The Impact of Batch Size on Skin Cancer Classification Using ResNet18

Ali M. Abdulameer , Raaed F. Hassan, Abbas F. Humadi

Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq

Article Info	ABSTRACT
<i>Article history:</i> Received April, 27, 2025 Revised May, 15, 2025 Accepted Jun, 01, 2025	Skin cancer is one of the most common types of cancer, often caused by prolonged exposure to ultraviolet (UV) radiation. While visual inspection can provide a preliminary diagnosis, accurate identification typically requires a biopsy, an invasive, costly, and time-consuming procedure. Convolutional neural networks (CNNs) offer a promising alternative by achieving high diagnostic accuracy with reduced time and cost. Despite
<i>Keywords:</i> Deep Learning Skin Cancer CNN Batch Size Artificial Intelligence	numerous studies in this area, the influence of training hyperparameters, particularly batch size, has not been extensively examined in the context of skin cancer detection. This study explores the effect of batch size on the performance of the ResNet18 network, known for its balance between accuracy and computational efficiency in medical imaging tasks. Batch size, defined as the number of training samples processed before weight updates, plays a critical role in model performance. Seven batch sizes (4, 8, 16, 32, 64, 128, and 256) were evaluated. Results revealed that smaller batch sizes led to lower performance, while a batch size of 32 achieved the highest multi-class accuracy. Batch sizes of 64 and 128 improved precision and recall. However, increasing the batch size between 32 and 128 are recommended for optimal results.
Corresponding Author:	

Alı M. Abdulameer Electrical Engineering Technical College, Middle Technical University Almasafi street, Baghdad, Iraq Email: bdc0063@mtu.edu.iq

1. INTRODUCTION

Skin cancer ranks as the fourth most prevalent cancer, with over nine thousand individuals diagnosed daily in the United States [1]. This cancer develops from the uncontrolled growth of skin cells, triggered by mutations caused by ultraviolet (UV) radiation, including sunlight [2]. Skin cancer is classified into benign and malignant tumors based on malignancy. Melanoma is considered the most dangerous type due to its high tendency to spread through the bloodstream to other body parts, causing severe damage to organs and even death if left untreated [2]. A major challenge in melanoma detection is its high visual similarity to melanocytic nevus, a non-cancerous and harmless mole, making early detection crucial for accurate differentiation. Besides melanoma, which originates from Melanocytes, the cells that produce melanin, the pigment responsible for skin color, there are two other malignant types of skin cancer: "Basal Cell Carcinoma" (BCC) and "Squamous Cell Carcinoma" (SCC). BCC, the most common skin cancer, originates from basal cells in the innermost layer of the epidermis. SCC, the second most common type, arises from squamous cells in the outermost layer of the epidermis and has a more aggressive behavior than BCC, making early diagnosis essential to prevent further complications [2].

Skin lesions are primarily diagnosed through naked-eye examination. Although this approach is convenient, its diagnostic accuracy is limited to about 60%, even for experienced dermatologists [3]. For better visualization, lesions can be examined under good lighting conditions and with magnifying equipment using a

Dijlah Journal of Engineering Sciences (DJES) V ISSN: **Printed: 3078-9656, E: 3078-9664**, paper ID: 39

technique called dermoscopy, which improves diagnostic accuracy to approximately 84% [4]. For a definitive diagnosis, a biopsy is required. This process requires the removal of a tissue sample from the lesion for microscopic analysis. However, this method requires well-equipped laboratories, making it both costly and time-consuming, which can delay treatment. Additionally, in remote or underserved areas, limited access to such facilities may prevent timely diagnosis, allowing the cancer to progress further before proper treatment can be administered [5].

To achieve diagnostic convenience with relatively high accuracy, many machine learning (ML) algorithms have been employed over the last two decades for this task. These include logistic regression (LR) [6], artificial neural networks (ANN) [7], decision trees (DT) [8], random forests (RF) [9], k-nearest neighbors (KNN) [10], and support vector machines (SVM) [11]. In these approaches, features were manually extracted based on the ABCDE criteria ("asymmetry, border irregularity, color variation, diameter, and evolution") and then fed into the algorithms [12]. However, this diagnostic approach did not significantly improve accuracy compared to visual inspection alone [13]. Moreover, it required manual feature extraction by specialized personnel. Later, researchers adopted convolutional neural networks (CNNs), an automated feature extraction technique that has proven successful in various fields, including medical imaging [14]. CNNs have achieved relatively high accuracy in diagnosing and classifying skin lesions [14]. The training process of CNNs depends on several parameters, such as network architecture, network depth, the loss function used to measure the error between predicted and true labels, the optimizer used to update network weights, and the batch size, which specifies the number of images processed in each iteration during training. Although these parameters have been comprehensively studied in many fields, research on skin cancer detection has largely focused on the impact of network architecture while neglecting other factors. Most studies emphasize the effect of architecture while omitting the influence of other training parameters.

Yilmaz et al. [15] performed a skin cancer classification task using three lightweight CNN architectures: MobileNet [16], MobileNetV2 [17], and NASNetMobile [18]. The study evaluated the effect of different batch sizes (16, 32, and 64) on the ISIC2017 dataset [19]. For the first architecture, MobileNet, batch sizes of 16 and 32 performed well, with a slight advantage for 32. However, a batch size of 64 resulted in significantly lower performance. For MobileNetV2, a batch size of 16 yielded the best performance, and the same was observed for NASNetMobile. Overall, the highest accuracy (approximately 82.00%) was achieved using NASNetMobile with a batch size of 16.

Islam et al. [20] conducted a binary classification task to differentiate benign from malignant skin lesions using a ResNet50 [21] CNN trained on a dataset from the ISIC archive [22]. They compared batch sizes of 32 and 64, finding that a batch size of 32 yielded the best accuracy (approximately 92.70%).

This research aims to perform a study about the effect of the batch size on skin cancer detection using the ResNet18 [21] architecture, as it is widely used for this task and offers an excellent trade-off between accuracy and computational complexity. Additionally, ResNet18 has low GPU VRAM requirements, making it feasible to experiment with larger batch sizes in this study.

2. METHODOLOGY

This section discusses the dataset, preprocessing techniques, CNN architecture, training process, and evaluation metrics in detail.

2.1. Dataset

The ISIC2018 [22] dataset is used to train the ResNet18 [21] network. The training subset consists of 10,015 dermoscopic images categorized into 7 classes: dermatofibroma (DF), "benign keratosis lesion" (BKL), "vascular lesion" (VASC), "actinic keratoses and intraepithelial carcinoma" (AKIEC), "melanocytic nevus" (NV), melanoma (MEL), and "basal cell carcinoma" (BCC). The validation and test subsets are provided separately, containing 193 and 1,512 images, respectively. For the training subset, the number of dermoscopic images per each class is represented in Table 1.

Table 1. Number of image samples in the ISIC2018 dataset per each class.

0 1	
Class Name	No. of image
BKL	1099
NV	6705
MEL	1113
VASC	142
DF	115
AKIEK	327
BCC	514

2.2. Preprocessing techniques

Input images are resized to 224×224 pixels before being fed into ResNet18, as this network requires a fixed input size. The images are also normalized by mapping pixel values from the [0,255] scale to the [0,1] range. This process helps prevent exploding gradients, stabilizes the network during training, and speeds up convergence. Furthermore, normalization reduces bias toward high-value pixels, ensuring that the CNN treats all inputs equally and focuses on patterns rather than absolute intensity differences. The normalization process is defined by Equation (1) [23].

(1)

$$x_{normalized} = \frac{x}{255}$$

Where, x is the pixel intensity of input image.

2.3. CNN architecture

Residual Network, commonly abbreviated as ResNet, is a popular CNN architecture introduced to address the "vanishing gradient" problem, which arises when a CNN becomes deeper, causing gradients to diminish in value. The reduction occurs due to the repeated multiplication of activation function outputs, each of which is less than one due to the normalization process. As the network depth increases, the gradients become even smaller. Since gradients are directly responsible for updating network weights, extremely small gradients lead to minimal weight updates, resulting in a very slow training process. ResNet overcomes this issue by incorporating shortcuts known as "residual connections," which allow feature maps to bypass certain layers and be passed directly to subsequent layers. This architecture facilitates the flow of information, leading to faster convergence and reduced training time. The basic residual block is shown in Figure 1, where the shortcut, denoted as "x identity," is added to the output side of the block [21].



Figure 1. The residual basic block structure from He et al. [21]

ResNet18 have 18 convolutional layers divided into 8 residual blocks. The entire architecture of this network is detailed in Table 2 [21].

Layer name	Output size	Details
conv1	112×112	7×7, 64, stride 2
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
-	1×1	Average pooling, Fully connected layer, Softmax

CNNs are used as automated feature extractors, where network weights are initially randomized and adjusted during training to minimize the loss function between the predicted probabilities and the true labels. In this study, the "Cross-Entropy Loss Function" is utilized, as it is commonly preferred for classification tasks. The loss function is described in Equation (2) [23].

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c}(\log \hat{y}_{i,c})$$

(2)

Here, *N* is the total number of samples. *C* is the number of classes. $y_{i,c}$ is the true label for the *i*th sample and *c*th class. $\hat{y}_{i,c}$ is the predicted probability for (i, c) indices.

The loss function is theoretically calculated for each data sample (image) to determine the gradient for each weight in the network. The arithmetic mean of all gradients is then computed to determine to update the network weights. The passage of the entire training dataset through the network during the learning process is referred to as an epoch. In practice, calculating loss and gradients for the entire dataset at once is infeasible due to the high GPU VRAM requirements. To address this, "Stochastic Gradient Descent" (SGD) is used. This optimization technique follows the same process but divides the training set into smaller batches, allowing computations to be performed in steps [23]. This study will examine the impact of batch size on training results.

2.4. Training process

The CNN model is customized and trained on the ISIC2018 [22] dataset using the "Deep Learning Toolbox" (version 14.3) in the MATLAB environment, running on a "Quadro RTX 5000" GPU with 16GB of VRAM. The customization involves replacing the fully connected layer with one that has seven outputs, corresponding to the number of classes in the dataset. The classification layer is also replaced to generate seven output predictions instead of the 1,000 outputs originally designed for ImageNet [24], the dataset on which the network was pretrained. To leverage the pretrained weights, the first three convolutional layers are frozen so that subsequent layers can train on these low-level features. The training process is conducted over 30 epochs, utilizing a scheduled learning rate, where the rate decreases every 10 epochs by a factor of 10. Training is performed using "Stochastic Gradient Descent with Momentum" (SGDM), with a momentum value of 0.9 to accelerate convergence. The batch size is varied to analyze its effect on model performance. All hyperparameters are detailed in Table 3.

Table 3. Training hyperparameters.

Hyperparameter	Value
Epochs	30
Optimizer	SGDM
Learning rate	(0.01), (0.001), (0.0001)
Batch size	4, 8, 16, 32, 64, 128, 256
Momentum	0.9
L2 Regularization	0.0001

To combat overfitting, data augmentation is applied during training by performing image rotation, flipping, rescaling, and shifting. Image rotation is applied at a random angle between -45° and 45° degrees. Rescaling is performed within a random range of 0.9 to 1.2. Flipping is carried out randomly along the x-axis and y-axis. Lastly, shifting is performed both vertically and horizontally with random values ranging from -30 to 30 pixels. Figure 2 shows an example of the augmentation process performed on original dermoscopic images.



Figure 2. Example of the augmenation techniques performed during the training process

2.5. Evaluation metrics

Evaluation metrics are essential for understanding the effectiveness of the CNN model in classifying skin cancer lesions and assessing the impact of batch size in this study. Multi-class accuracy is the most commonly used metric to evaluate CNN model performance. It measures the ratio of correct predictions to the total number of data samples, as expressed in Equation (3) [23].

$$Multi Class Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$
(3)

Here, TP represents true positive predictions, TN represents true negative predictions, FP represents false positive predictions, and FN represents false negative predictions.

Accuracy could be calculated for every class individually, as expressed in Equation (4).

$$Accuracy_{class(k)} = \frac{TP_{class(k)}}{Total number of prediction per class(k)}$$
(4)

Precision is another important metric to evaluate the model in the classification task, as it measures the proportion of correctly predicted positive cases (e.g., malignant lesions) out of all predicted positive cases. It is expressed in Equation (5) [23].

$$Precision = \frac{TP}{TP+FP}$$
(5)

Since this study focuses on malignant lesions, accuracy alone is not sufficient to assess model performance. Recall (sensitivity) is introduced to measure how well the model identifies actual positive cases. It is expressed in Equation (6) [23].

$$Recall = \frac{TP}{TP + FN}$$
(6)

The F1-score represents the harmonic mean of precision and recall, as expressed in Equation (7) [23].

$$F1 Score = 2 \times \frac{Precision \cdot Recall}{Precision + Recall}$$
(7)

All the previously mentioned metrics can be calculated using a table called the confusion matrix, which illustrates the overall performance of the network by detailing the true positives, true negatives, false positives, and false negatives for each class in the dataset [25].

On the other hand, the "Receiver Operating Characteristic" (ROC) curve graphically represents the relationship between the true positive rate and the false positive rate. Another important metric, the "Area Under the Curve" (AUC), measures the area under the ROC curve. A higher AUC value indicates better model performance [25].

3. **RESULTS**

In this section, a comprehensive performance comparison is presented for the ResNet18 model trained with seven different batch sizes: 4, 8, 16, 32, 64, 128, and 256. The main objective of the analysis is to conclude the most appropriate batch size for optimizing the accuracy and efficiency of skin cancer detection. To ensure a thorough evaluation, the model's performance is assessed using key metrics such as accuracy, Recall, Precision, and F1-score. The results corresponding to each batch size are systematically summarized in Table 4, providing insights into the trade-offs associated with different batch configurations. This analysis aims to identify an optimal batch size that balances accuracy and training stability, ultimately contributing to improved classification performance in real-world skin cancer detection applications.

Table 4. The evaluation metrics for ResNet18 model with multiple batch sizes.

Batch size	Multi-class accuracy	Accuracy	Precision	Recall	F1-Score	AUC
4	76.59%	93.31%	61.98%	55.18%	57.74%	93.37%
8	76.65%	93.33%	62.27%	55.32%	57.77%	93.30%
16	78.84%	93.95%	69.97%	59.39%	63.49%	95.36%
32	83.73%	95.35%	78.50%	68.93%	73.00%	96.66%
64	82.34%	94.95%	76.29%	69.39%	72.24%	96.07%
128	83.07%	95.16%	79.49%	69.22%	73.20%	96.59%
256	82.21%	94.92%	77.65%	66.82%	70.07%	96.38%

Table 4 shows that the CNN model performed poorly with low batch sizes but improved progressively until reaching a batch size of 32, where it achieved the highest multi-class accuracy. Other metrics showed further improvements with batch sizes of 64 and 128. However, increasing the batch size beyond this did not provide additional benefits. The training accuracy and loss curves for each batch size are illustrated in Figure 3.



Figure 3. The accuracy and loss curves for batch sizes of (a) 4 (b) 8 (c) 16 (d) 32 (e) 64 (f) 128

Dijlah Journal of Engineering Sciences (DJES) ISSN: **Printed: 3078-9656, E: 3078-9664**, paper ID: 39



Continuation of Figure 3. The accuracy and loss curves for batch sizes of (g) 256



Figure 4 shows the confusion matrices for the designated batch sizes.

Figure 4. The confusion matrices for batch sizes of (a) 4 (b) 8 (c) 16 (d) 32

81

Dijlah Journal of Engineering Sciences (DJES) ISSN: **Printed: 3078-9656, E: 3078-9664**, paper ID: 39



Continuation of Figure 4. The confusion matrices for batch sizes of (e) 64 (f) 128 (g) 256

Figure 5 shows the ROC curves for earch batch size.



Figure 5. The ROC curves for batch sizes of (a) 4 (b) 8

Dijlah Journal of Engineering Sciences (DJES) ISSN: **Printed: 3078-9656, E: 3078-9664**, paper ID: 39



Continuation of Figure 5. The ROC curves for batch sizes of (c) 16 (d) 32 (e) 64 (f) 128 (g) 256

4. CONCLUSION

This study analyzed the effect of batch size on the performance results of the ResNet18 CNN for skin cancer classification using a seven-class dataset. Seven batch sizes were evaluated: 4, 8, 16, 32, 64, 128, and 256. The model performed poorly with smaller batch sizes, while a batch size of 32 yielded the highest multi-class accuracy. Batch sizes of 64 and 128 resulted in improved precision and recall. However, further increasing the batch size beyond 128 did not enhance performance. Based on these findings, limiting the batch size between 32 and 128 is recommended for optimal skin cancer classification using the ResNet18 CNN model. Future work could explore

the impact of batch size on other deep learning architectures or investigate adaptive batch size techniques to further enhance performance.

ACKNOWLEDGEMENTS

We sincerely thank everyone who supported us throughout this research. We are especially grateful to Middle Technical University for its academic support and to the ISIC Archive for providing the dataset. Lastly, we appreciate our colleagues, friends, and families for their encouragement.

REFERENCES

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians*, vol. 69, no. 1, pp. 7-34, 2019.
- [2] J. L. Bolognia, J. L. Jorizzo, and J. V. Schaffer, *Dermatology e-book*. Elsevier Health Sciences, 2012.
- [3] H. Haenssle *et al.*, "Reader study level-I and level-II Groups. 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.
- [4] M. Vestergaard, P. Macaskill, P. Holt, and S. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting," *British Journal of Dermatology*, vol. 159, no. 3, pp. 669-676, 2008.
- [5] T. J. Hieken, R. Hernández-Irizarry, J. M. Boll, and J. E. Jones Coleman, "Accuracy of diagnostic biopsy for cutaneous melanoma: implications for surgical oncologists," *International journal of surgical oncology*, vol. 2013, no. 1, p. 196493, 2013.
- [6] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *The journal of educational research*, vol. 96, no. 1, pp. 3-14, 2002.
- [7] J. Zupan, "Introduction to artificial neural network (ANN) methods: what they are and how to use them," *Acta Chimica Slovenica*, vol. 41, no. 3, p. 327, 1994.
- [8] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," in 2011 IEEE control and system graduate research colloquium, 2011: IEEE, pp. 37-42.
- [9] L. Breiman, "Random forests," Machine learning, vol. 45, pp. 5-32, 2001.
- [10] Z. Zhang, "Introduction to machine learning: k-nearest neighbors," Annals of translational medicine, vol. 4, no. 11, p. 218, 2016.
- [11] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).*[Internet], vol. 9, no. 1, pp. 381-386, 2020.
- [12] L. Ferrante di Ruffano et al., "Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults," Cochrane Database of Systematic Reviews, vol. 2018, no. 12, 1996.
- [13] K. T. Clebak, L. Helm, and M. Helm, "Accuracy of dermoscopy vs. visual inspection for diagnosing melanoma in adults," *American family physician*, vol. 101, no. 3, pp. 145-146, 2020.
- [14] M. Naqvi, S. Q. Gilani, T. Syed, O. Marques, and H.-C. Kim, "Skin cancer detection using deep learning—a review," *Diagnostics*, vol. 13, no. 11, p. 1911, 2023.
- [15] A. Yilmaz, M. Kalebasi, Y. Samoylenko, M. E. Guvenilir, and H. Uvet, "Benchmarking of lightweight deep learning architectures for skin cancer classification using ISIC 2017 dataset," arXiv preprint arXiv:2110.12270, 2021.
- [16] A. G. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

- [17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4510-4520.
- [18] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 8697-8710.
- [19] N. C. Codella *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), 2018: IEEE, pp. 168-172.
- [20] M. S. Islam and S. Panta, "Skin Cancer Images Classification using Transfer Learning Techniques," arXiv preprint arXiv:2406.12954, 2024.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.
- [22] "International Skin Imaging Collaboration (ISIC)." https://www.isic-archive.com (accessed 1 Nov. 2024).
- [23] I. Goodfellow, "Deep learning," ed: MIT press, 2016.
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25, 2012.
- [25] T. Fawcett, "An introduction to ROC analysis," Pattern recognition letters, vol. 27, no. 8, pp. 861-874, 2006.

BIOGRAPHIES OF AUTHORS

	Eng. Ali Maan Abdulameer received his BSc degree in Electrical Engineering from the Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq. He is currently pursuing his MSc degree in Digital medical instruments at the same institution. His research interests include deep learning, image processing, and medical image analysis, with a focus on skin cancer detection and classification. He can be contacted at: bdc0063@mtu.edu.iq.
	Prof. Dr. Raaed Faleh Hassan received his BSc in Electrical Engineering from Baghdad
-	University in 1986, his MSc in Control Engineering in 1999, and his PhD in Electronics Engineering in 2006 from the University of Technology / Al-Rasheed College, Iraq. He is
	currently a professor in the Department of Control and Automation Engineering
	Techniques at the Electrical Engineering Technical College, Middle Technical University,
	Baghdad. He has held several academic leadership roles, including Vice Dean and Dean of
	the college. His research interests include power electronics drives, control systems, signal
	processing, and FPGA-based systems. He served as an editor for the Iraqi Journal of Computers Communications Control and Systems Engineering (2020–2022)
	He can be contacted at: drraaed alanbaki@mtu.edu.ig.
	Prof. Asst. Dr. Abbas Fadhal Humadi received his Bachelor's degree in General
	Medicine and Surgery from the University of Zagreb, Yugoslavia, in 1989. He earned his
1001	MSc in Occupational and Environmental Medicine from the College of Medicine,
en b	University of Baghdad, in 1996, and his PhD in Community Medicine from the College of
	Medicine, Al-Mustansiriyah University, in 2013. He is currently an assistant professor in the Medical Instrumentation Engineering Techniques Department at the Electrical
	Engineering Technical College Middle Technical University Baghdad Irag His research
	interests include public health, dermatology, skin cancer, medical devices, and
	diabetology.
	He can be contacted at: drabbas@mtu.edu.iq.