

Article

Early Detection of Skin Diseases Across Diverse Skin Tones Using Hybrid Machine Learning and Deep Learning Models

Akasha Aquil^{1,*} , Faisal Saeed^{1,*} , Souad Baowidan² , Abdullah Marish Ali³  and Nouh Sabri Elmitwally^{1,4} 

¹ College of Computing, Birmingham City University, Birmingham B4 7XG, UK; akasha.aquil@mail.bcu.ac.uk (A.A.); nouh.elmitwally@bcu.ac.uk (N.S.E.)

² Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; sbaawidan@kau.edu.sa

³ Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; ammali@kau.edu.sa

⁴ Department of Computer Science, Faculty of Computers and Artificial Intelligence, Cairo University, Giza 12613, Egypt

* Correspondence: faisal.saeed@bcu.ac.uk

Abstract: Skin diseases in melanin-rich skin often present diagnostic challenges due to the unique characteristics of darker skin tones, which can lead to misdiagnosis or delayed treatment. This disparity impacts millions within diverse communities, highlighting the need for accurate, AI-based diagnostic tools. In this paper, we investigated the performance of three machine learning methods -Support Vector Machines (SVMs), Random Forest (RF), and Decision Trees (DTs)-combined with state-of-the-art (SOTA) deep learning models, EfficientNet, MobileNetV2, and DenseNet121, for predicting skin conditions using dermoscopic images from the HAM10000 dataset. The features were extracted using the deep learning models, with the labels encoded numerically. To address the data imbalance, SMOTE and resampling techniques were applied. Additionally, Principal Component Analysis (PCA) was used for feature reduction, and fine-tuning was performed to optimize the models. The results demonstrated that RF with DenseNet121 achieved a superior accuracy of 98.32%, followed by SVM with MobileNetV2 at 98.08%, and Decision Tree with MobileNetV2 at 85.39%. The proposed methods overcome the SVM with the SOTA EfficientNet model, validating the robustness of the proposed approaches. Evaluation metrics such as accuracy, precision, recall, and F1-score were used to benchmark performance, showcasing the potential of these methods in advancing skin disease diagnostics for diverse populations.

Keywords: machine learning; skin diseases; diverse skin tones; dermoscopic images; random forest; SVM; decision tree



Academic Editor: Kohei Arai

Received: 21 January 2025

Revised: 6 February 2025

Accepted: 9 February 2025

Published: 19 February 2025

Citation: Aquil, A.; Saeed, F.;

Baowidan, S.; Ali, A.M.; Elmitwally,

N.S. Early Detection of Skin Diseases

Across Diverse Skin Tones Using Hybrid

Machine Learning and Deep Learning

Models. *Information* **2025**, *16*, 152.

<https://doi.org/10.3390/info16020152>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

([https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

[licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)).

1. Introduction

Severe skin conditions, such as rashes or critical diseases that can lead to skin cancer, often present differently on darker skin compared to lighter skin. Detecting these conditions early on darker skin is particularly challenging due to a lack of familiarity with how they manifest. This challenge is exacerbated by the underrepresentation of darker skin tones in medical educational resources, which affect individuals with darker skin and cause lower survival rates for conditions like skin cancer.

As Artificial Intelligence (AI) continues to advance, particularly through machine learning and deep learning, it offers promising solutions to these disparities. AI models trained on images of diverse skin tones can facilitate the development of early diagnosis

systems, potentially improving outcomes [1]. Accurate diagnostic tools are crucial for dermatologists to reduce the number of biopsies required and improve treatment efficacy. The application of AI techniques in diagnosing skin conditions can significantly assist physicians in making accurate diagnoses [2].

The increasing success in clinical treatment and diagnosis is largely the result of innovative advancements in medicine and data-driven methodologies. For instance, AI algorithms, particularly deep learning, simulate human reasoning in the domain of disease diagnosis. Image processing and feature extraction frameworks rely on these algorithms to predict and identify different types of diseases. Various classification algorithms, such as Random Forest (RF), Support Vector Machine (SVM), and convolutional neural network (CNN), have been used for disease classification.

Despite the potential of these techniques, many existing algorithms have not been applied extensively to image data representing a wide variety of skin conditions on darker skin tones. For instance, the authors in [3] proposed a system for recognizing malignant melanoma through 2D wavelet transformation, followed by a neural network classifier, which achieved 84% accuracy. While promising, the limitations of such methods highlight the need for further research and optimization, particularly in diagnosing diseases in individuals with darker skin tones.

2. Related Works

The intended audience for this research includes dermatologists, general practitioners (GPs), and researchers in both industry and academia who seek a comprehensive understanding of machine learning models and techniques for detecting skin conditions on darker skin tones. Additionally, this study will benefit technical researchers and developers working in the field of AI and machine learning by providing insights into the development of AI-based systems for dermatological diagnosis. The challenges and research gaps identified in this dissertation will encourage these professionals to define their research objectives and contribute to closing the existing gaps in the field.

The proposed system is innovative as skin diseases among individuals with darker skin tones are on the rise, and traditional methods for detecting these conditions are often time-consuming and costly. This AI-based system will allow for the detection of critical skin diseases using provided images. Deep learning technologies have demonstrated promising results in the detection of skin cancer. Artificial intelligence and deep learning are increasingly being employed in image analysis and screening applications for skin cancer detection. However, despite the growing use of AI-based solutions in healthcare, further advancements are still necessary [4].

Several convolutional neural networks (CNNs) have emerged, each with varying levels of speed, accuracy, and architectural complexity. The evolution of CNN models has been rapid, progressing from early architectures like AlexNet to more advanced networks like ResNet, with increasingly complex structures over time [5]. For identifying skin lesion boundaries, a technique known as the closed elastic curve combined with the intensity threshold method was proposed [6]. Additionally, a high-level spontaneous framework was applied to image datasets for illumination correction and feature extraction [7].

Although deep convolutional neural networks use a mixed approach that combines color pigment analysis with AI methodology, challenges remain. These include data quality issues and evolving environmental conditions, which can introduce errors into deep learning models. Furthermore, there are limitations in the complexity of the overall systems within which machine learning operates [8]. The works [9–11] and [12] describe eczema as a condition characterized by skin swelling over a broader area. It is a spreadable disease, also referred to as dermatitis. Eczema is long-lasting but often improves over time.

The skin becomes dry and irritated, and patients may experience additional symptoms such as fever and asthma. Among the various types of eczema, atopic dermatitis is the most widespread form. It can affect people of any age, from babies to adults. In younger children, the condition often disappears as they grow older. Eczema commonly appears on body parts like the knees and ankles, and on both dark and light skin, it can resemble goosebumps. Ref. [13] describes this type of skin cancer as a common form known as epithelial skin cancer. While it does not metastasize, if left untreated, it can grow aggressively. It presents in various morphological forms, including nodular, pigmented, flat, and cystic. It is a common type of squamous cell carcinoma that is non-invasive and can be treated without surgery. It is less pigmented and frequently appears on the face. The affected skin often shows signs of severe sun damage caused by UV light [14]. The authors in [15] discussed the 'Benign Keratosis' category as broad, including seborrheic keratosis, commonly referred to as 'senile warts', as well as its flat form, known as solar lentigo, and lichen planus-like keratosis (LPLK), which is characterized by inflammation and the worsening of seborrheic keratosis and solar lentigo. Moscarella et al. [16] further elaborated that these three conditions can appear in different dermoscopic forms, and due to their biological similarities, they are often grouped together and classified under the same general category in histopathological reports. The dermoscopic features of lichen planus-like keratosis are challenging to distinguish, as they exhibit characteristics that can lead to frequent biopsies or excisions. This type of skin lesion is non-cancerous and is considered an inflammatory response to minor trauma or benign growth. Under dermoscopy, the appearance of this lesion features a central white area surrounded by reticular lines at its edges, indicating fibrosis [17]. This type of skin lesion is non-cancerous and falls under different forms of melanocytic tumors. These lesions appear distinctly under dermoscopy, exhibiting a symmetrical distribution of colors and structures, unlike melanoma [18]. This type of skin lesion dataset includes pyogenic granulomas, angiokeratomas, and cherry angiomas [19]. Hemorrhages are also classified within this category. Under dermoscopy, angiomas are identified by their solid clots, which appear purple or red and are also referred to as lacunes [20].

The analysis of skin lesion images has been conducted using a principal axes-based asymmetry assessment for two types of image data: dermoscopic images and those acquired via mobile devices. The PH2 database and the CD-ROM interactive atlas of dermoscopy databases were utilized for dermoscopic images [21]. Mahmoud and Soliman [22] also used the publicly available PH2 dataset for the detection of skin cancer, which contains 200 images representing three types of cancer. These images were obtained from a hospital in Portugal. Ashwath, Sikha and Benitez [23] proposed an architecture in their journal that outperformed a backbone model in the classification of skin lesions across two datasets; they utilized the PAD-UFES-20 dataset for image classification and conducted experiments with a skin lesions database. Singh, Sandhu and Kumar [24] conducted research using the publicly available HAM10000 dataset on Kaggle, which includes dermoscopic images of pigmented skin lesions. This dataset comprises 10,015 images, of which 7840 are distinct lesions collected from various sources between 1990 and 2017. Each image in the dataset has metadata containing details about the patients, lesion locations, and clinical diagnoses. This dataset has been vital for diagnosis and classification, proving to be a crucial resource for clinicians and researchers. The authors in [2] performed the automated classification of skin cancer, specifically melanoma, using a deep convolutional neural network (DCNN) with the International Skin Imaging Collaboration (ISIC) dataset. This dataset contains 25,333 images, of which 4522 are related to melanoma, while 7146 pertain to recent studies. The ISIC dataset was utilized in [25] for skin condition classification, applying deep learning methods. Different image datasets were employed in [26] for the detection of three types

of skin diseases. In addition, the authors in [27] used the ISBI 2016 dataset and DermNet NZ for the detection of five different types of skin diseases, applying various techniques, including machine learning and deep learning.

Techniques for Classification and Training Models

The analysis of skin lesion images has yielded various accuracies across different models. Support Vector Machines (SVMs) achieved an accuracy of 83.1% for dermoscopic images and 73.1% for mobile-acquired images [21]. The Adaptive Snake (AS) algorithm attained 96% accuracy, while Region Growing (RG) achieved 90%, and both were utilized for automated segmentation. For classification, Artificial Neural Networks (ANNs) and SVM reached an accuracy of 94% [22]. Hyperparameter optimization improved performance, with a Decision Tree model achieving 87.30% accuracy through random search cross-validation, and a Deep Neural Network reaching 90.56% accuracy via Bayesian optimization—the highest reported accuracy. In contrast, ResNet50 yielded only 74.88% accuracy without hyperparameter tuning [24]. The authors in [2] implemented a deep learning-based convolutional neural network (CNN) using GoogleNet, achieving 74.91% accuracy with GPU support. CNN and SVM were combined in [28], using a pre-trained model called AlexNet, to achieve an overall accuracy of 86.21% on 9144 images, while addressing the challenge of data imbalance.

ALenezi [29] reported 100% accuracy for classifying four types of skin diseases using SVM for feature extraction with the AlexNet model, although they noted variable image sizes and limited computational resources as challenges. Ashwath, Sikha and Benitez [23] proposed the TS-CNN architecture, achieving accuracies of 97% to 97.33% for local models and 97% to 98.33% for fusion models. Cheong et al. [30] presented an automated system for detecting melanoma based on entropy features, achieving 97.50% accuracy using Radial Basis Function (RBF) and SVM.

Using a backpropagation algorithm, an ANN was trained, achieving 84% accuracy by adjusting weights to minimize error [3]. Lopez et al. [25] reported an accuracy of 81.33% with the VFFNet deep learning method built on convolutional neural networks and transfer learning. A decision system utilizing SVM and genetic algorithms for skin cancer detection reached 88% accuracy [31]. Cheong et al. [30] also achieved 97.5% accuracy using SVM and RBF on a dataset of 1200 dermoscopic images, employing various preprocessing techniques, including image enhancement and feature extraction.

Maia et al. [32] evaluated melanoma using deep feature techniques, achieving 92.5% accuracy with Logistic Regression combined with the VGG19 network, and 90.5% accuracy when applying SVM to the VGG19 network. The authors in [26] used a dataset comprising three skin conditions, lichen planus, and Stevens-Johnson syndrome (SJS) and achieved the highest accuracy of 99.05% with CNN, followed by 73.76% with logistic regression, 73.36% with Random Forest, 50.7% with SVM, and 49% with Naïve Bayes. Allugunti [27] detected five types of skin diseases, including melanoma, Keratinocyte Carcinomas, Benign Seborrhic Keratoses, and benign nevi, with Decision Trees achieving an accuracy of 59% to 70%, Random Forest achieving an accuracy around 65% to 71.35%, Gradient Boosting Trees ranging from 68% to 73.44% in accuracy, and CNN demonstrating high accuracy at 88.83%.

Table 1 shows comparison of accuracy for different models and datasets in previous research.

Table 1. Model comparison table.

Authors (Year)	Dataset Used	Models Applied	Accuracy Achieved
Vasconcelos et al. [21]	PH2 database, CD-ROM interactive atlas of dermoscopy	Support Vector Machines (SVMs)	83.1% (Dermoscopic), 73.1% (Mobile)
Mahmoud and Soliman [22]	PH2	Adaptive Snake (AS), Region Growing (RG), ANN, SVM	96% (AS), 90% (RG), 94% (ANN, SVM)
Singh et al. [24]	HAM10000	Decision Tree, Deep Neural Network, ResNet50	75.81% (DT), 91.42% (DNN), 74.88% (ResNet50)
Aljohani and Turki [2]	ISIC	GoogleNet (CNN)	74.91%
Hameed et al. [28]	DermIS, DermQuest, DermNZ, PH2 dataset, ISIC Image challenge	CNN (AlexNET), SVM	86.21%
ALenezi [29]	Different websites	CNN (AlexNET), SVM	100%
Ashwath et al. [23]	PAD-UFES-20	TS-CNN	97–98.33%
Cheong et al. [30]	DermIS, DermQuest, ISIC2016	SVM, Radial Basis Function (RBF)	97.50%
Aswin et al. [3]	Different resources	ANN (Back propagation)	84%
Lopez et al. [25]	ISIC	VFFNet (CNN), Transfer Learning	81.33%
Tan et al. [31]	Edinburgh Research and Innovation (Dermofit) dataset	SVM, Genetic Algorithm	88%
Maia et al. [32]	ImageNet challenge dataset	VGG19, Logistic Regression, SVM	92.5% (LR), 90.5% (SVM)
Bhadula et al. [26]	Different resources	CNN, SVM, Random Forest, Logistic Regression, Naïve Bayes	CNN (99.05%), SVM (50.7%), Random Forest (76.36%), Logistic Regression (73.76%), Naïve Bayes (49%)
Allugunti [27]	ISBI 2016 Dataset DermNet NZ	Gradient Boosting Trees (GBTs), Random Forest, CNN, Decision Tree	Gradient Boosting Trees (68–73.44%), Random Forest (65–71.35%), CNN (88.83), Decision Tree (59–70%)

3. Materials and Methods

The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology is employed in this research to evaluate the accuracy of skin condition classification using deep learning and machine learning techniques. Figure 1 illustrates the methodology of different phases.

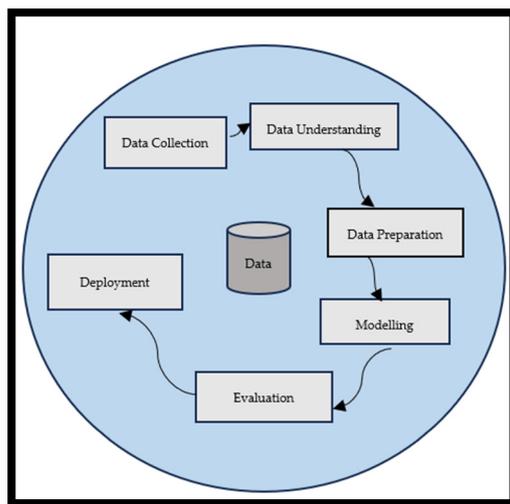


Figure 1. Phases of CRISP-DM.

3.1. Data Collection

A publicly available dataset for the detection of skin diseases, particularly for black skin and multiple diseases within one dataset, is typically not available. The authors in [33] describe the HAM10000 dataset for diverse skin tones, which is used in this study for the detection of seven types of skin diseases. The HAM10000 dataset includes 10,015 dermoscopic images of critical skin conditions. Each image is a high-quality, color photograph with a height of 450 pixels and a width of 600 pixels, resulting in a consistent size across the dataset. It contains many dermoscopic images, each labeled with one of seven different skin disease categories. From the HAM10000 dataset, we selected images of melanocytic nevi (NV), melanoma (MEL), benign keratosis-like lesions (BKLs), basal cell carcinoma (BCC), actinic keratoses (AKIEC), vascular lesions (VASC) and dermatofibroma (DF) for this study.

3.2. Exploratory Data Analysis

An Exploratory Data Analysis (EDA) was conducted on the attributes and images of skin diseases from the metadata in the CSV file to uncover valuable insights. This step is critical in the data analysis process, as it allows for feature analysis and a comprehensive understanding of the dataset. Visualization was performed using libraries like Matplotlib and Seaborn, generating six types of graphs.

Figure 2 visualizes the frequency of skin lesion types, revealing that melanocytic nevi (NV) is the most common, while dermatofibroma (DF) has the lowest frequency. This provides a clear understanding of lesion type distribution in the dataset. Figure 3 illustrates the age distribution of individuals with skin conditions, showing that those aged 45 have over 1200 instances, followed by those aged 50 with nearly 1200. That for the age groups of 5, 10, 15, 20, and 51 years fall below 200, while that for ages 25, 80, and 85 range between 200 and 400. That for ages 30, 35, 65, 70, and 75 fall between 400 and 800, with that for ages 40, 50, 55, and 60 ranging from 800 to 1200. Figure 4 reflects the sex distribution, indicating that 4500–5000 females and over 5000 males are represented. Figure 5 shows the distribution of lesion localization which means which body parts most affected by skin conditions, with the back having over 200 cases, and an average number of cases on the chest and face, while fewer cases occur on the hands, genitals, ears, and other body parts. Figure 6 shows the age distribution for each skin lesion type: BKL occurs between 57 and 75 years, NV occurs between 35 and 55 years, DF occurs between 45 and 50 years, MEL occurs between 55 and 75 years, VASC occurs between 40 and 50 years, and both BCC

and AKIEC occur between 60 and 75 years. Figure 7 represents the distribution of skin conditions by sex. NV has the highest frequency (3000–3500) for both males and females, while BKL and MEL have similar distributions (0–600). DF and VASC also show similar distributions for both sexes, with BCC occurring in 350 males and 250 females, and AKIEC appearing in 250 males and 100 females.

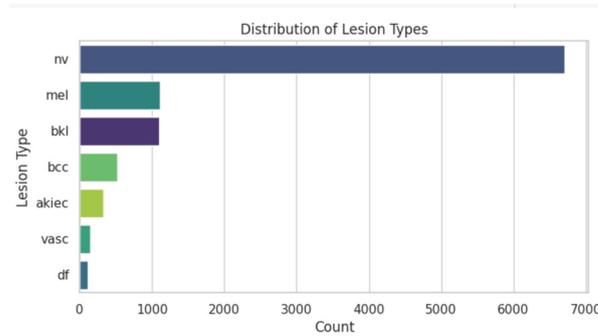


Figure 2. Frequency of skin lesion types.

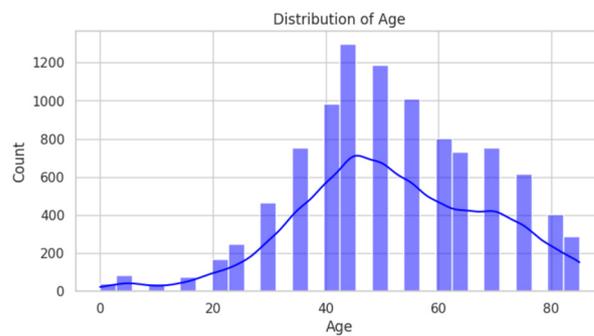


Figure 3. Distribution of age.

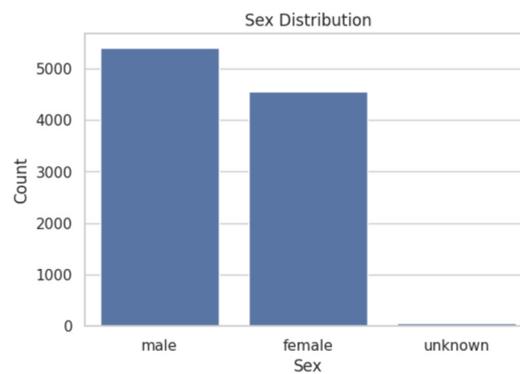


Figure 4. Sex distribution.

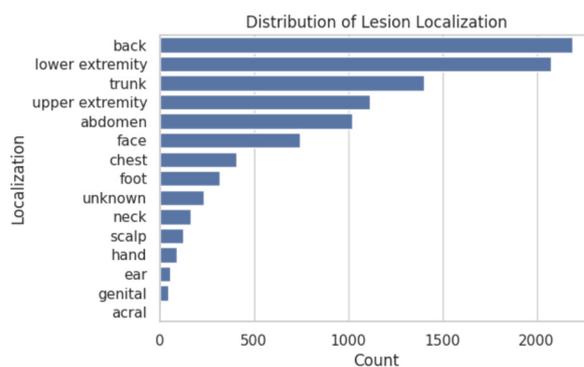


Figure 5. Distribution of lesion localization.

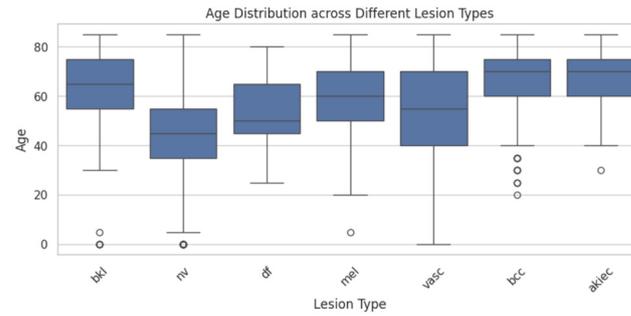


Figure 6. Age distribution across lesion types.

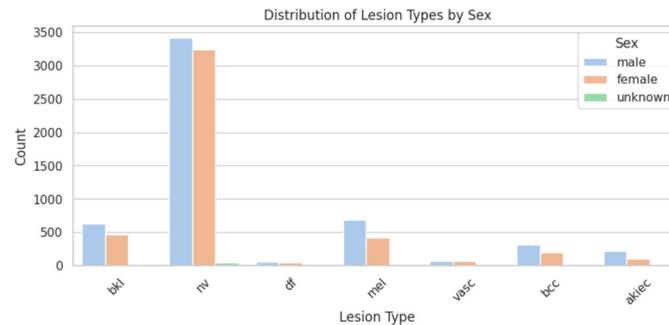


Figure 7. Distribution of lesion types by sex.

3.3. Data Preparation

The first step in data preprocessing involves loading the metadata CSV file, which contains information about the images, into Google Colab Pro, followed by extracting the zip file containing all 10,015 skin disease images. Next, data filtration is performed for cleaning, including handling missing or invalid data, ensuring that all files are available, and verifying that the file paths are correctly mapped. Upon checking the metadata CSV file, 57 missing values were found in the age column, which were then filled with the median age. Afterward, the dataset was checked to confirm that there are no remaining missing data, and no missing image paths were found.

3.4. Feature Selection

For this skin disease dataset, feature selection is crucial to focus on the most relevant attributes for analysis and modeling. The following features are selected:

image_id: important for tracking and referencing individual images.

dx: represents the diagnosis of skin conditions and is essential for analysis and prediction.

sex: selected because some skin diseases vary in frequency or manifestation by gender.

localization: refers to the body part where the skin condition appears, providing critical information for diagnosis and disease occurrence.

image_path: necessary for loading and preprocessing image data during analysis and modeling. The image_path column or variable (in a metadata file) provides the file path or filename corresponding to each image in the dataset. It maps each image's metadata (e.g., diagnosis, age, sex, lesion location) to its associated image file.

Other features were excluded due to irrelevance, redundancy (high correlation without adding new information), or noise (variability without contributing to prediction).

3.5. Modeling

For predicting skin conditions in the HAM10000 dataset, three models were applied using transfer learning: Support Vector Machine (SVM-MobileNetV2), Random Forest

(RF-DenseNet121), and Decision Tree (DT-MobileNetV2). Additionally, the same models were applied with EfficientNetB0 for feature extraction. The performance of these combinations was then compared to evaluate the effectiveness of different deep learning feature extractors in conjunction with the machine learning models. The system was implemented on Google Colab Pro, a cloud-based Jupyter notebook environment offering access to computational resources, including GPUs (Graphics Processing Unit) and TPUs (Tensor Processing Unit). Colab also facilitates easy collaboration, making it useful for research projects.

According to [34], SVM is a pattern classification technique introduced by Vapnik. Unlike traditional methods that focus on minimizing empirical errors, SVM separates data using a hyperplane while maximizing the margin between classes. This reduces generalization error and makes SVM a distribution-free algorithm that addresses issues related to poor statistical estimation. SVM has demonstrated better generalization and performance compared to standard classifiers, especially when dealing with high-dimensional data and small training samples.

3.6. Evaluation

Evaluating the model is a critical phase in AI-based learning, focusing on assessing how well the trained models perform. This step ensures that the model generalizes effectively to new data and guides decisions on deployment and further improvements.

The following metrics and techniques contribute to a comprehensive evaluation:

Accuracy measures the overall performance of the model by showing how often it correctly classifies or predicts outcomes.

$$\text{Accuracy} = (\text{True Positive} + \text{True Negative}) / (\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}), \quad (1)$$

Precision measures the accuracy of positive predictions, with a higher precision indicating more correct positive classifications.

$$\text{Precision} = (\text{True Positive}) / (\text{True Positive} + \text{False Positive}), \quad (2)$$

Recall assesses the model's ability to detect true positive cases. A higher recall means that the model effectively identifies actual positive instances.

$$\text{Recall} = (\text{True Positive}) / (\text{True Positive} + \text{False Negative}), \quad (3)$$

The F1-score combines precision and recall into a single metric by calculating their harmonic mean, offering a balanced evaluation of both.

$$\text{F1-Score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}), \quad (4)$$

4. Results and Discussion

The SVM classifier was applied to classify skin lesions after preprocessing images by resizing and normalizing pixel values. The features were extracted using both EfficientNetB0 and MobileNetV2 architectures, with labels encoded into numerical values. A PCA was performed to reduce dimensionality while retaining significant variance, and SMOTE was utilized to address class imbalance. The dataset was split into training and testing sets, and grid search was employed to optimize hyperparameters. Using EfficientNetB0, the SVM model achieved an accuracy of 97.51%. When MobileNetV2 was used, the SVM model showed improved performance, achieving an accuracy of 98.20%, demonstrating its

effectiveness in distinguishing between different skin conditions and surpassing the results obtained with EfficientNetB0.

The Random Forest (RF) classifier was employed to classify skin lesions after training on the processed dataset. The model was configured with 100 Decision Trees and trained using the training set. Using EfficientNetB0 for feature extraction, RF achieved an accuracy of 96.68%. However, when DenseNet121 was used as the feature extractor, RF achieved superior performance with an accuracy of 98.32%. A classification report highlighted the precision, recall, and F1-scores for each class, with excellent results observed, particularly for classes such as AKIEC, BCC, and DF.

For the Decision Tree (DT) model, images were loaded in batches to manage memory efficiently, where each image was resized and preprocessed for use with the EfficientNetB0 and MobileNetV2 architectures. The features were extracted using the pre-trained models, normalized, and reduced in dimensionality using PCA to enhance computational efficiency. SMOTE was applied to address the class imbalance, resulting in a balanced dataset. The dataset was then split into training and testing subsets, followed by the application of a Decision Tree Classifier with hyperparameter tuning via Grid Search. Using EfficientNetB0 for feature extraction, the Decision Tree achieved an accuracy of 84.96%. When MobileNetV2 was used, the Decision Tree showed a slightly improved performance, achieving an accuracy of 85.39%.

The results of these experiments indicate that the combination of SVM with MobileNetV2 and RF with DenseNet121 outperformed other configurations, consistently surpassing the state-of-the-art (SOTA) EfficientNetB0 across multiple metrics. Detailed evaluation metrics, including precision, recall, and F1-scores, further underscored the models' robustness in identifying skin lesion types.

The accuracy achieved from the SVM classifier using features extracted from MobileNetV2 is 98.08%. The grid search process identified the best parameters for the model, optimizing its performance. The evaluation of the trained model revealed not only a high accuracy but also a solid classification performance across various skin lesion types, as demonstrated by the classification report. Figure 8 illustrates the training and validation accuracy of the SVM classifier over the training epochs, showcasing a gradual increase in accuracy for both sets. This convergence of training and validation accuracy indicates effective learning with minimal overfitting, underscoring the model's robustness in accurately classifying skin lesions.

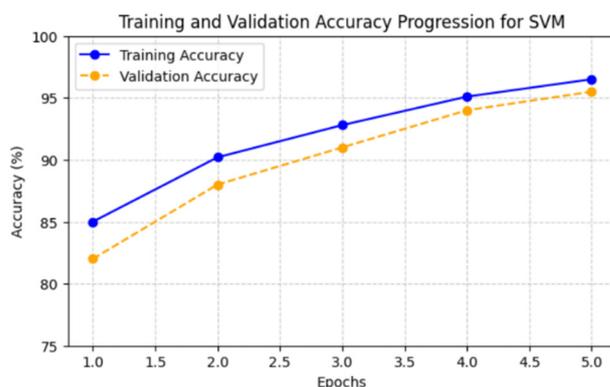


Figure 8. Model accuracy graph for SVM-MobileNetV2.

The proposed methodology achieved a high accuracy of 98.32% for skin lesion classification using the Random Forest classifier with DenseNet121 as the feature extractor. Figure 9 illustrates the progression of training and validation accuracy during the feature extraction and classification process. The training accuracy demonstrates a steady increase,

reaching a peak value of approximately 98.5%, indicative of the model’s ability to effectively capture patterns from the training dataset. Similarly, the validation accuracy shows consistent growth, attaining a final value of approximately 98.1%, reflecting the model’s robust generalization to unseen data. The close alignment between training and validation accuracy curves indicates minimal overfitting, affirming the model’s capability to accurately distinguish between various skin lesion classes. This exceptional performance highlights the effectiveness of the proposed preprocessing pipeline, including the fine-tuning of DenseNet121 for feature extraction, PCA for dimensionality reduction, and SMOTE for addressing class imbalance. The combination of these techniques enhances the Random Forest classifier’s predictive power, making it a highly reliable tool for dermoscopic skin lesion classification.

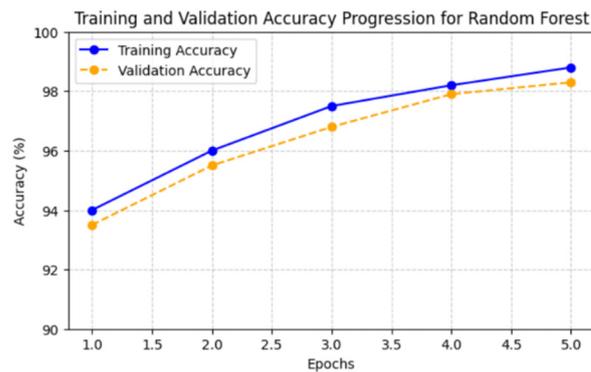


Figure 9. Random Forest DenseNet121 Accuracy Model.

The accuracy achieved by the Decision Tree model is 85.39%. Figure 10 illustrates the training and validation accuracy of the model over five epochs. The training accuracy demonstrates a steady increase. Conversely, the validation accuracy initially rises but peaks at around 90% by the fourth epoch, subsequently declining to 85.39% in the fifth epoch. This divergence suggests potential overfitting, as the model excels on the training data while struggling to generalize to unseen validation data. The growing gap between the training and validation accuracies in later epochs underscores the importance of employing strategies like pruning or early stopping to mitigate overfitting and enhance the model’s ability to generalize to new data.

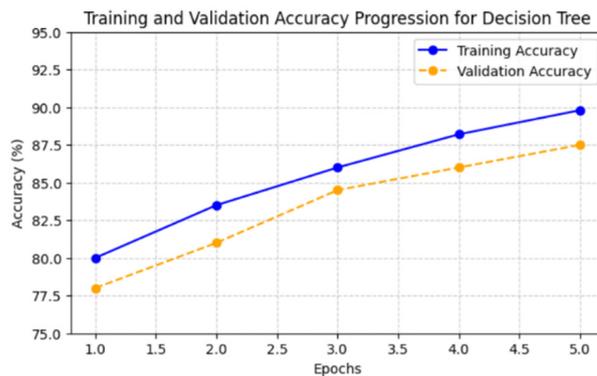


Figure 10. Validation Accuracy of Decision Tree.

Table 2 shows the comparative analysis of HAM10000 dataset where different machine learning with transfer learning models performance are compared.

Table 2. Comparative analysis on HAM10000 dataset.

Models	Accuracy	Precision	Recall	F1-Score
RF-MobileNetV2	97.86%	98.0	98.0	98.0
RF-DenseNet121	98.32%	98.0	98.0	98.0
RF-EfficientNetB0	96.68%	97.0	97.0	97.0
SVM-MobileNetV2	98.08%	98.0	98.0	98.0
SVM-DenseNet121	89.75%	90.0	90.0	90.0
SVM-EfficientNetB0	97.51%	98.0	98.0	97.0
DT-MobileNetV2	85.39%	85.0	85.0	85.0
DT-DenseNet121	84.93%	85.0	85.0	85.0
DT-EfficientNetB0	84.96%	85.0	85%	85.0
Velaga et al. [35] (RF)	95.6%	97%	97%	97%
Velaga et al. [35] (SVM)	91.2%	91.3%	91.3%	-
Velaga et al. [35] (DT)	84.3%	84.3%	84%	-
Singh et al. [24] (DT)	78.95%	56.71%	70.4%	56.28%

The results presented in the table compare the performance of various machine learning models (Random Forest, Support Vector Machine, and Decision Tree) integrated with three different deep learning feature extraction architectures (MobileNetV2, DenseNet121, and EfficientNetB0) and the previous studies [24,35] on the HAM10000 dataset. Overall, Random Forest combined with DenseNet121 achieves the highest accuracy (98.32%) compared to all the other methods along with excellent precision, recall, and F1-score (98%), demonstrating its robust ability to classify dermatological images. SVM with MobileNetV2 closely follows, achieving 98.20% accuracy and maintaining high precision, recall, and F1-score values (98%), showcasing the effectiveness of MobileNetV2 for feature extraction. Interestingly, Random Forest with MobileNetV2 also performs competitively (97.86% accuracy). However, Decision Tree models exhibit a noticeably lower performance across all feature extractors, with the highest accuracy being 85.39% when paired with MobileNetV2. This highlights the limitations of simpler models like Decision Trees for complex datasets, even with advanced feature extraction. EfficientNetB0, although effective, slightly underperforms compared to MobileNetV2 and DenseNet121 in most scenarios, with its best accuracy being 97.51% when paired with SVM. These results underscore the importance of choosing the right combination of feature extractor and classifier, with DenseNet121 and MobileNetV2 proving to be particularly effective for dermatological image analysis.

5. Conclusions and Future Scope

This research reviewed existing models to understand their strengths and limitations, applying advanced preprocessing techniques to enhance image quality and developing models to detect skin diseases across diverse skin tones. Due to the unavailability of a black skin dataset, the HAM10000 dataset was utilized in this study. Effective preprocessing methods, including resizing, normalization, and dimensionality reduction using PCA, along with model fine-tuning, such as dropout layers and hyperparameter optimization, helped mitigate overfitting. The SMOTE technique was applied to address the data imbalance, improving the model's generalization capabilities. The results demonstrated that Random Forest, when paired with DenseNet121, achieved the highest accuracy of 98.32%, followed closely by SVM with MobileNetV2 at 98.20%. The Decision Tree model with MobileNetV2 achieved an accuracy of 85.39%. In comparison, models utilizing EfficientNetB0

for feature extraction performed slightly lower, with SVM, Random Forest, and Decision Tree achieving accuracies of 97.51%, 96.68%, and 84.96%, respectively. These findings highlight the superior performance of the proposed models over existing predictions on the HAM10000 dataset, emphasizing their potential for dermoscopic applications. Future research will explore advanced architectures, such as transformers, to further enhance model performance. Techniques like Grad-CAM will be investigated for improving the interpretability and explainability of model predictions. Additionally, integrating multimodal data—such as patient demographics, medical history, and genetic information—has the potential to enhance model accuracy and enable personalized treatment recommendations. Further preprocessing methods, including Histogram Equalization, Gray-scale Conversion, Color Space Transformation, and Edge Detection, could also improve the detection of skin diseases, particularly in melanin-rich skin, where conditions are often harder to detect due to a lower contrast. Expanding datasets to include images of black and diverse skin tones will remain a key focus for improving generalization across populations.

Author Contributions: Conceptualization, A.A. and F.S.; methodology, A.A., F.S. and N.S.E.; software, A.A.; validation, S.B., A.M.A.; formal analysis, A.A., F.S., N.S.E., S.B. and A.M.A.; investigation, S.B. and A.M.A.; resources, S.B. and A.M.A.; data curation, A.A.; writing—original draft preparation, A.A. and F.S.; writing—review and editing, A.A., F.S., N.S.E., S.B. and A.M.A.; visualization, A.A.; supervision, F.S. and N.S.E.; project administration, F.S. and S.B.; funding acquisition, F.S. and S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (GPIP: 1907-612-2024). The authors, therefore, give their acknowledgement with thanks to DSR for their technical and financial support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets are available upon request.

Acknowledgments: This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (GPIP: 1907-612-2024). The authors, therefore, give their acknowledgement with thanks to DSR for their technical and financial support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rezk, E.; Eltorki, M.; El-Dakhkhni, W. Leveraging artificial intelligence to improve the diversity of dermatological skin color pathology: Protocol for an algorithm development and validation study. *JMIR Res. Protoc.* **2022**, *11*, e34896. [[CrossRef](#)]
2. Aljohani, K.; Turki, T. Automatic classification of melanoma skin cancer with deep convolutional neural networks. *Ai* **2022**, *3*, 512–525. [[CrossRef](#)]
3. Aswin, R.B.; Jaleel, J.A.; Salim, S. Implementation of ann classifier using matlab for skin cancer detection. *Int. J. Comp. Sci. Mob. Comput.* **2013**, *1002*, 87–94.
4. Stafford, H.; Buell, J.; Chiang, E.; Ramesh, U.; Migden, M.; Nagarajan, P.; Amit, M.; Yaniv, D. Non-melanoma skin cancer detection in the age of advanced technology: A review. *Cancers* **2023**, *15*, 3094. [[CrossRef](#)]
5. Kumari, S.; Umrao, S.; Kushwaha, D. Decoding the Skin with AI: A Review of Cutting-Edge Technologies and Applications. In Proceedings of the 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 15–16 March 2024; pp. 1276–1281.
6. Xu, L.; Jackowski, M.; Goshtasby, A.; Roseman, D.; Bines, S.; Yu, C.; Dhawan, A.; Huntley, A. Segmentation of skin cancer images. *Image Vis. Comput.* **1999**, *17*, 65–74. [[CrossRef](#)]
7. Amelard, R.; Glaister, J.; Wong, A.; Clausi, D.A. Melanoma decision support using lighting-corrected intuitive feature models. In *Computer Vision Techniques for the Diagnosis of Skin Cancer*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 193–219.
8. Senthil, V.; Shreyaa, V.; Kothandapany, V. Deep learning and rules-based hybrid approach to improve the accuracy of early detection of skin cancer. *Authorea Prepr.* **2022**. [[CrossRef](#)]

9. Kassani, S.H.; Kassani, P.H. A comparative study of deep learning architectures on melanoma detection. *Tissue Cell* **2019**, *58*, 76–83. [[CrossRef](#)] [[PubMed](#)]
10. Hekler, A.; Utikal, J.S.; Enk, A.H.; Solass, W.; Schmitt, M.; Klode, J.; Schadendorf, D.; Sondermann, W.; Franklin, C.; Bestvater, F.; et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *Eur. J. Cancer* **2019**, *118*, 91–96. [[CrossRef](#)] [[PubMed](#)]
11. Chattopadhyay, S.; Dey, A.; Singh, P.K.; Sarkar, R. DRDA-Net: Dense residual dual-shuffle attention network for breast cancer classification using histopathological images. *Comput. Biol. Med.* **2022**, *145*, 105437. [[CrossRef](#)] [[PubMed](#)]
12. Jahanbani, S.; Hansen, P.S.; Blum, L.K.; Bastounis, E.E.; Ramadoss, N.S.; Pandrala, M.; Kirschmann, J.M.; Blacker, G.S.; Love, Z.Z.; Weissman, I.L.; et al. Increased macrophage phagocytic activity with TLR9 agonist conjugation of an anti-Borrelia burgdorferi monoclonal antibody. *Clin. Immunol.* **2023**, *246*, 109180. [[CrossRef](#)]
13. Lallas, A.; Apalla, Z.; Argenziano, G.; Longo, C.; Moscarella, E.; Specchio, F.; Raucci, M.; Zalaudek, I. The dermatoscopic universe of basal cell carcinoma. *Dermatol. Pract. Concept.* **2014**, *4*, 11. [[CrossRef](#)]
14. Akay, B.N.; Kocyigit, P.; Heper, A.O.; Erdem, C. Dermatoscopy of flat pigmented facial lesions: Diagnostic challenge between pigmented actinic keratosis and lentigo maligna. *Br. J. Dermatol.* **2010**, *163*, 1212–1217. [[CrossRef](#)] [[PubMed](#)]
15. Zaballos, P.; Salsench, E.; Serrano, P.; Cuellar, F.; Puig, S.; Malvehy, J. Studying regression of seborrheic keratosis in lichenoid keratosis with sequential dermoscopy imaging. *Dermatology* **2010**, *220*, 103–109. [[CrossRef](#)] [[PubMed](#)]
16. Moscarella, E.; Zalaudek, I.; Pellacani, G.; Eibenschutz, L.; Catricalà, C.; Amantea, A.; Panetta, C.; Argenziano, G. Lichenoid keratosis-like melanomas. *J. Am. Acad. Dermatol.* **2011**, *65*, e85–e87. [[CrossRef](#)]
17. Zaballos, P.; Puig, S.; Llambrich, A. and Malvehy, J. Dermoscopy of dermatofibromas: A prospective morphological study of 412 cases. *Arch. Dermatol.* **2008**, *144*, 75–83. [[CrossRef](#)] [[PubMed](#)]
18. Rosendahl, C.; Cameron, A.; McColl, I.; Wilkinson, D. Dermatoscopy in routine practice: ‘Chaos and clues’. *Aust. Fam. Physician* **2012**, *41*, 482–487.
19. Zaballos, P.; Daufí, C.; Puig, S.; Argenziano, G.; Moreno-Ramírez, D.; Cabo, H.; Marghoob, A.A.; Llambrich, A.; Zalaudek, I.; Malvehy, J. Dermoscopy of solitary angiokeratomas: A morphological study. *Arch. Dermatol.* **2007**, *143*, 318–325. [[CrossRef](#)] [[PubMed](#)]
20. Zaballos, P.; Carulla, M.; Ozdemir, F.; Zalaudek, I.; Bañuls, J.; Llambrich, A.; Puig, S.; Argenziano, G.; Malvehy, J. Dermoscopy of pyogenic granuloma: A morphological study. *Br. J. Dermatol.* **2010**, *163*, 1229–1237. [[CrossRef](#)]
21. Vasconcelos, C.N.; Vasconcelos, B.N. Experiments using deep learning for dermoscopy image analysis. *Pattern Recognit. Lett.* **2020**, *139*, 95–103. [[CrossRef](#)]
22. Mahmoud, N.M.; Soliman, A.M. Early automated detection system for skin cancer diagnosis using artificial intelligent techniques. *Sci. Rep.* **2024**, *14*, 9749. [[CrossRef](#)]
23. Ashwath, V.A.; Sikha, O.K.; Benitez, R. TS-CNN: A three-tier self-interpretable CNN for multi-region medical image classification. *IEEE Access* **2023**, *11*, 78402–78418. [[CrossRef](#)]
24. Singh, J.; Sandhu, J.K.; Kumar, Y. An analysis of detection and diagnosis of different classes of skin diseases using artificial intelligence-based learning approaches with hyper parameters. *Arch. Comput. Methods Eng.* **2024**, *31*, 1051–1078. [[CrossRef](#)]
25. Lopez, A.R.; Giro-I-Nieto, X.; Burdick, J.; Marques, O. Skin lesion classification from dermoscopic images using deep learning techniques. In Proceedings of the 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), Innsbruck, Austria, 20–21 February 2017.
26. Bhadula, S.; Sharma, S.; Juyal, P.; Kulshrestha, C. Machine learning algorithms based skin disease detection. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *9*, 4044–4049. [[CrossRef](#)]
27. Allugunti, V.R. A machine learning model for skin disease classification using convolution neural network. *Int. J. Comput. Program. Database Manag.* **2022**, *3*, 141–147. [[CrossRef](#)]
28. Hameed, N.; Shabut, A.M.; Hossain, M.A. Multi-class skin diseases classification using deep convolutional neural network and support vector machine. In Proceedings of the 2018 12th International Conference on Software, Knowledge, Information Management & Applications (SKIMA 2018), Phnom Penh, Cambodia, 3–5 December 2018.
29. ALenezi, N.S.A. A method of skin disease detection using image processing and machine learning. *Procedia Comput. Sci.* **2019**, *163*, 85–92. [[CrossRef](#)]
30. Cheong, K.H.; Tang, K.J.W.; Zhao, X.; Koh, J.E.W.; Faust, O.; Gururajan, R.; Ciaccio, E.J.; Rajinikanth, V.; Acharya, U.R. An automated skin melanoma detection system with melanoma-index based on entropy features. *Biocybern. Biomed. Eng.* **2021**, *41*, 997–1012. [[CrossRef](#)]
31. Tan, T.Y.; Zhang, L.; Jiang, M. An intelligent decision support system for skin cancer detection from dermoscopic images. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2016), Changsha, China, 13–15 August 2016.

32. Maia, L.B.; Lima, A.; Pereira, R.M.; Junior, G.B.; de Almeida, J.D.; de Paiva, A.C. Evaluation of melanoma diagnosis using deep features. In Proceedings of the 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), Maribor, Slovenia, 20–22 June 2018.
33. Schandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [[CrossRef](#)]
34. Kuo, B.C.; Ho, H.H.; Li, C.H.; Hung, C.C.; Taur, J.S. A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *7*, 317–326. [[CrossRef](#)]
35. Velaga, N.; Vardinani, V.R.; Tupakula, P.; Pamidimukkala, J.S. Skin cancer detection using the HAM10000 dataset: A comparative study of machine learning models. In Proceedings of the 2023 Global Conference on Information Technologies and Communications (GCITC), Bangalore, India, 1–3 December 2023; IEEE: New York, NY, USA, 2023; pp. 1–7.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.