

Training CNN-based Model on Low Resource Hardware and Small Dataset for Early Prediction of Melanoma from Skin Lesion Images

Ivan Halim Parmonangan^{1*}, Marsella², Doharfen Frans Rino Pardede³,
Katarina Prisca Rijanto⁴, Stephanie⁵, Kreshna Adhitya Chandra Kesuma⁶,
Valentina Tiara Cahyaningtya⁷, Maria Susan Anggreainy⁸

¹⁻⁸ Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480

ivan.parmonangan@binus.ac.id; marsella005@binus.ac.id; doharfen.pardede@binus.ac.id;
katarina.rijanto@binus.ac.id; stephanie019@binus.ac.id; kreshna.kesuma@binus.ac.id;
valentina.cahyaningtyas@binus.ac.id; maria.anggreainy@binus.edu

*Correspondence: ivan.parmonangan@binus.ac.id

Abstract – Melanoma is a kind of rare skin cancer that can spread quickly to the other skin layers and the organs beneath. Melanoma is known to be curable only if it is diagnosed at an early stage. This poses a challenge for accurate prediction to cut the number of deaths caused by melanoma. Deep learning methods have recently shown promising performance in classifying images accurately. However, it requires a lot of samples to generalize well, while the number of melanoma sample images is limited. To solve this issue, transfer learning has widely adapted to transfer the knowledge of the pretrained model to another domain or new dataset which has lesser samples or different tasks. This study is aimed to find which method is better to achieve this for early melanoma prediction from skin lesion images. We investigated three pretrained and one non-pretrained image classification models. Specifically, we choose the pretrained models which are efficient to train on small training sample and low hardware resource. The result shows that using limited sample images and low hardware resource, pretrained image models yield better overall accuracy and recall compared to the non-pretrained model. This suggests that pretrained models are more suitable in this task with constrained data and hardware resource.

Keywords: Melanoma; CNN; Pretrained

I. INTRODUCTION

Cancer is a disease caused by uncontrolled abnormal cell growth. In 2008, cancer contributes to 13% of the deaths in the world. One of the most severe cancer types that occurs in humans is skin cancer (Sinaga, 2018). As the outermost layer of the body which covers and protect the

muscles and the organs underneath (Dildar et al., 2021), skin also needs to be protected from ultraviolet (UV) radiations and pollutants (Vijayalakshmi, 2019).

Human skin consists of two main layers: i) epidermis and ii) dermis. Epidermis layer is made of three types of cells, namely squamous, basal that is actively splitting creating keratinocyte, and melanocyte. The most common type of skin cancer is basal cell carcinoma (BCC), followed by squamous cell carcinoma (SCC), and melanoma (Al-Zou et al., 2019). The main cause of most skin cancer cases is overexposure to the sun. Hence, skin cancer will most likely happen on the skin areas that are exposed to the sun extensively such as head area and arm (Dildar et al., 2021).

Although BCC and SCC are the most common types of skin cancer, they are not aggressively spreading to the other organs and relatively easier to cure. In the other hand, while being the less common type of skin cancer, once melanoma occurs it will quickly spread to the other organs if it is not treated in the early stage. In the later stages, melanoma is difficult to cure and potentially fatal (Dildar et al., 2021) (Vijayalakshmi, 2019). Melanoma is the 9th leading cause of cancer-related deaths (Lu & Zadeh, 2022).

Diagnosing melanoma conventionally starts from looking for unusual skin lesions by the doctor, taking skin lesion samples, and observing it under the microscope for further biopsy. This process can be painful for the patients and taking a lot of time. Moreover, the accuracy totally depends on the expertise of the doctor whose average accuracy is less than 80% (Kadampur et al., 2020). Fortunately, recent image classification technology has allowed quick and better diagnosis by utilizing deep learning methods (Kadampur et al., 2020) (Lu & Zadeh, 2022).

One of the most popular deep learning (LeCun et al., 2015) methods to perform image classification is convolutional neural network (CNN). It has the ability to efficiently learn complex features of the image input in the subsequent layers and yield high performance among image classification tasks (Alzubaidi et al., 2021). However, deep learning methods require many labeled samples for training and labeled medical data is scarce. Therefore, this issue limits its usage in the medical setting.

In order to train the model to generalize to new data, we can train a new model to our data or we can use transfer learning on pre-trained model. A pre-trained model is a model which has been trained on larger dataset that may or may not have relation to the target domain. Transfer learning could yield better prediction performance on small datasets. Nevertheless, transfer-learning method yields different result depending on the dataset size and how close the samples is to the dataset used for pretraining the model. For some domains, training the model from scratch yield better prediction accuracy than finetuning a pretrained model (Rezaoana et al., 2022).

A previous study (Kalaiyarivu & Nalini, 2022) compared CNN, Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Light Gradient Boosting Machine (LGBM) for classifying seven skin diseases namely actinic keratosis, basal cell carcinoma, benign lesions of the keratosis, melanoma, melanocytic nevi, dermatofibroma, and vascular lesions from skin images. Their result shows that CNN that was trained from scratch yield highest accuracy of 84% compared to the other machine learning methods. However, one issue of their study is that the data is imbalanced for each dataset. Similar CNN-based method also proposed albeit using different image preprocessing method to classify skin cancer types on small dataset (Aima & Sharma, 2019). However, their model struggles to generalize on such small dataset.

Alwakid et al. (2022) (Alwakid et al., 2022) proposed on using Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) (Wang et al., 2018) to improve the image quality. The enhanced image is segmented to acquire the Region of Interest (ROI) which already provided in the dataset used (Tschandl et al., 2018) and augmented to fix the data imbalance. They used ResNet50 (He et al., 2016) as the classifier. Their result shows that using pretrained ResNet50 model yield significantly better result compared to their own proposed CNN architecture. Another research proposed on classifying benign and malignant skin lesion using Mask-RCNN (He et al., 2017) to select the skin lesion area and using pretrained ResNet152 to perform the binary classification (Acosta et al., 2021).

A study by Muhaba et al. (2022) (Muhaba et al., 2022) utilized pretrained MobileNetV2 to diagnose five types of common skin diseases. The dataset used contain only 200-300 image samples per skin disease type. Their result shows that even only using smartphone camera, the model can achieve 97.5% accuracy, 97.7% precision, and 97.7% recall. This finding proves that the MobileNetV2 is capable to perform well under limited dataset and hardware resource.

This study is aimed on finding out whether training from scratch is more suitable for early melanoma detection than transfer learning from the pretrained model. Furthermore, instead of using deep architecture such as VGG (Simonyan & Zisserman, 2014) and ResNet (He et al., 2016), we proposed using more resource-efficient architectures that can run on non-high-performance computers and mobile devices. We experimented with our own defined CNN based model and three ImageNet (Deng et al., 2009) pretrained models namely DenseNet121 (Huang et al., 2017), MobileNetV2 (Sandler et al., 2018), and EfficientNetV2 (Tan & Le, 2021). The task used in this study is binary classification between melanoma and non-melanoma skin conditions.

II. METHODS

As illustrated in Figure 1, the experiment in this study begin with the preprocessing of the image samples in the dataset. After the images are preprocessed, we experimented with four different CNN-based method. We also evaluated the accuracy, recall, and F1-score from the test result for each model used.

2.1 Dataset

We use HAM10000 (Tschandl et al., 2018) for training and testing. The dataset HAM10000 contains a collection of images of benign and malignant pigmented lesions. These pictures were taken of lesions on an Austrian patient who had very severe chronic sun damage over a period of 20 years. It is also known that people with darker skin color have lower risk of suffering from melanoma (Gloster and Neal, 2006).

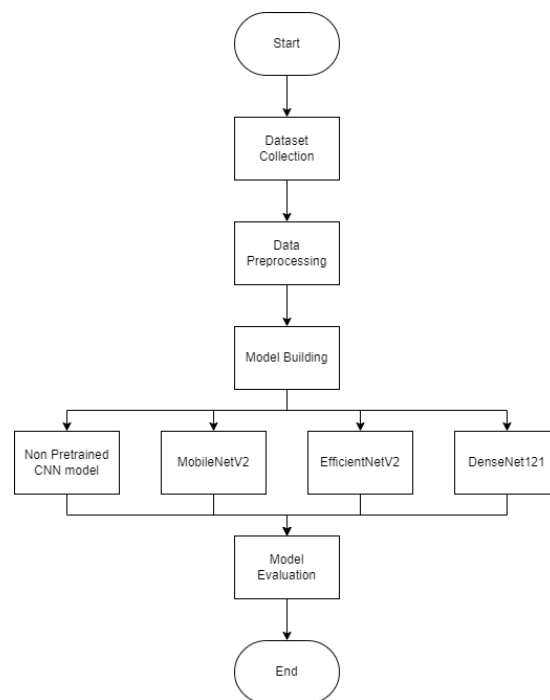


Figure 1. The proposed methodology

The dataset contains 17805 skin images in total where 8903 images are of skins with melanoma. Figure 2 shows an example of each class in the dataset. We split the dataset into 6:2:2 for training, validation, and test respectively.

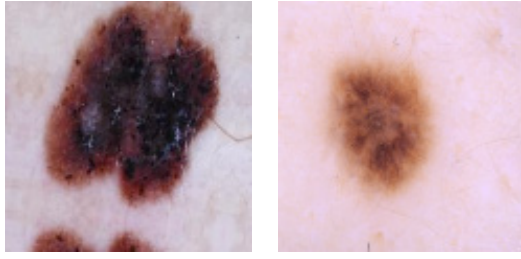


Figure 2. The example of melanoma skin lesion (left) and non-melanoma skin lesion (right).

2.2 Data Preprocessing

The preprocessing steps are done to let the model learn optimally from available samples (Tabik et al., 2017). More samples will lead to better generalization and less prone to overfitting (Alwakid et al., 2022). Each image sample in the dataset is resized into 224*224 pixel. The samples in the training set are augmented to generate more samples. The augmentation process done include image rotation, horizontal image mirroring, and changing the contrast. Lastly, all the images are converted into grayscale. The augmentation results from one of the training sets can be seen in Figure 3.

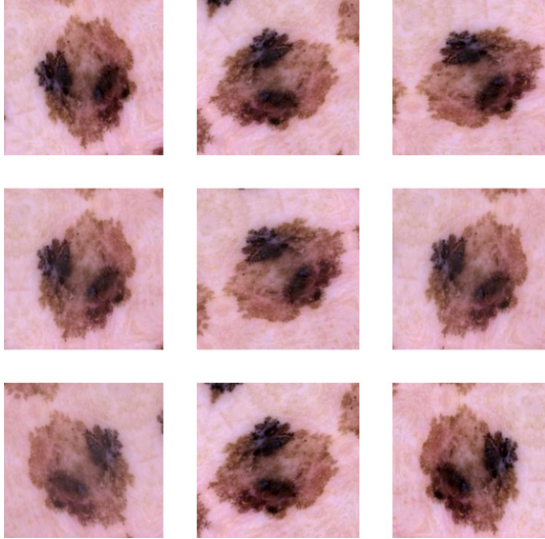


Figure 3. The sample of the augmented image

2.3 Model

We compared our own non-pretrained CNN-based model and three pretrained CNN-based model. For fair comparison, those four models were trained on the same dataset with the same training, validation, and test data. Furthermore, all models used in this study were trained on 30 epochs and batch size of 32. The models were optimized using Adam optimizer (Kingma & Ba, 2014) with the learning rate of 0.0001. Since our task is binary classification, we used the binary cross entropy to calculate the loss during training.

2.3.1 Non Pretrained CNN Model

Our non-pretrained CNN model consists of four convolutional blocks which consists of the sequence of 3*3 sized kernel convolutional layer followed by ReLU activation function and 2*2 max pooling layer. These blocks are then followed by a block that consists of another convolution layer with 3*3 kernel size, ReLU activation function, and a global average pooling layer with 2*2 kernel

size. The features then fed into a series of fully connected layers with dropout layers in between. The full architecture is illustrated in Figure 4.

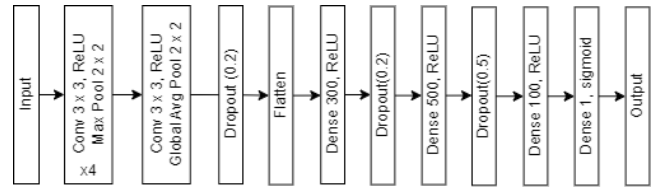


Figure 4. The architecture of the proposed non-pretrained CNN-based model.

2.3.2 Pretrained CNN Model

In this study, we used ImageNet pretrained models. We specifically choose the three models which were claimed to perform well with hardware constraint. The pretrained models used in this study are MobileNetV2, EfficientNetV2, and DenseNet121.

MobileNetV2 is a small-sized model which is originally designed to run on mobile devices. Its low complexity has also become an advantage in this study since the training data is relatively small (Nur et al., 2022). EfficientNetV2 is created for faster training by reducing the model size, thus reducing the number of parameters. It utilizes neural architecture search and progressive learning to further optimize the model (Saragih et al., 2022). DenseNet is a neural network model that is proposed to solve the problem with vanishing gradient. Vanishing gradient often happens with deep architectures, DenseNet solved this by densely connecting layers in order to prevent the gradient loss during backpropagation in the early layers (Bozkurt, 2021). Figure 5 show the architectures of the model used in this study.

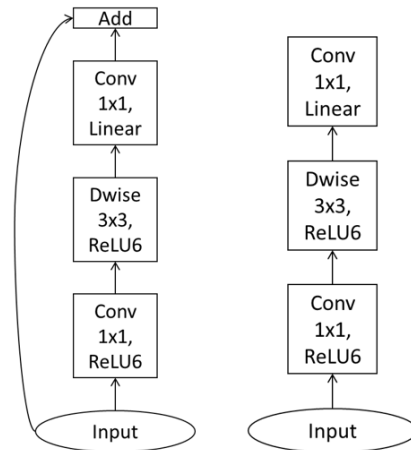


Figure 5. MobileNetV2 architecture. The left side image shows the single stride block with residual connection. The right side image shows dual stride block without residual connection. (Sandler et al., 2018)

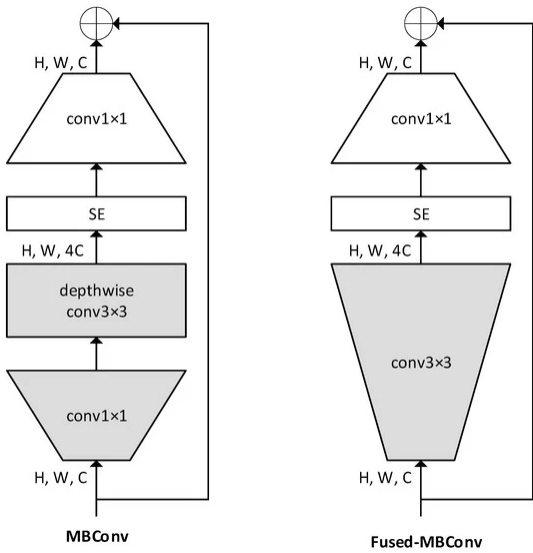


Figure 6. The structures of MBConv (the left side) and Fused-MBConv (the right side) as a fundamental building blocks of EfficientNetV2 (Gomroki et al., 2023)

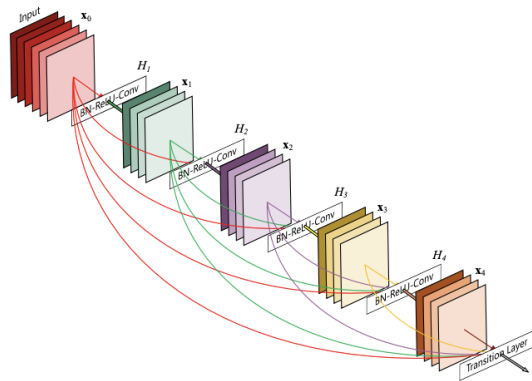


Figure 7. The layers of one dense block from the DenseNet architecture. Each layer takes all the preceding layers as the input (Huang et al., 2017)

2.4 Evaluation

Each model was evaluated by calculating the accuracy, recall, precision, and F1-score. We decided that recall is the most important metric since the task is to detect melanoma in the early stage. It would be fatal for the actual melanoma patients who are predicted as non-melanoma than the other way back. Therefore, it would be better for the model to have lower number of incorrectly actual positives as a negative.

III. RESULTS AND DISCUSSION

The test results from each model are shown in Table I. It was calculated from the confusion matrix produced by each of the models shown in Figure 8. The overall result shows that all the models achieve above 90% accuracy and above 96% precision. However, none of the models achieve higher than 89% recall.

Table I. The calculated accuracy, precision, recall, and F-score yield from the test set of each model architecture.

Model	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
Non-pretrained	91.35	99.53	83.10	90.58
MobileNetV2	93.77	97.56	89.78	93.51
DenseNet121	93.60	99.81	87.37	93.18
EfficientNetV2	93.34	98.43	88.10	92.98

Compared to the pretrained models, the result of the non-pretrained model is lower. This may be caused by the number of samples of which considered small and the regularization technique used. Furthermore, proper parameter tuning would improve the general performance.

Among the pretrained models used in this study, the highest accuracy is achieved by the MobileNetV2 model with 93.77%. However, DenseNet121 achieved the highest precision of 99.81%, while our non-pretrained model achieved only 0.28% lower than this. The highest recall score is 89.78% using MobileNetV2, show in Figure 8.

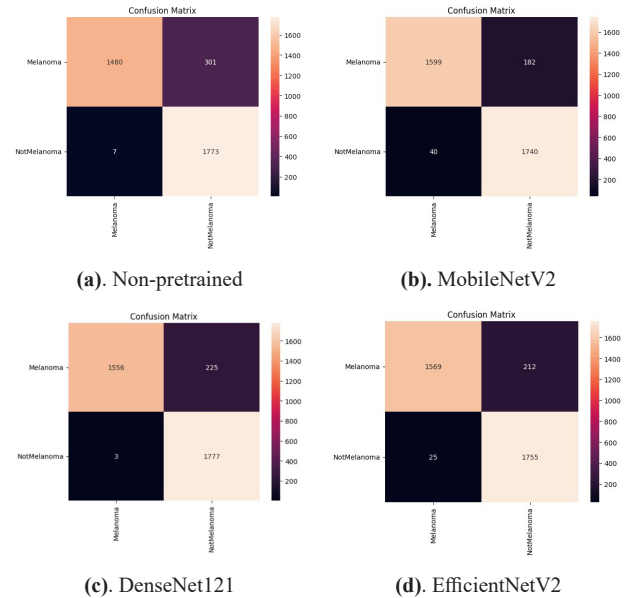


Figure 8. The confusion matrix of the test result from each of the models used. The rows represent the ground truth labels and the columns represent the predicted labels.

Looking at Figure 9, which illustrates the loss curve plotted during the training and validation, we can see that the pretrained model smoothly converges while our non-pretrained model fluctuates significantly. Our small batch size during training might cause this phenomenon, since some difficult samples may not be constantly included in every mini batch thus creating such sharp increases in loss curve.

Furthermore, we can see that pretrained models start from lower loss compared to the non-pretrained model. This suggests that pretrained models already learned the main image features from the dataset they previously trained on. This means that using less epochs, the pretrained models can achieve lower loss, thus reducing the training time.

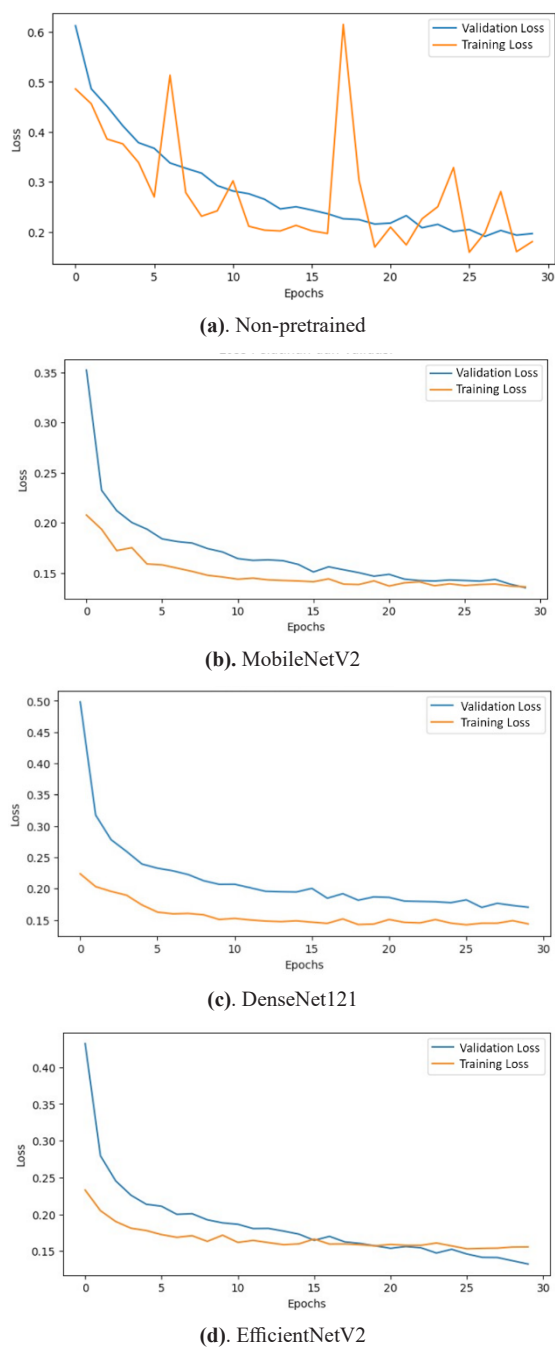


Figure 9. The loss plot for each of the models used within 30 epochs.

IV. CONCLUSION

This study is aimed to find out whether pretrained model is suitable model for detecting whether the given skin lesion image has melanoma or not. Furthermore, we picked the models which are claimed to be efficient to train with less hardware and data resource. Since the implementation is for early detection of melanoma skin lesion, we regard recall score as more important than the other evaluation metric values. The highest recall value of this study is achieved by using pretrained mobileNetV2 model. The result of non-pretrained model also shows that in this task, it is better to use pretrained model than training a model from scratch.

There are several things that are interesting to explore in the future. For instance, the result in this study is

not tested on real world patient data. This poses a challenge as the picture taken from various devices may affect the prediction result. Furthermore, the recall result in this study shows that further studies are necessary to reduce the number of false negatives.

REFERENCES

- Aima, A. and Sharma, A. (2019). Predictive approach for melanoma skin cancer detection using cnn. SSRN Electronic Journal.
- Al-Zou, A., Thabit, M., Al-Sakkaf, K., and Basaleem, H. (2016). Skin cancer: Clinico-pathological study of 204 patients in southern governorates of yemen. *Asian Pacific journal of cancer prevention: APJCP*, 17:3195–3199.
- Alwakid, G., Gouda, W., Humayun, M., and Sama, N. U. (2022). Melanoma detection using deep learning-based classifications. In *Healthcare*, volume 10, page 2481. MDPI.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74.
- Bozkurt, F. (2021). Classification of blood cells from blood cell images using dense convolutional network. *Journal of Science, Technology and Engineering Research*, 2(2):81–88.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (09). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.
- Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., and Mahnashi, M. H. (2021). Skin cancer detection: a review using deep learning techniques. *International journal of environmental research and public health*, 18(10):5479.
- Gloster, H. M. and Neal, K. (2006). Skin cancer in skin of color. *Journal of the American Academy of Dermatology*, 55(5):741–760.
- Gomroki, M., Hasanlou, M., and Reinartz, P. (2023). Stcd-effv2t unet: Semi transfer learning efficient-netv2 t-unet network for urban/land cover change detection using sentinel-2 satellite images. *Remote Sensing*, 15(5).
- He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings*

- of the IEEE conference on computer vision and pattern recognition, pages 770–778
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4700–4708.
- Jojoa Acosta, M. F., Caballero Tovar, L. Y., Garcia-Zapirain, M. B., and Percybrooks, W. S. (2021). Melanoma diagnosis using deep learning techniques on dermatoscopic images. *BMC Medical Imaging*, 21(1):1–11.
- Kadampur, M. A. and Al Riyae, S. (2020). Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Informatics in Medicine Unlocked*, 18:100282.
- Kalaiyarivu, M. and Nalini, N. (2022). Classification of skin disease image using texture and color features with machine learning techniques. *Mathematical Statistician and Engineering Applications*, 71(3s2):682–699.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.
- Lu, X. and Firoozeh Abolhasani Zadeh, Y. (2022). Deep learning-based classification for melanoma detection using xceptionnet. *Journal of Healthcare Engineering*, 2022.
- Muhaba, K. A., Dese, K., Aga, T. M., Zewdu, F. T., and Simegn, G. L. (2022). Automatic skin disease diagnosis using deep learning from clinical image and patient information. *Skin Health and Disease*, 2(1):e81.
- Nur Aziz Thohari, A., Triyono, L., Hestingsih, I., Suyanto, B., and Yobioktobera, A. (2022). Performance evaluation of pre-trained convolutional neural network model for skin disease classification. *JUITA: Jurnal Informatika*, 10:9.
- Rezaoana, N., Hossain, M. S., and Andersson, K. (2020). Detection and classification of skin cancer by using a parallel cnn model. In 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pages 380–386. IEEE.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4510–4520.
- Saragih, R. E., Roza, Y., Purnajaya, A. R., and Kaharuddin, K. (2022). Ambarella fruit ripeness classification based on efficientnet models. *Journal of Digital Ecosystem for Natural Sustainability*, 2(2):55–60.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Sinaga, D. (2018). The evaluation of skin cancer profile in fatmawati hospital centre. *Journal of Education and Practice*, 9(4):21–28.
- Tabik, S., Peralta C´amara, D., Herrera-Poyatos, A., and Herrera, F. (2017). A snapshot of image pre-processing for convolutional neural networks: case study of mnist. *INTERNATIONAL JOURNAL OF COMPUTATIONAL INTELLIGENCE SYSTEMS*, 10(1):555–568.
- Tan, M. and Le, Q. (2021). Efficientnetv2: Smaller models and faster training. In International conference on machine learning, pages 10096–10106. PMLR.
- Tschandl, P., Rosendahl, C., and Kittler, H. (2018). The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9.
- Vijayalakshmi, M. (2019). Melanoma skin cancer detection using image processing and machine learning. *International Journal of Trend in Scientific Research and Development (IJTSRD)*, 3(4):780–784.
- Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C. (2018). Esrgan: Enhanced super resolution generative adversarial networks. In Proceedings of the European conference on computer vision (ECCV) workshops, pages 0–0.