



IJECBE

International Journal of Electrical, Computer and Biomedical Engineering

IJECBE (2023), 1, 2, 136–146
Received (14 June 2023) / Revised (21 June 2023)
Accepted (04 July 2023) / Published (30 December 2023)
<https://doi.org/10.62146/ijecbe.v1i2.13>
<https://ijecbe.ui.ac.id>
ISSN 3026-5258

RESEARCH ARTICLE

Web Application Development Skin Lesion Classification Using Transfer Learning Architecture InceptionResNet-v2

Nanda Ilham Harahap and Fitri Yuli Zulkiffi*

Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Depok, 16424, Indonesia

*Corresponding author. Email: yuli@eng.ui.ac.id

Abstract

The development of machine learning continues from various domains where automation systems are needed. Advanced learning models, such as Convolutional Neural Networks (CNNs) in deep learning, can classify and identify objects even beyond human capabilities. One application is the classification of medical images skin cancer. Automatic diagnosis of skin cancer images is still challenging for CNNs. The use of transfer learning on classification has been leveraged for mobile, accurate, and fast automatic diagnosis. However, such models are imperfect in the categorization of skin lesions. Therefore, this study developed a web application for multiclass classification of 7 classes of disease through Streamlit and HuggingFace, with datasets from HAM10000 using TF Lite-conversion InceptionResNetV2. TF Lite-converted and the model's classification reports were analyzed. The results on EarlyStopping overall accuracy were 87.56%, top-2 95.05%, and top-3 97.46%. Moreover, latency and classification duration were measured on Streamlit Share and HuggingFace Spaces. The findings are Streamlit has a faster average latency (1.17 ms) than HuggingFace (1.49 ms). The latency standard deviation on HuggingFace less consistent (0.49 ms) than Streamlit (0.10 ms). The HuggingFace classification average duration and standard deviation is 116 ms and 5 ms, while Streamlit is better at 97 ms and 2 ms respectively.

Keywords: Skin cancer, Image Classification, Deep Learning, InceptionResNet-v2, Streamlit, HuggingFace, Latency

1. Introduction

The development of applications in the medical field has advanced rapidly, particularly in the sector of cancer diseases. Cancer is a disease that involves the growth of atypical cells with rapid spread or invasion into other parts of the body. Among other types of cancer, skin cancer is a harmful and dangerous type. Skin cancer can be cured only if it is identified at an early stage. In the human body, the skin plays a crucial role in covering the entire body, including muscles and bones. If there are any slight changes in the skin's function, it will affect the entire body system and thus play an important role. The most deadly form of skin cancer is melanoma. It is estimated that the burden of skin cancer will increase due to the gradual thinning of the Earth's ozone layer, which allows more ultraviolet radiation to pass through. According to data by D. O'Sullivan et al. [1] in Canada, one of the main causes of melanoma skin cancer is UV (ultraviolet) radiation exposure, accounting for approximately 62.3% in 2015. In 2018, an estimated 287.7 thousand people worldwide were diagnosed with melanoma [2]. In the US, approximately 9,500 people are diagnosed with skin cancer every day, and in 2019, an estimated 192,310 new cases of melanoma were reported. In the past decade, there has been a 50% increase in skin cancer cases, resulting in 60,000 deaths per year [3]. The five-year survival rate is higher for patients whose melanoma is detected early. The survival rate drops to 71% when the disease reaches the lymph nodes and 32% when it metastasizes to other organs. One of the main strategies to reduce melanoma is prevention through early detection or screening to identify it at the earliest stage, either through self-examination or with the help of a medical doctor, to improve survival by providing timely and appropriate treatment.

The developing diagnostic aid is computer-aided diagnosis (CAD), which can help accurately diagnose skin cancer based on dermoscopy images, thereby improving diagnostic accuracy. With the advancement of technology, CAD has utilized deep learning to achieve more accurate diagnoses compared to previous conventional methods [4]. Deep learning can autonomously learn from a large dataset by exploiting complex architectures through the adjustment of weight and bias values in an artificial neural network (ANN) that simulates the human brain. One such ANN method is the convolutional neural network (CNN). In recent years, Convolutional Neural Networks (CNNs) have outperformed dermatologists in differentiating various types of skin cancer. Therefore, efforts have been made to improve the accuracy of skin cancer classification. There are many CNN architecture models that can be used to classify dermoscopy images. One of the current classification models is the MobileNet [5] architecture, which can generate classification reports with top-3 accuracy, including an overall accuracy of 83.1%, Top 2 accuracy of 91.36%, and Top 3 accuracy of 95.34% on the HAM 10000 dataset. Research conducted by R. Dini et al. [6] on the same dataset, with the EfficientNet architecture, using data augmentation, resulted in classification reports of 85% accuracy, 76% precision, 68% recall, and 71% F1-score, respectively. Another study by S. S. Chaturvedi et al. [7] using the same dataset with pretrained weights from ImageNet improved the classification report using MobileNet with precision, recall, and F1-score of 89%, 83%, and 83%, respectively. However, this research from J. Frederich et al. [8] has not developed a practical application yet, as it still uses a GUI that requires users to understand the installation process, despite

achieving optimal accuracy, precision, recall, and F1-score with the EfficientNet architecture, namely 91%, 76%, 68%, and 71% for EfficientNetB0, and 89%, 78%, 73%, and 73% for EfficientNetB1.

This study aims to leverage the deep layers of InceptionResNet-V2 for the solution, it is an architecture that combines the Inception and ResNet networks [9]. The Inception network tends to be very deep, so it is necessary to replace the filter merging stage of the Inception architecture with residual connections. This combination allows Inception to benefit from the residual architecture while maintaining computational efficiency. The development of deep learning architectures has increased the potential of machine learning, where InceptionResNetV2 is used to redefine pixels in the HAM10000 database [10]. This architecture excels in accuracy on the ImageNet dataset and can also improve accuracy on skin lesion datasets when using pretrained weights from ImageNet, thus assisting in classifying and making efficient diagnostic decisions, especially for human skin lesions.

As per the problem, the detection of skin lesions requires additional resources and costs for remote areas to perform these checks, such as consulting hospitals that have these facilities. Additionally, since the outbreak of COVID-19, the entire sector has shifted towards minimal physical contact with the implementation of Large-Scale Social Restrictions (PSBB). This restricts certain activities of the population in a region suspected of being infected with COVID-19 to prevent the spread of the virus and optimize remote telecommunications for work and daily needs. Therefore, as an alternative, the development of mobile web applications can help address this issue by utilizing technology and advancements in the field of machine learning.

One solution that can be explored is the use of the Streamlit framework to develop an easy-to-use and accessible skin classification system for individuals anywhere. This classification efficiency achieved using TensorFlow Lite [11], which reduces the size of the entire training model to a very small size, from the dataset that is about 6GB to around 50MB file, which can remember the dataset images, through quantization conversion, without significantly changing the classification report, in which accelerates classification in the web application.

With this web application, specialized dermatologists and individuals potentially suffering from skin diseases can use the machine learning model to perform initial examinations from the comfort of their homes. This web application will allow users to upload images of skin lesions and receive information about possible related skin diseases. However, these results do not replace professional medical diagnosis but can provide useful initial insights and help individuals decide whether further consultation with a doctor is necessary. In this study, combining technologies such as TensorFlow, Keras, InceptionResNet-V2, and Streamlit, this web application becomes an effective and efficient solution for detecting and addressing skin diseases, particularly in remote areas or when facing social limitations caused by a pandemic.

2. Dataset and Pre-processing

HAM10000 dataset was chosen for this study, in which HAM10000 (Human Against Machine with 10000 training images) an open access data obtained from the International Skin Imaging Collaboration (ISIC) 2018, that contains various types of skin

cancer [10]. In this dataset there are 7 class labeled, squamous cell carcinoma (actinic keratoses dan intraepithelial carcinoma) as akiec, basal cell carcinoma as bcc, benign keratosis (serborrheic keratosis) as bkl, dermatofibroma as df, melanocytic nevi as nv, melanoma as mel, and vascular skin lesions (Cherry Angiomas, Angiokeratomas, Pyogenic Granulomas) as vasc which shown in Figure 1.

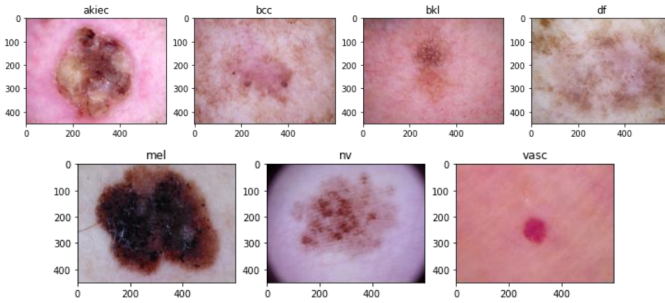


Figure 1. Labeled skin lesion from HAM 10000 ISIC 2018

The exploratory data has been done as shown in Figure 2. The image shown to explain the imbalanced dataset from the HAM1000. In which, the difference between the major data, the larger proportion of the data set, and the minor data, the smaller proportion of the data set, which was almost extreme. This arise a problem for classification to produce lower precision as there are low true lesion. Therefore the pre-processed is initialised to create a more proportional data.

Data preprocessing is the initial step in developing a skin cancer classification system using the InceptionResNetV2 model. This stage involves techniques such as image scaling and normalization to optimize the image data before using it in the model. Image scaling was performed to resize the images to match the expected input size of the model. These images are resized to 150x150 through the rescaling process. Next, the images undergo normalization or pixel intensity normalization to equalize the pixel intensities to a range of 0-1. This normalization process was also applied to augmented images generated using the imagedatagenerator in the Keras library, in which an API that runs deep learning from programming language Python for machine learning purposes through the platform TensorFlow [11]. The augmentation techniques include image rotation (random rotation), image shifting (horizontal and vertical offset), vertical and horizontal flipping (horizontal and vertical flip), and image zooming (random zoom). Furthermore, the dataset, which includes the duplicated images from augmentation, was used in the model training stage. The data was split for validation during the training process, with a split of 0.15 for testing and 0.85 for training. The combination of augmentation, pre-trained weights, and normalization was performed to achieve the best accuracy during testing or model training. This data preprocessing step was crucial for improving the efficiency and accuracy of the machine learning model in skin cancer classification. As a result, the total sample size was combined around 35000 for training, and less than 3000 for testing due to removal of duplicates.

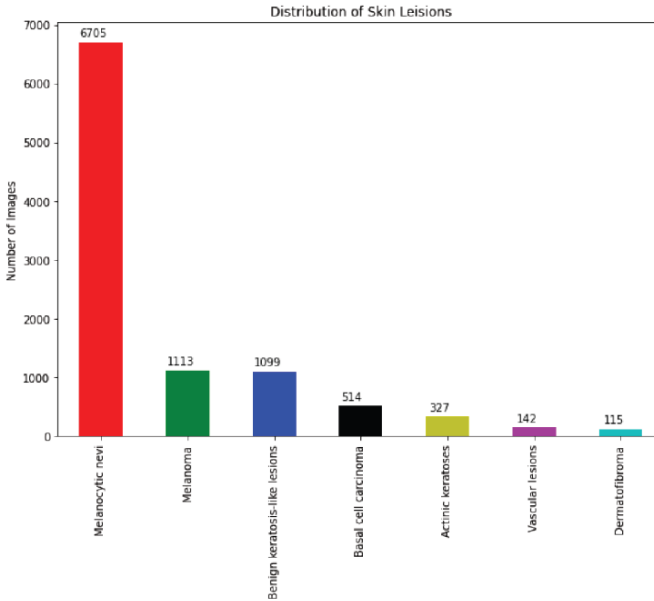


Figure 2. Exploratory data of the HAM 10000 ISIC 2018

3. Implementation and result

Model InceptionResNetV2 was used for the classification model. It is a newer model compared to Inception and ResNet, representing its initial development. According to Szegedy et al. [9], this model can outperform previous CNN models in terms of accuracy. When compared to previous models based on accuracy, it is quite high with a significant number of parameters, and it has successfully classified 1000 objects in the ImageNet database. Transfer learning remains the main approach in other image classification tasks, including the CIFAR-10 dataset and other medical datasets. The testing with the InceptionResNetV2 model has shown its superiority over its predecessors and existing models due to its features, such as the residual connection built on top of the Inception model. It incorporates residual connections instead of the filter concatenation stage of the Inception model, which consists of modules A, B, and C. This allows the model to enhance its learning capability while enjoying the benefits of the residual approach and maintaining computational efficiency. InceptionResNetV2 introduces special "Reduction Blocks" that reduce the size of the feature maps by performing max pooling in each reduction block, followed by feature concatenation or concatenation of Reduction A and B.

The testing in the research, with scheme of study in Figure 3 was initiated by modifying the deep learning layer of InceptionResNetV2, adding layers on top of the architecture, and a layer to rescale the input images to 255 using a sequential model. Below that, a flatten layer was added to reshape the vectors, and a dense layer was added to refine the model for classification. The dense layer with a SoftMax activation function was used to determine the probabilities for the 7 classes, corresponding to

the number of classes in the dataset. It is important to have features that guide model training due to the typically large computational units in deep learning.

The Adam optimizer was used. It is a stochastic gradient descent method based on adaptive estimation of first-order and second-order moments. The Adam optimizer is computationally efficient, requires minimal memory, is invariant to diagonal gradient scaling and is suitable for large-scale problems in terms of data/parameters. The optimizer can be set with a learning rate of 0.01, which is not adaptive as the epoch increases. It starts with a high learning rate and decreases towards the end. The epsilon value of the Adam optimizer is also adjusted to maintain stability in the ImageNet database, set to 0.1. This is important because stability can be disrupted, and the loss value may not decrease if the default value of 10^{-7} is used. Additionally, the research utilizes EarlyStopping and Checkpoint from Keras. EarlyStopping is used to stop the training when the validation loss reaches its lowest point. Checkpoint was used to save the model at the last epoch with the highest accuracy. The trained model was saved in a folder on the device so that it can be accessed later for model improvement. In this training, the maximum epoch was set to 200, with a batch size of 64, which adjusts the computational strength to achieve efficient timing, and early stopping was set to 20 epochs of patience. Categorical cross-entropy was used as the loss function for the imbalanced dataset condition.

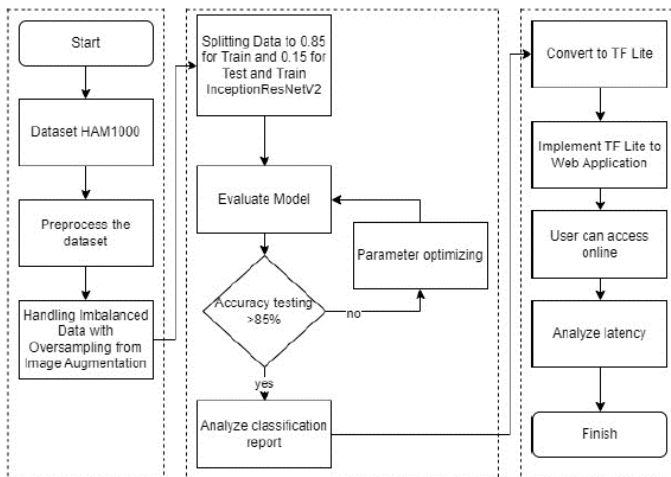


Figure 3. Flowchart of applying TL to pre-train InceptionResNet-v2 for skin cancer classification in web app

After the model has been evaluated, the next step was to convert the model to TensorFlow Lite (TF Lite) for easy implementation on mobile devices or devices with limited resources. With the converted model, the next step was to create a web application hosted by Streamlit and Hugging Face. The web application as shown in Appendix A will be used for real-time classification of skin lesion images. The built web application will allow users to upload skin lesion images and receive classification results from the trained InceptionResNetV2 model. During this process,

the application measures latency and the time required to generate classification results, which was important in assessing the practical performance of the system. Once the web application is successfully developed and tested, analysis on latency will be conducted to evaluate the application’s performance in terms of speed and efficiency. This analysis will help compare and identify areas that need improvement and ensure that the web application can be effectively used by users. The result was then evaluated from both models. The classification results shown in Table I, Table II and the result of top-n accuracy compared to referenced models shown in Table III, also Table IV shows the comparison of classification report between referenced models. In addition to that, the result of the loss and accuracy shown in Figure 4 and Figure 5. Based on the results, it shows that there is a good precision for nevi class, however not for the other classes. This is due to the processing of the false positive that is rather many by the augmented data that has been produced. This means that every duplicate that has been produced by augmentation could not create a good feature, therefore causing a lower precision on classes other than nevi. However, the process of fine tuning of the model has been done using pre-trained weight and other parameters. This resulted in good accuracy for the classification of the lesion because of the extraction feature that has been learned previously.

The implemented model achieved top-1 accuracy: 87,560%, top-2 accuracy: 95,048%, dan top-3 accuracy: 97,464%, the validation accuracy of 85.63% at epoch 24, while the training accuracy at the same epoch reached 93.63%, optimized at 87.56% accuracy. The next research should increase the number of classes for skin diseases worldwide and improve the accuracy and precision using the latest image processing models. The latest models can be developed using deep-layered or parameter-rich Keras models, incorporating augmentation techniques on the latest datasets, and utilizing more powerful devices for improved time efficiency and better accuracy.

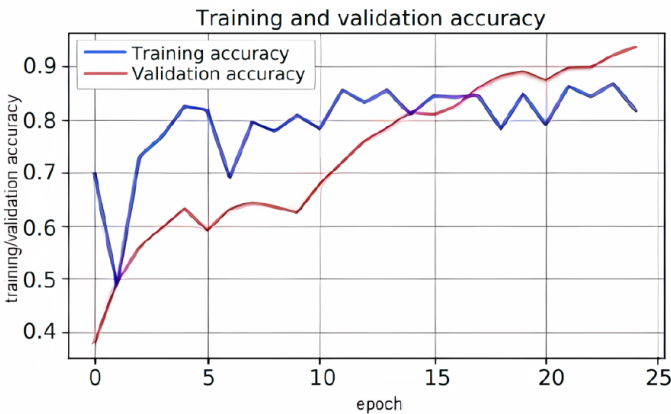


Figure 4. Training and validation loss of the model

The result of classification has also been calculated as shown in Table V showing the durations. In addition to that, when conducting experiments overall, the average duration from sending image data until receiving the inference results showed a result

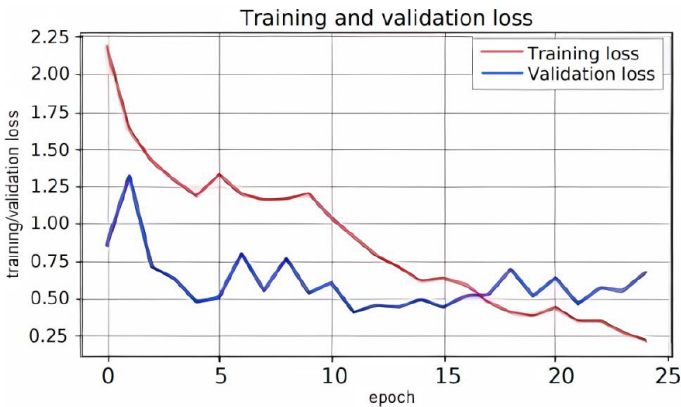


Figure 5. Training and validation loss of the model

Table 1. Classification Report of The Model

Class	Precision	Recall	F1-score	Accuracy
<i>Actinic Keratosis</i>	0.53	0.43	0.48	0.88
<i>Basal Cell Carcinoma</i>	0.59	0.62	0.62	
<i>Benign Keratosis</i>	0.6	0.55	0.57	
<i>Dermatofibroma</i>	0.56	0.83	0.67	
<i>Melanoma</i>	0.46	0.62	0.52	
<i>Nevi</i>	0.96	0.95	0.95	
<i>Vascular Lesions</i>	1	0.8	0.89	
Macro Average	0.67	0.69	0.67	
Weighted Average	0.88	0.88	0.88	

of 132 ms for Streamlit and 147 ms for HuggingFace. These values were obtained using the Python time library by measuring the time from the start line to the end line. In comparing the performance of the two web applications that uses the TensorFlow Lite (TFLite) Interpreter from 30 tests, the average classification time for HuggingFace and Streamlit was also measured using the same method, but specifically at the line of classification. The results showed that the HuggingFace application had an average classification time of 116 ms with a standard deviation of 5 ms. On the other hand, the Streamlit application had a slightly lower average classification time of 97 ms with a standard deviation of 2 ms.

When considering the overall duration, there is an additional time of approximately 33 ms for Streamlit and 30 ms for HuggingFace, apart from the web application's classification duration. This is because both web applications run Python code, which is an interpreted programming language, processing each line of code, and perform

Table 2. Classification Report of The TF Lite Model

Class	Precision	Recall	F1-score	Accuracy
<i>Actinic Keratosis</i>	0.53	0.43	0.48	0.88
<i>Basal Cell Carcinoma</i>	0.59	0.65	0.62	
<i>Benign Keratosis</i>	0.61	0.52	0.56	
<i>Dermatofibroma</i>	0.56	0.83	0.67	
<i>Melanoma</i>	0.45	0.65	0.53	
<i>Nevi</i>	0.96	0.95	0.95	
<i>Vascular Lesions</i>	1	0.8	0.89	
Macro Average	0.67	0.69	0.67	
Weighted Average	0.88	0.88	0.88	

Table 3. Top-N Result of The Model

Model	Top-1 accuracy	Top-2 accuracy	Top-3 accuracy
<i>MobileNet</i> [7]	83,1%	91,36%	95,34%
(Our model) InceptionResNetV2	87,56%	95,05%	97,46%

Table 4. Comparison of Classification Report of Referenced Model

Model	Precision	Recall	F1-score
<i>MobileNet</i> [7]	0.66~0.85	0.51~0.69	0.55~0.77
<i>MobileNetV2</i> [7]	0.32~0.71	0.54~0.81	0.51~0.8
<i>InceptionV3</i> [7]	0.37~0.73	0.52~0.75	0.5~0.76
<i>EfficientNetB0</i> [8]	0.76	0.68	0.71
<i>EfficientNetB1</i> [8]	0.78	0.73	0.73
<i>MobileNet</i> [6]	0.89	0.83	0.83
InceptionResNetV2	0.45~1	0.43~0.95	0.48 ~0.95

library downloads each time the test is conducted.

Furthermore, based on the analysis of latency data, several important points can be concluded regarding the performance of Streamlit and HuggingFace web applications. However, the Streamlit web application excels in terms of ping latency, offering faster response times and smaller variations compared to the HuggingFace web application, as shown in Tabel VI, in which shows that Streamlit has averaged at 1.17 ms, than HuggingFace averaging at 1.49 ms. In addition to that, standard deviation has shown less time for Streamlit at 0.10 ms, than HuggingFace at 0.50 ms. This is due to the sophisticated server-side processing that occur, also the optimization of web cache therefore resulting the number above better for Streamlit.

Table 5. The Total Classification Time Testing

Web Application	Average (ms)	Std (ms)	Min (ms)	Max (ms)
Streamlit	97	2	93	102
HuggingFace	116	5	113	138

Table 6. The Latency Time Testing

Data Latency Web App	Average (ms)	Std (ms)	Min (ms)	Max (ms)
Streamlit	1,17	0,10	0,98	1,40
HuggingFace	1,49	0,50	1,12	3,99

4. Conclusion

In this paper, classification of the 7 multi class dataset of HAM10000 was explained, showing great capability of the model based on the result. Based on the tests that have been carried out, several conclusions are obtained as follows. The model achieved top-1 accuracy: 87,560%, top-2 accuracy: 95,048%, dan top-3 accuracy: 97,464%. The model also achieved precision ranging from 0.43 to 1 for both TF Lite and original model. The model has low latency for Streamlit and HuggingFace, with Streamlit for a bit faster. The next research is expected to increase the number of classes for skin diseases worldwide and improve the accuracy and precision using the latest image processing models. The latest models can be developed using deep-layered or parameter-rich Keras models, incorporating augmentation techniques on the latest datasets, and utilizing more powerful devices for improved time efficiency and better accuracy.

References

- [1] Dylan E O’Sullivan et al. “Estimates of the current and future burden of melanoma attributable to ultraviolet radiation in Canada”. In: *Preventive medicine* 122 (2019), pp. 81–90.
- [2] Jacques Ferlay et al. “Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods”. In: *International journal of cancer* 144.8 (2019), pp. 1941–1953.
- [3] Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. “Cancer statistics, 2019”. In: *CA: a cancer journal for clinicians* 69.1 (2019), pp. 7–34.

- [4] Andre Esteve *et al.* “Dermatologist-level classification of skin cancer with deep neural networks”. In: *nature* 542.7639 (2017), pp. 115–118.
- [5] Jun Gao *et al.* “Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview”. In: *Mathematical Biosciences and Engineering* 16.6 (2019), pp. 6536–6561.
- [6] Harits Abdurrohman, Robih Dini, and Arief Purnama Muharram. “Evaluasi Performa metode Deep Learning untuk Klasifikasi Citra Lesi Kulit The HAM10000”. In: *PROSIDING SEMINAR NASIONAL INSTRUMENTASI, KONTROL DAN OTOMASI*. 2018, pp. 63–68.
- [7] Saket S Chaturvedi, Kajol Gupta, and Prakash S Prasad. “Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet”. In: *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*. Springer. 2021, pp. 165–176.
- [8] J. Himawan J. Frederich and M. Rizkinia. “APLIKASI DIAGNOSIS LESI KULIT OTOMATIS DENGAN MENGGUNAKAN TRANSFER LEARNING ARSITEKTUR EFFICIENTNET”. In: *Asian Congress on Biotechnology* (2022).
- [9] Christian Szegedy *et al.* “Inception-v4, inception-resnet and the impact of residual connections on learning”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 1. 2017.
- [10] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. “The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions”. In: *Scientific data* 5.1 (2018), pp. 1–9.
- [11] *Tensorflow*. https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator. Accessed: 2023-02-15.

Appendix A. Web Application Interface

