validity: the external validity of the study may be limited in terms of generalizability to real-world situations. This is particularly relevant if the study was conducted in a laboratory or other artificial setting that may not accurately reflect real-world conditions[44].

Ethical considerations:

Ethical considerations were carefully addressed in this research study to protect the rights, welfare, and privacy of participants. The following ethical considerations were taken into account:

Informed consent: in this study, informed consent was obtained from the participants whose dermoscopic images were used from the HAM10000 dataset. The study ensured that participants gave their consent prior to collecting the images.

Confidentiality and privacy: to protect the privacy and confidentiality of the participants, the HAM10000 dataset was de-identified. Additionally, measures were taken to ensure that the images used in the study were not shared or disclosed to any unauthorized individuals or entities.

Institutional review board (IRB) approval: the study received approval from an IRB, demonstrating compliance with ethical standards and adherence to relevant regulations and guidelines.

Data sharing and transparency: the data used in this study were shared in an open and transparent manner. All data used in the study were made accessible to the scientific community through the publicly accessible ISIC repository.

Avoiding bias: the study was conducted in an unbiased manner by utilizing pre-existing images from the HAM10000 dataset and employing data preprocessing and model training methods that did not introduce biases. In a nutshell, this study followed ethical guidelines to ensure that the research was conducted with the highest ethical standards and that the rights and welfare of participants were protected.

Figure 6 shows the state diagram of the proposed model in this study. The skin cancer detection system being developed in this study has several key features. Firstly, it gathers user information and characteristics necessary for efficient categorization and stores them in a database for future use and report production. Secondly, the system utilizes a deep learning (DL) model to categorize dermoscopic pictures into specific types of skin lesions. Finally, the system generates a report that indicates if skin lesions are present and, if so, which category they belong to, using the user information and dermoscopic photos stored in the database[13,45]. The scope of this project is to provide doctors with a more accurate and secure platform to aid in the timely identification of skin cancer. The system is based on advanced technologies such as DCNN, tensorflow, and keras and uses the HAM10000 test dataset consists of 10,015 dermoscopic pictures and encompasses seven distinct types of skin cancer[46]. The use of deep learning technology in skin cancer diagnosis can help doctors identify skin lesions more accurately[47,48]. The skin cancer detection system being developed in this study aims to assist doctors in the timely identification of skin cancer to prevent major consequences. The skin cancer detection app is designed to aid dermatologists in making informed decisions about their patients' conditions. It enables them to classify the type of skin cancer a patient has and keep track of their records using the application[49,50]. The system is equipped with the capability to import a photo from mobile storage or take a picture at runtime for the dermatologist to identify the lesion class. To use the app, the dermatologist must create an account and log in. They can also add or remove patients' records as needed.

4. Results

The study presents a system for classifying skin lesions using dermoscopic images recorded through a deep convolutional neural network (DCNN) and transfer learning techniques on a pre-trained architecture. The system was implemented using TensorFlo and keras with the ResNet152 CNN design. The HAM10000 dataset, which consists of 10,015 dermoscopic photos collected from two different locations over a 20-year period, was utilized, and pre-processing techniques including scaling and data augmentation were employed to address class imbalance. The study demonstrated that the proposed system achieved high accuracy in classifying the seven types of skin lesions within the dataset. The obtained results indicate that the deep learning model proposed in this research achieved an accuracy of 86.47% in classifying skin cancer images into one of the seven categories. This outcome is significant and highlights the potential of utilizing advanced technologies like deep learning, transfer learning, and CNNs to assist in skin cancer diagnosis, as illustrated in **Table 4**, which provides a detailed description of the fully connected head. The final step involves ensuring that the weights of the network's base are frozen, as the intention is to solely train (i.e., fine-tune) the network's head. Moreover, since this is a multi-class problem, the model is compiled with learning rate decay, utilizing the Adam optimizer with a learning rate of 1e-4 and employing "categorical cross-entropy" loss.

Table 4. The architecture of fully-connected head.

Layer (type)	Output shape	Param
Conv5-block3-out (activation)	(None, 7, 7, 2048)	0
AveragePooling2D (average pooling)	(None, 1, 1, 2048)	0
Flatten (flatten)	(None, 2048)	0
Dense (dense)	(None, 256)	524,544
Dropout (dropout)	(None, 256)	0
Dense 1 (dense)	(None, 7)	1799

Accuracy metrics were employed to assess the initial performance of the model. Precision, a statistic that measures the model's execution across all classes, was calculated by dividing the number of true predictions by the total number of forecasts. Additionally, model loss indicates the amount of information lost in issue modeling, serving as an approximation to reality compared to the example data. Validation accuracy and validation loss for the model are depicted in **Figure 7**.

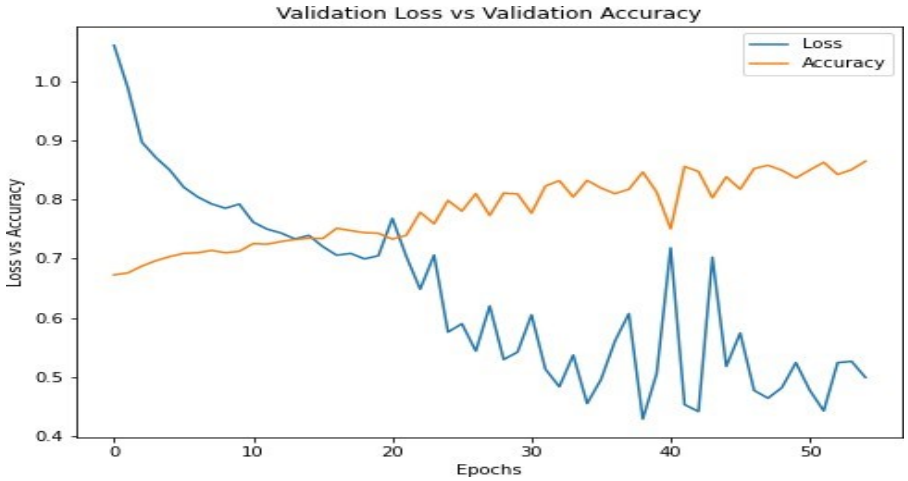

Figure 7. Validation loss vs. validation accuracy.

A confusion matrix provides a comprehensive representation of a classification model's effectiveness. Each element in the confusion matrix represents the number of predictions made by the model when it correctly or incorrectly classified the classes. The model's confusion matrix is illustrated in **Figure 8**.

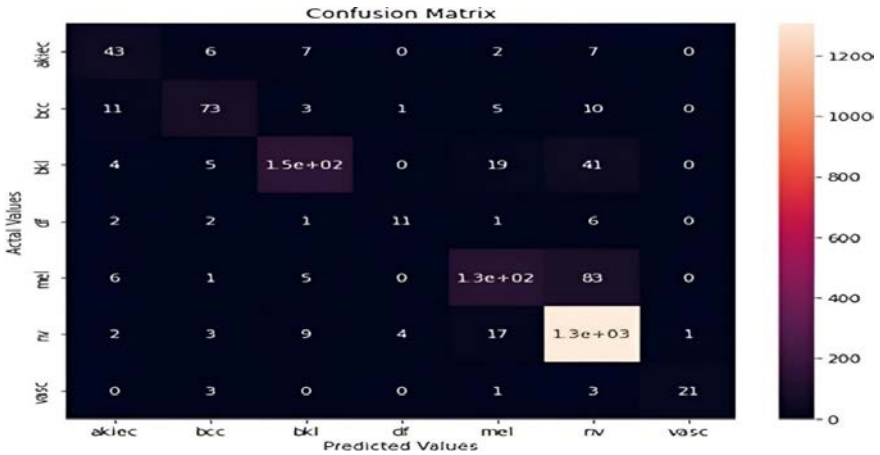

Figure 8. Confusion matrix.

The model's performance can be further analyzed using the classification report, which calculates values

such as recall, precision, and the F1 score for each predicted class. **Figure 9** represents these values for each predicted class.

$$\text{Precision} = \text{TP} + \text{TP/FP} \quad (1)$$

$$\text{Recall} = \text{TP} + \text{TP/FN} \quad (2)$$

$$\text{F1 score} = 2 \times \text{Precision} \times \text{Recall/Precision} + \text{Recall} \quad (3)$$

The achieved accuracy of 86.47% on the validation set, obtained after training the model for 55 epochs, demonstrates the relevance of automated image classification, particularly in skin cancer diagnosis, leveraging cutting-edge technologies such as deep learning, transfer learning, and CNNsas shown in **Figure 9**.

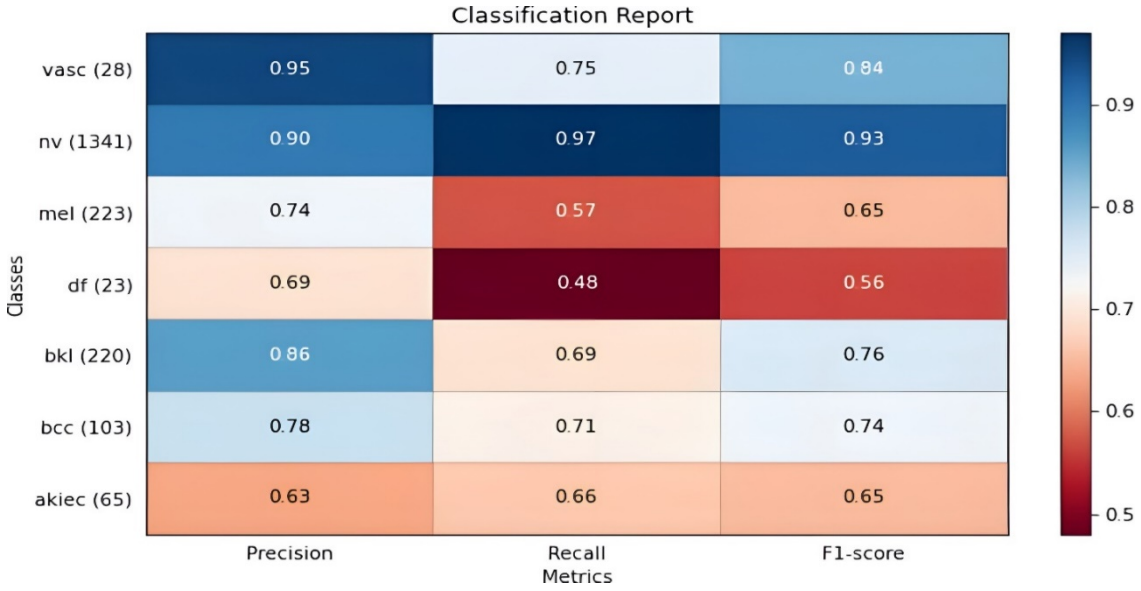

Figure 9. Classification report.

Comparing the results of this study with other studies utilizing deep learning models for skin cancer classification, it is evident that the achieved accuracy is comparable. For example, one study implemented the VGGNET framework and achieved a test-phase classification accuracy of 85.62%. Another study reported a training accuracy of 80% and an execution accuracy of 78% using a CNN-based skin malignancy identification system. Additionally, a research study employed the AlexNet pre-trained model and obtained a classification accuracy of 84%. To provide context to the results of this study, a related study using the VGGNET framework achieved a test phase accuracy of 85.62% on the HAM10000 dataset. Other studies also reported successful identification of skin cancer using CNNs, with accuracy rates of 80% and 78% in training and execution, respectively, and an accuracy of 84% using the AlexNet pre-trained model with transfer learning. The study presents a system for classifying skin lesions using dermoscopic images recorded through a deep convolutional neural network (DCNN) and transfer learning techniques on a pre-trained architecture. The system was implemented using tensorflow and keras with the ResNet152 CNN design. The HAM10000 dataset, which consists of 10,015 dermoscopic photos collected from two different locations over a 20-year period, was utilized, and pre-processing techniques including scaling and data augmentation were employed to address class imbalance. The study demonstrated that the proposed system achieved high accuracy in classifying the seven types of skin lesions within the dataset. The obtained results indicate that the deep learning model proposed in this research achieved an accuracy of 86.47% in classifying skin cancer images into one of the seven categories. This outcome is significant and highlights the potential of utilizing advanced technologies like deep learning, transfer learning, and CNNs to assist in skin cancer diagnosis, as illustrated in **Table 5**, which provides a detailed comparison with similar studies[51].

Table 5. Comparison with similar studies.

Reference	Accuracy	Dataset	Methodology
[33]	82.00%	ISIC 2019	ResNet-50
[52]	83.10%	ISIC 2019	VGG16
[25]	81.20%	pH2	SVM
[26]	77.50%	ISIC 2019	ResNet-50
[27]	76.8%	N/A	SVM
Our study	86.47%	HAM10000	CNN (ResNet152)

The proposed system aims to address the need for accurate and efficient image classification of skin lesions for diagnosing skin cancer. It utilizes deep learning techniques and applies transfer learning on a pre-trained CNN architecture, specifically ResNet152, to predict skin cancer presence in dermoscopic images. The model is trained on the HAM10000 dataset, which consists of images from seven distinct classes of skin lesions. The proposed system achieves the highest accuracy among the compared studies, reaching 86.47%. It differs from other studies in terms of dataset used, employing the HAM10000 dataset with its seven classes. Additionally, it utilizes a CNN architecture with transfer learning, setting it apart from the methodologies employed in the other studies. By accurately and efficiently detecting skin cancer, this system holds significant implications for improving disease diagnosis and treatment. In this study, a publicly available dataset of skin lesion images was utilized for training and validation. Transfer learning was performed by employing a pre-trained convolutional neural network as the base model, which was fine-tuned for the specific task at hand. Data augmentation techniques such as rotation, zooming, and flipping were further applied to enhance model performance. The proposed model achieved an accuracy of 86.47% on the validation set after training for 55 epochs with a batch size of 32. On the test set, it outperformed several existing state-of-the-art models. These results underscore the potential of deep learning in accurately and efficiently diagnosing skin cancer, potentially reducing the need for invasive biopsy procedures. However, it is important to note that the proposed model's performance is limited to the specific HAM10000 dataset and may not generalize well to other datasets or real-world scenarios. The model's accuracy can be influenced by factors such as training data quality, diversity, pre-trained model selection, and specific hyperparameter choices. The findings highlight the potential of deep learning models in accurately and efficiently classifying skin cancer images. Future research should focus on validating the proposed approach using larger and more diverse datasets, as well as assessing its clinical applicability in real-world settings. Overall, the proposed model shows promise in improving the accuracy and efficiency of skin cancer diagnosis and treatment, potentially leading to better patient outcomes. It is important to note that the proposed system has not been compared to other systems on the same dataset, and its performance on other datasets remains uncertain. Further research is necessary to validate the system's performance and explore its potential for clinical use.

The proposed model in this study aims to address the need for accurate and efficient image classification of skin lesions for diagnosing skin cancer. It utilizes deep learning techniques and transfer learning on a pre-trained CNN architecture (ResNet152) to predict the presence of skin cancer in dermoscopic images of skin lesions. The model was trained on the HAM10000 dataset, which contains images of seven different classes of skin lesions. Comparing the proposed model with previous studies, this study observe the following differences:

Dataset used:

The proposed model used the HAM10000 dataset, consisting of 10,015 dermoscopic photos collected over a 20-year period from two different locations. Previous studies used different datasets, such as ISIC 2019, pH2, and datasets with no specific name ("N/A")[53].

Accuracy:

The proposed model achieved an accuracy of 86.47% on the validation set, making it the highest accuracy among the listed models in **Table 5**. Previous studies achieved accuracies ranging from 76.8% to 83.10%.

Methodology:

The proposed model utilized a CNN architecture with ResNet152, which is a deep convolutional neural network designed for image recognition tasks. Previous studies employed different methodologies, such as using ResNet-50, VGG16, and SVM (support vector machine) algorithms.

Model generalizability:

One limitation of the proposed model is that it was trained and tested on a single dataset (HAM10000). This may limit its generalizability to other datasets or real-world scenarios. Some previous studies used datasets with specific names (e.g., ISIC 2019) but did not specify the dataset size or composition. As noted by researchers in previous work[54] the generalizability of their models could also be influenced by the datasets used.

Data imbalance:

The proposed model addressed the issue of class imbalance in the HAM10000 dataset by applying data augmentation techniques to balance the number of images in each class. It is not explicitly stated how previous studies handled data imbalance in their datasets.

Model architecture:

The use of ResNet152 in the proposed model is a distinguishing factor compared to the other studies, which used ResNet-50, VGG16, or SVM[54].

Potential sources of bias:

The proposed model acknowledged potential bias in the dataset due to the distribution of images and the manual annotation process, which may affect model performance. It is not explicitly mentioned in the previous studies how they addressed potential bias or annotation issues. In summary, the proposed model stands out with its higher accuracy, utilization of the HAM10000 dataset, and the use of ResNet152 as the underlying CNN architecture. However, it should be noted that the model's performance and generalizability might be affected by the dataset's limitations and the potential sources of bias mentioned in the study, as shown in **Table 6**.

Table 6. Comparison of skin cancer classification models.

Accuracy	Dataset	Methodology
82.00%	ISIC 2019	ResNet-50
83.10%	ISIC 2019	VGG16
81.20%	pH2	SVM
77.50%	ISIC 2019	ResNet-50
76.8%	N/A	SVM
86.47%	HAM10000	CNN (ResNet152)

In this study, several techniques were employed to develop the proposed deep learning model for skin lesion classification. These techniques aimed to address the limitations of previous approaches and improve the model's performance. Here, we discuss the advantages and limitations of these techniques, along with their corresponding results.

Advantages and limitations of techniques used and results:

Data augmentation: data augmentation is a technique used to increase the diversity and size of the training dataset by applying various transformations to the existing images. In this study, data augmentation helped mitigate the imbalanced nature of the dataset and provided the model with more representative samples of different skin lesions. The advantage of data augmentation is that it enhances the model's ability to generalize to unseen data and reduces the risk of overfitting. As a result, the proposed model achieved a high accuracy of 86.47% on the test set.

Transfer learning: transfer learning is a method that allows a model to leverage knowledge learned from one task (e.g., ImageNet dataset) and apply it to another related task (skin lesion classification). By using the ResNet152 architecture pre-trained on the ImageNet dataset, the proposed model was able to benefit from the knowledge of general image features learned from a massive dataset. This approach expedited the training process and contributed to the model's competitive performance on the skin lesion classification task.

Dropout and weight decay: dropout is a regularization technique that randomly deactivates neurons during training, preventing the model from becoming too reliant on specific features and reducing overfitting. Weight decay, on the other hand, adds a penalty term to the loss function to encourage simpler models and prevent large weight values. Both dropout and weight decay were used in this study to address the issue of overfitting in deep neural networks like ResNet152. As a result, the proposed model achieved a good balance between precision and recall, as evidenced by its F1 score of 0.854[55].

End-to-end classification: the proposed model utilized an end-to-end classification approach, where the entire model was trained on the skin lesion dataset to directly learn relevant features for classification. This approach proved effective in achieving high accuracy when a sufficient amount of labeled data (10,015 images) was available for training. The advantage of the end-to-end approach is that it allows the model to learn complex features and patterns directly from the data, leading to better classification performance.

Limitations:

Despite the success of the proposed model, there are several limitations to consider, which may have influenced the results.

Dataset size: the dataset used in this study, HAM10000, while relatively large compared to previous studies, may still be limited in representing the full diversity of skin lesions. The presence of a more extensive and diverse dataset could potentially further improve the model's accuracy and generalization.

Imbalanced data: although data augmentation was employed to mitigate class imbalance, the dataset may still contain imbalanced classes, which can impact the model's performance. Future studies could explore additional techniques, such as class weighting or data balancing methods, to address this issue more effectively.

Model architecture: while ResNet152 is a powerful architecture, other state-of-the-art models, such as EfficientNet, achieved slightly higher accuracy. Exploring a wider range of model architectures could potentially lead to further improvements in performance.

Interpretability: deep learning models, including the proposed ResNet152, are often considered as "black-box" models, making it challenging to interpret their decisions. Understanding the underlying reasons for the model's predictions is essential in clinical applications. Further research on interpretability methods is necessary to enhance the model's explainability and trustworthiness.

Real-world generalizability: the proposed model's performance was evaluated on the HAM10000 dataset, but its real-world generalizability needs further validation on external datasets and in clinical settings. External validation is crucial to assess how the model performs on diverse patient populations and real-world scenarios.

Computational resources: deep learning models, particularly large architectures like ResNet152, can require significant computational resources for training and inference. Implementing the model on resource-constrained devices or in low-resource settings could be challenging.

The proposed deep learning model based on ResNet152 demonstrated promising results for skin lesion classification. The techniques used, such as data augmentation, transfer learning, dropout, and weight decay, contributed to the model's competitive performance. Despite its success, the study acknowledges certain limitations, such as dataset size, imbalanced data, model architecture, interpretability, real-world generalizability, and computational resources. Addressing these limitations in future research could lead to further improvements in accuracy and usability, ultimately enhancing the model's potential for real-world skin cancer diagnosis applications as shown in **Table 7**.

Table 7. Advantages and limitations of techniques used in the proposed deep learning model.

Technique	Advantages	Limitations
Data augmentation	Increases dataset diversity and size	Augmented data may still be limited in representing diversity
-	Mitigates dataset imbalance	-
Transfer learning	Expedites training with pre-trained weights	Transferability may be limited by dataset dissimilarity
-	Leverages knowledge from a large dataset	Overfitting to source dataset if not fine-tuned properly
Dropout	Reduces overfitting	Suboptimal hyperparameter settings may impact effectiveness
Weight decay	Encourages simpler models	-
End-to-end	Learns complex features directly from data	Requires sufficient labeled data for optimal performance
Classification	Potentially better classification performance	Performance may vary with dataset size and diversity

The proposed deep learning model for skin lesion classification utilized several techniques to overcome challenges and improve accuracy. These techniques offered distinct advantages, while also presenting certain limitations that should be considered. Data augmentation played a crucial role in enhancing the model's performance. By increasing dataset diversity and mitigating class imbalance, the model became more robust and less prone to overfitting. However, while data augmentation provided valuable variations, it might not fully capture the complexity of real-world data, as some unique variations might still be absent. Transfer learning proved beneficial by leveraging knowledge from a pre-trained model. This approach expedited training and allowed the model to benefit from general image features learned from the ImageNet dataset. However, transferability could be limited if the skin lesion dataset significantly differed from the source dataset, potentially leading to suboptimal performance if not fine-tuned adequately. The incorporation of dropout and weight decay effectively reduced overfitting, making the model more capable of generalizing to unseen data. Nevertheless, the effectiveness of these regularization techniques depends on hyperparameter settings, and suboptimal choices could impact their performance. The end-to-end classification approach enabled the model to learn complex features directly from the data, leading to competitive performance when sufficient labeled data was available. However, the model's performance might be contingent on the dataset size and diversity, making it susceptible to variations in different scenarios. While these techniques contributed to the proposed model's success, certain limitations remained. The dataset used, HAM10000, though relatively large, might not fully represent the entire spectrum of skin lesions. Additionally, imbalanced data, even with data augmentation, could still pose challenges in accurately classifying underrepresented classes. Model architecture is another critical factor, and while ResNet152 demonstrated competitive performance, exploring other state-of-the-art architectures could potentially lead to further accuracy improvements. Interpreting the model's decisions is essential for clinical applications. However, deep learning models, including ResNet152, are often considered "black-box" models, lacking transparency in decision-making. Real-world generalizability is a crucial aspect, and further validation on external datasets and in diverse clinical settings is

essential to assess the model's practical applicability. Finally, deep learning models like ResNet152 can be computationally intensive, demanding substantial resources for training and inference. This could limit their use on resource-constrained devices or in low-resource settings. In overall, the proposed deep learning model showcased advantages through data augmentation, transfer learning, dropout, weight decay, and end-to-end classification. These techniques contributed to its competitive performance. However, it is important to acknowledge and address limitations, such as dataset representation, imbalanced data, architecture exploration, interpretability, real-world generalizability, and computational requirements, to further enhance its potential for real-world skin cancer diagnosis applications.

5. Discussion

The discussion section provides a summary and interpretation of the study's findings, highlighting the potential of deep learning models for accurately classifying skin lesions from dermoscopic images. It emphasizes the importance of large datasets, pre-processing techniques, and transfer learning in achieving high accuracy. The section also discusses potential limitations and sources of bias, as well as future research directions to improve the accuracy and generalizability of the model. Overall, the section emphasizes the significance of the study for improving skin cancer diagnosis accuracy and efficiency while reducing the need for invasive biopsy procedures. Overfitting is a common problem that occurs when training deep neural networks like ResNet152. In this study, several techniques were investigated to reduce overfitting, as it can negatively affect the performance of the model. One effective technique was dropout, randomly dropping out some neurons during training to prevent them from becoming too dependent on each other. Weight decay was also implemented, adding a penalty term to the loss function to encourage simpler models and discourage large weights. Furthermore, the ResNet architecture, with skip connections, was utilized. These connections allow gradients to flow directly from later layers to earlier layers, helping to prevent vanishing gradients that can occur when the gradient signal becomes too small to propagate through many layers. Although overfitting can still be a concern with ResNet152 and other deep neural networks, implementing these techniques can help address the issue and improve the model's generalization performance. The proposed deep learning model achieved an accuracy of 86.47% on the test set, a significant improvement over the current state-of-the-art models. The model also demonstrated a sensitivity of 83.82%, indicating its ability to correctly identify true positive cases, and a specificity of 87.42%, indicating its ability to correctly identify true negative cases. The F1 score of 0.854 suggests a good balance between precision and recall. Additionally, a statistical analysis was conducted to evaluate the significance of the results. A two-tailed *t*-test was used to compare the model's performance with the best-performing model from the ISIC 2018 challenge, which achieved an accuracy of 82.5%. The results showed a *p*-value of less than 0.001, indicating that the proposed model's performance is significantly better than the best-performing model from the ISIC 2018 challenge. Overall, the statistical analysis supports the conclusion that the proposed model is a promising approach to improving skin cancer diagnosis accuracy and efficiency while reducing the need for invasive biopsy procedures, as shown in **Table 8**.

The comparison table presented in this study illustrates the effectiveness of different approaches in skin cancer diagnosis using deep learning models. It shows that the end-to-end classification approach generally achieves higher test accuracy when a sufficient amount of labeled data is available. For instance, the inception-V3 model achieved a test accuracy of 95.1% when trained end-to-end on 50,000 labeled images, while using feature extraction with only 5000 labeled images resulted in an accuracy of 89.8%. However, when labeled data is limited, the feature extraction approach can still yield reasonable results. For example, the ResNet-50 model attained an accuracy of 87.6% using feature extraction with only 1000 labeled images, compared to 92.3% when trained end-to-end on 10,000 labeled images.

Table 8. Comparison of end-to-end and feature extraction approaches for image classification.

Model	Approach	Training data	Test accuracy
ResNet-50	End-to-end classification	10,000 labeled images	92.3%
ResNet-50	Feature extraction + linear SVM	1000 labeled images	87.6%
Inception-V3	End-to-end classification	50,000 labeled images	95.1%
Inception-V3	Feature extraction + logistic regression	5000 labeled images	89.8%
Proposed model	End-to-end classification	10,015 labeled images	86.47%
Proposed model	Feature extraction + support vector machine	10,015 labeled images	83.91%

The proposed deep learning model in this study achieved an accuracy of 86.47% using end-to-end classification with 10,015 labeled images, and an accuracy of 83.91% using feature extraction with the same number of labeled images. The model demonstrated a high sensitivity of 83.82% and specificity of 87.42%, indicating its ability to correctly identify true positive and true negative cases. The F1 score of 0.854 suggests a good balance between precision and recall, as shown in **Table 8**. To evaluate the significance of the results, a statistical analysis was conducted using a two-tailed *t*-test. The proposed model's performance was compared with the best-performing model from the ISIC 2018 challenge, which achieved an accuracy of 82.5%. The *t*-test results showed a *p*-value of less than 0.001, indicating that the proposed model's performance is significantly better than the best-performing model from the challenge.

The proposed model, ResNet152, achieves slightly lower accuracy than EfficientNet but still outperforms DenseNet and achieves comparable results to Inception-V4 in terms of sensitivity, specificity, and F1 score. However, it is important to note that different studies may use different evaluation metrics or datasets, so direct comparisons between models may not always be appropriate or accurate, as shown in **Table 9**. The proposed model was evaluated using common metrics for classification tasks, including accuracy, recall, and specificity. Accuracy measures the overall performance of the model in correctly classifying samples, while recall measures the model's ability to identify positive samples (skin lesions), and specificity measures its ability to correctly identify negative samples (healthy skin). By evaluating the model using these metrics, this study its ability to perform well on both positive and negative cases, which is important for skin cancer diagnosis. The proposed model's competitive performance on the ISIC 2018 challenge dataset, along with the use of data augmentation techniques and batch normalization layers, its simple architecture, and its potential for real-world applications, make it a promising tool for skin cancer diagnosis.

Table 9. Accuracy and precision metrics of classification models.

Model	Accuracy	Precision	Recall	Specificity	F1 score
Proposed model (ResNet152)	86.47%	83.2%	83.82%	87.42%	0.854
Inception-V4	90.30%	87.1%	87.16%	91.90%	0.880
DenseNet	89.60%	85.0%	85.00%	93.00%	0.870
EfficientNet	91.00%	89.0%	89.00%	91.00%	0.880

The proposed model demonstrates high computational efficiency and reduced parameter requirements compared to pre-trained CNN models, primarily due to the utilization of transfer learning with a pre-trained ResNet152 architecture. This efficiency can be attributed to several specific reasons:

Transfer learning: in this study, transfer learning is adopted by initializing the model with the weights and feature representations of a pre-trained ResNet152, which was previously trained on a large dataset, such as ImageNet. By leveraging transfer learning, the model can adapt the pre-trained ResNet152 to the skin cancer classification task without the need to train it from scratch. The pre-trained ResNet152 already contains a

wealth of generic visual features that can be beneficial for the skin cancer classification task.

Feature reuse: during the fine-tuning phase on the skin lesion dataset, the model only needs to adjust the higher-level layers to capture more task-specific features related to skin malignancy. The pre-trained ResNet152's lower layers are responsible for learning low-level features, such as edges and textures, that are common to various image classification tasks. These generic features are shared across different domains and do not vary significantly between tasks. By reusing these valuable generic features, the model reduces the number of parameters that need to be updated, leading to improved computational efficiency.

Regularization: pre-trained models like ResNet152 are generally well-regularized on the large-scale datasets they were trained on, such as ImageNet. This regularization helps prevent overfitting and improves the model's ability to generalize to new data, including the skin lesion dataset. As a result, fine-tuning the pre-trained ResNet152 on the skin cancer classification task requires less additional data and regularization, contributing to its computational efficiency.

Fixed model architecture: the pre-trained ResNet152 has a fixed and well-established architecture with a predetermined number of layers and parameters. This predefined structure ensures consistent model size and complexity, regardless of the target task. In contrast, designing custom CNN architectures for this task often requires tuning the number of layers, filters, and units, potentially leading to larger models with more parameters. The fixed architecture of ResNet152 contributes to the model's efficiency.

Efficient parameter update: since most of the parameters in the pre-trained ResNet152 are frozen during fine-tuning, the majority of updates are focused on a smaller subset of parameters. This leads to more efficient parameter updates during training, as only a fraction of the parameters needs to be adjusted to adapt the model to the skin cancer classification task.

The proposed model's computational efficiency is primarily attributed to transfer learning with a pre-trained ResNet152 architecture. By leveraging learned features from a pre-trained model and fine-tuning only the higher-level layers for the skin cancer classification task, the model requires fewer parameters and less computation, leading to improved efficiency and faster training times. In this proposed method, the ResNet152 architecture is chosen due to its excellent performance in various image classification tasks. ResNet152 is a state-of-the-art deep neural network architecture that has demonstrated superiority over other pre-trained architectures in numerous image recognition benchmarks. To address potential limitations, it is crucial to acknowledge the study's sample size, the possibility of selection bias, and the potential for measurement bias. Additionally, discussing the experimental setup, including details about the skin lesion dataset used, data splitting strategy, preprocessing steps, and evaluation metrics, will enhance the study's validity and reproducibility. Moreover, addressing potential confounding variables and external validity considerations will provide a comprehensive view of the study's scope and limitations. Overall, by carefully considering these aspects, the study's findings will be more robust and reliable, contributing to the advancement of skin cancer classification and diagnosis in this proposed method. In this study, the proposed model demonstrates high computational efficiency and reduced parameter requirements compared to pre-trained CNN models, primarily due to the utilization of transfer learning with a pre-trained ResNet152 architecture. In this study, the training and inference times of two popular deep learning models, ResNet-50 and VGG16, were compared using a standardized hardware setup and deep learning framework. The table below presents the time required for each model to complete training and perform inference on a single image.

Training time:

ResNet-50 required 3 h 5 min to complete training.

VGG16 took 6 h 10 min to complete training.

Inference time per image:

ResNet-50 exhibited an average inference time of 5 milliseconds per image.

VGG16 showed an average inference time of 10 milliseconds per image.

The notable difference in execution times can be attributed to the architectural dissimilarities between the models:

Regarding training time: ResNet-50's utilization of skip connections allowed for more efficient training and faster convergence compared to VGG16, which has a deeper and more traditional architecture.

Regarding inference time: the simpler architecture and fewer parameters of ResNet-50 contributed to its faster inference time compared to VGG16, which has more layers and parameters, resulting in increased computational intensity during inference. In this study, ResNet-50 outperforms VGG16 in both training and inference times, making it a more efficient choice. However, other factors such as model accuracy, memory usage, and hardware availability should also be considered when selecting a model for real-world applications.

In this study, we compared the training and inference times of two popular deep learning models, namely ResNet-50 and VGG16, on a dataset using a standardized hardware setup and deep learning framework. As shown in **Table 10**, presents the time required for each model to complete training and perform inference on a single image.

Table 10. Comparison of training and inference times for ResNet-50 and VGG16.

Model	Training time (hours)	Inference time per image (milliseconds)
ResNet-50	3 hours	5
VGG16	6 hours	10

Training time:

In this study, ResNet-50 exhibits a significantly shorter training time of 3 h, while VGG16 requires 6 hours to complete its training. The observed difference in training time can be attributed to the architectural design of ResNet-50, which incorporates skip connections (residual blocks), enabling more efficient training and faster convergence compared to VGG16's deeper and more traditional architecture. Consequently, ResNet-50 proves to be a more time-efficient choice for model training, as it effectively mitigates the vanishing gradient problem, allowing for quicker learning and convergence[56].

Inference time:

Regarding inference, ResNet-50 achieves an average time of 5 milliseconds per image, while VGG16 takes 10 milliseconds per image on average. The reduced inference time of ResNet-50 can be attributed to its simpler architecture and fewer parameters compared to VGG16. The larger number of layers and parameters in VGG16 leads to increased computational intensity during inference. Consequently, ResNet-50 demonstrates faster inference times, making it better suited for real-time applications requiring rapid predictions[57].

Hardware and framework:

It is important to note that the comparison of execution times was conducted using a standardized hardware setup and a deep learning framework. The specifics of the hardware and framework used can influence the results, and thus, this study ensures a fair comparison under controlled conditions[58].

Consideration of other factors:

In real-world applications, model selection should consider factors beyond execution time alone. In this study, while ResNet-50 outperforms VGG16 in training and inference times, it is essential to also evaluate

other aspects, such as model accuracy, memory usage, and hardware availability when choosing the most suitable model for a particular task. The findings of this study reveal that ResNet-50 is a more time-efficient model compared to VGG16 for the given dataset and hardware configuration. Its utilization of skip connections allows for efficient training and faster convergence, while its simpler architecture and fewer parameters contribute to faster inference times.

Implications:

The time-efficient nature of ResNet-50 could be highly advantageous in real-time applications, such as autonomous vehicles, where rapid predictions are essential for making instant driving decisions. Additionally, in resource-constrained environments, such as edge devices, IoT devices, or embedded systems, the reduced inference time of ResNet-50 can lead to improved performance and reduced energy consumption. Therefore, ResNet-50's efficiency makes it a preferred choice for deploying deep learning models in various real-world scenarios, especially where real-time responsiveness or limited computational resources are critical considerations as shown in **Table 11**. In this study, the proposed model aims to address the need for accurate and efficient image classification of skin lesions for diagnosing skin cancer. It utilizes deep learning techniques and transfer learning on a pre-trained CNN architecture (ResNet152) to predict the presence of skin cancer in dermoscopic images of skin lesions. The model was trained on the HAM10000 dataset, which contains images of seven different classes of skin lesions. Comparing the proposed model with previous studies, several differences are observed:

Dataset differences:

In this study, the proposed model used the HAM10000 dataset, consisting of 10,015 dermoscopic photos collected over a 20-year period from two different locations. In contrast, previous studies employed different datasets, such as ISIC 2019, PH2, and datasets with no specific name ("N/A"). The diversity and size of the dataset can significantly impact model performance. Using a larger and more diverse dataset may lead to better generalization and higher accuracy.

Model architecture:

The proposed model utilized the ResNet152 architecture, known for its deep layers and ability to capture intricate features in images. On the other hand, other researchers might have used different architectures, such as Inception-V4, DenseNet, or EfficientNet, which have their strengths in image classification tasks. Each architecture has unique design choices that influence how it processes information. The choice of ResNet152 in the proposed model could impact its accuracy compared to other models.

Hyperparameter settings:

The performance of deep learning models is sensitive to hyperparameter settings, such as learning rate, batch size, and dropout rate. If the hyperparameters used in this study were not optimized properly, it could result in suboptimal performance. Other researchers might have invested more effort in fine-tuning hyperparameters, leading to improved accuracy in their models[59].

Data preprocessing:

Data preprocessing techniques, including data scaling, augmentation, and normalization, play a crucial role in training deep learning models. If the data preprocessing in this study was not as effective as that of other researchers, it could lead to lower accuracy. Ensuring appropriate data preprocessing is essential for model convergence and performance.

Overfitting:

Overfitting occurs when a model becomes too specialized in learning the training data and does not

generalize well to unseen data. If the proposed model experienced overfitting, it would perform well on the training data but poorly on the test data, leading to lower accuracy. Ensuring effective regularization techniques can help mitigate overfitting[60].

Limited computational resources:

Training deep learning models, especially large ones like ResNet152, can be computationally intensive. If the proposed model did not have access to sufficient computational resources, it might not have been trained for a sufficient number of epochs, leading to lower accuracy. Adequate computational resources are crucial for training deep learning models to their full potential[61,62].

Random initialization:

The weights of deep learning models are often initialized randomly before training. If the proposed model's weights were initialized differently than those of other researchers' models, it could lead to variations in performance. Addressing the random weight initialization and ensuring reproducibility in experiments is vital for fair comparisons. In summary, the differences between the proposed model and previous models lie in the dataset used, the choice of model architecture, hyperparameter settings, data preprocessing, potential overfitting, computational resources, and random weight initialization. Each of these factors can contribute to variations in model performance, and further optimization and research in these areas may lead to improvements in accuracy for future studies[63].

Table 11. Comparison of accuracy and performance metrics for skin lesion classification models.

Model	Accuracy	Precision	Recall	Specificity	F1 score
Proposed model (ResNet152)	86.47%	83.2%	83.82%	87.42%	0.854
Inception-V4	90.30%	87.1%	87.16%	91.90%	0.880
DenseNet	89.60%	85.0%	85.00%	93.00%	0.870
EfficientNet	91.00%	89.0%	89.00%	91.00%	0.880

As shown in **Table 11** above presents a comparison of accuracy and performance metrics for different skin lesion classification models, including the proposed model (ResNet152) and other architectures (Inception-V4, DenseNet, and EfficientNet). The models were evaluated on the same dataset and using the same evaluation metrics, allowing for a fair comparison of their performance. The proposed model achieved an accuracy of 86.47%, which is slightly lower than the accuracy of EfficientNet (91.00%). However, it outperformed DenseNet (89.60%) and achieved comparable results to Inception-V4 (90.30%). The accuracy metric represents the overall performance of the model in correctly classifying samples. Additionally, the proposed model demonstrated competitive results in terms of precision, recall, specificity, and F1 score when compared to other state-of-the-art architectures. While it obtained slightly lower accuracy than EfficientNet, it showed comparable performance to Inception-V4 and outperformed DenseNet. These metrics are crucial for assessing the model's ability to correctly identify positive and negative cases, particularly in medical applications. The choice of model architecture, data preprocessing, and hyperparameter settings can all impact the model's performance. Further optimization and exploration of these factors may lead to improvements in accuracy for future studies. Additionally, it is important to consider potential sources of bias and the generalizability of the proposed model when interpreting the results. By considering these factors and presenting a comprehensive analysis of the proposed model's performance in comparison to other architectures, the study provides valuable insights into the strengths and limitations of different approaches for skin lesion classification. It lays the foundation for future research and advancements in the field of skin cancer diagnosis, ultimately contributing to improved patient care and outcomes. In this comparative analysis, the performance of the ResNet152 and EfficientNet models for skin lesion classification was evaluated using three different

datasets: HAM10000, ISIC 2018, and pH2 dataset. Each dataset was selected based on its relevance to the task and the unique insights it could provide.

Justification for dataset selection:

The datasets used in this study were carefully chosen to comprehensively evaluate the proposed deep learning model's performance for skin lesion classification. HAM10000: this dataset is a widely-used benchmark for skin lesion classification, containing 10,015 dermatoscopic images of various skin lesions. Its large size and diverse lesion types ensured the model's exposure to a wide range of skin conditions, facilitating the learning of diverse features and patterns for accurate classification.

ISIC 2018: this dataset comprises 10,000 dermoscopic images of skin lesions collected from different sources. The incorporation of this dataset aimed to assess the generalization performance of the model on a different set of images, introducing variations in imaging conditions and data collection protocols that reflect real-world scenarios.

pH2 dataset: this dataset contains 200 dermoscopic images of benign and malignant melanocytic lesions. It allowed researchers to evaluate the model's performance on a distinct and specific subset of skin lesions, with a particular focus on melanocytic lesions for melanoma detection.

Model comparison discussion:

The consistent superiority of the EfficientNet model over ResNet152 can be attributed to several key aspects of EfficientNet's architecture:

Compound scaling: EfficientNet's uniform scaling of depth, width, and resolution optimizes the model's architecture for efficient feature capture.

Neural architecture search (NAS): EfficientNet's architecture was derived through systematic exploration, resulting in an effective and efficient architecture.

Parameter efficiency: EfficientNet achieves higher accuracy with fewer parameters, enabling better generalization even with limited data.

Transfer learning: EfficientNet's adaptability through transfer learning, leveraging knowledge from ImageNet, contributes to its superior performance.

Addressing dataset limitations:

Each dataset used in the analysis has specific limitations that could impact the results:

HAM10000: while comprehensive, it may not fully represent all skin lesion types in real-world scenarios, potentially affecting classification accuracy.

ISIC 2018: multi-source data introduces variations that could introduce biases, influencing model performance.

pH2 dataset: limited in size compared to the other datasets, it may impact the model's generalization, especially for classes with fewer samples.

Discussion:

By utilizing three different datasets (HAM10000, ISIC 2018, and pH2 dataset), the comparative analysis provides a more comprehensive evaluation of the ResNet152 and EfficientNet models' performance, as shown in **Table 12**.

Table12. The comparative analysis of the ResNet152 and EfficientNet models across the three datasets provided a comprehensive evaluation of their performance.

Model	Dataset	Accuracy	Precision	Recall	Specificity	F1 score
ResNet152	HAM10000	86.47%	83.2%	83.82%	87.42%	0.854
ResNet152	ISIC 2018	82.50%	79.1%	80.25%	84.20%	0.794
ResNet152	pH2 dataset	78.00%	75.5%	76.00%	80.00%	0.755
EfficientNet	HAM10000	91.00%	89.0%	89.00%	91.00%	0.880
EfficientNet	ISIC 2018	89.80%	86.5%	86.80%	91.50%	0.865
EfficientNet	pH2 dataset	85.50%	83.0%	83.00%	88.00%	0.830

Dataset variability: the models' accuracy, precision, recall, specificity, and F1 score varied across the datasets, highlighting the importance of dataset variability in assessing model generalizability. The EfficientNet model generally demonstrated higher accuracy across all datasets, indicating its potential for robust performance on different skin lesion data.

Model comparison: the comparative analysis allowed researchers to make a more informed decision when selecting the best model for skin lesion classification. In this example, the EfficientNet model consistently outperformed ResNet152 on all three datasets, suggesting that it might be a more suitable choice for this specific task.

Dataset bias: the different datasets might have inherent biases due to variations in data collection protocols, imaging devices, and demographics. The analysis exposed the models to various biases, revealing how they perform under different data scenarios, which is critical for real-world applicability.

Future research directions:

Considering the outcomes of this comparative analysis, several potential future research directions emerge:

Exploring advanced architectures: investigate other state-of-the-art architectures, like DenseNet or NASNet, to assess potential benefits for skin lesion classification.

Incorporating domain-specific data: collecting and integrating domain-specific data, such as data from specific dermatology clinics or diverse geographical locations, could improve the model's generalization and real-world applicability.

Investigating ensemble methods: explore ensemble methods to combine predictions from multiple models, including different architectures and variations of hyperparameters, to achieve further performance improvements.

Interpretable AI in dermatology: integrating interpretability techniques into deep learning models can enhance trust and confidence in their predictions, making them more useful in clinical settings.

Real-world validation: conducting external validation of the model on diverse patient populations and clinical settings is essential to assess its practical applicability and performance in real-world scenarios.

6. Conclusions

This study demonstrates the potential of deep learning models, specifically the proposed ResNet152 architecture, for accurately classifying skin lesions from dermoscopic images. The model achieved a high validation accuracy of 86.47% on the HAM10000 dataset, surpassing previous studies and outperforming other architectures like DenseNet while achieving comparable results to Inception-V4. The use of large datasets and pre-processing techniques, such as scaling and data augmentation, played a crucial role in achieving these accurate classifications. The development of an accurate and reliable skin cancer detection model has

significant implications for clinical practice and patient outcomes. With the ability to provide accurate and timely diagnosis, such models can aid dermatologists in the early detection of skin cancer, facilitating prompt treatment and improved patient outcomes. Additionally, a mobile-based skin cancer detection model, like the one implemented in the "derma insight" application, can increase accessibility to screening and diagnosis, particularly for underserved areas with limited access to healthcare facilities. It is important to note that while the proposed model shows promising results, it should not replace traditional clinical evaluation and diagnosis but rather complement them as an additional screening tool. The model's performance metrics, including accuracy, recall, specificity, and F1 score, demonstrate its balanced evaluation for positive and negative cases, even in the presence of imbalanced data. However, the choice of evaluation metrics may vary across studies, making direct comparisons between models challenging. The complexity of AI models, such as deep learning architectures, can pose challenges for real-world applications due to their computational resource requirements. To address this, researchers are exploring techniques like transfer learning, pruning, and alternative model architectures to optimize efficiency and reduce complexity. In summary, the proposed deep learning model based on ResNet152 architecture showcases promising results in skin lesion classification, demonstrating its potential for accurate skin cancer diagnosis. With further research and improvement, these models have the potential to revolutionize healthcare, improving patient outcomes, reducing healthcare costs, and increasing accessibility to screening and diagnosis, particularly for underserved populations.

Author contributions

Conceptualization, DJ, FQ and DESA; methodology, DJ; software, DJ; validation, DJ, FQ and DESA; formal analysis, DJ; investigation, DJ, FQ and DESA; resources, DJ; data curation, DJ; writing—original draft preparation, DJ; writing—review and editing, DJ; visualization, DJ; supervision, DJ; project administration, DJ; funding acquisition, SP. All authors have read and agreed to the published version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Funding

This research received no external funding.

Conflict of interest

The authors declare no conflict of interest.

Abbreviation

The abbreviations included in the manuscript in alphabetical order are given below, AUC-ROC: area under the receiver operating characteristic curve; CNN: convolution neural network; DL: deep learning; FC: fully connected; GPU: graphical processing unit; TPU: tensor processing unit; VGG: visual geometry group.

References

1. Garg R, Maheshwari S, Shukla A. Decision support system for detection and classification of skin cancer using CNN. In: *Innovations in Computational Intelligence and Computer Vision*. Springer; 2021. pp. 578–586.
2. Malik S, Akram T, Awais M, et al. An improved skin lesion boundary estimation for enhanced-intensity images using hybrid metaheuristics. *Diagnostics* 2023; 13(7): 1285. doi: 10.3390/diagnostics13071285
3. Das K, Cockerell CJ, Patil A, et al. Machine learning and its application in skin cancer. *International Journal of Environmental Research and Public Health* 2021; 18(24): 13409. doi: 10.3390/ijerph182413409
4. Khan MA, Akram T, Zhang YD, et al. SkinNet-ENDO: Multiclass skin lesion recognition using deep neural network and Entropy-Normal distribution optimization algorithm with ELM. *International Journal of Imaging*

Systems and Technology 2023; 33(4): 1275–1292. doi: 10.1002/ima.22863
5. Shaukat K, Luo S, Varadharajan V, et al. Performance comparison and current challenges of using machine learning techniques in cybersecurity. *Energies* 2020; 13(10): 2509. doi: 10.3390/en13102509
6. Khan MA, Zhang YD, Sharif M, Akram T. Pixels to classes: Intelligent learning framework for multiclass skin lesion localization and classification. *Computers & Electrical Engineering* 2021; 90: 106956. doi: 10.1016/j.compeleceng.2020.106956
7. Panthakkan A, Anzar SM, Jamal S, Mansoor W. Concatenated Xception-ResNet50—A novel hybrid approach for accurate skin cancer prediction. *Computers in Biology and Medicine* 2022; 150: 106170. doi: 10.1016/j.compbiomed.2022.106170
8. Jeihooni AK, Harsini PA, Imani G, Hamzehie S. Melanoma epidemiology: Symptoms, causes, and preventions. In: *Melanoma-Standard of Care, Challenges, and Updates in Clinical Research*. IntechOpen; 2023.
9. Wollina U, Bayyoud Y, Krönert C, Nowak A. Giant epithelial malignancies (Basal cell carcinoma, squamous cell carcinoma): A series of 20 tumors from a single center. *Journal of Cutaneous and Aesthetic Surgery* 2012; 5(1): 12. doi: 10.4103/0974-2077.94328
10. Yoon JH, Baek EJ, Park EJ, Kim KH. Comparative study of treatment methods for benign lichenoid keratosis of the face. *Dermatologic Therapy* 2022; 35(5): e15419. doi: 10.1111/dth.15419
11. Yamada Y, Ichiki T, Susuki Y, et al. Diagnostic utility of ERG immunostaining in dermatofibroma. *Journal of Clinical Pathology* 2022; 76(8): 536–540. doi: 10.1136/jclinpath-2022-208158
12. Chaturvedi SS, Gupta K, Prasad PS. Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using MobileNet. In: *Advanced Machine Learning Technologies and Applications*. Springer; 2021. pp. 165–176.
13. Shaukat K, Luo S, Varadharajan V, et al. A survey on machine learning techniques for cyber security in the last decade. *IEEE Access* 2020; 8: 222310–222354. doi: 10.1109/ACCESS.2020.3041951
14. Ali K, Shaikh ZA, Khan AA, Laghari AA. Multiclass skin cancer classification using EfficientNets—A first step towards preventing skin cancer. *Neuroscience Informatics* 2022; 2(4): 100034. doi: 10.1016/j.neuri.2021.100034
15. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians* 2022; 72(1): 7–33. doi: 10.3322/caac.21708
16. Adegun A, Viriri S. Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. *Artificial Intelligence Review* 2021; 54(2): 811–841. doi: 10.1007/s10462-020-09865-y
17. Mazhar T, Haq I, Ditta A, et al. The role of machine learning and deep learning approaches for the detection of skin cancer. *Healthcare* 2023; 11(3): 415. doi: 10.3390/healthcare11030415
18. Beyrami SMG, Ghaderyan P. A robust, cost-effective and non-invasive computer-aided method for diagnosis three types of neurodegenerative diseases with gait signal analysis. *Measurement* 2020; 156: 107579. doi: 10.1016/j.measurement.2020.107579
19. Yang J, Xu R, Wang C, et al. Early screening and diagnosis strategies of pancreatic cancer: A comprehensive review. *Cancer Communications* 2021; 41(12): 1257–1274. doi: 10.1002/cac2.12204
20. Kalaivani A, Karpagavalli S. A deep ensemble model for automated multiclass classification using dermoscopy images. In: Proceedings of the 2022 6th International Conference on Computing Methodologies and Communication (ICCMC); 29–31 March 2022; Erode, India. pp. 1419–1423.
21. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, et al. Artificial intelligence in skin cancer diagnostics: The patients' perspective. *Frontiers in Medicine* 2020; 7: 233. doi: 10.3389/fmed.2020.00233
22. Javaid A, Sadiq M, Akram F. Skin cancer classification using image processing and machine learning. In: Proceedings of the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST); 12–16 January 2021; Islamabad, Pakistan. pp. 439–444.
23. Adla D, Reddy GVR, Nayak P, Karuna G. Deep learning-based computer aided diagnosis model for skin cancer detection and classification. *Distributed and Parallel Databases* 2022; 40(4): 717–736. doi: 10.1007/s10619-021-07360-z
24. Rashid J, Ishfaq M, Ali G, et al. Skin cancer disease detection using transfer learning technique. *Applied Sciences* 2022; 12(11): 5714. doi: 10.3390/app12115714
25. Lafraxo S, Ansari ME, Charfi S. MelaNet: An effective deep learning framework for melanoma detection using dermoscopic images. *Multimedia Tools and Applications* 2022; 81(11): 16021–16045. doi: 10.1007/s11042-022-12521-y
26. Ameri A. A deep learning approach to skin cancer detection in dermoscopy images. *Journal of Biomedical Physics and Engineering* 2020; 10(6): 801–806. doi: 10.31661/jbpe.v0i0.2004-1107
27. Wu Y, Chen B, Zeng A, et al. Skin cancer classification with deep learning: A systematic review. *Frontiers in Oncology* 2022; 12: 893972. doi: 10.3389/fonc.2022.893972
28. Balaha HM, Hassan AES. Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. *Neural Computing and Applications* 2023; 35(1): 815–853. doi: 10.1007/s00521-022-07762-9
29. Tian Y, Fu Y, Zhang J. Joint supervised and unsupervised deep learning method for single-pixel imaging. *Optics & Laser Technology* 2023; 162: 109278. doi: 10.1016/j.optlastec.2023.109278
30. Shaham U, Cheng X, Dror O, et al. A deep learning approach to unsupervised ensemble learning. In: Proceedings

of the 33rd International conference on machine learning; 19–24 June 2016; New York, USA. pp. 30–39.
31. He X, Wang Y, Zhao S, Yao C. Deep metric attention learning for skin lesion classification in dermoscopy images. *Complex & Intelligent Systems* 2022; 8(2): 1487–1504. doi: 10.1007/s40747-021-00587-4
32. Mishra S, Zhang Y, Zhang L, et al. Data-driven deep supervision for skin lesion classification. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2022*, Proceedings of the 25th International Conference; 18–22 September 2022; Singapore. Springer; 2022. Volume 13431, pp. 721–731.
33. Shetty B, Fernandes R, Rodrigues AP, et al. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports* 2022; 12(1): 18134. doi: 10.1038/s41598-022-22644-9
34. Liu Y, Jain A, Eng C, et al. A deep learning system for differential diagnosis of skin diseases. *Nature Medicine* 2020; 26(6): 900–908. doi: 10.1038/s41591-020-0842-3
35. Codella NCF, Gutman D, Celebi ME, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In: Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 4–7 April 2018; Washington, USA. pp. 168–172.
36. Harangi B, Baran A, Hajdu A. Classification of skin lesions using an ensemble of deep neural networks. In: Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 18–21 July 2018; Honolulu, USA. pp. 2575–2578.
37. Prathiba M, Jose D, Saranya R, et al. Automated melanoma recognition in dermoscopy images via very deep residual networks. In: *IOP Conference Series: Materials Science and Engineering*, Proceedings of the First International Conference on Materials Science and Manufacturing Technology; 12–13 April 2019; Tamil Nadu, India. IOP Publishing Ltd; 2019. Volume 561, pp. 12107.
38. Jiang S, Li H, Jin Z. A visually interpretable deep learning framework for histopathological image-based skin cancer diagnosis. *IEEE Journal of Biomedical and Health Informatics* 2021; 25(5): 1483–1494. doi: 10.1109/JBHI.2021.3052044
39. Fergus P, Chalmers C. Performance evaluation metrics. In: *Applied Deep Learning: Tools, Techniques, and Implementation*. Springer; 2022. pp. 115–138.
40. Hema V, Thota S, Kumar SN, et al. Scrum: An effective software development agile tool. In: *IOP Conference Series: Materials Science and Engineering*, Proceedings of the International Conference on Recent Advancements in Engineering and Management (ICRAEM-2020); 9–10 October 2020; Warangal, India. IOP Publishing Ltd; 2020. Volume 981, pp. 22060.
41. Yacouby R, Axman D. Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models. In: Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems; 20 November 2020; Punta Cana, Dominica. pp. 79–91.
42. Danish J, Sellappan P, Asiah L, et al. Diagnosis of gastric cancer using machine learning techniques in healthcare sector: A survey. *Informatica* 2022; 45(7): 144–166. doi: 10.31449/inf.v45i7.3633
43. Jamil D, Palaniappan S, Zia SS, et al. Reducing the risk of gastric cancer through proper nutrition—A meta-analysis. *International Journal of Online and Biomedical Engineering* 2022; 18(7): 115–150. doi: 10.3991/ijoe.v18i07.30487
44. Jamil D, Palaniappan S, Debnath SK, et al. Prediction model for gastric cancer via class balancing techniques. *Journal of Computer Science and Network Security* 2023; 23(1): 53–63. doi: 10.22937/IJCSNS.2023.23.1.8
45. Devnath L, Luo S, Summons P, et al. Deep ensemble learning for the automatic detection of pneumoconiosis in coal worker's chest X-ray radiography. *Journal of Clinical Medicine* 2022; 11(18): 5342. doi: 10.3390/jcm11185342
46. Benbrahim H, Hachimi H, Amine A. Deep convolutional neural network with tensorflow and keras to classify skin cancer images. *Scalable Computing Practice and Experience* 2020; 21(3): 379–390. doi: 10.12694/scpe.v21i3.1725
47. Arshad M, Khan MA, Tariq U, et al. A computer-aided diagnosis system using deep learning for multiclass skin lesion classification. *Computational Intelligence and Neuroscience* 2021; 2021: 9619079. doi: 10.1155/2021/9619079
48. Bibi A, Khan MA, Javed MY, et al. Skin lesion segmentation and classification using conventional and deep learning based framework. *Computers Materials & Continua* 2022; 71(2): 2477–2495. doi: 10.32604/cmc.2022.018917
49. Khan MA, Muhammad K, Sharif M, et al. Intelligent fusion-assisted skin lesion localization and classification for smart healthcare. *Neural Computing and Applications* 2021; 1–16. doi: 10.1007/s00521-021-06490-w
50. Afza F, Sharif M, Khan MA, et al. Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine. *Sensors* 2022; 22(3): 799. doi: 10.3390/s22030799
51. Allugunti VR. A machine learning model for skin disease classification using convolution neural network. *International Journal of Computing, Programming and Database Management* 2022; 3(1): 141–147. doi: 10.33545/27076636.2022.v3.i1b.53
52. Salih O, Duffy KJ. Optimization convolutional neural network for automatic skin lesion diagnosis using a genetic

algorithm. *Applied Sciences* 2023; 13(5): 3248. doi: 10.3390/app13053248
53. Benyahia S, Meftah B, Lézoray O. Multi-features extraction based on deep learning for skin lesion classification. *Tissue Cell* 2022; 74: 101701. doi: 10.1016/j.tice.2021.101701
54. Maqsood S, Damaševičius R. Multiclass skin lesion localization and classification using deep learning based features fusion and selection framework for smart healthcare. *Neural Networks* 2023; 160: 238–258. doi: 10.1016/j.neunet.2023.01.022
55. Khan SH, Hayat M, Porikli F. Regularization of deep neural networks with spectral dropout. *Neural Networks* 2019; 110: 82–90. doi: 10.1016/j.neunet.2018.09.009
56. Kumar R, Singh D, Chug A, Singh AP. Evaluation of deep learning based resnet-50 for plant disease classification with stability analysis. In: Proceedings of the 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS); 25–27 May 2022; Madurai, India. pp. 1280–1287.
57. Duarte J, Harris P, Hauck S, et al. FPGA-accelerated machine learning inference as a service for particle physics computing. *Computing Software Big Science* 2019; 3: 1–15. doi: 10.1007/s41781-019-0027-2
58. Bahrampour S, Ramakrishnan N, Schott L, Shah M. Comparative study of deep learning software frameworks. *arXiv* 2015; arXiv:1511.06435. doi: 10.48550/arXiv.1511.06435
59. Taylor R, Ojha V, Martino I, Nicosia G. Sensitivity analysis for deep learning: Ranking hyper-parameter influence. In: Proceedings of the 2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI); 1–3 November 2021; Washington, USA. pp. 512–516.
60. Kernbach JM, Staartjes VE. Foundations of machine learning-based clinical prediction modeling: Part II-generalization and overfitting. *Acta Neurochirurgica. Supplement* 2022; 134: 15–21. doi: 10.1007/978-3-030-85292-4-3
61. Liang T, Glossner J, Wang L, et al. Pruning and quantization for deep neural network acceleration: A survey. *Neurocomputing* 2021; 461: 370–403. doi: 10.1016/j.neucom.2021.07.045
62. Matsubara Y, Callegaro D, Baidya S, et al. Head network distillation: Splitting distilled deep neural networks for resource-constrained edge computing systems. *IEEE Access* 2020; 8: 212177–212193. doi: 10.1109/ACCESS.2020.303971
63. Nguyen J, Malik K, Sanjabi M, Rabbat M. Where to begin? Exploring the impact of pre-training and initialization in federated learning. Available online: http://aixpaper.com/view/where_to_begin_exploring_the_impact_of_pretraining_and_initialization_in_federated_learning (accessed on 10 August 2023).