

## Article

# Deep Learning-Based High-Frequency Ultrasound Skin Image Classification with Multicriteria Model Evaluation

Joanna Czajkowska <sup>1,\*</sup>, Paweł Badura <sup>1</sup>, Szymon Korzekwa <sup>2</sup>, Anna Płatkowska-Szczerek <sup>3</sup>  
and Monika Słowińska <sup>4</sup>

<sup>1</sup> Faculty of Biomedical Engineering, Silesian University of Technology, 41-800 Zabrze, Poland; pawel.badura@polsl.pl

<sup>2</sup> Department of Temporomandibular Disorders, Division of Prosthodontics, Poznan University of Medical Sciences, 60-512 Poznań, Poland; korzekwas@gmail.com

<sup>3</sup> Anclara sp. z o.o., 02-624 Warszawa, Poland; dr.platkowska@gmail.com

<sup>4</sup> Department of Dermatology, Military Institute of Medicine, 01-755 Warszawa, Poland; mslowinska@wim.mil.pl

\* Correspondence: joanna.czajkowska@polsl.pl; Tel.: +48-322-774-67

**Abstract:** This study presents the first application of convolutional neural networks to high-frequency ultrasound skin image classification. This type of imaging opens up new opportunities in dermatology, showing inflammatory diseases such as atopic dermatitis, psoriasis, or skin lesions. We collected a database of 631 images with healthy skin and different skin pathologies to train and assess all stages of the methodology. The proposed framework starts with the segmentation of the epidermal layer using a DeepLab v3+ model with a pre-trained Xception backbone. We employ transfer learning to train the segmentation model for two purposes: to extract the region of interest for classification and to prepare the skin layer map for classification confidence estimation. For classification, we train five models in different input data modes and data augmentation setups. We also introduce a classification confidence level to evaluate the deep model's reliability. The measure combines our skin layer map with the heatmap produced by the Grad-CAM technique designed to indicate image regions used by the deep model to make a classification decision. Moreover, we propose a multicriteria model evaluation measure to select the optimal model in terms of classification accuracy, confidence, and test dataset size. The experiments described in the paper show that the DenseNet-201 model fed with the extracted region of interest produces the most reliable and accurate results.

**Keywords:** high-frequency ultrasound; inflammatory skin diseases; skin lesions; image classification; deep learning; transfer learning; Grad-CAM



**Citation:** Czajkowska, J.; Badura, P.; Korzekwa, S.; Płatkowska-Szczerek, A.; Słowińska, M. Deep Learning-Based High-Frequency Ultrasound Skin Image Classification with Multicriteria Model Evaluation. *Sensors* **2021**, *21*, 5846. <https://doi.org/10.3390/s21175846>

Academic Editor: Yu-Dong Zhang

Received: 27 July 2021

Accepted: 27 August 2021

Published: 30 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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/licenses/by/4.0/>).

## 1. Introduction

Ultrasound (US) is described in Reference [1] as a powerful and ubiquitous screening and diagnostic imaging technique, which has found applications in different fields. Being non-invasive, convenient and safe, it is widely used in multiple organ examinations, prenatal screening, or guided biopsies. Its fast development, resulting in high-frequency ultrasound (HFUS, between 20 and 30 MHz) and ultra high-frequency ultrasound (UHFUS, >30 MHz), has opened up new opportunities for medical applications [2]. It is gaining popularity in dermatology, dermatological oncology, ophthalmology, cosmetology, and aesthetic medicine. US at frequencies below 100 MHz is now commonly used in medical practice [3], and the >100 MHz probes are constantly being designed [4]. Improving spatial resolution of acquired images, the higher frequency of the US probe enables clear visualization of superficial structures such as the fat layer, the muscle layer, blood vessels, hair follicles, and skin appendages [2,5–7]. HFUS is used for healthy skin analysis, where skin thickness is inversely proportional to age, and due to increased collagen production connected with aging, echogenicity tends to increase too [2]. It can be used to estimate the hair

follicle growth phase, identify their inflammation and early signs of adnexal pathologies [2]. HFUS is a reliable method with outstanding intrahand interreproducibility for measuring melanoma depth *in vivo* and may enable single-step surgical excision [8]. The use of ultrasound helps in therapeutic decisions and surgical planning for non-melanoma and melanoma skin cancers and detecting early neoplasms [2]. Kleinerman et al. [9] demonstrate its potential for differentiation between melanoma, benign nevi, and seborrheic keratoses, along with the monitoring of inflammatory conditions and photo damage.

One of the first HFUS applications in dermatology can be found in inflammatory skin diseases [2], where it has opened up new opportunities in diagnosis and treatment monitoring. Both HFUS and UHFUS enable reliable, accurate, and fast skin layer analysis. As reported in [2], the first implementation in inflammatory skin diseases was scleroderma, whereas recent studies describe US as a helpful objective marker in patients with psoriasis. The treatment effect in atopic patients observed using HFUS (20 MHz) was first described by [10]. The presence of a subepidermal low echogenic band (SLEB) in visually healthy skin enables skin lesion differentiation. Its thickness correlates with the histological degree of epidermal hyperkeratosis and intensity of inflammatory infiltrates [2,7,10,11]. Although different studies [2,11] examine the validity, repeatability, and reliability of skin measurements, which correlate with histological analysis, none of them analyze the HFUS capability in skin lesion differentiation. Such a possibility can be essential in the case of neoplastic or inflammatory skin diseases. The US-based assessment of the therapeutic decision is especially important in the latter, where the early differentiation between atopic dermatitis (AD) and psoriasis is crucial for the applied therapy. The classification step can also be significant for further image processing algorithms in image segmentation and image-based measurements.

Standalone visual analysis of HFUS images does not provide reliable results. The problem with HFUS skin image classification is its interpretation, usually followed and supported by visual or dermatoscopic analysis of the affected skin combined with medical history and interview. By analyzing the HFUS image only, the expert can rarely choose between AD, psoriasis, or healthy skin. On the other hand, the computer-aided diagnosis (CAD) system, which enables automated HFUS image classification, can support the physicians in their final decision. Moreover, further accurate image data analysis in different skin pathologies can strongly rely on the initial classification. Individual diseases require dedicated analysis, for example, the segmentation of particular structures, for example, tumors, the SLEB layer, epidermis layer, and so forth. The HFUS classification step should be considered an integral part of CAD in dermatology. Therefore, we decided to introduce a deep learning-based approach for this; the first in this field. Possible reasons for the absence of HFUS skin classification can be identified due to the relative novelty of HFUS in dermatology, a limited number of available datasets and experts able to analyze and annotate the data, or difficulties in the clear translation of the classical US interpretation into HFUS-based diagnosis.

Despite the lack of algorithms for HFUS image classification, different methods have appeared for automated and semi-automated HFUS (or UHFUS) image analysis. Pereyra et al. [12] first described the segmentation of skin lesions in a 25 MHz US using a generalized Rayleigh mixture model, then References [7,13–16] presented automatic skin layer segmentation algorithms. They employed classic segmentation approaches such as level sets and active contour models [14–16], or the most recent deep learning-based solutions [7]. The latter is now widely used in image segmentation, mainly through convolutional neural networks (CNNs), consisting of different layers responsible for dedicated data processing [17]. Skin layers were also segmented in optical coherence tomography (OCT) data [18–22], where the latest approach [22] involved the deep U-Net architecture, followed by the Savitzky–Golay filter and Fourier domain filtering.

Deep neural networks are also now common in the conventional US (<20 MHz) image analysis. Huang et al. [23] described an application for breast and liver lesion diagnosis, fetal and cardiovascular image classification, or as a thyroid nodule diagnosis support.

Liu et al. [1] added kidney, bone, prostate, and brain US image processing to this list, with CNNs mainly used for classification tasks. Cai et al. [24] underlined the importance of deep learning in US image analysis of breast nodules. The most applicable architectures in US data classification are GoogLeNet, introduced by Han et al. [25] to classify breast lesions and by Chi et al. [26] for thyroid nodules, VGGNet and fully-connected networks (FCN) to differentiate the level of liver fibrosis [27], or Inception-v3, ResNet-101, and DenseNet-169, achieving the best performance in the automatic classification of common maternal–fetal ultrasound planes [28].

Due to the problem with the access to the training data, the models above cannot mostly generalize knowledge to the unknown clinical data (acquired in different medical fields, conditions, acquisition protocols, etc.). Hence, various ideas were introduced to handle generalization and overfitting issues, for example, semi-supervised learning, transfer learning, learning from noisy labels and learning from computer-generated labels [29]. Transfer learning (TL) seems to be the most common. Van Opbroek et al. [30] reported that TL improves supervised image segmentation across imaging protocols in medical applications. It employs a model pre-trained in a source domain and then fine-tunes it to the target task [29]. Morid et al. [31] provided the summary of methods employing TL to medical image analysis using the CNNs trained on the non-medical ImageNet dataset [32]. Ayana et al. [33] limited the review to US imaging of breast cancer.

The fast development of DL approaches enables the extraction of discriminative image features to classify different image modalities. The most common solution, CNN, automatically learns a representation of visual features from the input image textures [34] providing superior performance compared to other conventional machine learning methods. However, the integration of intelligent systems into clinical practice requires a visual explanation for their decision [35]. Therefore, Zhou et al. [36] proposed class activation mapping (CAM) to indicate discriminative image regions used by the CNN in image data classification. Selvaraju et al. [35] extended this work, proposing a Grad-CAM method applicable to a broader range of CNN models, including CNNs with fully connected layers. An application of Grad-CAM to medical images was given in [34], where the authors proposed class-selective relevance mapping (CRM) for localizing and visualizing discriminative regions of interest (ROIs) in the multimodal images.

Due to the absence of classification frameworks for HFUS skin image data, we present such a solution in our work. The HFUS images of the skin are similar in different pathologies, so it is difficult to define a set of features that would enable their classification. Since CNNs reach for an extensive collection of primitive features generalized in subsequent layers, they seem well suited to this problem. This is the first work addressing HFUS skin image classification. The proposed framework starts with the epidermis segmentation step to reduce the region of interest, followed by estimating a skin layer map. Then, the classification step is verified with a dedicated procedure using the Grad-CAM algorithm and the previously obtained skin layer map. Our novel approach combines segmentation and classification, providing the final results with the estimated classification confidence level. The method was verified with a database consisting of 631 HFUS images of AD (303), psoriasis (77), or neoplastic diseases (200), with a control group of 51 subjects, all described by an expert in a multi-stage diagnosis. The experiments included the study of the segmentation step, the influence of ROI reduction with different parameters, CNN architecture selection, and Grad-CAM-based model verification.

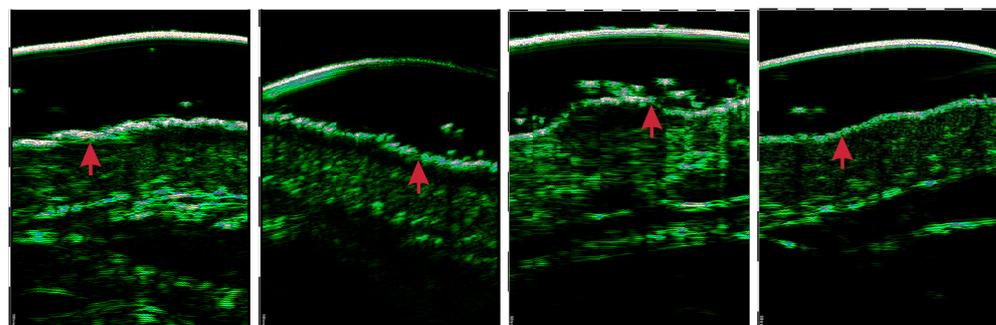
The datasets used for training and validation are specified in Section 2.1. The classification, epidermis segmentation, and classification confidence estimation are presented in detail in the remainder of Section 2. Section 3 describes the training and validation experiments along with the results. Due to the lack of state-of-the-art methods we can directly compete with, we assess our model in various architecture modes and flowchart parameter settings. The study is discussed in Section 4, and Section 5 concludes the paper.

## 2. Materials and Methods

### 2.1. Materials

To design a classification algorithm of different skin diseases, we combined three HFUS datasets. The first one, further called benchmark, consists of 380 HFUS images of 380 patients with atopic dermatitis (303 images) or psoriasis (77). The dataset is publicly available [37] along with the pre-trained SegUnet model for skin layer segmentation [7]. The second dataset consists of 200 images of 32 patients with different non-melanocytic skin tumors: BCC (143 images), fibroma (32), skin metastasis of breast cancer (10), keratofibroma (9), superficial BCC (3), and squamous cell carcinoma (3), whereas the third data set includes 51 images of 51 patients with healthy skin. All patients were Caucasian, men and women aged 18–86. The age in individual groups corresponds to the frequency of particular dermatological problems in specific age groups: mainly young adults in AD and psoriasis (18–45) and the elderly in skin tumors (69–86), with a broader range in the control group (26–65). The medical diagnosis for all patients was provided by an expert in dermatology, based on visual analysis of the skin, dermatoscopic image analysis, medical history, and medical interview. Thus, the methodology was designed, trained, and tested using a total of 631 HFUS images in four groups: AD, psoriasis, skin tumor, and control group.

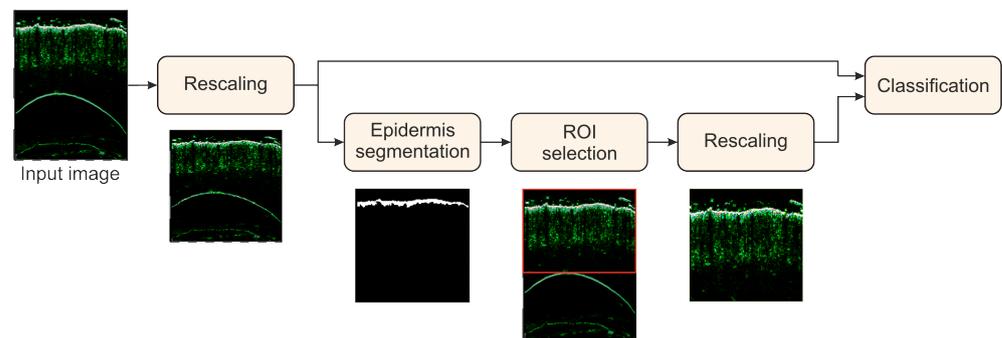
All the analyzed HFUS image data were acquired using DUB SkinScanner75 with a 75 MHz transducer. The images have the same size of  $2067 \times 1555$  pixels, but different resolutions (lateral  $\times$  axial):  $0.0019 \times 0.085$ ,  $0.0024 \times 0.085$ ,  $0.0031 \times 0.085$  mm/pix. The images were acquired at the affected skin area or forearm, in the case of the control group. The images were captured perpendicularly to the skin surface ( $\pm 10$  degrees) with minimal pressure from the physician (the DUB SkinScanner75 forces the user to position the probe to the right angle). The epidermal layer was delineated by one expert and verified by another in all 580 images of AD, psoriasis, and skin lesions. Exemplary images of four considered classes are shown in Figure 1.



**Figure 1.** Exemplary HFUS images of four classes under consideration: (from left) AD, psoriasis, skin tumor, and control group.

### 2.2. HFUS Skin Image Classification

The overall goal of the study was to develop and evaluate the HFUS image classification framework, which provides the most reliable classification results. The general scheme of the classification algorithm with two processing paths is shown in Figure 2. The first (top) path is based on the pre-trained and fine-tuned CNN model, leading directly to the final classification result. During the experiments, we considered five CNN models: DenseNet-201 [38], GoogLeNet [39], Inception-ResNet-v2 [20], MobileNet [40], and ResNet-101 [41], all pre-trained on the ImageNet dataset, and fine-tuned with our HFUS images of skin diseases. Moreover, two different augmentation setups were introduced to secure classification robustness.



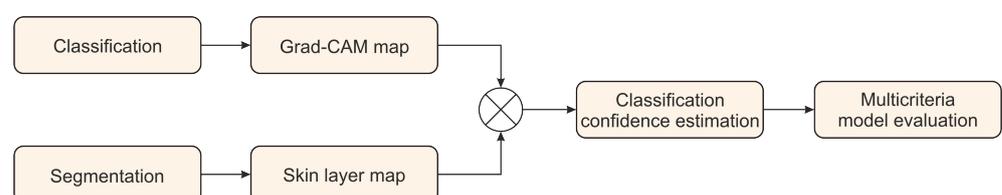
**Figure 2.** General scheme of the skin classification method.

In the second (bottom) path, we extract the ROI from the image to feed the classifier with the skin-layers-related content. For this, we first introduce the epidermis segmentation to our framework, as the epidermis layer is rather clearly visible in each HFUS image (see Figure 1). The epidermis segmentation is performed with the DeepLab v3+ network built on a pre-trained backbone for feature extraction [42,43]. Different models (trained via transfer learning) were tested to serve as a backbone: ResNet-18, ResNet-50 [41], and Xception [44]. Originally, the models were pre-trained using the ImageNet database [32,45]. Moreover, the model described in [7] trained for the benchmark database was considered as the possible choice at this step. Finally, we selected the DeepLab v3+ model with an Xception as the backbone, pretrained on ImageNet database, with the stochastic gradient descent with momentum optimizer (SGDM) as the most efficient in a series of experiments. Quantitative inter-model comparison can be found in Section 3. We used data augmentation through random geometric transformations (horizontal reflection,  $\pm 10$ -pixel translation in both directions). The batch size for training was set to 8, and the maximum number of epochs to 200. The loss function we employed was weighted generalized Dice loss [46], able to handle possible *epidermis/background* pixel distribution imbalance.

Based on the epidermis segmentation mask, the ROI for further analysis is selected. This procedure is an approach to unify the images before the main classification procedure. It addresses the gel and probe membrane appearance and standardizes the positioning of skin layers in the image. Depending on the experiment, we reduce the image area above and below the epidermis. The considered variants of ROI include: (1) removing the upper part of the image 30 pixels above the top pixel of the segmented epidermis layer; (2) removing the upper part of the image 1 mm above the top pixel of the epidermis; (3,4) removing the upper part as in (1) and (2), respectively, and removing the lower part 2 mm below the bottom pixel of the epidermis. Additionally, we set to 0 all pixels in all the areas in the ROI, denoted as avoided in the skin layer map (see Section 2.3.2).

### 2.3. Classification Confidence Estimation

In our study, we designed a methodology to assess the reliability of a classification model (Figure 3). It relies on two components: the Grad-CAM deep learning interpretability technique developed by Selvaraju et al. [35] and a skin layer map estimate based on our epidermis segmentation (Figure 2). The method produces a classification confidence level (CCL) measure to assess individual classification results quantitatively. Moreover, we introduce a dedicated multicriteria model evaluation procedure to evaluate the model robustness and reliability.



**Figure 3.** General scheme of the classification model confidence estimation method.

### 2.3.1. Grad-CAM Map

The Grad-CAM deep learning interpretability technique was proposed by Selvaraju et al. [35] to provide a visual explanation of the deep model decision. It employs the gradient of the classification score with respect to the final feature map. Grad-CAM enables the identification of parts of the input image with the greatest impact on the classification results. According to [35], the Grad-CAM map can be estimated as:

$$L^c = ReLU\left(\sum_k \alpha_k^c A^k\right), \quad (1)$$

where  $c$  is a target class (we use the class resulting from the CNN-based classification),  $\alpha_k^c$  are neuron 'importance' weights of a feature map  $k$  for a target class  $c$ , calculated as:

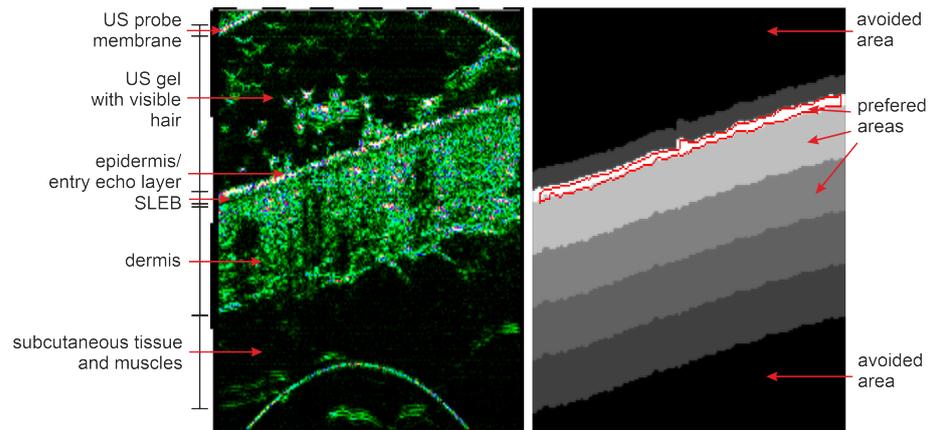
$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}, \quad (2)$$

where  $A^k$  stands for the feature map of the final convolutional layer,  $Z$  comes from its size, and  $\frac{\partial y^c}{\partial A^k}$  is the gradient of the score for class  $c$ ,  $y^c$  (before the softmax layer), with respect to the feature map  $A^k$  of the final convolutional layer.

### 2.3.2. Skin Layer Map

In their original work, Selvaraju et al. [35] provide a meaningful example of the classified objects (cat and dog), which, combined with a Grad-CAM map, enables visual evaluation of the network focus in the image. The problem with HFUS image classification is its visual interpretation. It is challenging to clearly define the ROI, which should be the network interest in target skin diseases. However, based on semantic interpretation and human understanding of HFUS images, we propose a skin layer map (SLM) to indicate regions, which are considered during medical diagnosis, and other regions that the network should not focus on to avoid overfitting.

Based on the epidermis segmentation outcome and skin layer thicknesses reported in [47,48] and measured in our database, we create the individual skin layer map for each input image. The map indicates the most relevant regions that should be considered in skin diagnosis and gain particular focus from the CNN. As the epidermis area results directly from the segmentation step, the dermis region location is estimated based on the thickness values given in [47,48] and measured in our dataset. The reported skin (epidermis and dermis) thickness in different interesting body parts varies from 1.5 mm to 2.5 mm [47]. Similar values can be observed in our dataset, where this thickness falls below <2 mm. On the other hand, the regions of the epidermis, possible SLEB, and upper dermis area are essential for our findings. Considering this and assuming that the skin layers are mostly parallel (with some exception in the skin tumor area), we generate the SLM as illustrated in Figure 4. Starting from the epidermis segmentation results, where the SLM has its maximum of 3, it decreases twice by 1 every 0.5 mm (half of the dermis layer thickness), and then twice by 0.5 every 0.5 mm until the dermis ends. Moreover, it is set to 0 in the US gel area located directly above (30 pixels) the epidermis and  $-1$  in the image regions, which should be avoided in further analysis. The avoided parts include the US probe membrane, remaining US gel area, or muscles.



**Figure 4.** Illustration of a skin layer map (right) obtained for the HFUS image (left) based on the epidermis segmentation results (delineated in red).

### 2.3.3. Classification Confidence Level

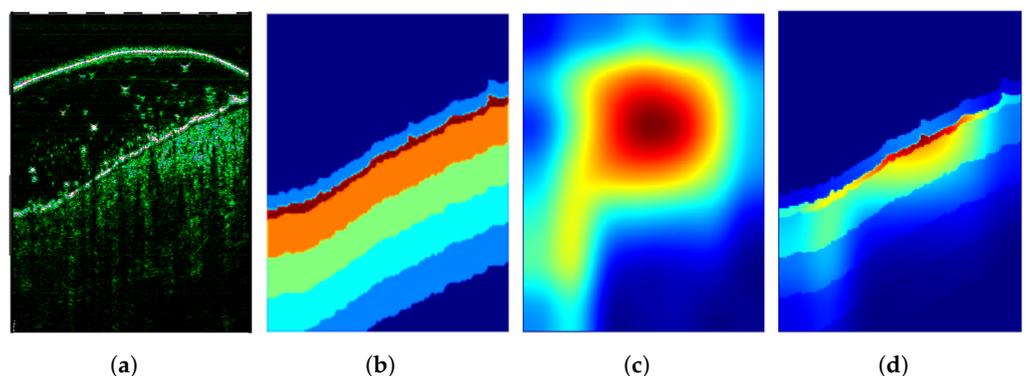
To assess the reliability of classification, we combine the Grad-CAM outcome and our SLM to get the classification confidence map and classification confidence level  $t_\kappa$  for a classification model  $\kappa$  (see Figure 5). Let  $L_\kappa(i, j)$  and  $H_\kappa(i, j)$  be the Grad-CAM map value and the SLM value, respectively, corresponding to the input image pixel  $I(i, j)$ . Assuming that  $L_\kappa$  is normalized for each image to the range of 0 to 1, the CCL is given as:

$$t_\kappa = \frac{1}{M} \sum_{k,l} G_\kappa(k, l), \quad (3)$$

where:

$$G_\kappa = \{L_\kappa(i, j)H_\kappa(i, j) : L_\kappa(i, j)H_\kappa(i, j) > 0, i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, n\}\}, \quad (4)$$

$m \times n$  is the input image size, and  $M$  stands for the cardinality of  $G_\kappa$ . Note that due to different sizes of the original matrices, the obtained Grad-CAM maps are rescaled to the original image size and eventually cropped to match the ROI before applying (3) and (4).



**Figure 5.** Illustration of the classification confidence map (d) based on the input image (a), skin layer map (b), and Grad-CAM map (c).

### 2.3.4. Multicriteria Model Evaluation

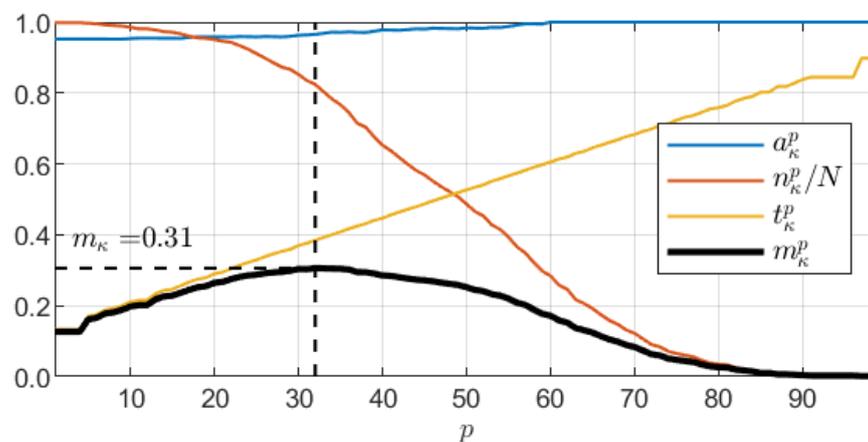
For the evaluation of a classification model  $\kappa$ , we propose a dedicated metrics  $m_\kappa$  based on three components: the CCL value  $t_\kappa$ , classification accuracy  $a_\kappa$ , and test dataset size  $n_\kappa$ . Each of them should be as large as possible to secure high classification accuracy and reliability. The concept constituting multicriteria model evaluation aims at achieving

high accuracy with high confidence over the largest possible dataset.  $m_\kappa$  is determined through maximization:

$$m_\kappa = \max_p \left( m_\kappa^p \right), \quad (5)$$

$$m_\kappa^p = a_\kappa^p \frac{n_\kappa^p}{N} t_\kappa^p, \quad (6)$$

where  $N = 631$  is the number of images in the entire dataset.  $p = 1, 2, \dots, 99$  is used to determine the CCL threshold  $t_\kappa^p$  as the  $p$ th percentile of CCL values produced by the model  $c$  over the entire dataset. In each iteration, we limit the dataset to  $n_\kappa^p$  samples with the CCL value greater than  $t_\kappa^p$  and calculate partial accuracy  $a_\kappa^p$ . Thus, with increasing  $p$ , the dataset size decreases, while  $t_\kappa^p$  increases. However, since such a procedure in general preserves strong samples, the classification accuracy over the limited dataset is also non-decreasing. Figure 6 shows sample charts of all four measures over a certain model  $\kappa$ .



**Figure 6.** Illustration of a multicriteria model evaluation measure  $m_\kappa$  with charts of  $m_\kappa^p$  and its components as a function of cutoff percentile number  $p$ .

### 3. Experiments and Results

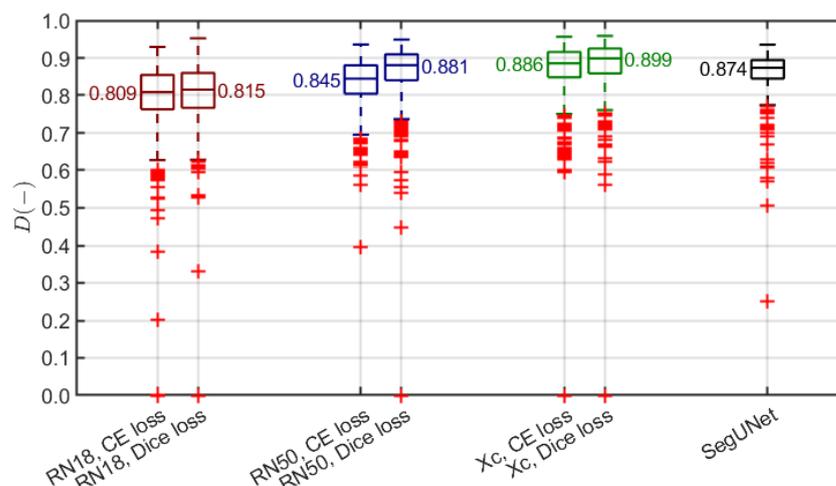
The training and experiments were performed in the Deep Learning Toolbox (version 14.2) of the Matlab software (9.10.0.1602886, R2021a) on a workstation with an 8-Core CPU @ 3.20 GHz, 64 GB RAM, and Nvidia Quadro RTX 6000 24 GB GPU.

#### 3.1. Epidermis Segmentation

As introduced in Section 2.3.2, the skin layer map is determined based on the initial epidermis segmentation. Thus, we verified different pre-trained (using ImageNet) models as a backbone for the DeepLab v3+ architecture: ResNet-18, ResNet-50, and Xception. We trained each using two different loss functions: cross-entropy and Dice loss. The data augmentation involved random geometric transformations (horizontal reflection,  $\pm 10$ -pixel translation in both directions). We also compared the obtained results with the SegUNet model described in [7]. To assess the segmentation, we employed the Dice index [49], commonly used to measure the spatial overlap in medical imaging studies, including HFUS analysis [7,15,50]. Since the epidermal layer was delineated by the expert in 580 images of AD, psoriasis, and skin tumors, the experiments were limited to this subset.

We used the external 10-fold cross-validation to assess the segmentation. The non-testing remaining data were divided into training and validation subsets (8:1 ratio) in each experiment. Moreover, we ensured that the image data of a single patient from the neoplastic dataset (200 images of 32 patients) were used only for either training or testing, never being shared by both subsets. The results are shown in Figure 7. Two observations are evident. The DeepLab v3+ model with the backbones pre-trained on ImageNet outperforms the SegUNet from [7], and the Dice-loss-based training is more efficient compared to the

cross-entropy loss. Hence, we selected the DeepLab v3+ model with an Xception backbone, SGDM optimizer, and Dice loss for the epidermis segmentation.

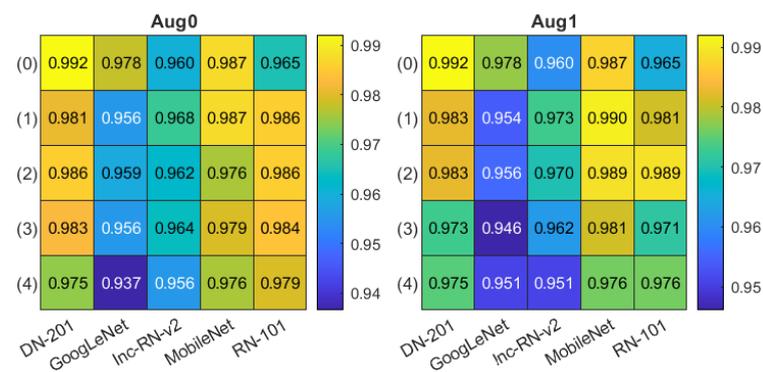


**Figure 7.** Summary of epidermis segmentation results (Dice index). Red, blue and green plots present our method with different pre-trained models used as a backbone of the DeepLab v3+ architecture (ResNet-18, ResNet-50, Xception) trained using cross-entropy loss and Dice loss. Black plot refers to the SegUNet model with the top performance in [7]. Each box covers 25th to 75th percentile confidence interval with median indicated by a central line. Whiskers refer to 1.5 times the interquartile range. Outliers are indicated with red +.

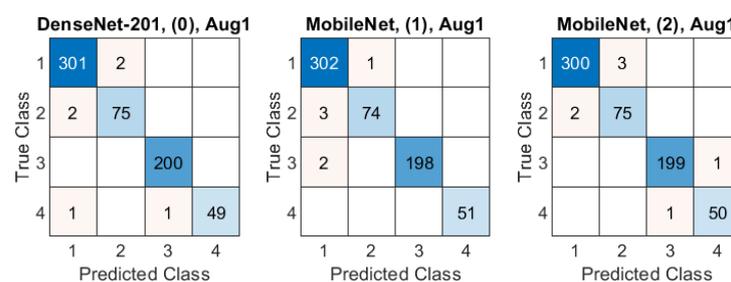
### 3.2. Classification

For skin classification, we considered five different CNN models, offering the most promising effectiveness in numerical experiments: DenseNet-201 [38], GoogLeNet [39], Inception-ResNet-v2 [20], MobileNet [40], and ResNet-101 [41], all pre-trained on the ImageNet database. Two modes of augmentation were used. In the first mode (denoted as Aug0), we applied random geometric transformations: horizontal reflection,  $\pm 10$ -pixel translation in both directions, similar to the one used in training the epidermis segmentation model. The second mode (Aug1) additionally employed random  $\pm 20^\circ$  rotation. Again, we used the external 10-fold cross-validation scheme with the non-testing data divided into training and validation subsets (8:1) in a patient-wise mode. The batch size for training was set to 8, and the maximum number of epochs to 50 with the SGDM optimizer and a cross-entropy loss. Additionally, we analyzed the ROI selection influence on the classification results. We considered all the four options for ROI extraction described in Section 2.2 with the variant numbers (1)–(4) given there used in all tables in this section. The results obtained by models fed with the original image are denoted as (0).

The classification accuracy obtained in all experiments are presented in Figure 8. The Aug0 and Aug1 in Figure 8 refer to data augmentation modes. Tables' rows refer to the ROI extraction modes (0)–(4) described in Section 2.2, whereas columns correspond to different deep models. The highest accuracies (0.992) obtained in both data augmentation modes by DenseNet-201 over the original-size images are highlighted in yellow. The confusion matrices for three best models (in terms of the highest accuracy) from Figure 8 are presented Figure 9: DenseNet-201, no ROI extraction (0), Aug0; MobileNet, ROI mode (1), Aug1, and MobileNet, ROI mode (2), Aug1.



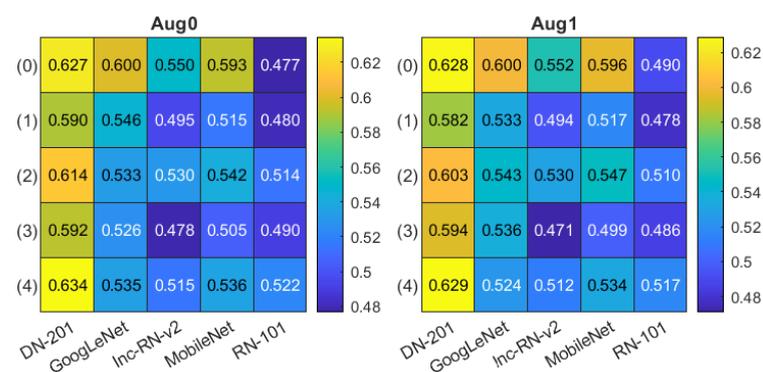
**Figure 8.** Classification accuracy summary in two data augmentation options Aug0, Aug1. Columns correspond to different deep models, rows refer to ROI extraction modes (described in the text).



**Figure 9.** Confusion matrices in three best-case classification models from Figure 8. Class IDs: 1–AD, 2–psoriasis, 3–skin tumor, 4–control group. Models are specified above each matrix in a format: model name, ROI extraction mode, data augmentation mode.

### 3.3. Classification Confidence Estimation

The mean classification confidence levels  $t_\kappa$  for each model investigated in Section 3.2 are given in Figure 10 (organized in correspondence with Figure 8). The highest  $t_\kappa$  value of 0.634 was obtained for DenseNet-201, for ROI in mode (4) with Aug0. The accuracy of that particular model was 0.975 (Figure 8), lower than the top accuracy (0.992).

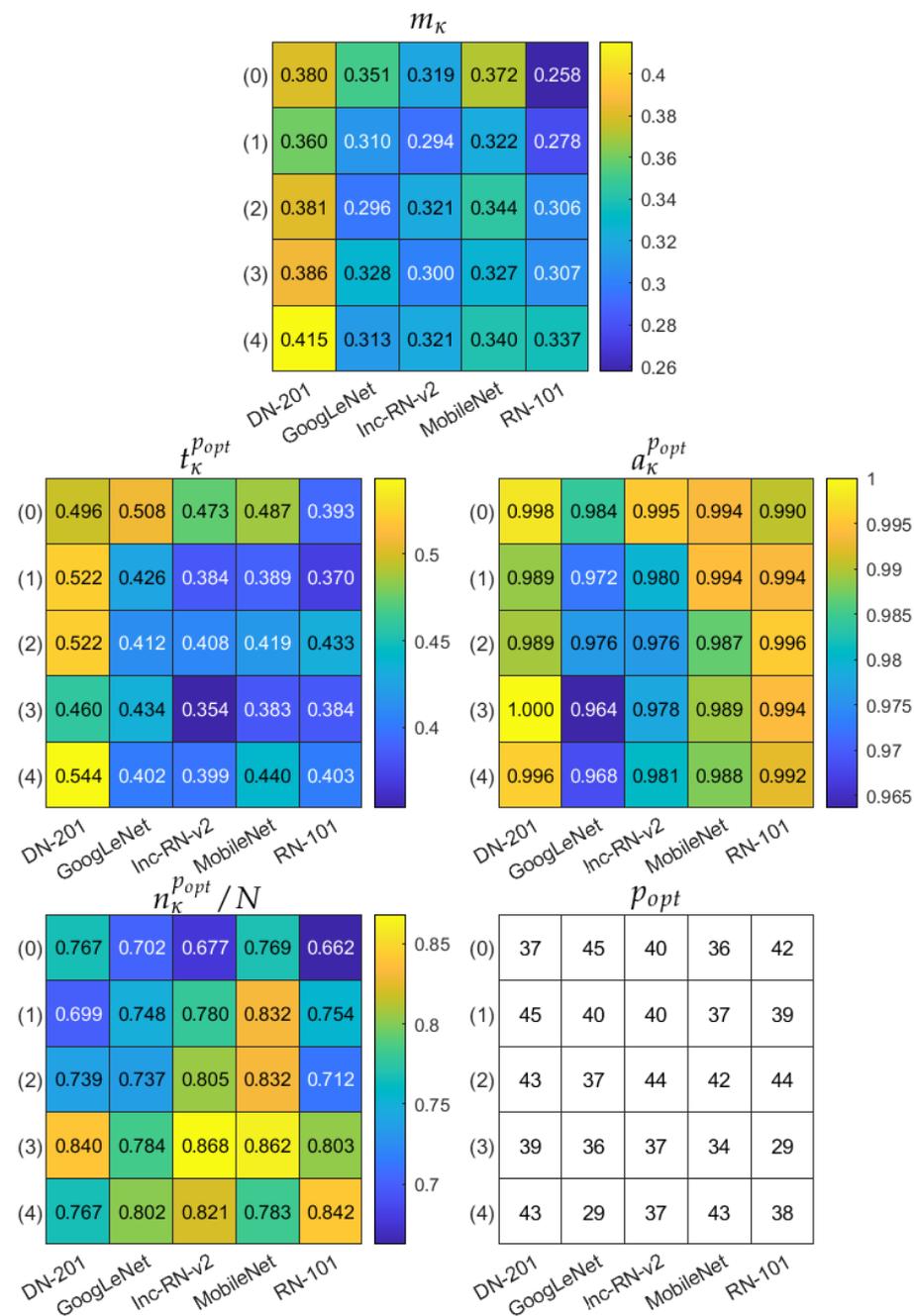


**Figure 10.** Mean classification confidence level  $t_\kappa$  summary in two data augmentation options Aug0, Aug1. Columns correspond to different deep models, rows refer to ROI extraction modes (described in the text).

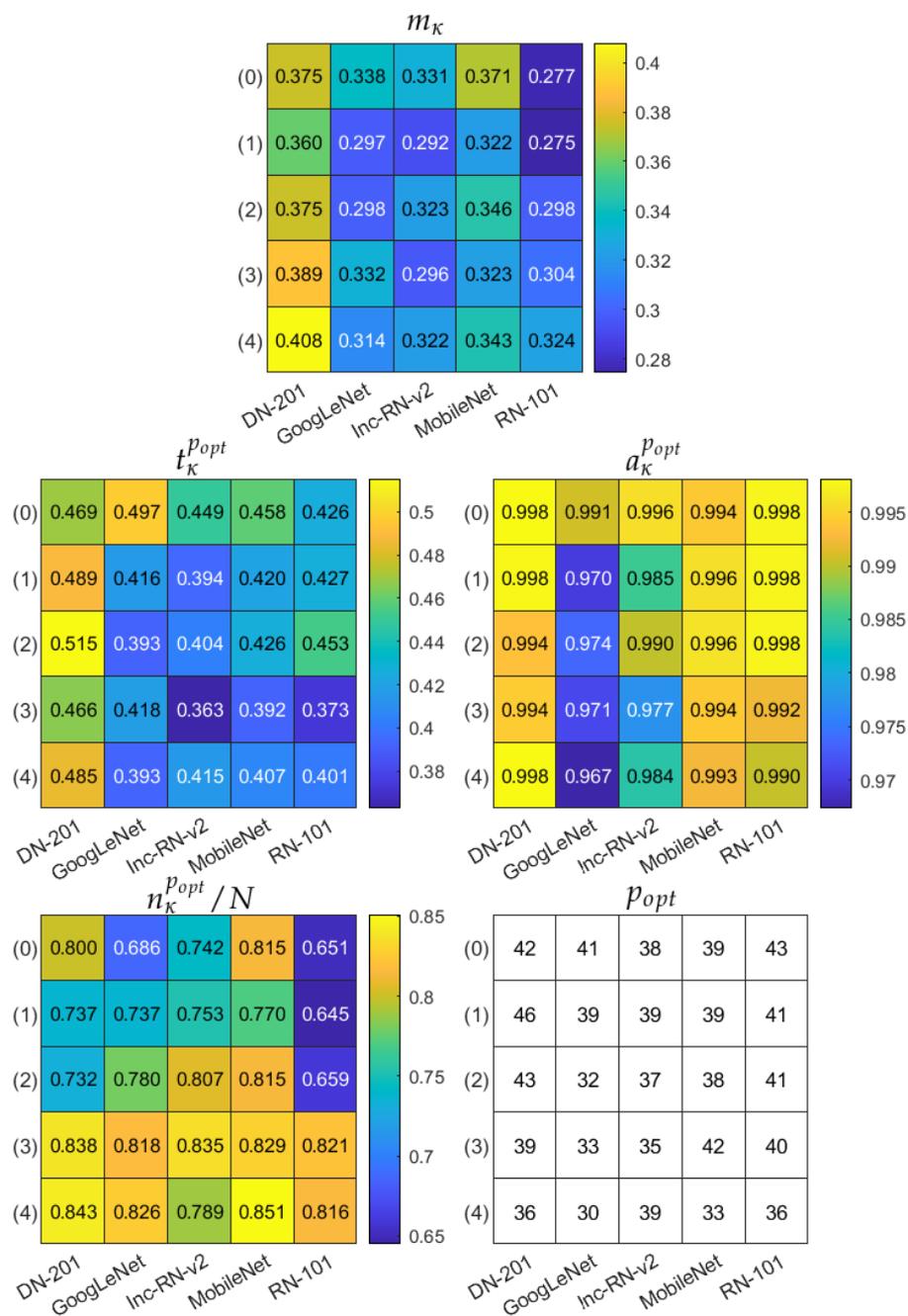
### 3.4. Multicriteria Model Evaluation

Finally, we apply the multicriteria model evaluation measure  $m_\kappa$  to the trained models. The results are gathered in Figures 11 and 12, corresponding to two augmentation modes Aug0 and Aug1, respectively. In both figures, the top table refers to the  $m_\kappa$  measure, and the remaining tables present the corresponding partial accuracy  $a_\kappa^{p_{opt}}$ , data subset size  $n_\kappa^{p_{opt}}/N$ , and classification confidence level  $t_\kappa^{p_{opt}}$ , all obtained for  $p_{opt} = \arg \max_p (m_\kappa^p)$ . The bottom right table shows the  $p_{opt}$  itself. The multicriteria model evaluation indicates DenseNet-201,

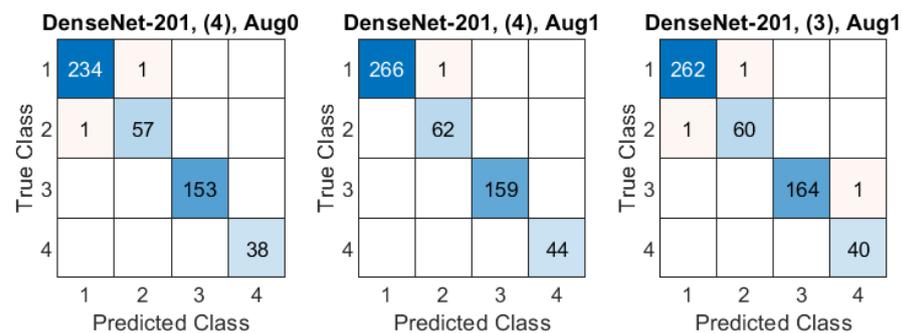
with ROI in mode (4) and Aug0 data augmentation as the most robust model, reaching 0.996 accuracy over 76.7% of data, for which the CCL is not less than 0.544. Confusion matrices for three best models in terms of  $m_\kappa$  are shown in Figure 13 over the subset limited by  $p_{opt}$ : DenseNet-201, ROI mode (4), Aug0; DenseNet-201, ROI mode (4), Aug1, and DenseNet-201, ROI mode (3), Aug1.



**Figure 11.** Multicriteria model evaluation summary in data augmentation option Aug0. Top table:  $m_\kappa$  measure. Remaining tables present components of  $m_\kappa$  obtained for  $p_{opt} = \arg \max_p (m_\kappa^p)$  and  $p_{opt}$  itself. In each table, columns correspond to different deep models, rows refer to ROI extraction modes (described in the text).



**Figure 12.** Multicriteria model evaluation summary in data augmentation option Aug1. Top table:  $m_\kappa$  measure. Remaining tables present components of  $m_\kappa$  obtained for  $p_{opt} = \arg \max_p (m_\kappa^p)$  and  $p_{opt}$  itself. In each table, columns correspond to different deep models, rows refer to ROI extraction modes (described in the text).



**Figure 13.** Confusion matrices in three best-case classification models from Figures 11 and 12 over a limited dataset with respect to the  $m_k$  value. Class IDs: 1–AD, 2–psoriasis, 3–skin tumor, 4–control group. Models are specified above each matrix in a format: model name, ROI extraction mode, data augmentation mode.

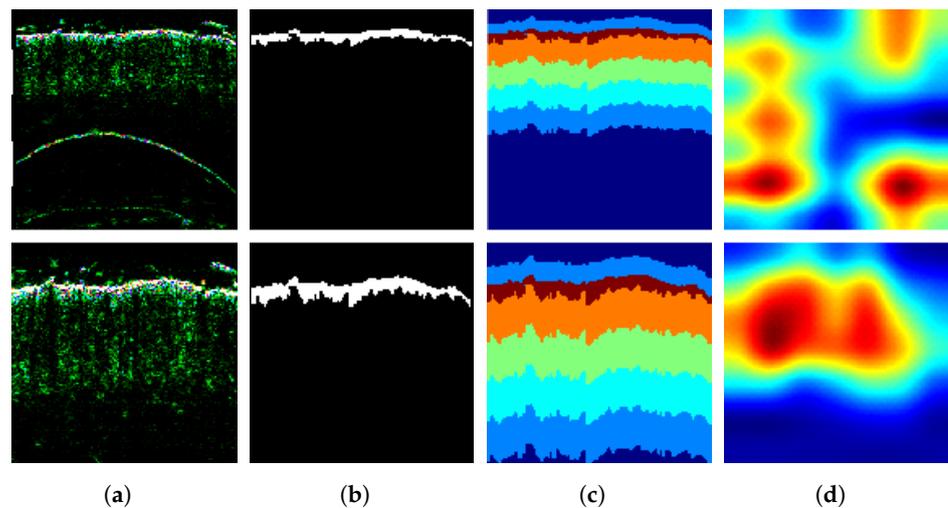
#### 4. Discussion

The methodology presented in this study mainly focuses on the classification of normative and pathological skin in HFUS images; yet, in the first step, we developed a method for the robust segmentation of the epidermal layer. The results were used for two purposes: to extract the meaningful ROI for classification and to enable reliable validation of classification confidence.

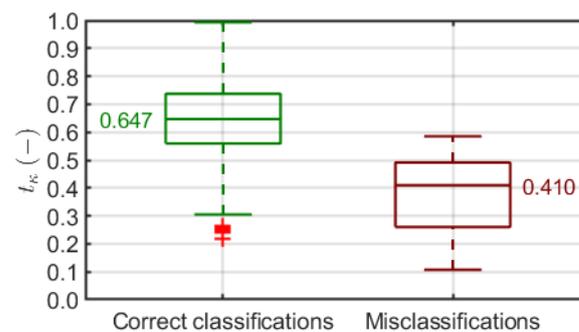
We prepared a deep model through transfer learning for segmentation, mainly due to the limited dataset size. With various deep semantic segmentation architectures under consideration, the experiments favored the DeepLab v3+ model with a pre-trained Xception backbone. The selection is justified by the results of Dice-index-based validation given in Figure 7. Since the Dice index is a reference measure in medical image segmentation, we were able to compare our results to rare previous research in this field. The transfer learning approach turned out to be more efficient than the best method presented in [7]. Thus, we find the transfer learning promising for possible future research on the segmentation of either epidermis or other skin layers.

Despite a generally good accuracy (over 0.93), the classification assessment over the entire dataset indicated some differences between models and ROI extraction or data augmentation modes (Figure 8), consistently favoring the DenseNet-201 ahead of the MobileNet. Most frequent misclassifications were related to AD and psoriasis cases—see examples in Figure 9. To ensure that the models focus on the meaningful regions of the image, we took a closer look at the classification process using the Grad-CAM tool. As a result, we could notice multiple cases where the networks focus on the gel layer or other parts instead of the skin. An example can be seen in Figure 14, where the same image was analyzed without (top) and with the ROI extraction (bottom). The Grad-CAM heatmap in the latter seems to cover the skin layers more adequately.

To address this matter, we introduced a classification confidence level measure to reward classifications where the Grad-CAM heatmap matches the skin region estimated layer-wise based on the epidermis segmentation outcome. The mean CCL distribution over all models and setups (Figure 10) did not necessarily correspond to the classification accuracy distribution (Figure 8). On the other hand, we noticed a significant correlation between the CCL value and classifier decision in individual cases. Figure 15 shows the distributions of  $t_k$  in correct and incorrect classification groups produced by one of the models. Low  $t_k$  is clearly more likely to lead to a wrong classification, whereas a high enough  $t_k$  guarantees the right decision. We observed such a relationship consistently in different models, and we find it a reliable justification for the use of Grad-CAM-based classification confidence assessment.



**Figure 14.** Illustration of Grad-CAM maps (d) obtained for the original image (top) and the extracted ROI (bottom) (a), with the corresponding epidermis segmentation results (b), and skin layer maps (c).



**Figure 15.** Distribution of classification confidence level  $t_k$  in correct and incorrect classification groups. Each box covers 25th to 75th percentile confidence interval with median indicated by a central line. Whiskers refer to 1.5 times the interquartile range. Outliers are indicated with red +.

Moreover, we extended our model evaluation methodology by introducing in Section 2.3.4 an additional multicriteria measure  $m_k$  that follows the increasing CCLs produced by the model over the test dataset. As we know from Figure 15, dropping off low-CCL cases should result in increasing accuracy over a limited dataset since the misclassifications should be excluded early. Our measure indicates the model that maximizes Equation (6), which balances classification accuracy, confidence, and dataset size. The multicriteria evaluation points out the DenseNet-201 model with the Aug0 data augmentation, fed by the extracted ROI in the (4) option (1 mm from the top, 2 mm dermis) with the top  $m_k$  value of 0.415. In this case, the classification accuracy  $a_k^{opt}$  reaches 0.996 with  $a_k^{opt} = 0.544$  over 76.7% of the dataset (Figure 11). Note that this model's accuracy over the entire dataset was 0.975 with a mean CCL of 0.634 (Figures 8 and 10).

The multicriteria model evaluation strategy attempts to handle outlier cases without manually pointing them out or removing them before the analysis. The latter would require defining additional image quality constraints, and so forth. Moreover, there is not much room to reduce datasets in deep learning: the smaller the dataset, the more questionable robustness of the model, which contradicts the original goal. In our approach, the number and character of outlier cases are automatically reflected after being processed by the model, enabling its quantitative assessment. We believe that such an idea can be widely used to design more reliable deep models not only for CAD applications.

Our results prove that the map of skin layers based on the initial epidermis segmentation combined with Grad-CAM can be used to evaluate the skin classification in HFUS. However, the individual layers' thickness may be affected by other factors, for example, the measurement location on the patient's body, gender, or body mass index. Hence,

a personalized skin map based on a medical interview or automated segmentation of all skin layers (which is rather challenging for now) is likely to improve the analysis.

## 5. Conclusions

This paper presents a novel CNN-based framework for the HFUS image classification, with a multicriteria model evaluation. It enables both classifying the HFUS data into four groups and evaluating the obtained results with a Grad-CAM-based tool. It is the first method targeting HFUS image classification combining inflammatory skin diseases and skin tumors and the first CNN application for this task. The obtained results prove that such an analysis is accurate and reliable, whereas the additional introduced procedures: epidermis segmentation, followed by skin layer map estimation and multicriteria model evaluation, enable the reliable validation of the method.

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

**Funding:** This research was funded partially by the Polish Ministry of Science and Silesian University of Technology statutory financial support No. 07/010/BK\_21/1006 (BK-296/RIB1/2021) and partially by the Silesian University of Technology grant No. 07/010/RGH20/1004.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the Military Institute of Medicine, Warsaw, Poland, No. No. 21/WIM/2021 from 19 May 2021.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

- Liu, S.; Wang, Y.; Yang, X.; Lei, B.; Liu, L.; Li, S.X.; Ni, D.; Wang, T. Deep Learning in Medical Ultrasound Analysis: A Review. *Engineering* **2019**, *5*, 261–275. [\[CrossRef\]](#)
- Levy, J.; Barrett, D.L.; Harris, N.; Jeong, J.J.; Yang, X.; Chen, S.C. High-frequency ultrasound in clinical dermatology: A review. *Ultrasound J.* **2021**, *13*. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bhatta, A.; Keyal, U.; Liu, Y. Application of high frequency ultrasound in dermatology. *Discov. Med.* **2018**, *26*, 237–242. [\[PubMed\]](#)
- Fei, C.; Chiu, C.T.; Chen, X.; Chen, Z.; Ma, J.; Zhu, B.; Shung, K.K.; Zhou, Q. Ultrahigh Frequency (100 MHz–300 MHz) Ultrasonic Transducers for Optical Resolution Medical Imagining. *Sci. Rep.* **2016**, *6*. [\[CrossRef\]](#) [\[PubMed\]](#)
- de Oliveira Barcaui, E.; Carvalho, A.C.P.; Lopes, F.P.P.L.; Piñeiro-Maceira, J.; Barcaui, C.B. High frequency ultrasound with color Doppler in dermatology. *An. Bras. Dermatol.* **2016**, *91*, 262–273. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bezugly, A.; Sedova, T.; Belkov, P.; Enikeev, D.; Voloshin, R. Nevus sebaceus of Jadassohn—High frequency ultrasound imaging and videodermoscopy examination. Case presentation. *Med. Pharm. Rep.* **2020**, *94*, 112–117. [\[CrossRef\]](#)
- Czajkowska, J.; Badura, P.; Korzekwa, S.; Płatkowska-Szczerek, A. Deep learning approach to skin layers segmentation in inflammatory dermatoses. *Ultrasonics* **2021**, *114*, 106412. [\[CrossRef\]](#) [\[PubMed\]](#)
- Heibel, H.D.; Hooley, L.; Cockerell, C.J. A Review of Noninvasive Techniques for Skin Cancer Detection in Dermatology. *Am. J. Clin. Dermatol.* **2020**, *21*, 513–524. [\[CrossRef\]](#)
- Kleinerman, R.; Whang, T.B.; Bard, R.L.; Marmur, E.S. Ultrasound in dermatology: Principles and applications. *J. Am. Acad. Dermatol.* **2012**, *67*, 478–487. [\[CrossRef\]](#)
- Polańska, A.; Silny, W.; Jenerowicz, D.; Kniola, K.; Molińska-Glura, M.; Dańczak-Pazdrowska, A. Monitoring of therapy in atopic dermatitis—Observations with the use of high-frequency ultrasonography. *Skin Res. Technol.* **2015**, *21*, 35–40. [\[CrossRef\]](#)
- Chen, C.; Cheng, Y.; Zhu, X.; Cai, Y.; Xue, Y.; Kong, N.; Yu, Y.; Xuan, D.; Zheng, S.; Yang, X.; et al. Ultrasound assessment of skin thickness and stiffness: The correlation with histology and clinical score in systemic sclerosis. *Arthritis Res. Ther.* **2020**, *22*. [\[CrossRef\]](#)
- Pereyra, M.; Dobigeon, N.; Batatia, H.; Tourneret, J. Segmentation of Skin Lesions in 2-D and 3-D Ultrasound Images Using a Spatially Coherent Generalized Rayleigh Mixture Model. *IEEE Trans. Med. Imaging* **2012**, *31*, 1509–1520. [\[CrossRef\]](#) [\[PubMed\]](#)

13. Gao, Y.; Tannenbaum, A.; Chen, H.; Torres, M.; Yoshida, E.; Yang, X.; Wang, Y.; Curran, W.; Liu, T. Automated Skin Segmentation in Ultrasonic Evaluation of Skin Toxicity in Breast Cancer Radiotherapy. *Ultrasound Med. Biol.* **2013**, *39*, 2166–2175. [CrossRef]
14. Sciolla, B.; Delachartre, P.; Cowell, L.; Dambry, T.; Guibert, B. Improved boundary segmentation of skin lesions in high-frequency 3D ultrasound. *Comput. Biol. Med.* **2017**, *87*, 302–310. [CrossRef]
15. Czajkowska, J.; Korzekwa, S.; Pietka, E. Computer Aided Diagnosis of Atopic Dermatitis. *Comput. Med. Imaging Graph.* **2020**, *79*, 101676. [CrossRef]
16. Marosán, P.; Szalai, K.; Csabai, D.; Csány, G.; Horváth, A.; Gyöngy, M. Automated seeding for ultrasound skin lesion segmentation. *Ultrasonics* **2021**, *110*, 106268. [CrossRef]
17. Hesamian, M.H.; Jia, W.; He, X.; Kennedy, P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J. Digit. Imaging* **2019**, *32*, 582–596. [CrossRef] [PubMed]
18. Roy, A.G.; Conjeti, S.; Karri, S.P.K.; Sheet, D.; Katouzian, A.; Wachinger, C.; Navab, N. ReLayNet: Retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed. Opt. Express* **2017**, *8*, 3627–3642. [CrossRef]
19. Kepp, T.; Droigk, C.; Casper, M.; Evers, M.; Hüttmann, G.; Salma, N.; Manstein, D.; Heinrich, M.P.; Handels, H. Segmentation of mouse skin layers in optical coherence tomography image data using deep convolutional neural networks. *Biomed. Opt. Express* **2019**, *10*, 3484. [CrossRef] [PubMed]
20. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
21. Gu, Z.; Cheng, J.; Fu, H.; Zhou, K.; Hao, H.; Zhao, Y.; Zhang, T.; Gao, S.; Liu, J. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–2292. [CrossRef]
22. del Amor, R.; Morales, S.; Colomer, A.; Mogensen, M.; Jensen, M.; Israelsen, N.M.; Bang, O.; Naranjo, V. Automatic Segmentation of Epidermis and Hair Follicles in Optical Coherence Tomography Images of Normal Skin by Convolutional Neural Networks. *Front. Med.* **2020**, *7*. [CrossRef] [PubMed]
23. Huang, Q.; Zhang, F.; Li, X. Machine Learning in Ultrasound Computer-Aided Diagnostic Systems: A Survey. *BioMed Res. Int.* **2018**, *2018*, 5137904. [CrossRef]
24. Cai, L.; Gao, J.; Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* **2020**, *8*, 713. [CrossRef]
25. Han, S.; Kang, H.K.; Jeong, J.Y.; Park, M.H.; Kim, W.; Bang, W.C.; Seong, Y.K. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys. Med. Biol.* **2017**, *62*, 7714–7728. [CrossRef]
26. Chi, J.; Walia, E.; Babyn, P.; Wang, J.; Groot, G.; Eramian, M. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. *J. Digit. Imaging* **2017**, *30*, 477–486. [CrossRef]
27. Meng, D.; Zhang, L.; Cao, G.; Cao, W.; Zhang, G.; Hu, B. Liver fibrosis classification based on transfer learning and FCNet for ultrasound images. *IEEE Access* **2017**, *5*, 5804–5810. [CrossRef]
28. Burgos-Artizzu, X.P.; Coronado-Gutiérrez, D.; Valenzuela-Alcaraz, B.; Bonet-Carne, E.; Eixarch, E.; Crispi, F.; Gratacós, E. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. *Sci. Rep.* **2020**, *10*. [CrossRef] [PubMed]
29. Karimi, D.; Warfield, S.K.; Gholipour, A. Critical Assessment of Transfer Learning for Medical Image Segmentation with Fully Convolutional Neural Networks. 2020. Available online: <https://arxiv.org/abs/2006.00356> (accessed on 30 May 2020).
30. van Opbroek, A.; Ikram, M.A.; Vernooij, M.W.; de Bruijne, M. Transfer Learning Improves Supervised Image Segmentation Across Imaging Protocols. *IEEE Trans. Med. Imaging* **2015**, *34*, 1018–1030. [CrossRef] [PubMed]
31. Morid, M.A.; Borjali, A.; Del Fiol, G. A scoping review of transfer learning research on medical image analysis using ImageNet. *Comput. Biol. Med.* **2021**, *128*, 104115. [CrossRef]
32. ImageNet. 2021. Available online: <http://www.image-net.org> (accessed on 8 April 2021).
33. Ayana, G.; Dese, K.; woon Choe, S. Transfer Learning in Breast Cancer Diagnoses via Ultrasound Imaging. *Cancers* **2021**, *13*, 738. [CrossRef]
34. Kim, I.; Rajaraman, S.; Antani, S. Visual Interpretation of Convolutional Neural Network Predictions in Classifying Medical Image Modalities. *Diagnostics* **2019**, *9*, 38. [CrossRef] [PubMed]
35. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626. [CrossRef]
36. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. [CrossRef]
37. Czajkowska, J.; Badura, P.; Płatkowska-Szczerek, A.; Korzekwa, S. Data for: Deep Learning Approach to Skin Layers Segmentation in Inflammatory Dermatoses. *Ultrasonics* **2021**, *114*, 106412. [CrossRef]
38. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. 2018. Available online: <https://arxiv.org/abs/1608.06993> (accessed on 28 January 2018).

39. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. 2014. Available online: <https://arxiv.org/abs/1409.4842> (accessed on 17 September 2014).
40. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2019. Available online: <https://arxiv.org/abs/1801.04381> (accessed on 21 March 2019).
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
42. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. 2017. Available online: <https://arxiv.org/abs/1706.05587> (accessed on 5 December 2017).
43. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. 2018. Available online: <https://arxiv.org/abs/1802.02611> (accessed on 22 August 2018).
44. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807. [[CrossRef](#)]
45. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; L., F.F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
46. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer International Publishing: Basel, Switzerland, 2017; pp. 240–248. [[CrossRef](#)]
47. Laurent, A.; Mistretta, F.; Bottiglioli, D.; Dahel, K.; Goujon, C.; Nicolas, J.F.; Hennino, A.; Laurent, P.E. Echographic measurement of skin thickness in adults by high frequency ultrasound to assess the appropriate microneedle length for intradermal delivery of vaccines. *Vaccine* **2007**, *25*, 6423–6430. [[CrossRef](#)]
48. Oltulu, P.; Ince, B.; Kokbudak, N.; Findik, S.; Kilinc, F. Measurement of epidermis, dermis, and total skin thicknesses from six different body regions with a new ethical histometric technique. *Turk. J. Plast. Surg.* **2018**, *26*, 56. [[CrossRef](#)]
49. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [[CrossRef](#)]
50. Sciolla, B.; Digabel, J.L.; Josse, G.; Dambry, T.; Guibert, B.; Delachartre, P. Joint segmentation and characterization of the dermis in 50 MHz ultrasound 2D and 3D images of the skin. *Comput. Biol. Med.* **2018**, *103*, 277–286. [[CrossRef](#)]