Hybrid Ensemble Learning with Wavelet Convolutional Model and EfficientNet for Accurate Skin Cancer Classification

Hind Fekak[1,] Nihad Essaadi[2], Oumaima Oukmour[2]
ENSA, University Hassan First, Berrechid, Morocco

Abstract— This study proposes a novel hybrid architecture that combines wavelet-based techniques with a convolutional neural network (CNN) featuring residual blocks and EfficientNet to achieve accurate skin cancer classification. The model is designed to diagnose and differentiate among similar skin diseases by capturing contextual information at various spatial scales. The integration of features extracted by convolutional blocks with wavelet coefficients obtained from wavelet decomposition enhances the model's capability to detect patterns and structures of different sizes in skin images. The study focuses on analyzing the effect of wavelet decomposition on skin disease images, developing an efficient CNN block based on dilated convolutions and channel concatenation, and conceiving a hybrid model that synergizes wavelets and CNN blocks.

Keywords— hybrid architecture, classification, deep learning, convolutional neural networks

1 Introduction

The proposed hybrid architecture strategically combines wavelet decomposition and convolutional neural networks (CNNs), particularly based on residual blocks and EfficientNet, to enhance the accuracy of skin cancer classification. By leveraging wavelet decomposition, the model captures contextual information at varying spatial scales, enabling the detection of patterns and structures of diverse sizes within skin images. This approach facilitates a multi-resolution specification of the images, achieved by seamlessly integrating wavelet coefficients with features extracted by convolutional blocks. The fusion of local details captured by wavelets and global contextual information from convolutional blocks empowers the model to obtain a comprehensive view of skin images. The primary objectives encompass analyzing the impact of wavelet decomposition, developing an efficient CNN block, and innovating a hybrid model that synergizes wavelets and CNNs for improved dermatological image classification. This holistic strategy aims to address the complexities of skin diseases by combining the strengths of wavelet-based techniques and convolutional networks, resulting in a more nuanced and accurate classification system.

2 Related work

In the preceding research, the model was crafted to comprehend contextual details across varied spatial scales, discerning patterns and structures of different sizes in skin images. To enhance its efficacy further, forthcoming enhancements may explore advanced attention mechanisms for an improved grasp of contextual information. Moreover, the optimization of utilizing wavelet coefficients alongside features from convolutional blocks could be achieved through innovative fusion techniques. The incorporation of state-of-the-art CNN architectures like EfficientNet has the potential to broaden the model's capabilities. Essential for transparent decision-making, the inclusion of interpretability methods is imperative. Expanding validation to encompass larger, more diverse datasets and conducting clinical trials involving medical professionals will augment the practical applicability and credibility of the model, positioning it as a more valuable tool in dermatological imaging and diagnosis.

Table 1. References of skin disease classification with deep learning

Inf. Time	Model	Accuracy	Précision	Recall	F1-score
40 ms/st	Naïve Wavelet-CNN model	86%	85%	86%	84%
76 ms/st	Wavelet Residual-Block model and attention	92.5%	93%	92%	91%
125 ms/st	Efficient network with noisy student weights	93.9%	94%	93%	93%
154 ms/st	Ensemble with average prediction	94.3%	95%	94%	94%

3 Dataset

Our dataset is focused on dermatology and aims to support the classification of skin lesions. Each entry in the dataset is uniquely identified by a combination of "lesion_id" and "image_id." The "dx" column categorizes the skin lesion diagnosis. The "dx_type" field indicates the diagnostic method employed. Additional information includes the age of the individuals with skin lesions, their gender (captured in the "sex" column), and the anatomical location of the skin lesions, specified by the "localization" column (e.g., "scalp," "ear").

This dataset is evidently designed for machine learning applications, providing comprehensive features for training models to accurately classify skin lesions based on diverse attributes, including age, gender, and lesion characteristics.

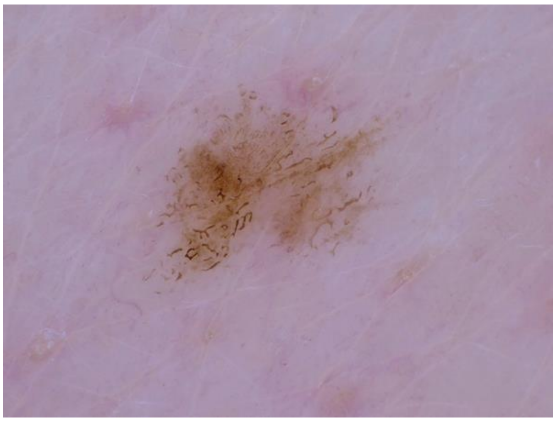


Fig. 1. Random samples of skin lesions from our dataset

our dataset contains different images of skin diseases, in JPEG extension, several images with lower quality were deleted in the checking of the dataset. After quality control step, the dataset was divided into 80% of the samples used for training and 20% of the samples are used for the validation of the network.

	lesion_id	image_id	dx	dx_type	age	sex	localization
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear

Fig. 2. Examples of our dataset

Table 2 shows the number of images in each of the 7 classes, each category name represents a class of skin disease, and each label is associated with its category.

Table 2. Labels of the 7 skin diseases class

Label	Class Name
0	akiec
1	bcc
2	bkl

3	df
4	mel
5	ny
6	yasc

4 Data augmentation

In deep learning models, the quantity and diversity of training data play a crucial role in enhancing performance. However, in scenarios where access to a large dataset is limited, image augmentation serves as a valuable technique to artificially expand the dataset. This involves applying various transformations to the existing images, such as rotation, flipping, and brightness adjustments. The augmentation values used are outlined in Table 3.

Table 3. Values used in data augmentation

Process	Value
Rescale	1./255
Random Contrast	Probability: 10%
Random Saturation	Probability: 10%
Random Crop	Probability: 10%
Gaussian Noise	Probability: 10%
Cutout	Probability: 10%

To implement this, the Keras ImageDataGenerator API was utilized. The custom_augmentation function employs a set of transformations, including random contrast adjustment, random saturation adjustment, random cropping within specified ranges, adding Gaussian noise, and applying cutout techniques. These transformations are randomly applied based on predefined probabilities to generate diverse training images and improve the model's robustness. The integration of these augmentation techniques contributes to the model's ability to generalize well and accurately classify skin disease images.

5 workflow

The workflow of the proposed hybrid architecture unfolds in a systematic manner, utilizing a multi-stage process to optimize the model for accurate skin cancer classification. Commencing with four levels of wavelet-decomposed inputs, each level undergoes independent processing through dedicated residual blocks. This staged decomposition ensures that the model captures intricate details at multiple scales. The extracted features from these wavelet decomposition levels are then synergistically combined with features obtained from the residual blocks. This fusion is a pivotal step in enhancing the model's capability to capture diverse and complementary information, ultimately contributing to a more holistic understanding of skin images.

The architecture further incorporates fully connected layers, integrating activations, batch normalization, and dropout mechanisms for regularization. These elements are crucial in fine-tuning the model's performance, preventing overfitting, and improving generalization. The final output layer is designed for multi-class classification, categorizing skin disease classes based on the amalgamated features from both the wavelet and residual block pathways.

Noteworthy features of the workflow include the incorporation of wavelet-based preprocessing techniques, which set the stage for effective decomposition and subsequent processing. Attention modules are strategically introduced, allowing the model to focus on informative or discriminating elements of the input while disregarding less relevant information. Additionally, the adoption of Swish activation functions adds a layer of enhanced non-linearity to the model, further enriching its capacity to capture intricate patterns and structures in dermatological images. Overall, the workflow is meticulously crafted to leverage the synergies between wavelet decomposition and convolutional blocks, ensuring an optimized and comprehensive approach to skin cancer classification.

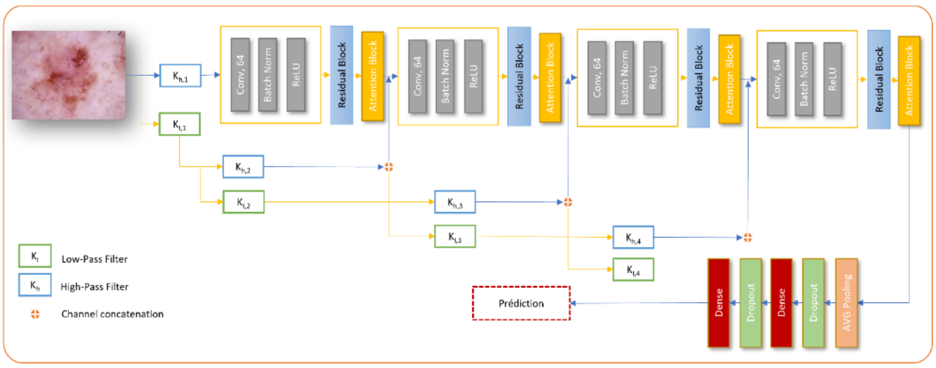


Fig. 3. Wavelet Attention-based CNN model.

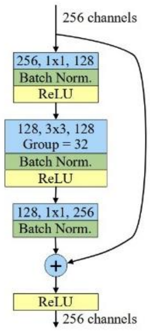
5.1 Wavelets

In the intricate realm of dermatological image analysis, this study harnesses the potent capabilities of wavelet transforms to unravel the complexities inherent in skin images. Wavelets, acting as sophisticated mathematical tools, take center stage by orchestrating the decomposition of images into multiple scales. This orchestrated decomposition is not merely a technical maneuver but a strategic approach to provide a hierarchical representation of image details. The multi-scale lens that wavelets offer becomes instrumental in capturing nuanced information at varying levels of granularity, facilitating a more comprehensive understanding of dermatological images. Beyond mere decomposition, wavelets play a pivotal role in elucidating the textural and structural properties embedded in these images. Skin texture, intricate patterns, subtle contours, and unique features associated with specific skin disorders are meticulously examined through the lens of wavelet analysis. The study recognizes and capitalizes on wavelets' knack for identifying and characterizing diverse textures, patterns, and contours, attributes critical for accurate diagnosis. Moreover, wavelets unfold their prowess in highlighting unique properties linked to specific skin conditions, contributing to the model's ability to discern between different diseases. One of the remarkable facets of wavelet employment lies in its efficacy in data dimensionality reduction. The coefficients derived through wavelet decomposition serve as compact descriptors, effectively mitigating the challenges posed by vast dermatological image databases. This reduction not only streamlines processing tasks but also enhances the efficiency of analyzing extensive datasets—a pivotal aspect in the realm of medical image research. In essence, wavelets transcend their role as mere decomposition tools; they emerge as comprehensive instruments for unraveling intricate details, textures, and distinctive features in dermatological images, enriching the proposed hybrid model's capacity for accurate skin cancer classification.

5.2 Residual Blocks

Critical to the advancement of dermatological image classification, residual blocks are instrumental in overcoming challenges inherent in training deep neural networks on complex skin images. These blocks enable the learning of residual mappings, simplifying the training of deeper architectures by focusing on the difference between input and expected output. Introducing short connections or skip connections within residual blocks serves as a key strategy, creating efficient pathways for gradient flow and mitigating the vanishing gradient problem. This problem is particularly pronounced in deep architectures, where gradients can diminish as they traverse multiple layers, hindering effective weight updates. Residual blocks, by maintaining a clear gradient flow, ensure that information can easily propagate through the network during backpropagation. Leveraging these blocks results in improved feature learning, allowing the extraction of hierarchical features that capture intricate patterns and structures present in dermatological images. The enhanced generalization achieved through residual blocks enables models to perform effectively on diverse and previously unseen images, contributing to higher accuracy in identifying various skin conditions. In summary, the incorporation of residual blocks in dermatological image classification significantly improves the efficiency of deep architectures, addressing key issues and advancing the capabilities of models for accurate and reliable skin disease identification.

Fig. 4. Residual Block



5.3 EfficientNEt

Contrastingly, the EfficientNet model, acknowledged for its outstanding performance across diverse tasks, serves as a sturdy foundation due to its impressive classification capabilities. A crucial element in this procedure entails initiating the EfficientNet model by incorporating weights obtained through the noisy student methodology. This pioneering technique is rooted in the noisy student pre-training method, where a student model undergoes training on augmented data and subsequently generates pseudo-labels for previously unseen training data. The iterative nature of this process, involving both teacher and student models, significantly contributes to refining the model's overall performance and enhancing its adaptability to the specific target task.

The noisy student pre-training method proves particularly advantageous as it introduces a form of self-training, allowing the model to learn from its own predictions. This approach empowers the model to become more resilient and versatile in addressing variations and complexities within the data. It is noteworthy that the effectiveness of the noisy student pre-training technique has been demonstrated in enhancing the performance of deep learning models across various domains, establishing it as a potent tool in model initialization and fine-tuning. By seamlessly integrating these nuanced details into the architecture, the EfficientNet model emerges as a dependable cornerstone, proficient in addressing challenges related to dermatological image classification with superior accuracy and adaptability.

6 Methodologies

The proposed hybrid architecture seamlessly integrates ensemble learning principles, combining the sophisticated capabilities of a wavelet-based network and the robust EfficientNet model, to establish a powerful and nuanced framework for accurate skin cancer classification. Ensemble learning, a cornerstone of this methodology, excels in amalgamating diverse models to elevate predictive performance and overall system resilience. Fig. 2 elucidates the intricate classification process, showcasing the innovative ensemble learning architecture that synergizes the unique strengths of the wavelet network and EfficientNet models while strategically mitigating their respective limitations. The wavelet network contributes its adeptness in capturing intricate patterns and nuanced features, thanks to its transformative capabilities through wavelet-based processing. On the other hand, the EfficientNet model, recognized for its exceptional performance, provides a robust foundation with strong classification abilities. The pivotal step of initializing the EfficientNet model with noisy student weights underscores adaptive pre-training, employing an iterative process involving teacher and student models for improved adaptability to the target task. To optimize training dynamics, a warmup phase gradually increases the learning rate, followed by a cosine learning rate scheduler, contributing to effective convergence. Learning rate adjustments dynamically fine-tune the model's adaptability, and the incorporation of batch normalization and regularization layers stabilizes learning, reducing undesirable effects such as over-fitting. The model's sophisticated understanding of neural network training strategies, epitomized by the use of callbacks, strikes a balance between exploration and exploitation in the learning process. Overall, this ensemble learning model represents a promising solution for accurate skin cancer classification, capitalizing on the synergistic potential of distinct neural network architectures and paving the way for advancements in medical image analysis.

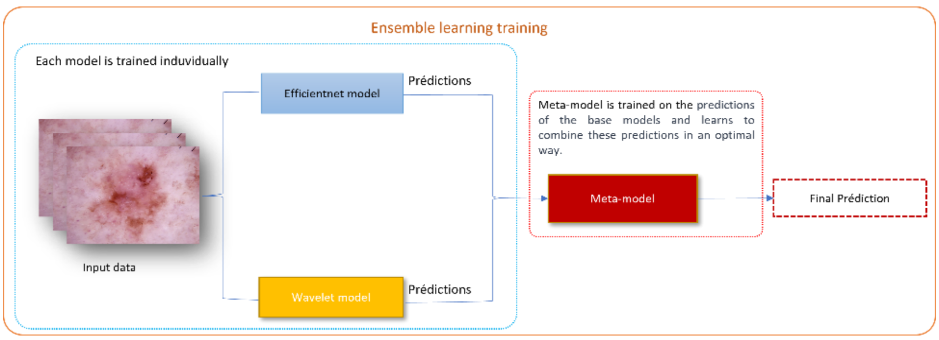


Fig. 5. General workflow of the proposed model.

6.1 Performance measures

Accuracy and error rate measures are very important to evaluate a model, however they may be deceptive in some situation as these are data dependent. In case of imbalanced dataset where the number of majority samples is too large compared to the minority samples, the classifier gets biased towards the majority samples. We can see in such cases that we have a high accuracy value and very low error rate. These results project as if the classifier is an ideal one which is not the real scenario. Indeed, even if these two metrics indicate that the classifier is efficient, the minority classes are not well classified. There are several alternatives to overcome such imbalanced dataset scenarios, we can rely on other metrics which may prove more useful and give more accurate insight into the performance of the classifier. These metrics are precision, recall and AUC. These measures are defined as follows:

Accuracy (ACC) quantifies the proportion of the labels that were correctly classified divided by all predictions that were made on the test set, which is formally expressed as Equation 1:

$$Accuracy = \frac{(TN+TP)}{(TN+FP+FN+TP)}$$

Precision also known as or positive predictive value quantifies the proportion of the labels properly classified that are truly positive, as represented formally in Equation 2:

$$Precision = \frac{TP}{(FP+TP)}$$

Recall quantifies the proportion of misclassified labels that are truly positive, as represented formally in Equation 3:

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

7 Results and discussion

7.1 Wavelets

In our pursuit of elevating model accuracy, a strategic initiative was undertaken to fine-tune the wavelet model's configuration. Meticulous modifications to parameters aimed at optimizing performance, including a refined batch size of 4 and an increased image size of 384 pixels, both contributing to enhanced training dynamics. The learning rate parameters were meticulously tuned, starting at 0.000008, reaching a maximum of 0.0000325, and a minimum of 0.000001. To facilitate smooth learning rate transitions, a warmup phase of 5 epochs was implemented, followed by a cosine learning rate scheduler with a decay factor of 0.6 for effective convergence. Data augmentation techniques, such as rotation, shearing, zooming, and shifting, were applied with a transform probability of 1.0, enriching the model's ability to generalize. These modifications, alongside parameters like label smoothing and 20 test-time augmentation (TTA) steps, collectively resulted in an optimized wavelet model configuration. This comprehensive approach underscores our commitment to pushing the boundaries of accuracy in dermatological image classification.

<table>
<tr><th>Time</th><th>#</th><th>Log Message</th></tr>
<tr><td>36546.55</td><td>1</td><td>Epoch 200: val_accuracy did not improve from 0.92004</td></tr>
<tr><td>36546.55</td><td>2</td><td>
9643/9643 [==============================] - 177s 18ms/step - loss:
accuracy: 0.9949 - val_loss: 0.5314 - val_accuracy: 0.8987 - lr: 1.0</td></tr>
<tr><td>36928.16</td><td>3</td><td>/opt/conda/lib/python3.10/site-packages/traitlets/traitlets.py:2930: FutureWarning: --Exporter.preprocessors=["remove_papermill_header.RemovePapermillHeader"] for containers is deprecated in traitlets 5.0. You can pass `--Exporter.preprocessors item` ... multiple times to add items to a list.</td></tr>
<tr><td>36928.16</td><td>4</td><td>warn(</td></tr>
<tr><td>36928.17</td><td>5</td><td>[NbConvertApp] WARNING | Config option `kernel_spec_manager_class` not recognized by `NbConvertApp`.</td></tr>
<tr><td>36928.19</td><td>6</td><td>[NbConvertApp] Converting notebook __notebook__.ipynb to notebook</td></tr>
</table>

7.2 EfficientNet

In the pursuit of refining the EfficientNet model, several strategic modifications were introduced. The model was compiled using the 'adam' optimizer with 'categorical_crossentropy' as the loss function and 'accuracy' as the evaluation metric. A learning rate callback was implemented, dynamically adjusting the learning rate during training based on a specified schedule. This scheduler ensures efficient convergence by carefully controlling the learning rate throughout the training process, starting at 0.000008, reaching a maximum of 0.0000325, a minimum of 0.000001, a ramp-up period of 5 epochs, and a decay rate of 0.6. A ModelCheckpoint was employed to save the best weights based on validation accuracy, enhancing the model's ability to generalize and improve overall performance. These thoughtful adjustments collectively contribute to the optimized training dynamics of the EfficientNet model, resulting in enhanced accuracy and robustness.

8 Conclusion

In conclusion, this study introduces a groundbreaking hybrid ensemble model, fusing wavelet-based techniques with the EfficientNet architecture, to address the intricate challenges of accurate skin cancer classification. Through a synergistic integration of diverse neural network components, including attention modules and wavelet decomposition, the proposed model exhibits a remarkable capacity to navigate the complexity inherent in dermatological images. Its proficiency in capturing complex patterns and structures is amplified by the multi-resolution perspective provided by wavelet decomposition, leading to enhanced pattern recognition. The model's ability to leverage both local and global context information significantly improves overall classification accuracy for different skin diseases. Valuable insights are offered into the impact of wavelet decomposition on skin disease identification, emphasizing the instrumental role of wavelets in capturing textural and structural properties. The integration of residual blocks plays a pivotal role in addressing challenges associated with dermatological image classification, facilitating gradient flow and enhancing feature learning. The success of this hybrid ensemble model not only provides a novel solution for accurate skin cancer classification but also underscores the potential of hybrid approaches in advancing medical image analysis. As the field progresses, the study's findings contribute to a deeper understanding of the synergies between wavelet decomposition, attention mechanisms, and convolutional architectures, offering new perspectives for tackling the challenges posed by intricate dermatological images and paving the way for further advancements in medical image analysis.

9 Authors

Hind FEKKAK is an Third-year engineering student, specializing in Information Systems Engineering and Big Data.

Nihad ESSAADI is anThird-year engineering student, specializing in Information Systems Engineering and Big Data.

Oumaima OUAKMOUR is anThird-year engineering student, specializing in Information Systems Engineering and Big Data.