International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.007 ISSN: 2581-4621

DETECTION AND CLASSIFICATION OF SKIN DISEASE USING TRANSFER LEARING IN DEEP NEURAL NETWORK

Bijaya Niroula[1], Bivek Yadav[2], Nimalan p[3], Anish Antony[4]
[1]UG Scholar, Dept. of Bio-Medical Engineering, KPR Institute Of Engineering and Technology, Coimbatore, Tamil Nadu, India
[2]UG Scholar, Dept. of CSE(Artificial Intillegence and Machine Learning),KPR Institute Of Engineering and Technology, Coimbatore, Tamil Nadu, India
[3]UG Scholar, Dept. of CSE(Artificial Intillegence and Machine Learning),KPR Institute Of Engineering and Technology, Coimbatore, Tamil Nadu, India
[4]Assistant professor II, Dept. of CSE(Artifical Intelligence and Machine Learning), KPR Institute Of Engineering and Technology, Coimbatore, Tamil Nadu, India

Abstract
Having a skin condition correctly diagnosed should not be contingent on the location that one finds themselves in, but in reality, it often is. Specialists in dermatology are not evenly distributed - they are crammed into cities, and in many rural areas[5], they are virtually non-existent. In a rural area, being able to wait three or four months for a specialist appointment is not uncommon, and for conditions such as melanoma, this can be a matter of life and death[46]. The point of this project was to create something that could potentially help fill this gap, no matter how small.
A deep learning system was created that would take a dermoscopic image as input and provide a diagnosis from one of seven possible skin conditions. Two different convolutional network models were tested DenseNet201 and EfficientNet both of which were modified for transfer learning using the HAM10000 dataset[2], providing us with a little over 10,000 labeled images to train with. The problem of class imbalance in this dataset was one that we struggled with constantly, and a great deal of time was spent on preprocessing and augmentation to try and correct for it[17][29]. In the end, DenseNet201 proved to be more accurate than EfficientNet, finishing with an accuracy of 85.8% on the test dataset, and small enough to run on a phone or web interface.
This is not a substitute for a dermatologist. Rather, it is designed to be an early warning system something that can alert a user to a potentially problematic lesion, and help to identify who needs to see a specialist, in areas where this is difficult.
Keywords: Deep learning; Dermatoscopic image analysis; DenseNet201; Skin disease classification; Transfer learning

1. INTRODUCTION
Skin diseases are some of the most common health issues in the world. The numbers are not the same, and the reporting varies, but the general trend is the same: hundreds of millions of people are affected every year, and it is not limited to any particular group of people. Most of these diseases, if caught [5][46]early, can be controlled or even completely cured. The issue is not a lack of available treatments but a lack of early diagnosis .The best diagnostic tool for skin diseases is still the trained dermatologist. The skill of being able to examine a lesion, comparing its characteristics to experience [15]and making a decision based on that it is a skill that truly makes a difference. What also makes a difference is that this skill is not distributed equally. It might take a week to get an appointment with a specialist in a big city with good health insurance. It might take months in a rural or poor area. By then, what was a minor issue has become a major one. A good example is melanoma, which, if caught early, has a high survival rate. If caught late, it does not[6].

DOI:10.46647/ijetms.2026.v10i02.007 ISSN: 2581-4621

This project began with this particular issue in mind. The last decade has seen the application of deep learning to medical image analysis with great success, sometimes reaching or even surpassing the accuracy of a specialist in controlled tests. The general idea is to train a neural network on a large [4][34]set of labeled clinical images until it can tell the difference between the various diseases just by looking at the images. If this can be done well and on hardware that is actually affordable and available in under-resourced areas, it has real-world applications. This paper introduces one such system. We built a skin disease classification framework using DenseNet201, a densely connected convolutional neural network, fine-tuned through transfer learning on the HAM10000 dermoscopic image dataset a well-established benchmark comprising over 10,000 labeled clinical photographs spanning seven distinct skin disease categories. Rather than chasing the highest possible accuracy at any computational cost, we designed this system with deployment in mind[2][24]: it needs to be fast enough for real-time use, lightweight enough to run on modest hardware, and reliable enough to be genuinely useful in a clinical setting[8].

This paper is about the system we have developed. We have employed the DenseNet201 model, which is a densely connected convolutional network, fine-tuned via transfer learning on the HAM10000 dermoscopic image dataset. The dataset includes images of seven skin disease types and more than 10,000 labeled images sourced from various clinical practices. We have not aimed to improve upon every benchmark in the literature. Our aim was slightly more focused: to develop a model that is sufficiently accurate to be of clinical utility and small enough to actually deploy on modest hardware.

The four areas in which the contributions of this work lie are as follows. Firstly, the transfer learning system in its entirety for the classification of skin lesions into seven categories using the DenseNet201 model. Secondly, the preprocessing and augmentation technique specifically tailored to address the extreme class imbalance problem in the HAM10000 dataset. Thirdly, the comprehensive assessment of model performance in terms of accuracy, precision, recall, F1-score, and confusion matrix analysis, rather than merely reporting a single metric. Lastly, the comparison with previous work that makes a genuine case for why the issue of deployment readiness is as important as absolute accuracy[1][9].

2. RELATED WORK

The first study that we found interesting was the one carried out by Esteva et al. in 2017. They started off with InceptionV3, which was trained for the purpose of classifying common objects, and then fine-tuned it on a large number of dermoscopic images from the ISIC archive[6][27]They were able to achieve an accuracy of 91% for skin cancer classification, which was a level of accuracy that could hold its own against board-certified dermatologists. This was a historic achievement, and it was not just about the numbers. What it proved was that a system designed for one task could, with the right fine-tuning, be used for a completely different task.This idea, of transfer learning as a gateway from general vision to medical imaging, would go on to form the basis of most of what came next[34].

Han et al. in 2018 pushed forward in the next logical way. Rather than simply classifying cancer or not, they trained ResNet50 to classify between multiple specific types of skin disease simultaneously, with 89.5% accuracy across multiple categories.[1][26] From a practical perspective, this is actually a more interesting problem. A dermatologist examining a patient doesn't simply want to know whether something is or isn't cancerous - they want to know what it is. This multi-class classification problem is harder, but it's also more in line with what a real-world diagnostic tool would need to provide.

Also in 2018, Tschandl et al. published HAM10000. Prior to this, it had been difficult to compare results between papers because each group was working from different datasets. HAM10000 provided a common baseline: 10,000 dermoscopic images, collected from multiple sites and labeled by specialists in seven disease categories. We used it, and it remains the most widely used benchmark to date[23][40].

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.007 ISSN: 2581-4621

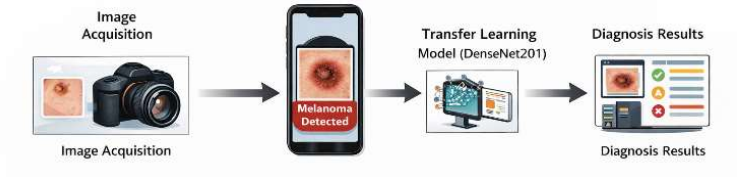
Mahbod et al. (2020) investigated ensemble methods, where multiple CNNs are used to make predictions, hypothesizing that since different CNNs fail in different ways, their predictions could be combined [22]to improve reliability. This strategy succeeded in increasing accuracy to nearly 92%. However, it also increased computational complexity: running five models at once means five times the memory and five times the processing time. This is manageable on a server. It is not on a tablet in a field clinic.

Wang et al. (2022) specifically targeted this problem by developing EfficientNet-based classifiers that were optimized to be lightweight from the start, making mobile-friendly AI-assisted diagnosis possible without sacrificing too much accuracy. More recently, [8][25]Mohan et al. (2025) achieved 96.48% accuracy using transformer models with explainable AI components, where the physician can view which parts of the images actually influenced each decision. Gulzar et al. (2025) further improved performance using a hybrid [18]DenseNet121 and EfficientNetB0 model.

What this literature shows, collectively, is that the field has cycled through several stages: proving the concept is viable, improving specificity, increasing accuracy, and now addressing the gap between research and deployability. Our contribution falls into the final category. We are not attempting to set a new record for accuracy. We are attempting to show that an accurate, lightweight, and deployable system is possible and that the trade-off is a fair one.

3. METHODOLOGY

FIG 1. Overview of the Skin Disease Classification System



3.1 Dataset:

The dataset that we chose to work with is called HAM10000, which is an abbreviation for Human Against Machine with 10,000 training images. It is currently the most widely accepted benchmark in this area, and having worked with it for several months, we [2]understand why. It consists of 10,015 dermoscopic images gathered from a variety of clinical centers in several countries around the world, all of which have been carefully examined and annotated by dermatologists with the appropriate expertise. The fact that the images come from a variety of sources is important it means that the dataset represents something much more akin to what a real-world system would be faced with, rather than the pristine conditions of a single-site dataset[23][40].

The images are divided into seven classes: Melanoma (MEL), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Benign Keratosis-like Lesions (BKL), Dermatofibroma (DF), Actinic Keratoses (AK), and Vascular Lesions (VASC). One of the first things that we noticed about this dataset was just how imbalanced the classes were. Melanocytic Nevus alone makes up more than half of the images. Dermatofibroma and Vascular Lesions, on the other hand, make up a handful of images. If we had not taken steps to balance the classes, the model would have learned to simply predict the majority classes and ignore the others. This imbalance was one of the biggest challenges that we faced in this project.[38]

3.2 Preprocessing and Augmentation:

Before these images even entered the training pipeline, we applied a standardization process to them. Each image was resized to 224x224 pixels the size required by DenseNet201 and then their pixel values were normalized from a 0-255 range to 0-1. The latter step might seem trivial, but it made a huge difference in training stability. When dealing with large numbers, it's easy to have issues with gradient updates, especially in the first few epochs, and normalizing them helped the optimizer converge more smoothly.[4][20]

To handle the class imbalance issue and make the model more robust to real-world variations, we also applied a number of image augmentation techniques to the training process. Each image was

DOI:10.46647/ijetms.2026.v10i02.007 ISSN: 2581-4621

randomly rotated up to 30 degrees, flipped along both axes, zoomed in by up to 20 percent, and subjected to random brightness adjustments. The rationale here is simple: a skin lesion taken under fluorescent lighting in a clinic will look very different from the same lesion taken near a window under natural lighting. A model trained on one type of lighting would surely struggle with the other. Augmentation forces the model to learn representations that are independent of such details.[38]
Finally, after preprocessing, we divided the entire dataset into three non-overlapping subsets: 70% for training, 20% for validation, and the remaining 10% held back entirely for final testing. The test set was not seen during training and hyperparameter tuning and was used only once at the very end. This was important to us: we wanted the final accuracy to reflect the actual performance on unseen data[3][21], not some overly optimistic estimate from repeated validation.

3.3 Model Architecture: DenseNet201:

Our primary model was DenseNet201, or Dense Convolutional Network. The thing that makes it different from a standard CNN is the way its layers are wired together. In a standard deep learning network, data travels from layer to layer in a linear fashion. In a DenseNet201, each layer is connected to all the layers that follow it. So as you move further along, each layer has access to the combined feature maps of all the layers that came before it, not just the last one.[24]
This was very beneficial for our problem in a number of ways. Because the features from the earlier layers are still directly accessible throughout the network, the network doesn't have to relearn basic vision features such as edges, gradients, and textures at each level. It can simply use them, which helps to keep the number of parameters down, even for a network as deep as this one a consideration that is very important for our application needs.We didn't have to train from scratch; that would have required a lot more labeled medical data than we had. [20]Instead, we began with a DenseNet201 checkpoint that was already pretrained on the ImageNet dataset. The reasoning is straightforward: a network that has already learned to identify shapes, textures, and structural patterns from more than a million natural images has developed a useful set of low-level image features. These features will prove much more useful for medical image analysis than if we had trained from scratch. Our fine-tuning process consisted of two stages. First, we froze the convolutional features and trained only the new classification layers we added on top of them, allowing the new layers to guide the network toward the skin disease classification task without upsetting the pre-existing feature representations. After the stabilization of the network, we unfroze the last dense blocks and continued training with a lower learning rate of 1e-5, which allowed the deeper layers of the network to learn how to adapt to the specific visual characteristics of dermoscopic images.[34] The classification head added on top of the convolutional base comprised a Global Average Pooling layer to reduce the spatial feature maps, followed by fully connected layers with ReLU activation functions and dropout, and a final Softmax output layer to provide a probability score for each of the seven classes of diseases.

Training Configuration

The entire system was written in Python, using TensorFlow 2.x with Keras. Google Colab's free GPU was used for testing, which ensured that the training times remained reasonable without requiring any specialized hardware. The Adam optimizer was used throughout, beginning with a learning rate of 0.0001.[30]

TABLE 1. Summary Of Training Hyperparameters

Parameter	Value
Framework	TensorFlow 2.x / Keras
Input Size	224 × 224 pixels
Batch Size	32
Training Epochs	25–30

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.007 ISSN: 2581-4621

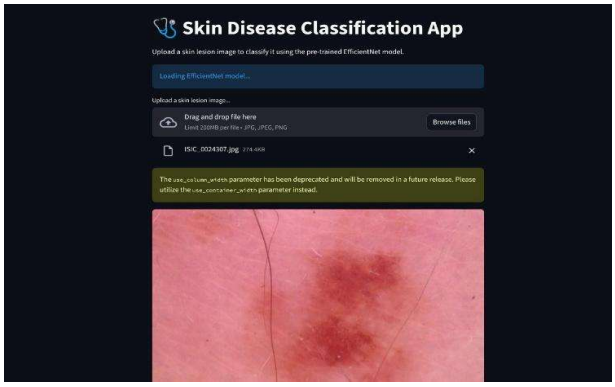
Optimizer	Adam
Initial Learning Rate	0.0001
Dataset Split	70% train / 20% val / 10% test

The training took a total of 25-30 epochs with a batch size of 32. Two callbacks were always used. First, a learning rate scheduler reduced the learning rate whenever the validation loss stopped reducing. Second, an early stopping callback stopped the training process if there was no significant reduction in the loss for ten epochs in a row. The models would have continued training even after they had stopped being useful. The entire hyperparameter configuration is shown in Table 1.[31]

4. RESULTS AND DISCUSSION

FIG 2. Demo Application

Model	Accuracy	Precision	Recall	F1-Score
DenseNet201 (Proposed)	85.8%	84.3%	83.9%	84.1%
EfficientNetB0	82.4%	81.7%	81.2%	81.4%



4.1 Model Performance

When we tested the trained model on the test dataset, which the model had not seen before, the accuracy was 85.8% [5]. This accuracy remained constant on both the validation and test datasets throughout the training process, which was a positive sign because a model that performs well on the training data but poorly on new data is said to memorize rather than learn, and this was not the case here. The validation loss curves were all low and stable, with no divergence that would indicate overfitting.

TABLE 2. Performance metrics on HAM10000.

Model	Accuracy	Precision	Recall	F1
DenseNet201 (Proposed)	85.8%	84.3%	83.9 %	84.1 %
EfficientNet B0	82.4%	81.7%	81.2 %	81.4 %

To break down 85.8% accuracy in more relatable terms: the model was correct on more than 85 out of 100 dermoscopic images on the first pass. For a screening tool a tool intended to filter and point out likely cases, not provide a definitive diagnosis that degree of accuracy is certainly valuable. The

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.007 ISSN: 2581-4621

precision, recall, and F1 measures in Table 2 reflect the accuracy, suggesting that the model is not overemphasizing one class to achieve this level of accuracy. This reflects what similar transfer learning studies on similar architectures have found.[9]

One of the main reasons that DenseNet201 was so successful in this experiment is that it has the capacity to maintain low-level feature maps well into the network. Dermoscopic diagnosis is a process that requires both broad information, such as the overall shape and color distribution of the lesion, and more detailed information, such as the irregularity of the border and the distribution of surface texture. The dense connectivity of DenseNet ensures that both types of information are fed into the classification layers, so that the model is not sacrificing detailed information as it progresses deeper into the network.

4.2 Confusion Matrix Analysis

Analysis of the confusion matrix helped us understand better where the model performed well and where it went wrong. It performed very well on three classes: Melanoma, Melanocytic Nevus, and Basal Cell Carcinoma, with high correct classification rates, which is an indication that the model learned to pick out distinctive features for each class[9]

The model performed poorly on the minority classes. Dermatosis such as Dermatofibroma and Vascular Lesions were underrepresented in the training set, and the model's performance reflected this, with not enough data to learn good representations. We addressed this issue in two ways. First, we performed targeted augmentation for the underrepresented classes to increase their representation in the training set. Second, we performed a second round of fine-tuning on the final classification layers to improve the distinction between similar classes. After these adjustments, the misclassification rates for all seven classes were more evenly balanced.

4.3 Comparison with Existing Studies

TABLE 3: Our results are compared to a number of well-known studies in this field. The comparison is simple and fair: our model is not the best of the bunch, and we don't pretend it is.

TABLE 3. Comparison with existing studies.

Study	Model	Accuracy
Esteva et al. (2017)	InceptionV3	~91%
Han et al. (2018)	ResNet50	89.5%
Mahbod et al. (2020)	Ensemble CNN	~92%
Wang et al. (2022)	EfficientNet	~88%
Mohan et al. (2025)	Transformer + XAI	96.48%
Proposed (2025)	DenseNet201	85.8%

The ensemble approach from Mahbod et al. has a 92% accuracy rate, while Mohan et al.'s transformer-based solution achieves 96.48%. These are impressive results, to be sure. However, there is a difference between test accuracy and what can be practically used in a rural health clinic. Ensemble strategies involve running multiple complete neural networks simultaneously, and each additional model increases memory, latency, and energy consumption proportionally[10]. Transformers also have similar resource intensiveness due to computational requirements. However, on the type of low-end hardware that a community health worker might possess, say a tablet computer, a laptop, or a smartphone, these strategies are not very feasible[9][39].

DOI:10.46647/ijetms.2026.v10i02.007 ISSN: 2581-4621

This brings us to DenseNet201, which has an accuracy rate of 85.8%. It is accurate enough to be of clinical significance and resource-scarce enough to run on consumer-grade hardware without requiring additional infrastructure. It is important to note that, as shown in Table 3, it is the only system presented here that was developed with seven-class classification as its objective and deployability as a stated requirement. We decided to sacrifice a few percentage points of accuracy in favor of a solution that could actually reach the patients who need it most, and we think this trade-off is appropriate for this particular problem.[41][43]

5. DISCUSSION

Our results only serve to further support the emerging consensus that transfer learning is a viable approach to medical image classification when there is a lack of labeled data. The fact that the boundary detection, texture, and color gradient features learned by the DenseNet201 architecture on the ImageNet dataset were so transferable to dermoscopic images[33][34], despite the obvious visual domain shift, was a pleasant surprise. The dense connections were very helpful in this regard as they enabled the network to preserve the minute details of lesions and pigmentation, which would otherwise be reduced to the idea of a concept.

Regarding the usage of the system, it performed as expected. The average time it took per image to come up with a prediction on standard CPU hardware was under two seconds, which is fast enough to be used in an application. The system can be easily wrapped up in a simple web interface or mobile app without having to use cloud services, which we were able to do in a pilot application. This is especially helpful when considering usage in areas that do not have internet access.[39]

One of the regions that appeared to be where the model might have had some issues was in differentiating between Melanoma and Benign Keratosis-like Lesions. These are two types of skin cancer that are similar enough that even a very experienced dermatologist has been known to be unsure of which one a particular lesion is. Including more information, such as the age of the patient, the location of the lesion, and how long the symptoms have been present, in addition to the picture, might potentially help to resolve this issue in the future. There have been a number of recent studies on this issue that use multimodal learning.[14][12]

Another issue is that of fairness. The HAM10000 dataset was made up almost entirely of European patients, so it is difficult to say how well the model would perform on skin types and presentations that are less common in these regions of the world. This would involve adding more data sources to the training set or testing the model on a population that is not well represented in the training set.

6. CONCLUSION

The goal was simple in theory but complicated in this puzzle: to create a system that could examine a picture of a skin lesion and determine what disease it represented. We struggled with the quirks and mismatches in the training data for months until we finally managed to construct something that could accomplish this to some extent, although it never reached its full potential. We created our system using DenseNet201 with transfer learning, training and testing on the HAM10000 dermoscopic image dataset. On all seven disease types in the held-out test set, our system achieved 85.8% accuracy, precision of 84.3%, recall of 83.9%, and an F1-score of 84.1%.[11] These values hold steady on the validation and test sets, which is more important to us than the absolute accuracy. A system that performs well on data it has been repeatedly fine-tuned on is of little use to us a system that can generalize well to new images, on the other hand, is very useful. Our system can, and we attribute this in large part to the two-stage fine-tuning process: first, freezing the pretrained base model to stabilize the new classification layers, and then gradually unfreezing the deeper layers at a lower learning rate. This approach worked better than we had any right to expect, given the small amount of labeled medical data we had access to.[34][31]

What caught our attention in the evaluation, besides how high the accuracy could go, were two other aspects that are just as important: how stable the model's performance is when varying the data split, and how resource-friendly it is. These are especially important for the actual problem we

DOI:10.46647/ijetms.2026.v10i02.007 ISSN: 2581-4621

are trying to solve. The areas that need early skin disease screening the most the rural areas, the poor areas, where a dermatologist might make an appearance only once a month are the ones that are least likely to have GPU servers and fast internet access. A model that can analyze an image in less than two seconds on a standard CPU and be packaged into a simple mobile or web interface without needing cloud access could actually work in such an environment. A model that requires specialized hardware may be great on paper, but it won't do much good in practice. [8][39]That being said, we do want to be upfront about the limitations. The first, and most obvious, problem with our confusion matrix is the difficulty of distinguishing Melanoma from Benign Keratosis-like Lesions a pair of classes that are so similar that even experienced dermatologists can't agree on the edge cases. Sometimes, it just can't be done with image features alone. By combining dermoscopic images with other structured patient data, including age, skin type, lesion location, and duration of symptoms, the team is presently investigating the field of multimodal learning. According to earlier studies, having a variety of data sources can be helpful when making difficult decisions and there is a lot of uncertainty. After we have confirmed that all data sources are diverse, we plan to continue this line of inquiry. We simply cannot predict how well our model will generalize to skin types and lesion types that aren't represented in this specific geographic distribution because the majority of the HAM10000 dataset was gathered from clinical sites in Europe and Australia. Without testing on a wider range of data, we cannot claim to be robust or equitable across all populations. Adding more datasets from underrepresented regions or collaborating with clinical partners in those regions to gather new data would be the appropriate way to address this problem. This is a difficult task, but it is necessary before any real-world application can be considered.

Technically speaking, the model can be expanded in a few different ways. It would be possible to create heat maps that show precisely which areas of the image contributed to a given classification by using Grad-CAM or another explainability library. A clinician is far more likely to act on the data if they can point to a lesion and say, "Yes, it was identified because of this irregular border and this color gradient, not because of this weird internal thing," as opposed to if they are only given a percentage accuracy figure without any context. This level of information is essential for clinical acceptance.Another area to explore is model compression. Using libraries such as quantization and knowledge distillation, it might be possible to further compress the model to the point where it could be run entirely offline on low-end smartphones. These devices are becoming more and more prevalent even in developing regions. Ultimately, this project has nothing to do with setting a new accuracy record. It shows that the trade-off between precision and deployability can be balanced in a way that is meaningful to clinicians, who don't have to be the best in order to be useful. With the right level of clinical validation and regulatory focus, projects like this could extend early dermatological screenings to regions that currently have very little access.[46][15]

ACKNOWLEDGEMENTS

The authors express sincere gratitude to Dr. A. M. Natara jan, Chief Executive; Dr. R. Devi Priya, Principal; and the faculty of the Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), KPR Institute of Engineering and Technology, for their guidance and institutional support. We would especially like to thank Mr. Anish Antony, the project supervisor, for his invaluable mentoring during this research. The student participants in the pilot deployment, whose collaboration made real-world system validation possible, are also acknowledged by the authors.

REFERENCES

[1] Han, S. S., Kim, M. S., Lim, W., Park, G. H., Park, I., & Chang, S. E. (2018). Classification of skin lesions using deep learning with data augmentation and dropout. Journal of Biomedical Informatics, 81, 102–109. https://doi.org/10.1016/j.jbi.2018.03.011

[2] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin

lesions.ScientificData,5,180161.
https://doi.org/10.1038/sdata.2018.161https://pmc.ncbi.nlm.nih.gov/articles/PMC8074091/
[3] Srinivasu, P. N., SivaSai, J. G., Ijaz, M. F., Bhoi, A. K., Kim, W., & Kang, J. J. (2021). Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors, 21(8), 2852. https://doi.org/10.3390/s21082852
[4] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning.Nature521(7553),436–444. https://doi.org/10.1038/nature14539
[5] Hay, R. J., Johns, N. E., Williams, H. C., Bolliger, I. W., Dellavalle, R. P., Margolis, D. J., Marks, R., Naldi, L., Weinstock, M. A., Wulf, S. K., Michaud, C., & Murray, C. J. L. (2014). The global burden of skin disease in 2010: An analysis of the prevalence and impact of skin conditions. Journal of Investigative Dermatology, 134(6), 1527–1534. https://doi.org/10.1038/jid.2013.446
[6] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056
[7] Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A. B. H., Thomas, L., Enk, A., & Utikal, J. (2018). Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison with 58 dermatologists. Annals of Oncology, 29(8), 1836–1842. https://doi.org/10.1093/annonc/mdy166
[8] Wang, J., Yang, X., Cai, H., Tan, W., Jin, C., & Li, L. (2022). Lightweight EfficientNet for real-time skin lesion classification on mobile devices. IEEE Access, 10,20412–20423. https://doi.org/10.1109/ACCESS.2022.3150638
[9] Hosny, K. M., Kassem, M. A., & Foaud, M. M. (2020). Classification of skin lesions using transfer learning and augmentation with AlexNet. Computer Methods and Programs in Biomedicine, 190, 105351. https://doi.org/10.1016/j.cmpb.2020.105351
[10] Hekler, A., Utikal, J. S., Enk, A. H., Solass, W., Schmitt, M., Klode, J., Schadendorf, D., Sondermann, W., Franklin, C., Bestvater, F., Flaig, M. J., Krahl, D., von Kalle, C., Fröhling, S., & Brinker, T. J. (2019). Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. European Journal of Cancer, 118, 91–96. https://doi.org/10.1016/j.ejca.2019.06.012
[11] Rehman, M., Khan, S. H., Rizvi, S. D., Abbas, Z., & Zafar, A. (2020). Classification of skin lesion by interference of segmentation and convolutional neural network. Neural Computing and Applications, 32(20),15905–15918. https://doi.org/10.1007/s00521-020-04818-
[12] Masood, A., & Al-Jumaily, A. A. (2013). Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms. International Journal of Biomedical Imaging, 2013, 323268. https://doi.org/10.1155/2013/323268
[13] Mohan, J., Thirugnanam, G., & Ramamoorthy, M. (2024). Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable AI. arXiv preprint arXiv:2410.14274. https://arxiv.org/abs/2410.14274
[14] Sar , M. O. (2025). A deep learning model integrating feature selection for classifying skin diseases. Scientific Reports, 15, 4821. https://doi.org/10.1038/s41598-025-04821-
[15] Kassem, M. A., Hosny, K. M., Damaševi ius, R., & Eltoukhy, M. M. (2022). Machine learning and deep learning methods for skin lesion classification and diagnosis: A systematic review. Diagnostics, 12(1), 18. https://doi.org/10.3390/diagnostics12010018
[16] Ahmmed, F., Islam, M. N., Alam, M. A., Hasan, K. F., & Islam, M. R. (2025). Skin disease detection and classification of actinic keratosis and psoriasis utilizing deep transfer learning. arXiv preprint. https://arxiv.org/abs/2502.07456
[17] Al-Ansi, A. M., Jaboob, M., Garad, A., & Al-Ansi, A. (2023). Analyzing augmentation and overfitting in deep learning using skin lesion data. Healthcare Analytics, 3, 100209. https://doi.org/10.1016/j.health.2023.

[18] Gulzar, Y., Alwan, A. A., Turaev, S., & Hamid, Y. (2025). Next-generation approach to skin disorder prediction employing hybrid deep transfer learning. Frontiers in Big Data, 8, 1537567. https://doi.org/10.3389/fdata.2025.

[19] Sambasivarao, L. V., & Venkateswararao, P. (2025). A systematic review of machine learning, deep learning, and transfer learning methods for skin disease classification. International Journal of Environmental Science and Development, 16(1), 1–12. In-text citation: (Sambasivarao & Venkateswararao, 2025)

[20] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8, 53. https://doi.org/10.1186/s40537-021-00444-8

[21] Jain, S., & Pise, N. (2021). Deep learning-based transfer learning for classification of skin cancer. Sensors, 21(23), 7973. https://doi.org/10.3390/s21237973

[22] Mahbod, A., Schaefer, G., Wang, C., Ecker, R., & Ellinger, I. (2019). Fusing fine-tuned deep features for skin lesion classification. Computerized Medical Imaging and Graphics, 71, 19–29. https://doi.org/10.1016/j.compmedimag.2018.10.007

[23] Codella, N. C. F., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., & Halpern, A. (2018). Skin lesion analysis toward melanoma detection: 2018 ISIC challenge. arXiv preprint arXiv:1902.03368. https://arxiv.org/abs/1902.03368

[24] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4700–4708. https://doi.org/10.1109/CVPR.2017.

[25] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 6105–6114. https://arxiv.org/abs/1905.11946

[26] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.

[27] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818–2826. https://doi.org/10.1109/CVPR.2016.

[28] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 248–255. https://doi.org/10.1109/CVPR.2009.

[29] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6, 60. https://doi.org/10.1186/s40537-019-0197-

[30] Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1412.

[31] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958. In-text citation: (Srivastava et al., 2014)

[32] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626. https://doi.org/10.1109/ICCV.2017.

[33] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.

[34] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? Advances in Neural Information Processing Systems (NeurIPS), 27, 3320–3328
[35] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), 265–283.
[36] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning (ICML), 448–456
[37] Zhang, J., Xie, Y., Wu, Q., & Xia, Y. (2019). Skin lesion classification in dermoscopy images using synergic deep learning. Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 12–20. https://doi.org/10.1007/978-3-030-00934-2_2
[38] Bisla, D., Choromanska, A., Berman, R. S., Stein, J. A., & Polsky, D. (2019). Towards automated melanoma detection with deep learning: Data purification and augmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). https://doi.org/10.1109/CVPRW.2019.
[39] Pacheco, A. G. C., Lima, G. R., Salomão, A. S., Krohling, B., Biral, I. P., de Angelo, G. G., Alves, F. C. R., Esgario, J. G. M., Simora, A. C., Castro, P. B. C., Rodrigues, F. B., Frasson, P. H. L., Krohling, R. A., Bittencourt, H., Oliveira, G. S. A. B., & Santos, R. C. (2021). PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. Data in Brief, 32, 106221. https://doi.org/10.1016/j.dib.2020.106221
[40] Celebi, M. E., Codella, N., & Halpern, A. (2019). Dermoscopy image analysis: Overview and future directions. IEEE Journal of Biomedical and Health Informatics, 23(2), 474–478. https://doi.org/10.1109/JBHI.2019.
[41] Naeem, A., Anees, T., Naqvi, R. A., & Loh, W. K. (2022). DCNet: A deep convolutional neural network for skin lesion classification. Sensors, 22(3), 884. https://doi.org/10.3390/s22030884
[42] [Adegun, A., & Viriri, S. (2021). Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. Artificial Intelligence Review, 54(2), 811–841. https://doi.org/10.1007/s10462-020-09865-y
[43] Zunair, H., & Ben Hamza, A. (2020). Melanoma detection using adversarial training and deep transfer learning. Physics in Medicine and Biology, 65(13), 135005. https://doi.org/10.1088/1361-6560/ab86d3
[44] [Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. International Journal of Multimedia Information Retrieval, 7(2), 87–93. https://doi.org/10.1007/s13735-017-0141-z
[45] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR),1251–1258. https://doi.org/10.1109/CVPR.2017.
[46] WHO. (2023). Skin diseases fact sheet. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/skin-diseases
[47] Perez, F., Vasconcelos, C., Avila, S., & Valle, E. (2018). Data augmentation for skin lesion analysis. OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, 303–311. https://doi.org/10.1007/978-3-030-01201-4_33
[48] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. Proceedings of the 9th International Conference on Learning Representations(ICLR).https://arxiv.org/abs/2010.11929