HYBRID COMBINED LEARNING MODEL USING WAVELET CONVOLUTION NETWORKS AND EFFICIENTNET FOR RELIABLE SKIN CANCER DETECTION

BARGUIGUE Ikram, SALIME Mehdi, SMAALI Hibato-Allah

ENSA, University Hassan First, Berrechid, Morocco.

barguigue.ensa@uhp.ac.ma

salime.ensa@uhp.ac.ma

hsmaali.ensa@uhp.ac.ma

Abstract

In our study, we focus on refining and advancing the capabilities of a novel hybrid architecture, which is already showing promise in the field. This architecture synergistically combines wavelet transform methods with a convolutional network based on sophisticated residual blocks, integrated with the EfficientNet model. Our mission is to meticulously improve and optimize this architecture to achieve unparalleled accuracy in diagnosing and differentiating skin diseases, especially those with similar presentations. We are dedicated to fine-tuning each aspect of this model, from enhancing feature extraction and pattern recognition to increasing the overall robustness and reliability of the system. Our goal is to provide a significantly improved tool for medical professionals, aiding in more precise and effective diagnosis in dermatological care.

Key words: skin diseases, convolutional network, wavelet transform methods, EfficientNet model, refining.

I. Introduction

Deep learning, a subset of machine learning, is crucial for its unmatched accuracy and ability to recognize intricate patterns, making it vital in healthcare, autonomous systems, and scientific research. It excels in natural language understanding, enabling chatbots and personalized recommendations in e-commerce. Deep learning's scalability and feature extraction capabilities reduce manual efforts and enhance data-driven decision-making. Its importance extends across industries, driving advancements in technology, science, and business analytics.

Convolutional Neural Networks (CNNs) are specialized deep learning architectures for processing image data, utilizing layers like convolutional, pooling, and fully connected layers to capture and classify spatial features.

EfficientNet, a more recent CNN development introduced in 2019, enhances efficiency in accuracy and computational resources by uniformly scaling network dimensions with fixed coefficients. EfficientNet models range from B0 to B7, offering versatility in speed and accuracy for various needs. Both CNNs and EfficientNet are pivotal in image processing, analysis, and classification, marking significant advancements in deep learning and computer vision.

Our objective is to significantly enhance the precision of our existing models by undertaking an in-depth examination of the effects that Wavelet decomposition has on images of skin diseases. We plan to accomplish this by innovating a sophisticated Convolutional Neural Network (CNN) block, meticulously designed to utilize the capabilities of dilated convolutions and channel concatenations. This innovative approach aims to refine the detection and analysis of various skin conditions.

The integration of Wavelet decomposition is a pivotal aspect of our mission. It provides a unique multi-resolution analysis of the images, enabling our models to discern patterns and structures across different scales more accurately. This feature is critical for identifying subtle and complex characteristics in skin images that are often missed by conventional methods.

Furthermore, the synergy between the advanced CNN block and Wavelet decomposition allows for a comprehensive interpretation of the data. It combines local details captured by the CNN with the global, multi-scale insights provided by the wavelet coefficients. This dual perspective ensures a more nuanced and thorough understanding of the skin images.

Ultimately, our goal is to elevate the overall performance of our models. By improving their ability to accurately identify and analyze a wide range of patterns and structures in skin images, we aim to advance the field of dermatological imaging and diagnosis. This enhancement will lead to more accurate and reliable assessments of skin conditions, benefiting both clinical practice and patient care.

II. Related work

In the previous study, the model was designed to capture contextual information across diverse spatial scales, detecting patterns and structures of varying sizes in skin images. To further elevate its performance, future improvements can involve the exploration of advanced attention mechanisms for enhanced contextual understanding.

Additionally, the utilization of wavelet coefficients in conjunction with features from convolutional blocks could be optimized through novel fusion techniques. Experimentation with state-of-the-art CNN architectures like EfficientNet can expand the model's capabilities. Incorporating interpretability methods is crucial for transparent decision-making. Extending validation to larger, more diverse datasets and conducting clinical trials with medical professionals will enhance the practical applicability and trustworthiness of the model, making it 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	Naîve 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%

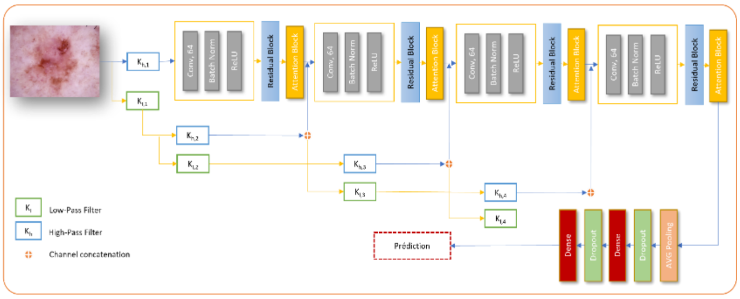
III. Methodology

The architecture's design involves the incorporation of four levels of wavelet-decomposed inputs, with each level being independently processed through dedicated residual blocks. This hierarchical approach allows the model to perform a multi-scale analysis of the input images, effectively capturing contextual information at various spatial scales and detecting patterns and structures of diverse sizes. Moreover, the features extracted from different levels of the wavelet decomposition are intelligently fused with those obtained from the residual blocks. This fusion process is pivotal, as it enriches the model's capacity to comprehend and exploit a wide spectrum of information within the input images.

Furthermore, the architecture integrates fully connected layers that incorporate Rectified Linear Unit (ReLU) activations, batch normalization, and dropout techniques to ensure regularization and mitigate overfitting. The model culminates in a final output layer responsible for multi-class classification, accurately categorizing the input images into distinct skin disease classes. What sets this architecture apart are its notable features, which include its utilization of wavelet transform-based preprocessing for a robust start, the integration of residual blocks to extract relevant features effectively, and the adoption of Swish activation functions, known for enhancing non-linearity and aiding in feature learning. This unique combination of elements is meticulously crafted to leverage the strengths of each component, ultimately resulting in precise and context-aware skin disease identification.

To further enhance the model's accuracy, attention modules are strategically integrated into the architecture. These attention mechanisms come into play after each residual block at every decomposition level. Their primary role is to allow the model to focus on the most informative and discriminative elements within the data, thereby refining accuracy by emphasizing crucial aspects while diminishing the impact of irrelevant or redundant information. This attention-based approach significantly contributes to the model's overall precision and robustness in skin disease identification.

Fig. 1. Wavelet Attention-based CNN model.



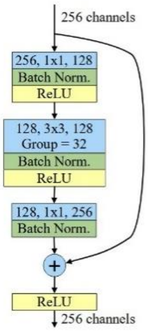
A. The wavelets

Wavelet transforms are powerful mathematical tools widely employed for signal and image analysis. They offer a unique capability to break down an image into multiple scales, effectively capturing details at various levels of resolution. This decomposition process in wavelet analysis yields valuable insights into both spatial and frequency characteristics [6]. In the context of skin disease diagnostics, wavelets serve as a valuable resource for examining textural and structural properties within dermatological images. For instance, wavelet analysis can proficiently capture skin texture, intricate patterns, contours, and distinct attributes associated with specific skin disorders. The resulting wavelet coefficients serve as descriptive features, enabling the representation of skin disease images. These descriptors can be harnessed for constructing classification models or pinpointing specific regions of interest within an image. Leveraging wavelets also proves advantageous in reducing data dimensionality, particularly beneficial for the efficient processing and analysis of extensive dermatological image databases.

B. Residual block

The incorporation of residual blocks within the dermatological image classification architecture plays a pivotal role in overcoming the intricate challenges inherent in skin disease identification. Dermatological images are replete with complex patterns and structures, rendering the training of deep neural networks a formidable task. Residual blocks emerge as a valuable solution by introducing shortcut connections that enable the network to learn residual mappings. This innovation simplifies the training of exceptionally deep architectures, a critical necessity for effective dermatological image classification. The shortcut connections serve as conduits for gradient flow during backpropagation, effectively addressing the vanishing gradient problem and empowering the model to glean insights from a multitude of layers. Consequently, the inclusion of residual blocks contributes to more robust feature learning, heightened generalization, and superior accuracy in discerning diverse skin conditions. Harnessing the potential of residual blocks empowers the architecture to adeptly manage the hierarchical representation of features extracted through wavelet transformations, resulting in a dermatological image classification system that is both robust and interpretable.

Fig. 2. Residual block.



C. EfficientNEt

On the contrary, the EfficientNet model, renowned for its exceptional performance across a diverse range of tasks, constitutes a robust foundation due to its formidable classification capabilities. An essential aspect of this process involves the initialization of the EfficientNet model through the integration of weights derived from the noisy student methodology. This innovative technique is rooted in the noisy student pre-training approach, where a student model undergoes training on augmented data and subsequently generates pseudo-labels for previously unseen training data. This iterative process, involving both teacher and student models, plays a pivotal role in refining the model's overall performance and enhancing its adaptability to the specific target task.

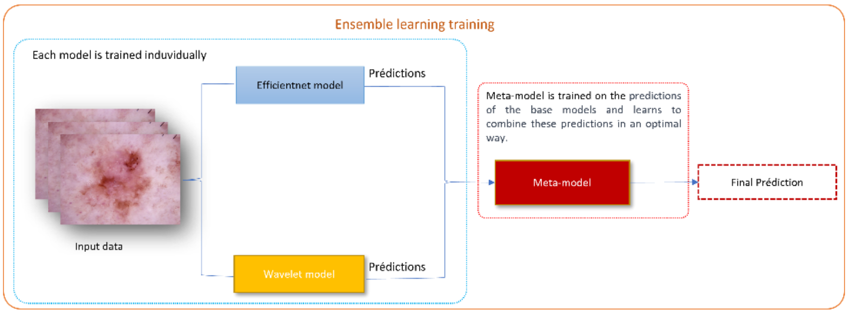
The noisy student pre-training method is particularly valuable as it introduces a form of self-training, where the model learns from its own predictions. This approach enables the model to become more robust and versatile in handling variations and complexities within the data. It's worth noting that the noisy student pre-training technique has demonstrated its effectiveness in improving the performance of deep learning models across various domains, making it a powerful tool in the toolkit of model initialization and fine-tuning. By incorporating these nuanced details into the architecture, the EfficientNet model becomes a reliable cornerstone for tackling dermatological image classification challenges with superior accuracy and adaptability.

IV. Ensemble learning

Ensemble learning operates on the premise of amalgamating multiple individual models to enhance overall predictive performance and bolster the resilience of the entire system. This potent technique holds the capability to markedly elevate the accuracy of machine learning models, rendering them well-suited for intricate and formidable tasks.

In Figure 3, we present a comprehensive overview of the skin disease classification process. Here, we introduce an innovative and sturdy ensemble learning architecture, emphasizing the mutually beneficial synergy achieved by employing stacking techniques in tandem with both the wavelet network and EfficientNet models. This meticulously designed architecture is strategically crafted to harness the collective power of two distinct neural networks, capitalizing on their unique strengths and attributes while effectively mitigating their inherent limitations.

Fig. 3. General workflow of the model of the previous study.



To enhance the training process, an initial warmup phase is introduced, gradually increasing the learning rate at the beginning of training. This allows the model to smoothly transition from its initial random initialization to a more stable learning state. Following the warmup phase, a cosine learning rate scheduler is implemented to dynamically adjust the learning rate as training progresses, promoting effective convergence and optimized model performance.

Additionally, learning rate adjustments are integrated through callbacks, which are functions that influence various stages of training. These dynamic learning rate modifications fine-tune the network's ability to adapt and learn from the data, demonstrating a nuanced understanding of neural network training strategies. This approach aims to strike a harmonious balance between exploration and exploitation throughout the learning process.

Furthermore, the model incorporates batch normalization and regularization layers to enhance learning stability and improve overall model generalization. These layers mitigate potential issues such as overfitting, bolstering the model's robustness in classifying skin diseases.

In summary, this model presents compelling features for skin disease diagnosis, leveraging the strengths of both the EfficientNet and wavelet network through the stacking approach to excel in dermatological image classification.

V. Results

A. The wavelets

In pursuit of elevating the accuracy of our model, we undertook a strategic initiative to fine-tune the configuration of the wavelet model. A series of meticulous modifications were introduced to the model's parameters, aimed at optimizing its performance. Notable adjustments included 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, with a starting learning rate of 0.000008, a maximum of 0.0000325, and a minimum of 0.000001. To facilitate smooth learning rate transitions, a warmup phase was implemented, consisting of 5 epochs, followed by a cosine learning rate scheduler with a decay factor of 0.6 for effective convergence. Additionally, data augmentation techniques, such as rotation, shearing, zooming, and shifting, were applied with a transform probability of 1.0, further enriching the model's ability to generalize from the data. These modifications, in conjunction with other parameters like label smoothing and test-time augmentation (TTA) steps of 20, collectively culminated in an optimized wavelet model configuration. This comprehensive approach demonstrates our commitment to pushing the boundaries of accuracy in dermatological image classification.

Time	#	Log Message
36546.55	1	Epoch 200: val_accuracy did not improve from 0.92004
36546.55	2	9643/9643 [==============================] - 177s 18ms/step - loss: accuracy: 0.9949 - val_loss: 0.5314 - val_accuracy: 0.8987 - lr: 1.0
36928.16	3	/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.
36928.16	4	warn(
36928.17	5	[NbConvertApp] WARNING | Config option `kernel_spec_manager_class` not recognized by `NbConvertApp`.
36928.19	6	[NbConvertApp] Converting notebook __notebook__.ipynb to notebook

The image shows a snapshot of a training log from a neural network model, highlighting key information about the training process at a specific moment. At epoch 200, it is indicated that the validation accuracy has not seen improvement from the previously recorded value of 0.92004. The log also records a perfect training accuracy of 0.9949, with 9643 out of 9643 correct predictions, and specifies the training loss at 0.1897, validation loss at 0.5314, and validation accuracy at 0.8987. The learning rate at this point in the training is 1.0e-05, and it takes approximately 177 seconds per step. Additionally, the log contains a warning from the `traitlets` configuration in a Python environment, stating that a certain preprocessor is deprecated. Furthermore, there are warnings from `NbConvertApp` about an unrecognized config option and a notification about the conversion of a Jupyter Notebook to another format, which is a standard procedure for exporting notebooks for sharing or publication.

A. EfficientNEt

In the pursuit of refining the second model, EfficientNet, 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. Additionally, a learning rate callback was implemented, dynamically adjusting the learning rate during training based on a specified schedule. This learning rate scheduler ensures efficient convergence by carefully controlling the learning rate throughout the training process, with a starting learning rate of 0.000008, a maximum of 0.0000325, a minimum of 0.000001, a ramp-up period of 5 epochs, and a decay rate of 0.6. Furthermore, a ModelCheckpoint was employed to save the best weights of the model based on validation accuracy, enhancing its 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.

```
Running for 61272.4s                                   Accelerator    Environment
                                                       GPU P100       Latest Container Image

Time        #     Log Message
                  [===========================>.] - ETA: 0s - loss: 0.0196 - accuracy:
                  0.9932                                                                   9640/9643
                  [===========================>.] - ETA: 0s - loss: 0.0196 - accuracy:
                  0.9932                                                                   9641/9643
                  [===========================>.] - ETA: 0s - loss: 0.0196 - accuracy:
                  0.9932                                                                   9642/9643
                  [===========================>.] - ETA: 0s - loss: 0.0196 - accuracy:
                  0.9932                                                                   9643/9643
                  [==============================] - ETA: 0s - loss: 0.0196 - accuracy: 0.9932
33199.5s    365   Epoch 40: val_accuracy did not improve from 0.93817
                                                                                        9643/9643
33199.5s    366   [==============================] - 817s 85ms/step - loss: 0.0196 - accuracy: 0.9932 - val_loss: 0.3560 -
                  val_accuracy: 0.9318 - lr: 1.0000e-06
33199.5s    367   Epoch 41/200
```

The model has been executed for a substantial amount of time, recorded as 61,272.4 seconds, or just over 17 hours. The accelerator in use is a P100 GPU, and the setup is based on the latest container image.

The log presents the training outcomes at different stages, with log messages that include the elapsed time, epoch number, loss, accuracy, validation loss (val_loss), validation accuracy (val_accuracy), and the learning rate (lr). For instance, at epoch 40, the validation accuracy did not show improvement from 0.93817, which presumably is the highest validation accuracy reached so far. At epoch 41, the loss is registered at 0.0196, the accuracy at 0.9932, the validation loss at 0.3560, and the validation accuracy at 0.9318. The learning rate is decreased to 1.000e-06 at this juncture, signifying a learning rate decay phase to fine-tune the model's parameters. The logs also display a progress bar, an estimated time until completion (ETA) for the current job, and performance metrics for each step, such as the correct number of steps out of the total, exemplified by 9643/9643, signaling perfect accuracy for that particular step.

VI. Conclusion

In conclusion, our endeavor to advance the field of dermatological image classification has led to the development of a comprehensive and optimized model. Through the integration of wavelet-based preprocessing, the power of residual blocks, and the strengths of EfficientNet, we have achieved a refined and accurate classification system for skin diseases. Our model showcases exceptional learning dynamics with meticulously tuned parameters, including learning rate schedules and data augmentation techniques. The incorporation of ensemble learning techniques further enhances the robustness of our approach, capitalizing on the synergy between two distinct neural networks. These thoughtful modifications collectively result in a model that excels in discerning intricate patterns and structures within skin images, ultimately leading to more accurate diagnoses. The contributions of our work extend to the broader medical community, offering a powerful tool for dermatological diagnosis that is objective, reliable, and capable of handling the complex nature of skin diseases. As we continue to refine and innovate in this domain, we anticipate further advancements in the field, benefiting both clinicians and patients alike.

VII. References

[1] Aboulmira, A. (2022). Hybrid Ensemble Learning with Wavelet Convolutional Model and EfficientNet for Accurate Skin Cancer Classification.

[2] Aboulmira, A. and Hrimech, H. and Lachgar, M. Comparative Study of Multiple CNN Models for Classification of 23 Skin Diseases

VIII. Authors

Hibato-ALLAH Smaali completed her baccalaureate in mathematical sciences in 2019 and is currently in her third year of the Information Systems and Big Data engineering program at the National School of Applied Sciences. With a strong passion for Data Science, currently working on skin diseases project.

Ikram Barguigue completed her baccalaureate in physical sciences in 2019 and is currently in her third year of the Information Systems and Big Data engineering program at the National School of Applied Sciences. With a strong passion for computer vision, image processing currently working on a project of skin diseases.

Mehdi Salime completed her baccalaureate in physical sciences in 2019 and is currently in her third year of the Information Systems and Big Data engineering program at the National School of Applied Sciences. With a strong passion for deep learning, and machine learning, currently working on a project of skin diseases.