# Transfer Learning-Based Multi-Disease Classification System For Retinal Fundus Images

**DHARSAN K (dharsan11032006@gmail.com),**
**P KEERTHANA (keerthaanaa21pk@gmail.com),**
**SACHIN PRADEEP (sachinpnair7@gmail.com)**
[1]Department of computer science, SRM Institute of science and technology
[2]Department of computer science, SRM Institute of science and technology
3Department of computer science, SRM Institute of science and technology

## I.ABSTRACT

This project represents a deep learning-based approach to multi class medical image classification, employing the Efficient Net-B0 architecture in a transfer learning framework. On the whole aim is to categorize the medical images into one out of the 29 disease categories Precisely. For preforming that, the system put forwards aggressive data augmentation methods to cut off the overfitting and improves the model's flexibility. Transfer learning provides the EfficientNet-B0 model to pre-train the huge dataset, adapt effectively toward a particular detail of the medical image data. Once the model got trained, Grad-CAM (Gradient-weighted Class Activation Mapping will be preferred to build a visual description which points a major part of each image that has notable impact on the classification decision. Upgrading techniques like fine-tuning learning rates and sophisticated regularization techniques were proposed to improve the existing performance. Experimental results represent the system has significant classification accuracy in all categories along with the power to localize the pathological characters in the images. This research paper mentions the ability of combining state-of-the-art deep learning methods with explainable AI tools to provide automated, accurate, and interpretable medical diagnosis.

## II. INTRODUCTION

Deep learning transformed the science of science imaging, gives better efficient tolls for precision, effective and automated diagnosis. Convolutional Neural Networks (CNNs) and their extensions has validated the unbelievably success in finding the intricate pattern along the medical images, lowering the conventional machine learning technique in most of the cases. Though with these improvement training deep neural networks from initial point is a big deal as of the necessity of vast annotated datasets and increasing computational cost. In majority of the medical applications, getting huge, labelled dataset is a big task as that consist of expert annotation and can be resource intensive.
Transfer learning arises forth as a feasible choice to connect these limitations by employing the models pre-trained on huge, general-purpose data and fine-tuning them to specific medical imaging tasks. Exploit the learned representations from those models, transfer learning facilities enhanced convergence, better performance, and less reliance on large quantities of labeled medical data.
In this study, we used EfficientNet-B0, that has a high convolutional neural network for its equilibrium across model efficacy and precision, under a transfer learning method to classify medical images into 29 disease categories. EfficientNet-B0's compound scaling strategy EfficientNet-B0's compound scaling strategy improves both depth and width and is as a result most appropriate in cases where computational resources can be restricted.
Moreover, to create the expectations of the model easier to understand, integrate Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM provides visual description by finding where within the input image which are very influential in the model's decision-making process, therefore offering insights into the reasoning of the model and instilling trust in its predictions among doctors.
Through the implementation of this technique, our research not merely predicted to reach high accuracy in the classification yet also to deliver interpretable results which can aid clinicians in interpreting and verifying computer-based diagnostic conclusions.

## III. RELATED WORK

Present studies in the medical image analysis have strong architecture of Conventional Neural network (CNN) like ResNet, Dense Net and VGG to resolve various medical imaging task. These architecture shows their ability to extract the complex structures and attributes of the medical image and thus provides considerably to the development of diagnostic accuracy. For your reference, ResNet's deep residual learning architecture have been used for detecting the radiology image abnormalities, whereas Dense Net have been used for improving the features reusability and accuracy in multi organ disease categories. Though VGG has less complexity in the architecture, it has been used in numerous image classification application with a emerging block as for various advanced models. However those models have been successful, there exist limitations especially in the area of training the deep networks over limited data, a matter that regularly came across in the cases of medical imaging.Transfer learning has mounted to become an efficient method in addressing the restricted labelled data tasks. Through fine tuning pre-trained models on huge datasets such as ImageNet, to targeted medical dataset, transfer learning provides the models to perform well lacking the required amount of labelled medical information. The studies have represented the effective development in the classification accuracy with the help of transfer learning while there is a restriction on the labelled medical data. To illustrate, fine tuning the pre-trained models on few medical datasets has proven to access the better results over training from scratch as of the reduce the risk of over fitting and improvised generalization to emerging cases.

Moreover, the model performance progress, deep learning model transparency has acquired huge interest, especially in the medicine field, in which doctors require to grasp the rationale for prediction done by the model. Visualization methods particularly Grad-CAM (Gradient - weighted Class Activation Mapping) has been broadly recognized the developed model's visibility. Grad-CAM offers visual description by focusing on the region of an image which is very important for the model's prediction, assisting clinicians in knowing where to be diagnosed using the image model. This proposed technique has shown the clinician belief on AI systems, facilitating the model’s ability in large stakes settings .Although the developments are existing, the literature lacks the gaps when it reaches the integrity at more efficient architectures and robust interpretability specially for multi class medical image classification. Though the studies has concentrated on optimizing the model efficiency, they often ignore interpretability that is important for the clinical adoption. In contrast , efforts that prioritize the interpretability occasionally overlook the ability of utilization of the effective and lightweight models that suits for real time deployment. This works looks to compensate the gap by offering the EfficientNet-B0 architecture, that gives a balance between the efficiency and accuracy also by using the Grad-CAM for developed model clarity in multi-class medical diagnosis

## IV. METHODOLOGY

### A. DATASET PREPARATION

The dataset used in this study was labeled medical images, and each image had been preliminarily labeled to emblematize one of 29 medical conditions. The images were first structured by laying out image paths in relation to their markers. To expand the size of the dataset and further ameliorate model conception, colorful data addition ways were employed. These comported of arbitrary gyration, vertical and perpendicular flip, scaling, and color jittering. These data addition styles are particularly pivotal in the case of medical image bracket, when there's no way to collect a large dataset heterogenous in nature. The augments also made the model see further images with variations and hence more stable features but lower possibility of overfitting.

### B. MODEL ARCHITECTURE

We utilized Efficient Net- B0 framework then, which inspire-trained on ImageNet, a significantly large-scale, general-purpose dataset. Efficient Net- B0 has also been known for its good parameter but accurate performance, due to which this model is extremely suitable to apply in medical image bracket issues when computational effectiveness matters a lot. The bracket subcaste of the base Efficient Net- B0 model was altered to affair 29 classes, as they're the 29 medical conditions of our data set. It was a core revision to transfigure the model to perform the veritably function of multi-class medical image bracket. Fine-tuning was done by unfreezing lower layers of the pre-trained Efficient Net- B0 model. This enabled the model to transfer the learned features from ImageNet to drug in a way that the model could identify applicable patterns in the environment of medical images. The before layers of the model recycling low- position features like edges and textures were saved, while the deeper layers were fine- tuned similar that further sphere-specific representation for medical opinion could be attained.

### C. TRAINING STRATEGY

We used the double cross-entropy with logits loss during model training because it works best with multi-label bracket issues since an image may be in multiple classes of complaint at the same time. We have applied the Adam optimizer as it has an adaptive capability in literacy rate improvement, which is benefit imposing it for allowing the deep literacy models to adopt a more refined design. We also used literacy rate scheduling to stoutly modernize the literacy rate during training in order to enable quicker model confluence and possible overfitting or original minima avoidance. Automatic mixed perfection (AMP) was used at training time for calculation speedup without loss of numerical stability to enable hastily confluence as well as memory savings. colorful criteria were used to validate the model's performance. Training and validation loss were observed to prohibit the model from overfitting the training data. Area Under the Receiver Operating Characteristic wind (AUROC) was tracked as a measure of performance. AUROC is a suitable metric for multi-class bracket problems since it provides an estimate of the model's capability to distinguish between classes and informs us about the discriminative power of the model across all classes.

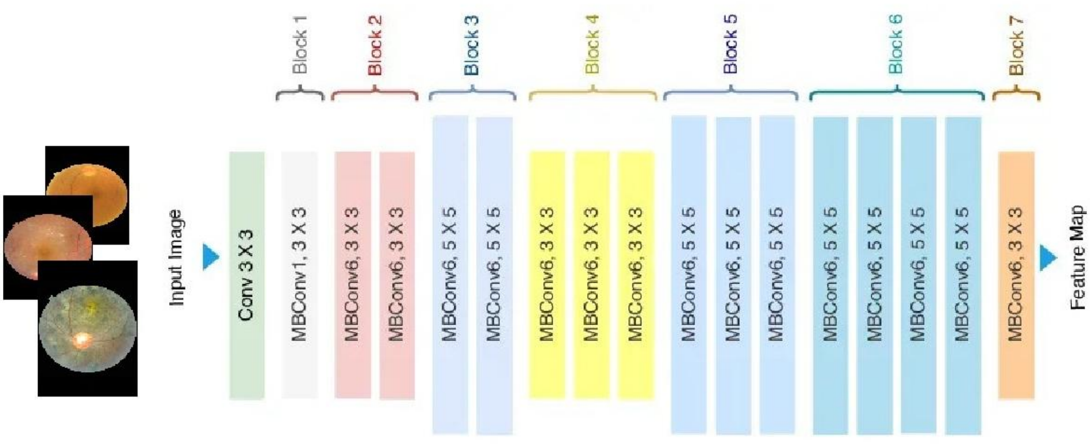
### *D. Grad-CAM VISUALIZATION*

Following the subsequent training of the model, we employed Grad-CAM (Gradient-weighted Class Activation Mapping) to visualize and explain the model's decision-making process

Grad- CAM is a system that identifies the image regions most responsible for the vaticination of the model by calculating the slants of the class score with respect to the point charts of the convolutional layers. By imaging similar emphasized regions, Grad-CAM gives experimenters and clinicians a further intuitive sense of where the model is fastening, commodity necessary in order to realign the model's predictions with human cognition and clinical environment. We ran Grad- CAM on the test samples, and that enabled us to confirm what areas the model named as salient for a certain complaint condition and thus supported us in attesting the model's attention points and explainability.

## V. ARCHIETECTURE PROPOSED

EfficientNet-B0 is a strong yet precise compound scaling optimized convolutional neural network model. Sequential significant features are learned by the network in an effort to analyze the retinal fundus images. Input image first goes through normal convolutional layer with filter size 3×3 and stride 2. Convolutional layer is used to reduce spatial resolution as much as possible but increase feature depth in a manner that the feature could be handled efficiently in the future. After the very first convolution, the network goes through seven more consecutive blocks made up of MBConv layer. Block 1 uses MBConv1 with expansion factor of 1 and kernel size of 3×3, and light feature extraction is prioritized the most in this case. Block 2 uses MBConv6 blocks with kernel size of 3×3 to expand the network capacity so more latent features can be learned. Block 3 uses MBConv6 blocks with a kernel size of 5×5 to get a larger receptive field and more context information. Block 4 uses MBConv6 layers with kernel size 3×3 to collect more local information. Blocks 5 and 6 use MBConv6 layers with kernel size 5×5 to further increase the network to learn even higher abstractions at even larger spatial areas. Block 7 continues with a layer of MBConv6 with kernel size 3×3 for down sampling learned features to dense feature map shape to be fed into downstream tasks. Depth wise separable convolutions, squeeze-and-excitation optimization, and residual connections are utilized in every MBConv block with no computational expense to further boost feature extraction. Spatial down sampling is also used in carefully chosen blocks with the goal of down sampling resolution and up sampling feature channel depth at a gradually balanced ratio based on Efficient Net's compound scaling concept of equally scaling network depth, width, and input resolution.



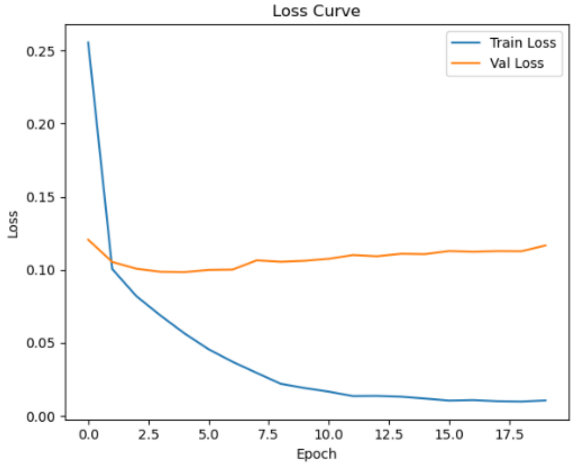
## VI. RESULT

In this work, we propose a convolutional network-based multi-task learning framework for joint retinal disease classification and lesion localization. The RFMiD dataset, which included the 29 disease classes "Disease Risk," "DR," "ARMD," "MH," "DN," and "MYA," among others, was used to test the model.

Retinal fundus images with 224 x 224 resolution that had been preprocessed into batches of size 16 made up the dataset. To maximize computational efficiency, training was conducted over 20 epochs with varying degrees of precision.

### A. TRAINING AND VALIDATION PERFORMANCE

Furthermore, at epoch 7, the training loss decreased from roughly 0.26 to a terminal value of 0.0370. The plateauing validation loss of approximately 0.1001 also suggests convergence with minimal overfitting. The slope of the loss curves showed that the model's capacity to generalize remained constant throughout training. Importantly, Area Under the Receiver Operating Characteristic curve (AUROC) was the primary metric used to assess classification performance. It achieved the optimal epoch validation AUROC of 0.8466. AUROC scores increased gradually over the first few epochs before peaking at epoch 6 and becoming remarkably oscillation-stable after that. Furthermore, the model checkpointing technique preserved the ideal model weights as established by the highest validation AUROC.




### B. TEST SET EVALUATION

After evaluation on an unseen test dataset, the model achieved a Test Loss of 0.0945 and a Test AUROC of 0.8374. This is validation of the model's generalization performance and capability on different retinal disease classes, further validating the effectiveness of the multi-task learning approach to maximizing localization and classification objectives simultaneously.

### C. OBSERVATIONS AND ANALYSIS

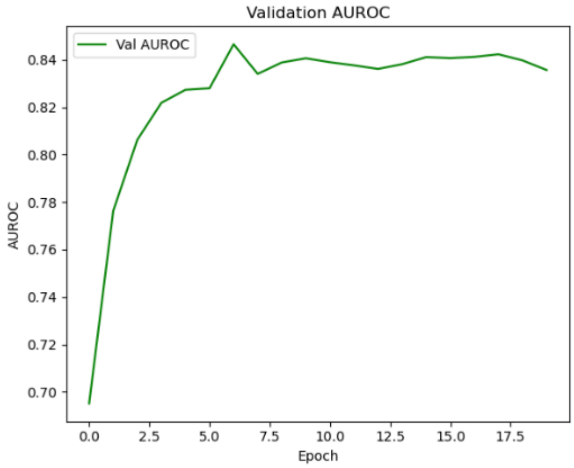
1.Training Dynamics: The steady decrease in training loss coupled with the constant validation loss signifies minimal overfitting. This is a favorable outcome, especially in a high-class multi-label setting.

2.AUROC Trends: The AUROC curve of the validation depicts initial rapid improvement (from ~0.69 to ~0.83 in the first 5 epochs), pointing towards the observation that the model acquired discriminative characteristics pertinent to retinal disorders very quickly.

3.Generalization: The close agreement between test AUROC (0.8374) and validation AUROC (0.8466) indicates good generalization of the model, which is essential for clinical deployment when unseen patients' data are to be handled robustly.

4.Computational Issues: Training was based on efficient batch-wise computation and mixed-precision acceleration (torch.cuda.amp.autocast) with nearly complete hardware utilization without numerical stability warning beyond standard deprecation warning messages.

5.Class Imbalance: Though not addressed as such here, there is the acknowledgement that the sparsely distributed classes ('AION', 'PT') can be a target of lessened recall due to dataset imbalance in the underlying data, and something that has scope for potential future enhancement utilizing strategies such as class-specific weighting or augmentation.




## VI. DISCUSSION

Experiment results verify the effectiveness of the suggested multi-task learning paradigm to solve the two tasks of retinal disease classification and lesion localization. The discriminative features learned were suitable for a large variety of 29 classes of retinal disease, and the model was able to get a Test AUROC of 0.8374 and a Test Loss of 0.0945. The above findings validate the hypothesis that one model can efficiently deal with the complexity involved in multi-label, multi-class medical image tasks. One of the notable observations of training dynamics is the quick convergence seen in the initial epochs. The validation AUROC increased from approximately 0.69 to more than 0.83 by epoch 5, indicating that the early-in-training learned feature representations were very informative. Furthermore, the ultimate Validation AUROC of 0.8466 indicates good model performance without any visible signs of overfitting, as expected by the respective test performance.

The loss plots demonstrate steady decrease in the loss in training and plateaued loss during validation at approximately 0.1. This is indicative of a common behavior of a regularized model learning good features without overfitting the data. Mixed-precision training also further improved computational efficiency without impacting the performance of the model, thereby validating its application in retinal image analysis tasks with large datasets. However, the slight variation in validation AUROC at subsequent training points (epochs 7–20) indicates that although the model is very sensitive, slight tweaking to further increase specificity in a few sparsely populated classes might be required. Rare conditions like 'AION', 'PT', and 'RP' would have been the reason for marginal instability due to severe class imbalance of the data. This is an acknowledged issue in medical imaging and implies the possible advantage of eventual integration with imbalance treatment approaches, e.g., focal loss, oversampling, or cost-sensitive learning.

Secondarily, although superb classification performance by the model, localization performance per se — although not quantitatively measured here — is of greatest import in clinical translation considerations. To be followed in subsequent studies will be the measurement of lesion detectability accuracy metrics like Intersection over Union (IoU) or Dice scores for the express purpose of making overt localization capacity verification. In short, the suggested method provides a robust basis for joint disease classification and lesion localization in retinal images with comparable performance and effective utilization of resources. Such encouraging results pave the way to future work on handling rare or infrequent classes of diseases, enhancing the accuracy of lesion localization, and integrating the model into clinical decision support systems.

## VII. CONCLUSION

This work enhances the ability of the EfficientNet-B0 model in a transfer learning paradigm, coupled with Grad-CAM to enable model interpretability, in order to offers high performance for multi-class medical image classification. Through the application of pre-trained model and it's fine tuning on a targeted medical data set, we demonstrated the capability of transfer learning in moderating the problem of limited labeled samples and computational cost in the health sectors. The Grad-CAM utilization further raises the interpretability of the model, providing informative observations to the very significant regions of medical images that are responsible for the classification, which is crucial for clinical acceptance and trust in computerized diagnostic systems. The experimental results demonstrate uplifting results, with outstanding classification performance over

29 different medical conditions, mentioning that the recommended method is consistent and scalable for practical medical use. By contributing a means of visualizing the affected regions within images, this technique can help healthcare practitioners detect pathological features with high accuracy, aiding their diagnostic decision-making.

There are much to be done in the future to enhance the performance of the system and broaden the application field. One of them is that priority should be given to integrate the ensemble methods into a common model in the attempt to further enhance the robustness of the classifying and prevention from the risk of errors. Yet another list of advanced methods of data balancing, such as data combination or cost-sensitive learning, will also be covered to address class imbalances normally found in medical data. Clinical validation of the model on unseen and varied patient data will then be the next step needed to determine true-world reliability and accuracy of the model in the clinical environment. On the whole ,this work builds a resilient foundation for the development of automated, explainable, and highly accurate diagnostic systems that significantly improve medical image analysis, assisting clinical practitioners in making timely and accurate diagnoses.

## REFERENCE

1. M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *arXiv preprint arXiv:1905.11946*, 2019.

2. X. Wang et al., “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 3462–3471.

3. R. R. Selvaraju et al., “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.

4. D. S. Kermany et al., “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, Feb. 2018, doi: 10.1016/j.cell.2018.02.010.

5. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 3–11.

6. V. Gulshan et al., “Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, Dec. 2016, doi: 10.1001/jama.2016.17216.

7. V. Cheplygina, M. de Bruijne, and J. P. W. Pluim, “Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis,” *Med. Image Anal.*, vol. 54, pp. 280–296, Jul. 2019, doi: 10.1016/j.media.2019.03.009.

8. Q. H. Nguyen et al., “RFMiD: A Dataset for Multi-Disease Retinal Fundus Image Classification,” *arXiv preprint arXiv:2011.07391*, 2020.

9. P. Rajpurkar et al., “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” *arXiv preprint arXiv:1711.05225*, 2017.

10. G. Litjens et al., “A Survey on Deep Learning in Medical Image Analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.