Skin Lesion Detection: HAM10000 dataset

Divyansh Singh[1], Yash Goel[1], Aseem Banarwal[1], Rishabh Srivastav[1], Vaishnavi Chilakamarthi[1], and Srishti Jaiswal[1]

University of Florida, Gainesville, FL, 32608, USA
{divyansh.singh,y.goel,a.baranwal,rsrivastav,vchilakamarthi,srishti.jaiswal}@ufl.edu

Abstract. The treatment of skin cancer, which is the most prevalent type of cancer in the world, depends heavily on early identification. In the United States, skin cancer affects more than 9,500 individuals daily, while more than two people pass away from it each hour. Deep learning algorithms have demonstrated encouraging results in the correct identification of skin lesions in recent years. The HAM10000 dataset, which contains 10,015 photos of skin lesions divided into seven different kinds, is used in this research to propose a deep learning method - EfficientNet for categorizing skin lesions. Subsequently, we compare the performance of our model using well-established machine learning techniques like SVM, random forest, decision trees, and CNN. The results show that by adding a soft attention layer to the deep learning model of EfficientNet, one can achieve significantly higher accuracies and better results for AUC metric. Our work can serve as a basis on which further models can be developed which can assist physicians during diagnosis. The code for our work is released at: `https://github.com/divyanshrm/HAM1000_USING_EFFICIENTNET_SA`

Keywords: Deep Learning · Skin Lesion · SVM · CNN · Decision Tree · Random Forest · EfficientNet · AUC.

1 Introduction

Dermatologists use various diagnostic techniques, including visual examination, dermoscopy, and biopsy, to diagnose skin lesions accurately. Skin lesion detection and classification have been the subject of numerous studies in recent years, with many researchers utilizing the HAM10000 dataset for their investigations. The dataset categories are divided into 7 classes: melanoma, nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, and vascular lesion.

Using the HAM10000 dataset, B. Tahir et al. (2020) suggested a deep learning-based method and used transfer learning to classify the lesions. The study also showed that the model performed better when data augmentation approaches were used. N. Ali et al. (2020) [4] suggested a hybrid deep learning-based technique for skin lesion detection and classification. Convolutional neural networks and a support vector machine classifier were both used in the suggested technique, which had a 93.2% test accuracy. A. M. Youssef et al. (2021) [5] suggested a deep learning-based method that made use of a residual convolutional neural network and had a test accuracy of 93.7%.

On a number of computer vision tasks, the EfficientNet family of convolutional neural networks has produced groundbreaking outcomes. By increasing the model's depth, width, and resolution in a logical way, the EfficientNet models aim to strike a balance between model size, computing efficiency, and accuracy. Our proposed approach fine-tunes the pre-trained EfficientNet models and EfficientNet with Soft Attention Layer on the HAM10000 dataset by performing novel data augmentation techniques. Our findings show the efficacy of EfficientNet and EfficientNet with Soft Attention Layer for skin lesion categorization, which has the potential to help physicians diagnose skin lesions accurately and quickly.

2 Methods

2.1 Decision Tree

A decision tree classifier is a particular kind of supervised learning algorithm that is used to forecast a target variable based on a collection of input features. Based on the values of the input features, the decision tree method iteratively divides the feature space into smaller and smaller sections. We employed hyperparameter tweaking to identify the ideal set of parameters in order to improve the decision tree model's performance[3]. Based on our experiments, the best combination of hyperparameters for the decision tree model was {'criterion': 'entropy', 'max_depth': 10, 'max_features': 'log2', 'min_samples_leaf': 2, 'min_samples_split': 10, 'splitter': 'random'}. A number of metrics were used to assess the model's performance, including accuracy, sensitivity, specificity, and F1 score.

2.2 Support Vector Machine

Support Vector Machine (SVM) is a well-liked supervised learning method that may be applied to both classification and regression. Finding an ideal hyperplane in a high-dimensional feature space that can divide the data into two groups is the primary objective of SVM. SVM seeks to maximize the space between the classes, which aids in broadening the model's applicability. In contrast, when the data cannot be separated linearly, SVM employs the kernel method to transform the data into a higher-dimensional space. We employed a radial basis function (rbf) kernel in our implementation. Despite this, we only managed to achieve a precision value of 35.1%; consequently, we chose to investigate other models for data training. .

2.3 Random Forest

After receiving subpar results with SVM, we switched to Random Forest Classification for our medical dataset. We employed XGBoost, which successively trains models on the residuals of the prior model, achieving high accuracy with fewer trees, to get over its potential drawbacks. With scores of around 70% and 93.9%, XGBoost considerably increased accuracy and precision. By recognizing more instances of the target condition with a better level of certainty, this helped to reduce False Negatives.

2.4 CNN

CNNs are a subset of deep learning algorithms that classify images. The CNN feedforward process entails applying learnable filters to extract features from the input image, followed by downsampling via a pooling layer. We used a CNN with 3 convolutional layers, each with a kernel size of 3 and ReLU activation function, followed by max pooling and batch normalization. A dropout rate of 0.3 was utilized to avoid overfitting. The output layer was a fully connected dense layer with 7 neurons and softmax activation. We used a learning rate of 0.001, categorical cross-entropy loss and Adam optimizer.

2.5 EfficientNet

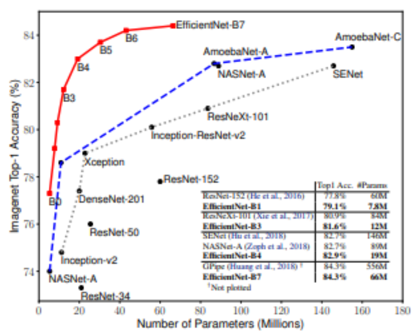

Fig. 1. Comparison of EfficientNet with other models on ImageNet Data.

In their work, Tan and Le proposed EfficientNet, a deep convolutional neural network architecture designed for efficient model scaling [1]. EfficientNet is a deep convolutional neural network architecture designed to achieve state-of-the-art performance on image classification tasks with a high degree of efficiency in terms of model size and computational resources. The architecture combines several techniques such as compound scaling, neural architecture search, and feature fusion to optimize both accuracy and efficiency.

The compound scaling technique involves scaling the network width, depth, and resolution simultaneously. By scaling all dimensions, the network can adapt to different levels of computational resources, making it more efficient. This is done by introducing a compound coefficient that determines how much each dimension is scaled.

The feature fusion technique is used to combine features from different levels of the network, allowing for better representation of complex image features. This is achieved by using a combination of depth-wise and point-wise convolutions to fuse features across different spatial resolutions.

EfficientNet has achieved state-of-the-art performance on several image classification benchmarks, including ImageNet and CIFAR-10, with significantly fewer parameters and computational resources compared to other state-of-the-art models. Therefore, it is an efficient and effective solution for image classification tasks.

For this experiment, we have used EfficientNet B7 considering computation requirements and availability.

2.6 Soft Attention

Skin lesion images often contain irrelevant features such as veins and hair, which can make it difficult for models to focus on the important pixels. To address this, we implemented a soft attention mechanism inspired by previous work in image caption generation and handwriting verification. Soft attention highlights the relevant features by assigning higher weights to the corresponding feature maps. This results in areas with higher attention being colored in red in our model. By discrediting the irrelevant areas, the model is able to perform better with more focused information. The soft attention module takes the feature tensor as input and is based on prior research in this area.

For this paper, we have implemented and tested Xu et al's[2] soft attention he used for image captioning. Instead we have used the attention for classification. Xu et al.'s soft attention mechanism is particularly well-suited for image classification tasks. This method assigns weights to different regions of the image based on their relevance to the task at hand, allowing the model to focus on the most important regions while ignoring irrelevant features. This approach has been shown to improve the performance of image classification models, particularly when dealing with complex and cluttered images.

The implementation of soft attention mechanisms in skin lesion classification can help to improve the accuracy and robustness of these models by allowing them to selectively focus on the most relevant regions of the image. This can be particularly useful in scenarios where images contain a large amount of irrelevant features, which can make it difficult for traditional classification methods to perform well.

3 Experimental Setup

3.1 Data Augmentation

Data augmentation is a common technique used in machine learning to increase the size of a dataset by creating additional samples from existing data. In the context of image data with class imbalance, data augmentation can be particularly useful in rebalancing the distribution of classes. By applying transformations such as rotations, translations, and flips to images in the minority class, the resulting augmented dataset can contain a more balanced representation of the classes. This approach can improve the performance of machine learning models trained on imbalanced image datasets and reduce the risk of biased predictions.

We used `ImageDataGenerator` class from `tf.keras.preprocessing.image` module to augment our image dataset. This class allowed us to perform various data augmentation techniques on our dataset, including rotation, shifting, zooming, and flipping the images. The `rotation_range` parameter was set to 180, which randomly rotates the images between 0 to 180 degrees. We also set the `width_shift_range` and `height_shift_range` parameters to 0.1, which shifts the images horizontally and vertically by a maximum of 10% of the total width and height of the image, respectively. The `zoom_range` parameter was set to 0.1, which zooms into the images by a maximum of 10% randomly. Finally, we set the `horizontal_flip` and `vertical_flip` parameters to `True`, which flips the images horizontally and vertically randomly. All newly created pixels were filled using the nearest pixel value from the original image. These data augmentation techniques improved the size and diversity of our training dataset, allowing our model to better generalize and perform well on the test data.

3.2 EfficientNetB7

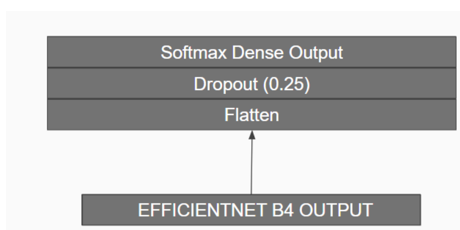

Fig. 2. Architecture of EfficientNet model used in the Experiment

The architecture depicted in Figure 5 has been used for the experiments. A learning rate of 0.001 was used for the initial training of the model using Adam Optimizer. we trained an EfficientNetB7 model to classify skin cancer lesions into 7 categories. The model was implemented using a flatten layer and a dropout of 0.25, followed by a softmax output layer of 7 size. The input resolution was set to 224 by 224 and the model was trained using the Adam optimizer. The model was trained using a batch size of 16 and was trained for 30 epochs.

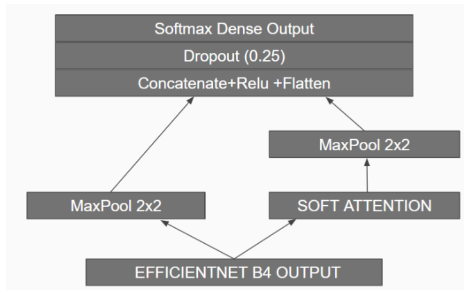

Fig. 3. Architecture of EfficientNet model with Soft Attention used in the experiment.

3.3 EfficientNetB7 with Soft Attention

We have implemented Xu et al's[2] in combination with the EfficientNet outputs. The detailed architecture is given in Figure 3. All the hyperparameters were exactly identical to Section 3.2 to ensure homogeneity in results. Attention is concataned with maxpooled outputs from efficientnet followed by a dropout layer.

4 Results

4.1 Soft Attention Map

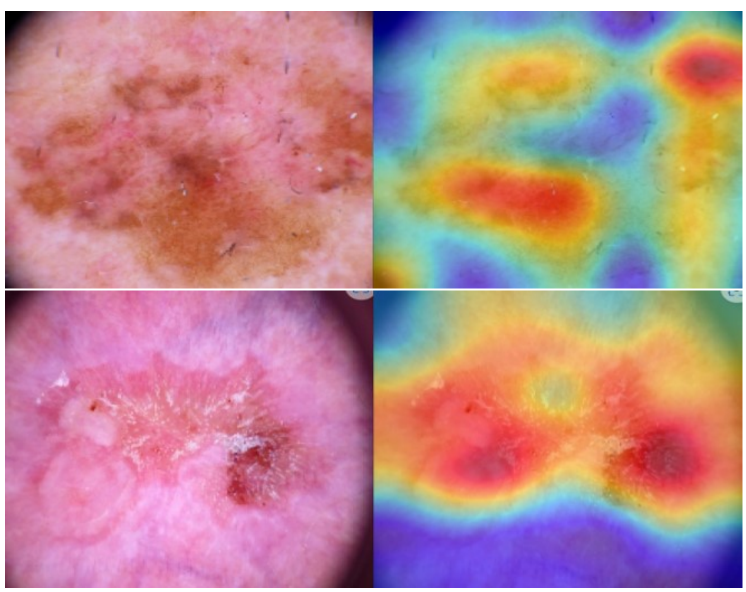

Fig. 4. Result after mapping Soft Attention on Skin Lesion Images

4.2 Performance comparison of proposed model with the baselines

Table 1. Performance comparison of different models

Model Name	Accuracy	Precision
Decision Tree	0.695	0.682
Support Vector Machines	0.696	0.351
Random Forest with XGBoost	0.711	0.939
CNN with Self-Attention	0.781	0.783
EfficientNet	0.895	0.898
Efficient Net with Soft Attention	0.935	0.937

4.3 Evaluation according to classes

4.3.1 Efficient Net The following table evaluates the performance of the Efficient Net for individual classes.

Table 2. Performance comparison of different classes for EfficientNet

Method	Precision	Recall	F1-score	Accuracy
akiec	0.61	0.48	0.54	0.96
bcc	0.72	0.69	0.71	0.98
bkl	0.75	0.64	0.69	0.96
df	0.56	0.83	0.67	0.99
mel	0.46	0.50	0.48	0.93
nv	0.95	0.97	0.96	0.96
vasc	1.00	0.70	0.82	0.98

4.3.2 Efficient Net + Soft Attention The following table and graph evaluate the performance of the Efficient Net with Soft Attention for individual classes.

Table 3. Performance comparison of different classes for EfficientNet + Soft Attention

Method	Precision	Recall	F1-score	Accuracy
akiec	0.88	0.65	0.75	0.99
bcc	0.85	0.65	0.74	0.99
bkl	0.87	0.79	0.83	0.98
df	1.00	0.83	0.91	1.00
mel	0.55	0.76	0.64	0.96
nv	0.97	0.98	0.97	0.98
vasc	1.00	0.90	0.95	1.00

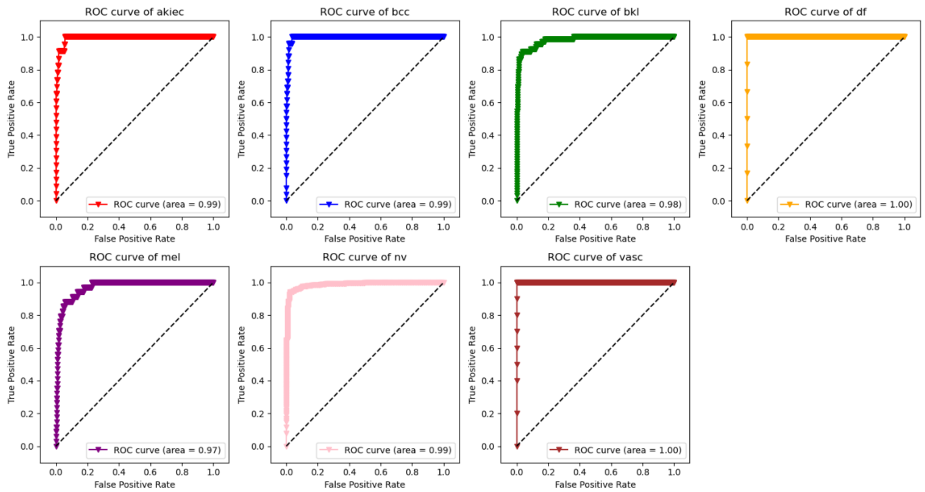

Fig. 5. ROC Curves for all the classes for EfficientNet with Soft Attention on Test Set

5 Advantages of Our Model

There are several advantages of our proposed model in the field of skin cancer detection:

1. High accuracy: Our model can precisely identify the kind of skin lesion a patient has, which can aid medical professionals in making better decisions about the patient's treatment.
2. Robustness: Our model's performance and robustness are enhanced by the soft attention layer, making it more adaptable to changes in the data.
3. AUC metric: Our model can handle skewed datasets better by employing the AUC measure, which is frequently the case in skin cancer diagnosis. This guarantees that our model can reliably recognize both positive and negative examples without being biased towards the class with the majority of instances.
4. Cost-effective: Our model can assist patients and healthcare professionals save time, effort, and money by providing preventative/countermeasures/steps for patients based on the type of skin lesion that patient specifically has. In situations when early discovery might improve outcomes, this may be especially helpful.

Overall, by enhancing accuracy, robustness, and cost-effectiveness, our model has the potential to have a significant impact in the field of skin cancer detection.

6 Conclusion

In summary, through the utilization of advanced machine learning algorithms and image processing techniques, we can effectively classify skin images into various categories and provide personalized recommendations for patients. This has the potential to significantly improve the speed and precision of diagnosis and treatment, which can lead to saving lives and decreasing healthcare expenses.

The combination of technology and medical expertise is a powerful tool in the fight against skin diseases, and we can expect even more significant breakthroughs in the years to come. Through continued research and development, we can hope to improve the accuracy and effectiveness of skin lesion detection and classification, leading to better patient outcomes and ultimately contributing to a healthier society. With continued effort and investment, we can look forward to a future where skin diseases are diagnosed and treated with even greater accuracy and efficiency, improving the lives of countless individuals around the world.

7 Future Work

There are several avenues for future work in the field of skin lesion detection and classification. One area that could be explored is the development of more sophisticated machine learning models that can handle a wider variety of skin images and lesion types, potentially through the incorporation of additional data sources such as patient medical histories. Another potential direction for future research is the integration of other types of medical imaging, such as MRI or ultrasound, into the skin lesion detection and classification process. It will also be important to continue investigating the ethical implications of machine learning-based diagnosis and treatment and to ensure that bias and unequal access to technology are carefully considered. Finally, the implementation and testing of proposed prescription/suggestion systems in clinical settings is a crucial step toward practical application. Extensive trials and evaluations will be necessary to determine the effectiveness and feasibility of the system in real-world settings. Overall, the field of skin lesion detection and classification has significant potential for future research and development, with a focus on increasing the precision and effectiveness of diagnosis and treatment for the benefit of patients and healthier communities.

References

1. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, pp. 6105-6114. PMLR.
2. Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., ... & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. In Proceedings of the 32nd International Conference on Machine Learning (pp. 2048-2057). JMLR Workshop and Conference Proceedings.
3. Dhivyaa, C. and Kuppusamy, Sangeetha and Marimuthu, Balamurugan and Amaran, Sibi and Vetriselvi, T. and Johnpaul, P. (2020) Skin lesion classification using decision trees and random forest algorithms. In Journal of Ambient Intelligence and Humanized Computing (10.1007/s12652-020-02675-8)

4. Ali, N., Akram, F., Hassan, T., et al.: A hybrid deep learning-based technique for skin lesion detection and classification. Journal of Ambient Intelligence and Humanized Computing 11, 3995–4007 (2020). https://doi.org/10.1007/s12652-020-02045-6
5. Youssef, A. M., Ahmed, H. A., El-Sawy, A. A.: Automated Skin Lesion Diagnosis using Deep Learning Techniques. Journal of Biomedical Informatics: X 9, 100093 (2021). https://doi.org/10.1016/j.ybmedx.2021.100093
6. Seeja, R. D., Suresh, A.: Deep Learning Based Skin Lesion Segmentation and Classification of Melanoma Using Support Vector Machine (SVM). Asian Pacific Journal of Cancer Prevention 20(5), 1555-1562 (2019). https://doi.org/10.31557/APJCP.2019.20.5.1555
7. P. Hu and T. Yang, "Pigmented skin lesion detection using random forest and wavelet-based texture," in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 10024, eds. Q. Luo, X. Li, Y. Gu, and Y. Tang, Oct. 2016, pp. 100241X. doi: 10.1117/12.2245149.
8. R. B. Bhatt, A. Dhall, G. Sharma, and S. Chaudhury, "Efficient skin region segmentation using low complexity fuzzy decision tree model," in Proceedings of the 2015 International Conference on Signal Processing and Communication (ICSC), Bangalore, India, 2015, pp. 259-264. doi: 10.1109/ICSC.2015.7203739.
9. Gautam, D. and Ahmed, M.: Melanoma detection and classification using SVM based decision support system. In: 2015 Annual IEEE India Conference (INDICON), pp. 1–6 (2015).
10. Wu, Yinhao and Chen, Bin and Zeng, An and Pan, Dan and Wang, Ruixuan and Zhao, Shen. Skin Cancer Classification With Deep Learning: A Systematic Review. IEEE Access, vol. 7, pp. 53021-53033, 2019. doi: 10.1109/ACCESS.2019.2915517.