Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Squeeze-and-Excitation Channel Attention in a Compact CNN for 24-Class Clinical Skin Disease Classification

Md. Ariful Islam

sm.islam.ariful@gmail.com

East Delta University

Research Article

Keywords: skin disease classification, squeeze-and-excitation, channel attention, convolutional neural network, ablation study, medical image analysis

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11656703/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Squeeze-and-Excitation Channel Attention in a Compact CNN for 24-Class Clinical Skin Disease Classification

Md. Ariful Islam
Department of Computer Science and Engineering, East Delta University
Chattogram, Bangladesh
sm.islam.ariful@gmail.com

***Abstract*—Access to dermatologists is limited in many low-resource regions, which creates a need for automated triage tools that are accurate yet light enough for modest hardware. This paper presents a compact convolutional neural network (CNN), trained from scratch, that places a Squeeze-and-Excitation (SE) channel-attention block in each of its four convolutional stages. From a public collection of clinical skin-disease photographs, a class-balanced dataset of 68,568 images in 24 classes (23 disease categories and healthy skin) was built by randomly undersampling every category to 2,857 images. On a held-out split of 13,728 images, the SE-CNN (5.01 M parameters) reached 88.55% top-1 accuracy and a macro F1 of 0.888, with a micro-averaged ROC AUC of 1.00. In a controlled ablation on a reduced subset, adding SE blocks to an otherwise identical network raised validation accuracy from 64.2% to 73.1% and test macro F1 from 0.49 to 0.59 while adding less than 1% to the parameter count. On the same subset, the SE-CNN also exceeded MobileNetV2 and DenseNet121 trained under the same protocol. The paper closes with an explicit account of the evaluation limits of this study, including image-level splitting and the reuse of the held-out split for early stopping, and the experiments needed before clinical use.**

***Index Terms*—Skin disease classification, squeeze-and-excitation, channel attention, convolutional neural network, ablation study, low-resource healthcare**

I. Introduction

Skin conditions are among the most common reasons people seek medical care, ranging from eczema and fungal infections to life-threatening melanoma. In rural and low-income regions, including much of Bangladesh, patients often wait a long time to see a dermatologist or never see one, and many self-treat with over-the-counter products. Automated image-based triage could shorten this path if it is accurate and can run on the hardware that small clinics actually have.

Convolutional neural networks (CNNs) have shown strong results in dermatology. Esteva *et al.* trained Inception-v3 on 129,450 clinical images and reported performance comparable to board-certified dermatologists [1], and later studies confirmed expert-level results on dermoscopic melanoma recognition [2], [3]. Most of these results, however, rely on large pretrained backbones. Attention mechanisms offer a cheaper route to better features: the Squeeze-and-Excitation (SE) block [4] learns a weight for every feature channel from global context and rescales the feature maps at negligible computational cost.

This work asks a narrow question: how much does SE channel attention contribute to a compact CNN trained from scratch on a broad, multi-class set of clinical skin photographs? The contributions are:

1) a compact SE-CNN (5.01 M parameters) for 24-class classification of clinical skin-disease photographs, trained without transfer learning;
2) a class-balanced 68,568-image benchmark built by undersampling, with no duplicated images;
3) a controlled ablation that isolates the effect of the SE blocks under identical data and training settings;
4) an explicit statement of the evaluation limits of the study and of the experiments required before clinical use.

II. Related Work

Large-scale CNNs. Following Esteva *et al.* [1], Han *et al.* fine-tuned ResNet-152 on 19,398 clinical images covering 12 diseases [5], Haenssle *et al.* compared an Inception-v4 model with 58 dermatologists [2], and Brinker *et al.* showed that a CNN trained on dermoscopic images performed on par with 145 dermatologists [3]. Classic architectures such as VGGNet [6], GoogLeNet [7] and ResNet [8] became standard backbones, and Lopez *et al.* showed that transfer learning with VGGNet is effective when medical datasets are small [9].

Lightweight and attention-based models. Srinivasu *et al.* combined MobileNetV2 with LSTM units for resource-constrained devices [10]. Nirupama and Virupakshappa built a hybrid of MobileNetV2, SE blocks, atrous spatial pyramid pooling and channel attention, evaluated on several benchmark datasets [11]. Demyanov *et al.* showed that even a five-layer CNN is useful for dermoscopy pattern classification [12].

Feature fusion and optimisation. Nasr-Esfahani *et al.* used preprocessing and a CNN for melanoma detection in clinical images [13]. Akram *et al.* [14], Bibi *et al.* [15] and Dillshad *et al.* [16] fused deep features from several networks and selected them with optimisation algorithms. These pipelines are effective but complex. Few studies isolate the contribution of SE attention inside a compact network trained from scratch on a

broad multi-class set of clinical photographs, which is the gap this paper addresses.

III. Dataset and Preprocessing

A. Data Source and Curation

The images come from the public *SkinDisease* collection on Kaggle [17], which aggregates clinical photographs from several sources with wide variation in resolution, lighting, framing and body site. The collection contains 35 folders. Eleven were excluded: the generic "Benign" and "Malignant" folders, umbrella folders that overlap finer classes (combined fungal, bacterial and viral groupings, and three separate fungal folders), and the tumour folders (melanoma, nevi and moles; actinic keratosis, basal cell carcinoma and other malignant lesions; seborrheic keratoses and other benign tumours). The 24 retained classes comprise 23 disease categories and one *healthy skin* class. The disease categories span inflammatory conditions (e.g., eczema, atopic dermatitis, psoriasis and lichen planus, contact dermatitis, urticaria), bacterial, viral and parasitic infections (e.g., cellulitis, impetigo, chickenpox, shingles, cutaneous larva migrans), connective-tissue, bullous and systemic disease, vascular lesions, pigmentation disorders, and hair and nail disease. Because the tumour folders were excluded, the model does not address skin cancer.

Every category was randomly undersampled without replacement to 2,857 images, giving 68,568 images in a perfectly balanced set. No image was duplicated or synthetically generated before splitting, so class balance cannot place copies of the same file in both partitions.

B. Split and Preprocessing

The balanced set was split at random, at the image level, into 80% for training (54,840 images) and 20% held out (13,728 images, 572 per class). Images were resized to 224 × 224 pixels and rescaled to [0, 1]. Data augmentation was applied on the fly to training batches only: random rotation (up to 30°), zoom (0.3), horizontal and vertical shifts (0.2), shear (0.2) and horizontal flipping, with nearest-neighbour filling of empty pixels.

IV. Method

A. Network Architecture

Fig. 1 shows the SE-CNN. It has four convolutional blocks with 64, 128, 256 and 512 filters. Each block applies a 3 × 3 convolution with ReLU, batch normalisation [18], an SE block, a second 3 × 3 convolution with ReLU, batch normalisation, and 2 × 2 max pooling, so the spatial size falls from 224 to 14 while the channel depth rises to 512. The classifier applies global average pooling, a 512-unit fully connected layer with ReLU, dropout with rate 0.5 [19], and a 24-way softmax. Table I summarises the layers. The full model has 5.01 M parameters.

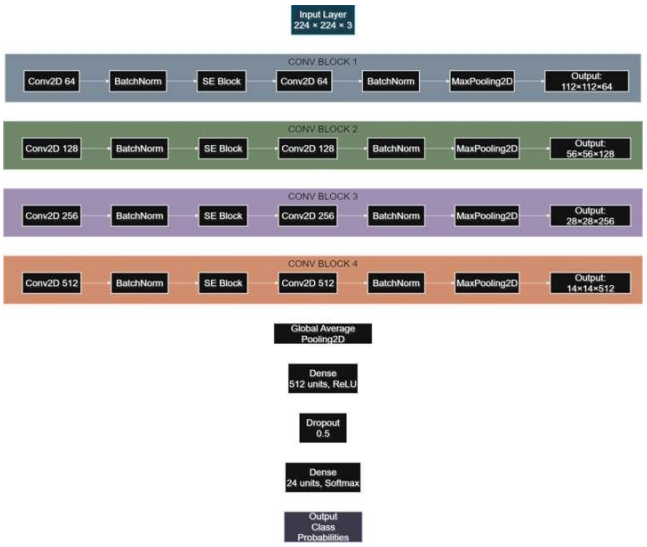


Fig. 1. Architecture of the proposed SE-CNN. Each of the four convolutional blocks contains an SE block after the first convolution and batch normalisation.

TABLE I
Layer Summary of the SE-CNN

Stage	Operations	Output
Input	RGB image	224 × 224 × 3
Block 1	2 × Conv 3×3 (64), BN, SE, MaxPool	112 × 112 × 64
Block 2	2 × Conv 3×3 (128), BN, SE, MaxPool	56 × 56 × 128
Block 3	2 × Conv 3×3 (256), BN, SE, MaxPool	28 × 28 × 256
Block 4	2 × Conv 3×3 (512), BN, SE, MaxPool	14 × 14 × 512
Head	GAP, FC 512 (ReLU), Dropout 0.5	512
Output	FC 24, softmax	24

B. Squeeze-and-Excitation Block

Given a feature map $\mathbf{U}$ with C channels of size $H \times W$, the SE block first *squeezes* each channel to a single value by global average pooling:

$$z_c = (1 / HW) \, \Sigma_{i,j} \, u_c(i, j). \quad (1)$$

It then *excites* the channels through a two-layer bottleneck with reduction ratio r,

$$\mathbf{s} = \sigma(\mathbf{W}_2 \, \delta(\mathbf{W}_1 \, \mathbf{z})), \quad (2)$$

where δ is ReLU, σ is the sigmoid, $\mathbf{W}_1 \in \mathbb{R}^{(C/r)\times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C\times(C/r)}$. Finally, each channel is *rescaled*: $\tilde{u}_c = s_c \cdot \mathbf{u}_c$. With $r = 16$, the first SE block ($C = 64$) adds only 580 parameters, and the four SE blocks together add about 44.5 k parameters, under 1% of the model.

C. Training

All models were trained with the Adam optimiser [20] at a learning rate of 10^{-4}, categorical cross-entropy loss and a batch size of 64, for up to 30 epochs. Early stopping monitored the validation loss with a patience of 5 epochs and restored the best weights. The held-out split was evaluated at the end of every epoch and served as the validation set, so the reported model is the epoch with the lowest held-out loss (epoch 29, training

accuracy 94.3%); Section VII discusses the consequence of this choice.

D. Baselines and Ablation

Two kinds of comparison were run on a reduced subset of the data to limit training cost. First, for the ablation, a *Baseline CNN* identical to the SE-CNN but with every SE block removed was trained from scratch with the same data, augmentation and hyperparameters. Second, four standard architectures, MobileNetV2, DenseNet121, ResNet50 and EfficientNetB0, were trained with the same pipeline. Performance is reported as top-1 accuracy, macro-averaged precision, recall and F1, the confusion matrix, and the micro-averaged ROC curve.

V. Results

A. Full Dataset

On the 13,728 held-out images, the SE-CNN reached **88.55%** top-1 accuracy, against a chance level of 4.2% for 24 balanced classes, with macro-averaged precision, recall and F1 of 0.893, 0.886 and 0.888 (Table II). Because the classes are balanced, balanced accuracy equals top-1 accuracy. The confusion matrix in Fig. 2 is strongly diagonal. The two weakest classes are eczema (F1 0.66) and psoriasis, lichen planus and related diseases (F1 0.61), which overlap visually with each other and with other inflammatory dermatoses. At the other extreme, six classes reach F1 of 0.99 or higher: healthy skin, bacterial cellulitis, impetigo, cutaneous larva migrans, chickenpox and shingles. The micro-averaged ROC curve (Fig. 3) gives an AUC of 1.00.

TABLE II
Full-Dataset Results on the Held-Out Split

Metric	Value
Top-1 accuracy	88.55%
Balanced accuracy	88.55%
Macro precision / recall / F1	0.893 / 0.886 / 0.888
Weakest classes (F1)	Psoriasis & lichen planus 0.61; eczema 0.66
Classes with F1 ≥ 0.99	6 of 24
Micro-averaged ROC AUC	1.00

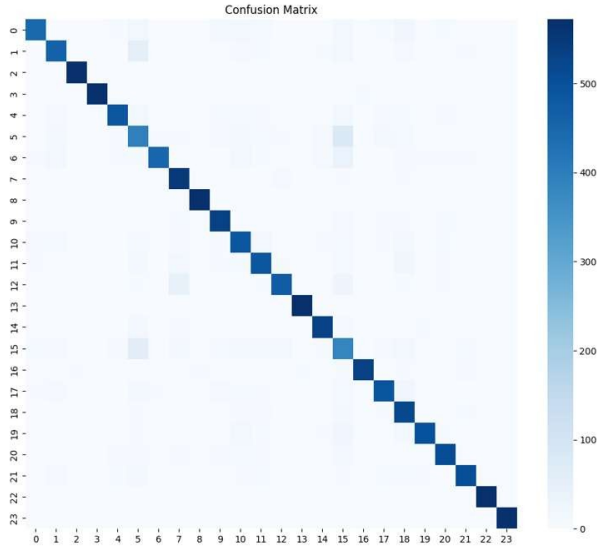


Fig. 2. Confusion matrix of the SE-CNN on the 13,728-image held-out split (24 classes, 572 images per class).

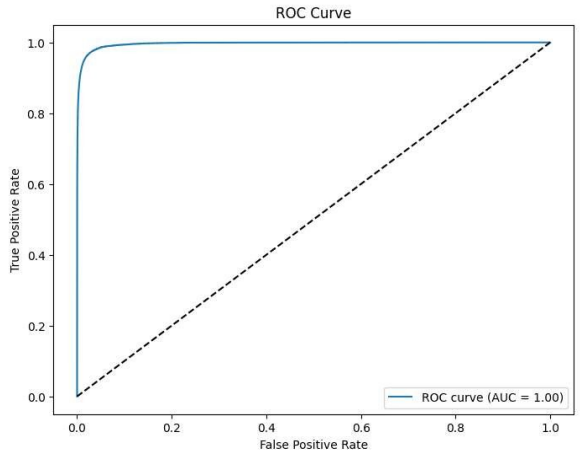


Fig. 3. Micro-averaged ROC curve of the SE-CNN on the held-out split (AUC = 1.00).

B. Comparison with Standard Architectures

Table III and Fig. 4 summarise the subset comparison. The SE-CNN reached 87.4% training and 72.9% validation accuracy, and 59.9% test accuracy, against 51.6% for the best standard architecture, MobileNetV2, and below 50% for DenseNet121. ResNet50 and EfficientNetB0 did not converge in this pipeline: EfficientNetB0 stayed at 4.3%, essentially the chance level of 4.2%. Such results point to an incompatibility between these models and the shared training pipeline (for example, in the input scaling each architecture expects) rather than to a weakness of the architectures, so they are reported for completeness but excluded from the conclusions.

TABLE III
Subset Comparison with Standard Architectures

Model	Params	Train acc.	Val. acc.
MobileNetV2	2.29 M	48.1%	50.8%
DenseNet121	7.06 M	42.7%	46.5%
ResNet50*	23.6 M	8.6%	10.9%

Model	Params	Train acc.	Val. acc.
EfficientNetB0*	4.08 M	4.3%	4.3%
SE-CNN (ours)	5.01 M	87.4%	72.9%

*Did not converge under the shared training pipeline; excluded from conclusions.

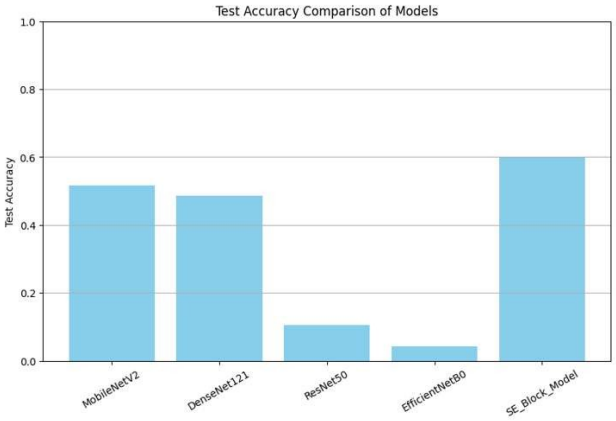


Fig. 4. Test accuracy of each model on the reduced subset.

C. Ablation of the SE Blocks

Table IV isolates the effect of channel attention. With all other factors fixed, adding SE blocks raised validation accuracy by 8.9 percentage points, lowered the loss from 1.32 to 0.89, and improved test macro precision, recall and F1 by 9 to 10 points each. For the SE-CNN, the test F1 of 0.59 agrees with its subset test accuracy of 59.9% in Fig. 4, as expected for balanced classes.

TABLE IV
ABLATION: EFFECT OF SE BLOCKS (REDUCED SUBSET)

Variant	Val. acc.	Loss	P	R	F1
Baseline CNN (no SE)	64.2%	1.32	0.52	0.50	0.49
SE-CNN	73.1%	0.89	0.61	0.60	0.59

Accuracy is measured on the validation split; P, R and F1 are macro-averaged on the subset test split. Both variants share data, augmentation and hyperparameters.

VI. DISCUSSION

The ablation is the most direct evidence in this study: in a controlled comparison, SE channel attention added about nine points of validation accuracy and ten points of test F1 for less than 1% more parameters. This supports the use of cheap attention modules when model size matters. At 5.01 M parameters, the SE-CNN is far smaller than ResNet50 (23.6 M) and DenseNet121 (7.06 M), comparable to EfficientNetB0 (4.08 M) and about twice the size of MobileNetV2 (2.29 M).

The gap between the subset (59.9% test accuracy) and the full dataset (88.55%) shows that a network trained from scratch benefits strongly from more data. The concentration of errors in a few categories suggests where clinical review effort, targeted data collection or hierarchical labelling would help most.

VII. LIMITATIONS AND FUTURE WORK

Several limitations bound how far these results can be taken:

1) **Image-level split.** The source collection provides no patient identifiers, so photographs of the same patient or lesion may fall in both partitions. Such overlap inflates held-out performance, and the AUC of 1.00 should be read with this in mind.

2) **Held-out split reused for early stopping.** Because the 20% split also guided early stopping, it is not a fully independent test set.

3) **Single split.** No cross-validation or confidence intervals are reported.

4) **Possible source shortcuts.** The near-perfect classes carry folder prefixes ("Ba", "Vi", "Pa") that suggest they came from a different source than the rest, so the model may partly recognise image source rather than disease.

5) **Baselines.** The standard architectures were not tuned individually, and two failed to converge.

6) **Population.** No skin-tone metadata is available, so performance on darker skin types, which are common in South Asia, is unknown.

7) **Interpretability.** No saliency analysis such as Grad-CAM is reported in this version.

Future work will (i) use a separate validation set and an untouched test set, (ii) report cross-validated results with confidence intervals, (iii) tune baselines fairly, including pretrained variants, (iv) evaluate on an external dataset with patient-level separation and skin-tone annotation, such as Fitzpatrick 17k [21], and (v) add saliency maps to check that the model attends to lesions rather than background.

VIII. CONCLUSION

A compact SE-CNN trained from scratch reached 88.55% accuracy on a 24-class, class-balanced set of clinical skin-disease photographs, and a controlled ablation attributed roughly nine points of validation accuracy to SE channel attention at under 1% parameter cost. The results support lightweight attention-based CNNs as a starting point for dermatology triage in low-resource settings, subject to the patient-level and external validation outlined above.

ACKNOWLEDGMENT

An AI-based writing assistant was used for language editing and formatting of this manuscript. All experiments, analyses and conclusions are the author's own, and the author takes full responsibility for the content.

REFERENCES

[1] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[2] H. A. Haenssle *et al.*, "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Ann. Oncol.*, vol. 29, no. 8, pp. 1836–1842, 2018.

[3] T. J. Brinker *et al.*, "A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task," *Eur. J. Cancer*, vol. 111, pp. 148–154, 2019.

[4] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.

[5] S. S. Han *et al.*, "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm," *J. Invest. Dermatol.*, vol. 138, no. 7, pp. 1529–1538, 2018.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.

[7] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 1–9.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[9] A. R. Lopez, X. Giro-i-Nieto, J. Burdick, and O. Marques, "Skin lesion classification from dermoscopic images using deep learning techniques," in *Proc. 13th IASTED Int. Conf. Biomed. Eng. (BioMed)*, 2017, pp. 49–54.

[10] P. N. Srinivasu *et al.*, "Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, 2021.

[11] Nirupama and Virupakshappa, "MobileNet-V2: An enhanced skin disease classification by attention and multi-scale features," *J. Imaging Inform. Med.*, 2024, doi: 10.1007/s10278-024-01271-y.

[12] S. Demyanov, R. Chakravorty, M. Abedini, A. Halpern, and R. Garnavi, "Classification of dermoscopy patterns using deep convolutional neural networks," in *Proc. IEEE 13th Int. Symp. Biomed. Imaging (ISBI)*, 2016, pp. 364–368.

[13] E. Nasr-Esfahani *et al.*, "Melanoma detection by analysis of clinical images using convolutional neural network," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2016, pp. 1373–1376.

[14] T. Akram *et al.*, "Dermo-optimizer: Skin lesion classification using information-theoretic deep feature fusion and entropy-controlled binary bat optimization," *Int. J. Imaging Syst. Technol.*, vol. 34, p. e23172, 2024.

[15] S. Bibi *et al.*, "MSRNet: Multiclass skin lesion recognition using additional residual block based fine-tuned deep models information fusion and best feature selection," *Diagnostics*, vol. 13, p. 3063, 2023.

[16] V. Dillshad *et al.*, "D2LFS2Net: Multi-class skin lesion diagnosis using deep learning and variance-controlled marine predator optimisation: An application for precision medicine," *CAAI Trans. Intell. Technol.*, 2023, doi: 10.1049/cit2.12267.

[17] SheikhResearch, "SkinDisease," Kaggle dataset. [Online]. Available: https://www.kaggle.com/datasets/sheikhresearch/skindisease (accessed Oct. 3, 2025).

[18] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, 2015, pp. 448–456.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.

[21] M. Groh *et al.*, "Evaluating deep neural networks trained on clinical images in dermatology with the Fitzpatrick 17k dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2021, pp. 1820–1828.