MULTICLASS SKIN LESION CLASSIFICATION USING EFFICIENTNETV2-SMALL

Serra Aksoy[a*]
[a] Ludwig Maximilian University of Munich (LMU), Munich, GERMANY
[*] *Corresponding Author:* serra.aksoy@campus.lmu.de

ABSTRACT

Skin cancer presents a formidable challenge in healthcare, consisting of various forms such as melanoma and benign lesions, which demand early detection and precise treatment. In this study, the application of an advanced deep learning model, EfficientNetV2-Small, for the automated classification of skin lesions using the HAM10000 dataset was proposed. This dataset is renowned for its diversity and comprehensive representation of dermoscopy images. This research addresses the critical need for accurate and timely diagnosis in dermatology. The proposed EfficientNetV2-Small model achieved a remarkable accuracy of 88.62% in distinguishing between different types of pigmented skin lesions, including melanoma, nevi, and other benign and malignant lesions. This performance emphasizes the importance of deep learning techniques in image analysis and diagnostic decision support. By using state-of-the-art computational tools, the aim of this study is to contribute to the advancement of clinical practices in dermatology, offering a reliable and objective tool for dermatologists to improve diagnostic accuracy and optimize patient care pathways.

Keywords: Skin Cancer. Melanoma. Benign Lesions. EfficientNetV2-Small. HAM10000 Dataset. Dermoscopy Images.

1. INTRODUCTION

Skin cancer is a major concern in healthcare, where early detection and accurate diagnosis are crucial for effective treatment. Whether it's melanoma or benign lesions, identifying the type of skin cancer accurately can significantly impact a patient's outcome [1]. Convolutional neural networks (CNNs) are essential in medical imaging, particularly for tasks like classifying and segmenting images [2-4]. Recent studies by Esteva et al. [5] and Haenssle et al. [6] have highlighted CNN models' superiority over certified dermatologists in accurately classifying skin lesions. The International Skin Imaging Collaboration (ISIC) [7] has played a crucial role by curating dermoscopic image datasets and organizing challenges to assess the efficacy of new algorithms and models in diagnosing skin cancer.

In recent literature, numerous deep learning methods have been proposed for diagnosing skin cancer. Kadampur et al. [8] introduced an application-based model using the Squeezenet architecture [9], achieving an impressive average AUC score of 0.99 when evaluated with 10% of the HAM10000 dataset [10]. Their work highlights the effectiveness of deep learning in achieving high accuracy in skin lesion classification tasks.

Rahman et al. [11] proposed an ensemble-based learning approach utilizing DenseNet, ResNet, Xception, SeResNeXt, and ResNeXt models. Their method achieved the highest micro average recall of 0.94 for multiclass classification, evaluated on both the ISIC 2019 [12,13] and HAM10000 [10] datasets. This approach demonstrates the robustness and versatility of combining multiple deep learning models to enhance diagnostic accuracy across different datasets.

Chaturvedi et al. [14] developed a modified ResNeXt-101 model specifically tailored for classifying skin lesions into seven categories. Fine-tuning their model on a substantial dataset of 89,213 dermoscopy images, they achieved an impressive accuracy of 93.20% on the HAM10000 dataset [10]. The research underscores the significant strides in deep learning applications for precise and efficient skin cancer diagnosis, showcasing advancements that hold promise for improving clinical outcomes.

These studies collectively illustrate how deep learning techniques continue to advance the field of dermatology, offering powerful tools for accurate and reliable diagnosis of skin conditions, ultimately benefiting both patients and healthcare providers.

Recent advancements in deep learning models, like EfficientNetV2-Small, offer promising potential in automating the classification of skin lesions. By utilizing datasets like HAM10000 [10], which are known for their rich and varied dermoscopy images, these models can accurately distinguish between different types of pigmented skin lesions, including melanoma, nevi, and other benign and malignant lesions.

In multiclass classification, it is evaluated how well a model distinguishes between different categories by using a confusion matrix. This matrix breaks down the model's predictions for each class, showing us the number of true positives (correctly identified instances), false positives (incorrectly identified instances), false negatives (missed instances), and true negatives (correctly rejected instances). In practical terms, a high F1 score suggests that the model strikes a good balance between accuracy and completeness in its predictions, making it a valuable tool for evaluating and refining multiclass classification models. By analyzing these metrics for each class, it can be determined areas where the model needs improvement. This process is crucial for refining the model's training and fine-tuning its performance to ensure accurate and reliable diagnoses.

The high accuracy achieved by these models highlights the power of deep learning in enhancing image analysis and supporting diagnostic decisions. Incorporating such advanced computational tools into dermatology can significantly improve clinical practices. These tools provide dermatologists with reliable and objective methods to enhance diagnostic precision and streamline patient care, ultimately leading to better outcomes in the battle against skin cancer.

2. MATERIAL AND METHOD

2.1. Data Preparation and Augmentation

The HAM10000 dataset [10] was utilized in this study, containing a diverse collection of dermatoscopic images of pigmented lesions. All significant diagnostic categories within the domain of pigmented skin lesions were represented in the dataset. Included categories were actinic keratoses and intraepithelial carcinoma/Bowen's disease (AKIEC), basal cell carcinoma (BCC), benign keratosis-like lesions (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), and vascular lesions (VASC). To enhance the model's generalization capability and to prevent overfitting, data augmentation techniques were applied to the training dataset. Variability in the training data was introduced by these augmentations, simulating real-world scenarios where skin lesions can appear in various orientations and under different conditions. Two sets of transformations were utilized: one for the training dataset and another for the validation and test datasets. For the validation and test sets, each image was resized to a fixed size of 224x224 pixels, converted to a tensor, and normalized. For the training dataset, several augmentation techniques were applied to introduce variability. Each image was resized to 224x224 pixels, with horizontal and vertical flips applied with a probability of 0.6, random rotations by up to 15 degrees, conversion to a tensor, and normalization.

The dataset was split into training, validation, and test sets using a stratified approach to ensure that each set maintained the same distribution of classes. The split was done with proportions of 70% for training, 15% for validation, and 15% for testing. Initially, image paths and their corresponding labels were extracted from a CSV file containing image filenames and class labels. The images were then divided into training and temporary sets, with the temporary set subsequently split into validation and test sets. This method ensured that the distribution of different skin lesion categories was consistent across all sets. After splitting, the images were organized into separate directories for training, validation, and test sets, with subdirectories for each class within these main directories. This structured organization facilitated efficient loading and processing of the data during model training and evaluation.

2.2. Proposed Model

The proposed model for the classification of melanoma and other skin lesions is EfficientNetV2-Small, an advanced version of the original EfficientNet architecture. EfficientNetV2-Small, developed to further enhance the performance and efficiency of convolutional neural networks, introduces several improvements over its predecessor. This architecture is designed to optimize both training speed and accuracy, making it well-suited for the task of skin lesion classification.

EfficientNetV2-Small employs a novel scaling method that balances the dimensions of depth, width, and resolution using a compound coefficient, resulting in an architecture that can be effectively scaled up or down depending on the specific requirements. This model leverages Fused-MBConv layers, which

combine the advantages of both Mobile Inverted Bottleneck Convolution (MBConv) layers and standard convolutions. These layers improve computational efficiency and reduce the number of parameters, while still maintaining high representational power.

One of the significant enhancements in EfficientNetV2-Small is the use of progressive learning, which adjusts the image resolution and network complexity during training. This approach speeds up the training process and improves overall accuracy by starting with smaller images and simpler models, and gradually increasing the resolution and complexity as training progresses. This progressive strategy allows the network to learn more effectively, capturing details and patterns crucial for accurate skin lesion classification.

EfficientNetV2-Small also incorporates squeeze-and-excitation (SE) optimization, a technique that adaptively recalibrates channel-wise feature responses by explicitly modeling the interdependencies between channels. This optimization enhances the network's ability to focus on the most informative features, thereby improving its classification performance.

In the final stage of the network, a softmax activation function is applied to the output layer. This function converts the logits into a probability distribution over the predefined classes. Each class is assigned a probability value, indicating the likelihood that the input image belongs to that class. The class with the highest probability is then selected as the model's prediction. This probabilistic approach enables precise and reliable classification of skin lesions, ensuring that the model provides confident and interpretable predictions. (Figure 1)

By using EfficientNetV2-Small, the proposed model benefits from state-of-the-art techniques and optimizations, achieving high accuracy and efficiency in the challenging task of melanoma classification. This makes it a powerful tool for assisting dermatologists and improving the early detection and treatment of skin cancer.

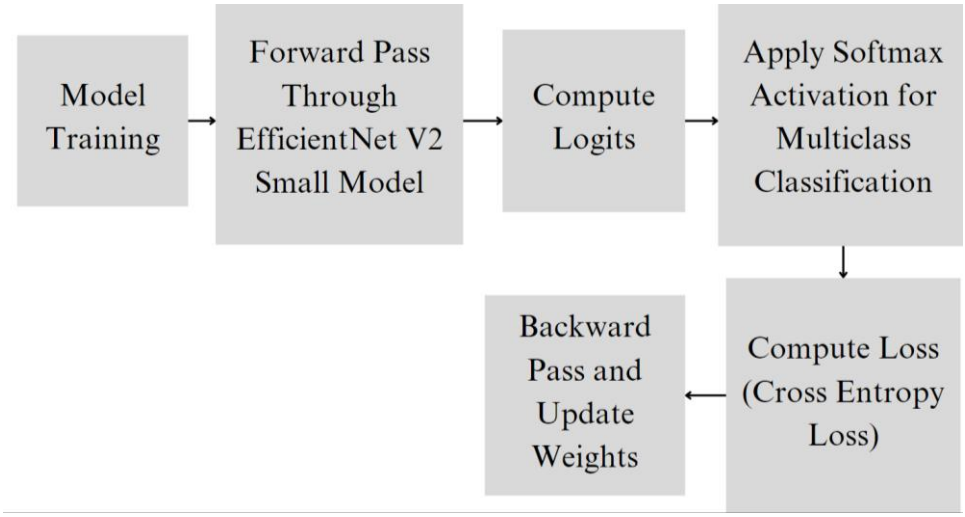

Figure 1. Training Flow Chart of the Proposed Model.

2.3. Experimental Setup

The training and evaluation of the model were performed using a high-performance computing setup. The system was equipped with an Intel Core i9 processor and an NVIDIA GeForce RTX GPU. This hardware configuration provided the necessary computational power for handling the intensive operations involved in training deep learning models. The model was implemented using the PyTorch deep learning framework, which offers flexibility and efficient GPU utilization for training complex neural networks.

The dataset was split into training, validation, and test sets using a stratified approach to ensure that each set maintained the same distribution of classes. The split proportions were 70% for training, 15% for validation, and 15% for testing. The training images were resized to 224x224 pixels, and augmentations such as horizontal and vertical flips with a probability of 0.6, and random rotations up to 15 degrees were applied. For the validation and test sets, the images were resized to 224x224 pixels and normalized to ensure uniformity.

The EfficientNetV2-Small model was employed for the classification of melanoma and other skin lesions. During the training phase, the model was trained on the training set and validated on the validation set with a batch size of 32 for 20 epochs. The training process involved iterative updates to the model's parameters to minimize the cross-entropy loss, which measures the discrepancy between the predicted and actual class labels. The Adam optimizer with a learning rate of 0.01 was used to update the model parameters. Mixed precision training, which uses both 16-bit and 32-bit floating point types to speed up computation and reduce memory usage, was used. This was managed using an automatic mixed precision (AMP) scaler, which ensured stable and efficient training.

The training process was monitored by evaluating the training and validation loss and accuracy at the end of each epoch. The training step involved forward propagation, loss calculation, and backpropagation to update the model weights. The validation step involved evaluating the model on the validation set without updating the weights, providing an estimate of the model's performance on unseen data (Figure 2).

For the evaluation, the training and validation loss and accuracy curves were plotted to visualize the model's learning progress and to detect any signs of overfitting or underfitting. After training, the model was tested on the test set to evaluate its performance on completely unseen data. The predictions made by the model on the test set were used to create a confusion matrix, which provides a detailed breakdown of the model's classification performance across all classes. The overall accuracy of the model was calculated using the ground truth of the test sets and the predictions on the test set, giving a measure of the model's effectiveness in correctly classifying the skin lesions (Figure 3).

This comprehensive experimental setup ensured that the model was robustly trained and thoroughly evaluated, providing reliable performance metrics for the classification of melanoma and other skin lesions.

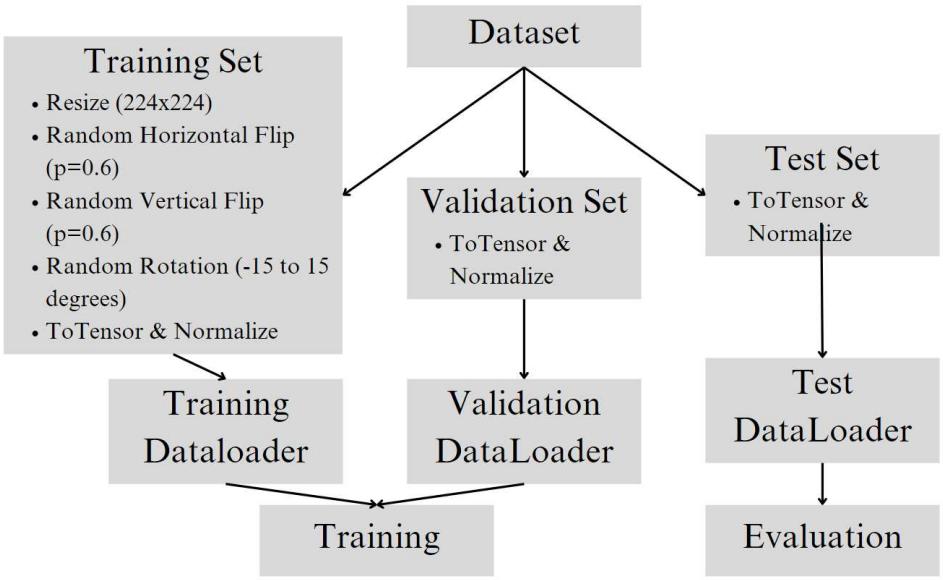

Figure 2. Experimental Setup.

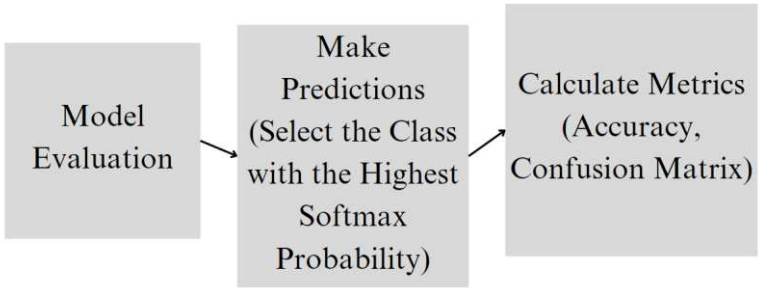

Figure 3. Evaluation Flow Chart of the Proposed Model.

3. RESULTS AND DISCUSSION

The proposed EfficientNetV2-Small model demonstrated promise in classifying skin lesions, achieving an overall accuracy of 88.62%. The training and validation curves show that while the model learned effectively, there was a tendency towards overfitting, as indicated by the divergence between training and validation metrics. The training loss consistently decreased, reaching below 0.1, while the validation loss fluctuated between 0.4 and 0.5 after initial epochs. The training accuracy steadily increased to approximately 0.98, whereas the validation accuracy plateaued around 0.85 (Figure 4). The confusion matrix provided detailed insights into the classification performance across different skin lesion categories. The model performed exceptionally well in classifying nevi, with 976 correct predictions and minimal misclassifications, and showed reasonable accuracy for melanoma, correctly identifying 108 cases but misclassifying 43 as nevi. Benign keratosis-like lesions and basal cell carcinoma were also well-classified, though there were notable misclassifications into other categories, particularly NV and MEL. Actinic keratoses and intraepithelial carcinoma had moderate classification performance, with misclassifications into NV and BKL. Dermatofibroma and vascular lesions showed lower accuracy, with confusion among several categories. The high accuracy in classifying nevi suggests effective feature learning, while the misclassification of melanoma as nevus indicates a need for improved differentiation (Figure 5).

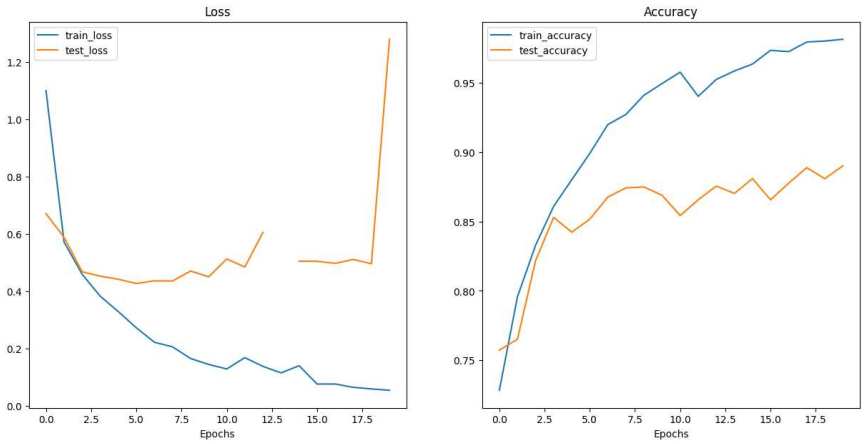

Figure 4. Training and Validation Loss and Accuracy Curves.

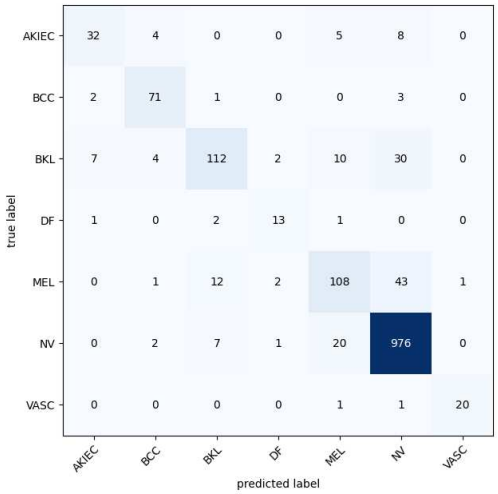

Figure 5. Confusion Matrix for Skin Lesion Classification.

Future work could focus on enhancing generalization capabilities through data augmentation, incorporating diverse training samples, and employing regularization techniques. Additionally, exploring advanced architectures or ensemble methods might further boost classification performance. In conclusion, the proposed EfficientNetV2-Small model offers a robust framework for skin lesion classification, demonstrating high accuracy and potential for clinical application, with further refinements likely to enhance its diagnostic capabilities for early detection and treatment of skin cancer.

4. CONCLUSION

In conclusion, the application of the EfficientNetV2-Small model in this study has shown promising results in the multiclass classification of skin lesions, achieving an impressive accuracy of 88.62% across various diagnostic categories. This research emphasizes the critical role of utilizing deep learning models for dermatological diagnostics.

Future efforts could focus on enhancing the model's ability to generalize across diverse datasets and improve differentiation between similar classes, particularly in distinguishing melanoma from benign nevi. Strategies such as increased data augmentation, regularization techniques, and exploring more advanced architectures or ensemble methods could further increase the model's performance. Moreover, integrating real-time clinical data and feedback loops from dermatologists could refine the model's accuracy and applicability in real-world settings.

Ultimately, the EfficientNetV2-Small model represents a significant step forward in using artificial intelligence for dermatological diagnostics, poised to contribute significantly to early detection and precise treatment of skin cancer. Continued advancements in this field hold great promise for improving healthcare outcomes and patient well-being in dermatology.

REFERENCES

1. Jasil, S.P.G.; Ulagamuthalvi, V. A Hybrid CNN Architecture for Skin Lesion Classification Using Deep Learning. *Soft Comput* **2023**, doi:10.1007/s00500-023-08035-w.
2. Jiang, H.; Diao, Z.; Shi, T.; Zhou, Y.; Wang, F.; Hu, W.; Zhu, X.; Luo, S.; Tong, G.; Yao, Y.-D. A Review of Deep Learning-Based Multiple-Lesion Recognition from Medical Images: Classification, Detection and Segmentation. *Computers in Biology and Medicine* **2023**, *157*, 106726, doi:10.1016/j.compbiomed.2023.106726.
3. Brinker, T.J.; Hekler, A.; Utikal, J.S.; Grabe, N.; Schadendorf, D.; Klode, J.; Berking, C.; Steeb, T.; Enk, A.H.; Von Kalle, C. Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. *J Med Internet Res* **2018**, *20*, e11936, doi:10.2196/11936.
4. Nazir, S.; Dickson, D.M.; Akram, M.U. Survey of Explainable Artificial Intelligence Techniques for Biomedical Imaging with Deep Neural Networks. *Computers in Biology and Medicine* **2023**, *156*, 106668, doi:10.1016/j.compbiomed.2023.106668.
5. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *542*, 115–118, doi:10.1038/nature21056.
6. Haenssle, H.A.; Fink, C.; Schneiderbauer, R.; Toberer, F.; Buhl, T.; Blum, A.; Kalloo, A.; Hassen, A.B.H.; Thomas, L.; Enk, A.; et al. Man against Machine: Diagnostic Performance of a Deep Learning Convolutional Neural Network for Dermoscopic Melanoma Recognition in Comparison to 58 Dermatologists. *Annals of Oncology* **2018**, *29*, 1836–1842, doi:10.1093/annonc/mdy166.
7. ISIC International Skin Imaging Collaboration Available online: https://www.isic-archive.com/.
8. Kadampur, M.A.; Al Riyaee, S. Skin Cancer Detection: Applying a Deep Learning Based Model Driven Architecture in the Cloud for Classifying Dermal Cell Images. *Informatics in Medicine Unlocked* **2020**, *18*, 100282, doi:10.1016/j.imu.2019.100282.
9. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <0.5MB Model Size 2016.
10. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Sci Data* **2018**, *5*, 180161, doi:10.1038/sdata.2018.161.
11. Rahman, Z.; Hossain, Md.S.; Islam, Md.R.; Hasan, Md.M.; Hridhee, R.A. An Approach for Multiclass Skin Lesion Classification Based on Ensemble Learning. *Informatics in Medicine Unlocked* **2021**, *25*, 100659, doi:10.1016/j.imu.2021.100659.
12. Codella, N.; Rotemberg, V.; Tschandl, P.; Celebi, M.E.; Dusza, S.; Gutman, D.; Helba, B.; Kalloo, A.; Liopyris, K.; Marchetti, M.; et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). **2019**, doi:10.48550/ARXIV.1902.03368.

13. Combalia, M.; Codella, N.C.F.; Rotemberg, V.; Helba, B.; Vilaplana, V.; Reiter, O.; Carrera, C.; Barreiro, A.; Halpern, A.C.; Puig, S.; et al. BCN20000: Dermoscopic Lesions in the Wild. **2019**, doi:10.48550/ARXIV.1908.02288.
14. Chaturvedi, S.S.; Tembhurne, J.V.; Diwan, T. A Multi-Class Skin Cancer Classification Using Deep Convolutional Neural Networks. *Multimed Tools Appl* **2020**, *79*, 28477–28498, doi:10.1007/s11042-020-09388-2.