

Using Ensemble of Multiple Fine-Tuned Efficient Net Models for Skin Cancer Classification

Mr. Rohit Daundkar¹, Mr. Kaustubh Shirke², Dr. Jasbir Kaur³, Assistant Professor Mr. Suraj Kanale⁴

Master's in Computer Applications (MCA) at Guru Nanak Institute of Management Studies,^{1,2}

Director GNIMS B-School, Head of Information Technology and HR at Guru Nanak Institute of Management Studies,³

Guru Nanak Institute of Management Studies, ⁴

Matunga, Mumbai, India

Abstract- Skin cancer is a prevalent form of cancer, and its early and accurate identification is critical for effective treatment. In this research paper, using an ensemble of fine-tuned Efficient Net models we proposed an improved approach for skin cancer classification. Our methodology incorporates data augmentation techniques to augment the dataset size, fine-tuning of the Efficient Net model by unfreezing the last few blocks, and employing an average ensemble for enhanced classification accuracy. The proposed approach when compared with other related work proved its effectiveness by outperforming them. Furthermore, our proposed ensemble method shows a precision value of 0.990, and accuracy of 0.988. Our findings demonstrate the effectiveness of the proposed methodology and its potential to significantly improve the diagnosis and treatment of skin cancer.

Keywords— efficient net, ensemble, skin cancer, transfer learning, fine-tuning

I. INTRODUCTION

With an increasing number of incidences in the past few decades, skin cancer has become a significant global health issue. It is difficult to estimate the true prevalence of skin cancer, due to underreporting and lack of diagnostic criteria. But as more epidemiologic investigations take place, it has been revealed that rates of both Melanoma skin cancer and Non-Melanoma Skin Cancer have been increasing for the past few decades [1].

Skin cancer is the most frequent kind of cancer, accounting for more than a million occurrences each year. In the United States alone, 1 in 5 people develops skin cancer in the course of their lifetime. It results in approximately 15000 fatalities and more than three billion dollars in medical expenses annually. Like many other cancers associated with environmental causes, the prevalence of skin cancer rises significantly with age. This reflects the long delay between development of cancer and carcinogen exposure. Skin Pigmentation and UV exposure also have a big impact on skin cancer [2].

The loss of the ozone layer over the past two decades has resulted in a greater exposure to UV radiation, raising the likelihood of various diseases in humans, including skin cancer [3]. Malignant melanomas have been estimated to account for 132,000 cases, whereas non-melanoma skin cancers account for 2 million cases globally [4].

Approximately 80% of occurrences of skin cancer are Basal Cell Carcinoma (BCC), making it one of the most prevalent kind of cancer. BCC usually starts as a little papule that develops slowly over a period of time, and eventually it

appears as a glossy papule having pearly edges. Approximately 16% of all cases are Squamous Cell Carcinoma (SCC) making it the second most prevalent form of skin cancer. The tumor can start as a scaly or rough area and progressively develop into a nodular growth with a plaque or warty surface. The cancer initially spreads locally, affecting the adjacent skin and lymph nodes, and can subsequently extend to nearby organs. Only 4% of skin cancer instances are malignant melanoma (MM), yet it causes 65% of deaths from the disease. Majority of Malignant Melanoma cells are detectable by visual examination as they start on skin's surface [5].

Typically to detect skin cancer, doctors use a biopsy approach. It is a slow and painful process in which a sample of potentially malignant skin lesions is removed and tested by the doctor. Thanks to the recent developments in computer-based technologies, skin cancer can be diagnosed more quickly, affordably and comfortably [6]. Therefore, accurate and timely detection of skin cancer is critical in determining the appropriate treatment and improving patient diagnosis outcomes. The increasing availability of digital imaging technology, coupled with recent advancements in machine learning and artificial intelligence algorithms, has presented new opportunities for developing automated systems to aid in the swift recognition and classification of skin cancer.

In this paper, we leveraged the power of ensemble learning and fine-tuned EfficientNet models to propose an improved approach for skin cancer classification. Ensemble learning techniques combine multiple individual models to create a stronger and more robust classifier, harnessing the diversity of their predictions and enhancing overall accuracy.

II. LITERATURE REVIEW

In [7], Adi Alhdhaif et al. used data balancing methods like SMOTE to propose a deep learning approach based on attention mechanism by choosing a soft attention-based module to obtain a feature map by focusing on input data features. The proposed method used the HAM10000 dataset to obtain an accuracy of 95.94% by using two convolutional layers, a soft attention module, followed by four convolution layers. Chao Xin et al. proposed a VIT based classifier that obtained an accuracy of 94.3% on the Ham10000 dataset. The author serialized the images using multi-scale and overlapping sliding windows along with multi-scale patch embedding [8]. It was followed with data augmentation and data normalization. Datta et al. compared the performance of four different CNN models (ResNet, Inception ResNetV2, DenseNet and VGG) with and without the use of Soft Attention mechanism. The best results were obtained when

Inception ResNet V2 was coupled with Soft-Attention, which achieved an accuracy of 93.4% on sampled and normalized HAM10000 dataset [9].

Saket S. Chaturvedi et al. performed fine-tuning of five pre-trained CNNs on the HAM10000 dataset. The best accuracy was 93.20% recorded for ResNetXt101 and InceptionResNetV2 each [10]. Karar Ali et al. trained multiple EfficientNet(B0-B7) models and achieved the best accuracy of 87.91% using EfficientNetB4. The author used data augmentation, noise-removal, resolution-scaling and fine-tuning on the HAM10000 dataset [11]. Namozo A. et al. used a four convolution Layer based CNN model that used an adaptive piecewise linear activation function. When evaluated on an augmented HAM10000 dataset, it achieved an accuracy of 95.86% [12].

Dataset which are imbalanced can produce problems in training a deep learning model. Using an imbalance dataset can produce skewed and biased predictions [14]. In real world problems, getting perfectly balanced data for each class in a dataset is difficult [15]. Thus, to tackle this problem, we have used Data Augmentation. It is used for increasing the amount of data by making minor adjustments to the training data already present or by generating new synthetic data from the existing data. Data Augmentation produces a balanced dataset by increasing the sample size for those imbalanced classes. As shown in Fig. 2., various data augmentation techniques were used like rotation, translation, zooming, and flipping etc. to create a balanced dataset.

III. PROPOSED SYSTEM

Data Preparation

In this research we used the HAM10000 dataset [13], which is a widely used and publicly available dataset in the field of dermatology research. The dataset contains seven classes of dermatoscopic images of skin lesions and a total of 10,015 images. The images of each class are shown in Fig. 1.



Fig. 1. Image classes in dataset [13]

Some of the classes in the dataset have a relatively higher number of images than other classes, thus making it highly imbalanced. Table 1 provides the details regarding the dataset.

Classes	Total Images
Melanocytic Nevus (NV)	6705
Melanocytic Melanoma (MEL)	1113
Benign Keratosis (BKL)	1099
Basal Cell Carcinoma (BCC)	514
Actinic Keratosis (AKIEC)	327
Vascular Lesion (VASC)	142
Dermatofibroma (DF)	115
Combined	10015

Table 1. Number of Images Per Classes

Dataset which are imbalanced can produce problems in training a deep learning model. Using an imbalance dataset can produce skewed and biased predictions [14]. In real world problems, getting perfectly balanced data for each class in a dataset is difficult [15]. Thus, to tackle this problem, we have used Data Augmentation. It is used for increasing the amount of data by making minor adjustments to the training data already present or by generating new synthetic data from the existing data. Data Augmentation produces a balanced dataset by increasing the sample size for those imbalanced classes. As shown in Fig. 2., various data augmentation techniques were used like rotation, translation, zooming, and flipping etc. to create a balanced dataset.



Fig.2. Examples of various data augmentation techniques.

Transfer Learning

Transfer learning is an effective technique that enables the transfer of knowledge from a source to a target domain, thereby enhancing the learning process and improving performance on the target task. By leveraging previously trained models that have been trained on large-scale datasets, transfer learning overcomes the limitations of insufficient labeled data. This approach allows the model to extract and generalize valuable features from the source domain to the target domain, leading to faster convergence and improved accuracy [16]. The ability of transfer learning to leverage prior knowledge has made it a widely adopted methodology across various domains. Its effectiveness lies in its capacity to harness the wealth of knowledge accumulated in one domain and apply it to a related task, thereby enabling more efficient and effective learning.

In our proposed method, we used transfer learning by including six models belonging to the family of EfficientNetV1 [17] and EfficientNetV2 [18].

Fine-Tuning

Fine-tuning is a widely employed approach in transfer learning, aimed at adapting a pre-trained model to a specific task or domain. Given that there is dissimilarity between the image dataset and the pre-trained weights, using the pre-trained weights would not result in an optimal solution. Thus, a fine-tuning process is conducted to refine the model's parameters and align them with the characteristics of the new image domain.

During fine-tuning, the pretrained model uses a new dataset to undergo further training, which enables it to learn domain-specific features and enhance its performance on the target task. By adjusting the model's parameters through this process, we can effectively adapt the model to the unique characteristics and intricacies of the medical image dataset. There are several ways to implement fine-tuning, this includes training some parameters of the last layers, to fully training all the layers.

We finetuned EfficientNetV1B3, EfficientNetV1B5, EfficientNetV1B7, EfficientNetV2B0, EfficientNetV2B2, EfficientNetV2B3. For EfficientNetV1B3 and EfficientNetV1B5, we unfreeze the last two blocks. For the models EfficientNetV1B7, EfficientNetV2B0, EfficientNetV2B2 and EfficientNetV2B3 we unfreeze the last block. To reduce the spatial dimensions of the feature maps, a global_average_pooling_2d layer was added to provide a fixed

size output. Further, multiple groups of dropout and dense layers were added. The dropout layer had a dropout rate of 0.2 and the dense layer had a ReLu activation function. Finally, a dense layer with a softmax [19] activation function was applied for multiclass classification.

Ensemble

Ensemble learning is a powerful technique that improves the performance and generalization by integrating the predictions of multiple models. Through the combination of diverse models, which may employ different algorithms, model architectures, or subsets of training data, ensembles overcome the limitations of individual models [20]. This is accomplished through voting, averaging, or weighted averaging of the constituent models predictions, resulting in a final prediction with enhanced accuracy and robustness. In our proposed model, we performed average ensemble learning using Average Layer. In this we assigned the same weights to all six models. Following formula is used to average the outputs of all the models.

$$\text{Prediction} = \frac{\sum P_i}{N}$$

Where P_i denotes the probability for each model and the number of models is denoted by N .

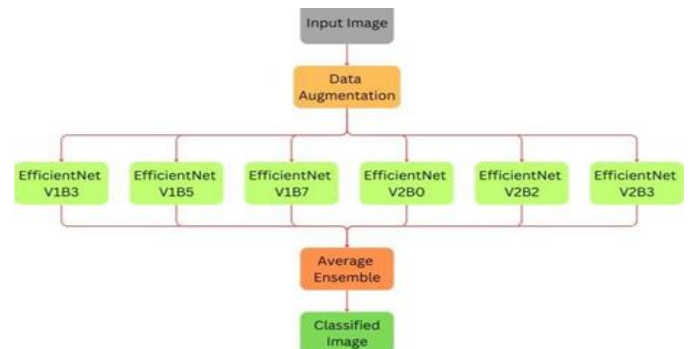


Fig. 3. Architecture of proposed method

IV. RESULTS

Setup

In our experiment, we used TensorFlow and Keras framework to build our neural network. We used a Kaggle environment to train and evaluate our model. Intel Xeon 2.20 GHz CPU (32 GB) and NVIDIA Tesla P100 GPU (16 GB) was equipped with the experimental hardware.

Evaluation

- **Precision:** It denotes the fraction of correctly predicted positive cases out of all predicted positive cases.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall:** It denotes the fraction of correctly predicted positive cases out of all actual positive cases.

$$\text{Recall} = \frac{TP}{\text{Actual Positives}}$$

- $$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy: It denotes the fraction of correct prediction out of all the cases.
- $$\text{F1 Score} = \frac{2PR}{P + R}$$

F1 Score: It is the harmonic mean of precision and recall and is a balanced evaluation of model's performance.

True Positive, True Negative, False Positive, and False Negative are represented by the letters TP, TN, FP, and FN respectively. Precision and Recall are denoted as P and R respectively.

The performance of the individual model and the proposed ensemble of fine-tuned EfficientNet models is shown in Table 2. Our proposed method outperforms the individual fine-tuned EfficientNet in precision, accuracy, recall and F1 score. Thus, assembling the individual model together improves their collective performance.

Model Name	Precision	Recall	F1 Score	Accuracy
EfficientNetV1B3	0.973	0.972	0.973	0.972
EfficientNetV1B5	0.969	0.967	0.969	0.968
EfficientNetV1B7	0.969	0.968	0.969	0.968
EfficientNetV2B0	0.971	0.967	0.971	0.970
EfficientNetV2B2	0.977	0.974	0.976	0.975
EfficientNetV2B3	0.971	0.968	0.971	0.971
Ensemble	0.990	0.986	0.989	0.988

Table2. Individual And Ensemble Model Metric

We compare the efficiency of our proposed model with that of various relevant and related approaches that uses skin cancer images to detect and classify skin cancer. We used accuracy and precision as the metrics to compare our proposed solution with that of other state-of-art methods as seen in Table (3). Thus, by outperforming other methods by a significant margin, our proposed method proved its effectiveness.

Author	Accuracy	Precision
Chao Xin et al. [8]	0.943	0.941
Datta et al. [9]	0.934	0.937
Saket S. Chaturvedi et al. [10]	0.932	0.870
Karar Ali et al. [11]	0.879	0.880

Kemal Polat et al. [21]	0.929	0.929
Zhiwei Qin et al. [22]	0.952	0.966
Our proposed method	0.988	0.990

Table 3. A Comparison of Existing Approaches With our Proposed Approach

V. CONCLUSION

In this research paper, we proposed an improved method that uses an ensemble of fine-tuned EfficientNet models for skin cancer classification. The HAM10000 dataset was used to assess the performance of our proposed method. However, the dataset was extremely imbalanced which could skew the result. To address these issues, we implemented various data augmentation techniques. We first individually trained and evaluated six different fine-tuned EfficientNet models. Additionally, we integrate the six individual models into a unified model, where their softmax outputs are averaged through the implementation of an average layer.

This ensemble method obtained an accuracy of 98.8% and precision of 99.0%, which outperforms not only our individual models, but also other state-of-art established methods by significant differences in these metrics. Thus, our proposed method will be helpful for dermatologists to make decisions in future. Our future work could focus on exploring additional ensemble techniques, incorporating other deep learning architectures, and evaluating the generalizability of the proposed approach on larger and more diverse skin cancer datasets. Ultimately, we aim to contribute to improved diagnostic accuracy, early detection, and treatment outcomes in the field of dermatology.

REFERENCES

- P. Gruber and P. M. Zito, Skin Cancer. StatPearls Publishing, 2023.
- J. D'Orazio, S. Jarrett, A. Amaro-Ortiz, and T. Scott, "UV Radiation and the Skin," International Journal of Molecular Sciences, vol. 14, no. 6. MDPI AG, pp. 12222–12248, Jun. 07, 2013.
- E. R. Parker, "The influence of climate change on skin cancer incidence – A review of the evidence," International Journal of Women's Dermatology, vol. 7, no. 1. Ovid Technologies (Wolters Kluwer Health), pp. 17–27, Jan. 2021.
- S. Mohapatra, N. V. S. Abhishek, D. Bardhan, A. A. Ghosh, and S. Mohanty, "Skin Cancer Classification Using Convolution Neural Networks," Lecture Notes in Networks and Systems. Springer Singapore, pp. 433–442, Jun. 12, 2020.

5. R. Gordon, "Skin Cancer: An Overview of Epidemiology and Risk Factors," *Seminars in Oncology Nursing*, vol. 29, no. 3. Elsevier BV, pp. 160–169, Aug. 2013.
6. M. Dildar et al., "Skin Cancer Detection: A Review Using Deep Learning Techniques," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10. MDPI AG, p. 5479, May 20, 2021.
7. A. Alhudhaif, B. Almaslukh, A. O. Aseeri, O. Guler, and K. Polat, "A novel nonlinear automated multi-class skin lesion detection system using soft-attention based convolutional neural networks," *Chaos, Solitons & Fractals*, vol. 170. Elsevier BV, p. 113409, May 2023.
8. A. Xin et al., "An improved transformer network for skin cancer classification," *Computers in Biology and Medicine*, vol. 149. Elsevier BV, p. 105939, Oct. 2022.
9. S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft Attention Improves Skin Cancer Classification Performance," *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data*. Springer International Publishing, pp. 13–23, 2021.
10. S. S. Chaturvedi, J. V. Tembhurne, and T. Diwan, "A multi-class skin Cancer classification using deep convolutional neural networks," *Multimedia Tools and Applications*, vol. 79, no. 39–40. Springer Science and Business Media LLC, pp. 28477–28498, Aug. 04, 2020.
11. K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4. Elsevier BV, p. 100034, Dec. 2022.
12. A. Namozov, D. Ergashev, and Y. I. Cho, "Adaptive Activation Functions for Skin Lesion Classification Using Deep Neural Networks," *2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS)*. IEEE, Dec. 2018.
13. P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1. Springer Science and Business Media LLC, Aug. 14, 2018.
14. A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," *2018 International Interdisciplinary PhD Workshop (IIPhDW)*. IEEE, May 2018.
15. J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1. Springer Science and Business Media LLC, Mar. 19, 2019.
16. A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung, and M. A. Azim, "Transfer learning: a friendly introduction," *Journal of Big Data*, vol. 9, no. 1. Springer Science and Business Media LLC, Oct. 22, 2022.
17. M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv*, 2019.
18. M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," *arXiv*, 2021.
19. C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv [cs.LG]*, 2018.
20. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10. Institute of Electrical and Electronics Engineers (IEEE), pp. 99129– 99149, 2022.
21. Polat and K. Onur Koc, "Detection of Skin Diseases from Dermoscopy Image Using the combination of Convolutional Neural Network and One-versus-All," *Journal of Artificial Intelligence and Systems*, vol. 2, no. 1. Institute of Electronics and Computer, pp. 80– 97, 2020.
22. Z. Qin, Z. Liu, P. Zhu, and Y. Xue, "A GAN-based image synthesis method for skin lesion classification," *Computer Methods and Programs in Biomedicine*, vol. 195. Elsevier BV, p. 105568, Oct. 2020.