Bachelor thesis

Artificial Intelligence

Radboud University

Usability of CNN and Attention Mechanisms for Classifying Melanoma Image

Author:
Nayeong Kim (s1006313)
1215kny@gmail.com

First supervisor:
dr. R.S. van Bergen
Artificial Intelligence
rubenvanbergen@gmail.com

Second supervisor:
dr. T.C. Kietzmann
Artificial Intelligence
t.kietzmann@donders.ru.nl

July 2, 2021

Abstract

Malignant melanoma accounts for about 2% of all malignancies in the Western countries, particularly in the United States, and is a disease that kills more than 9,000 people each year. In general, skin lesions are difficult to detect accurately through visual criteria, but if they are detected well at an early stage, unnecessary time and cost for additional diagnosis can be reduced. This study proposes a solution using a deep learning-based CNN to solve the problem of skin cancer classification. Preprocessing to solve the class imbalance problem is performed and transfer learning architecture to select a backbone architecture model and train it successfully. Furthermore, we apply one of the new deep learning techniques, so-called 'Attention', to the existing model to find out whether the model architecture replaced by the attention layer has better performance. As a result, it is expected that several proposed artificial intelligence algorithms will be utilized to build better computer-aided diagnostic algorithms, which will help early detection of malignant melanoma.

Contents

Chapter 1

Introduction

Skin cancer is one of the most common types of cancer, and melanoma in particular accounts for the majority of deaths from skin cancer, despite it accounts for a small percentage of skin cancers. Melanoma is a fatal disease, but if detected early, most can be treated with mild surgery. Traditionally, doctors have diagnosed skin cancer by looking at all suspicious moles on a person's body with the naked eye. However, this causes many mistakes and is insufficient for accurate detection, because even experts have a hard time identifying cancer as malignant especially when it's in its early progress.

As with other skin cancers, if melanoma can be detected early and accurately using deep learning-based CNN models with the help of data science, this will make treatment much more effective. In the United States, an international skin imaging collaboration (ISIC) focused on automatic analysis of skin lesions has been established to collect related data and lay the foundation for research. ISIC provides more than 2,000 melanoma images collected from clinical centers around the world along with expert diagnosis results. In this study, therefore, using dermatoscope images and diagnostic data provided by ISIC in 2020 to automate the entire pipeline using computer vision. Using a deep neural network such as AlexNet that can be trained on thousands of images of benign and malignant categories, the model determines for itself whether a new image belongs to the benign or malignant class. Then, to see if the attention mechanism, one of the emerging deep learning techniques, can help improve diagnostic performance, we conduct further research using a relatively simple model called an All Convolutional Net(AllConvNet) to confirm the results.

This technique can prevent misdiagnosis, increase the likelihood of early detection, and provide higher efficiencies in treatment, such as reducing unnecessary biopsies. In particular, since this model is non-invasive and has a very low cost compared to other scrutiny, it can be utilized for human health in places such as developing countries in Africa where there are no specialized doctors.

Chapter 2

Preliminaries

Malignant melanoma refers to a malignant skin cancer composed of melanocytes. Melanocytes are normal cells in the skin or mucous membranes, and cells called melanin made by these cells determine skin color. For example, sunbathing or excessive sun exposure cause the melanocytes to make more melanin, which leads the skin to darken. Because melanocytes are congenitally large, they often tend to develop into malignant melanoma. Cancer that occurs in normal melanocytes is called melanoma.

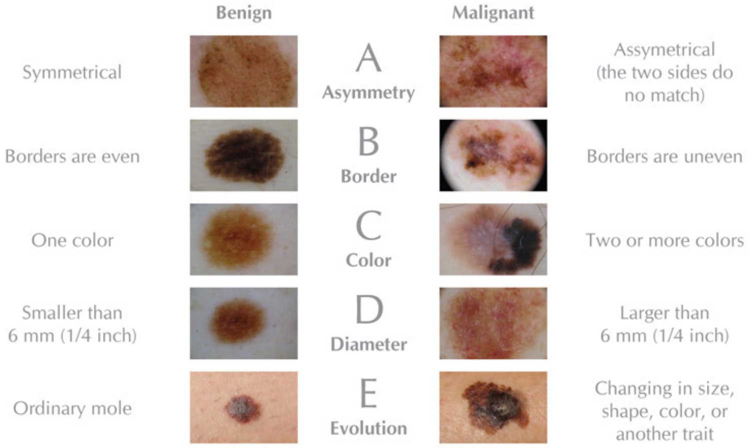

Benign
Malignant
Symmetrical
A
Asymmetry
Assymetrical (the two sides do no match)
Borders are even
B
Border
Borders are uneven
One color
C
Color
Two or more colors
Smaller than 6 mm (1/4 inch)
D
Diameter
Larger than 6 mm (1/4 inch)
Ordinary mole
E
Evolution
Changing in size, shape, color, or another trait

Figure 2.1: Melanoma ABCDE criteria, adapted from ”HC Marbella International Hospital”

Malignant melanoma is a disease with a high cure rate if detected early, so it is important to detect it as quickly and accurately as possible. Diagnosis of malignant melanoma is often observed with normal macroscopic observation or using images acquired using a dermoscope, which is often confused with a normal black mole. A visual criterion exists to distinguish melanoma from a normal black mole [8]. This clinical assessment is often referred to as ABCDE [12]. See figure 2.1.

However, even if a diagnosis is made according to the above ABCDE rules, this task is not easy to achieve with high accuracy. Therefore, this study proposes an algorithm that can assist medical experts in diagnosis using deep neural networks. The image used in the study is a medical image of an unspecified patient, and it is difficult to apply criteria such as diameter(D) and size change(E) because the distance taken is different depending on the image. Therefore, this study is based on a model that has learned only the features of A, B, and C criteria of the lesion area.

Chapter 3

Related Work

Many studies have been published on the task of classifying skin cancer using deep learning and computer vision techniques. These tasks use different approaches, including classification, segmentation and detection, and image processing using different types of filters. The work related to the presented study is:

Almansour et al. developed a melanoma classification algorithm using k-means clustering and a Support Vector Machine [2]. Esteva et al. developed a melanoma diagnostic algorithm [3]. Dorj et al. developed a convolutional neural network with more than 50 layers in the ISBI 2016 challenge dataset for the classification of malignant melanoma [18]. In 2018 Brinker et al. used deep convolutional neural networks to classify binary class problems in dermatoscopy images [17]. In 2019, Yiqi Yan et al. proposed an attention-based method for melanoma recognition [19]. 2020 Qishen Ha et al. proposed ensemble of CNN models with different backbones and input sizes [15].

Next, it contains attention function, which has been widely used in the design of attention models in deep learning. In a separate problem involving handwriting generation for a given sequence of text, Graves designed a differentiable attention model that aligns text characters with much longer pen traces. Here the sort only goes in one direction [10]. Inspired by the idea of learning alignment, Bahdanau et al. proposed a differentiable attention model without severe unidirectional alignment restrictions [6]. Ashish Vaswani et al. proposed Transformer, a new simple network architecture based only on the attention mechanism, completely eliminating recursion and convolution [5].

This research is also based on the aforementioned approach. This research deals with binary type skin cancer classification, and uses oversampling and data augmentation techniques as preprocessing steps for the task. It compares the performance of AlexNet and AllConvNet [1][14]. Afterwards, by replacing the first convolution layer of the AllConvNet with an attentional layer, it checks whether the newly constructed model can have as much classification performance as the existing AlexNet by improving the diagnostic performance.

Chapter 4

Research

In this study, the deep learning training and test data sets are performed through the procedure shown in Figure 4.1. The research was carried out in the Python environment of google colab, and the use of GPU is required because CUDA-based libraries of pytorch and torchvision packages are used.

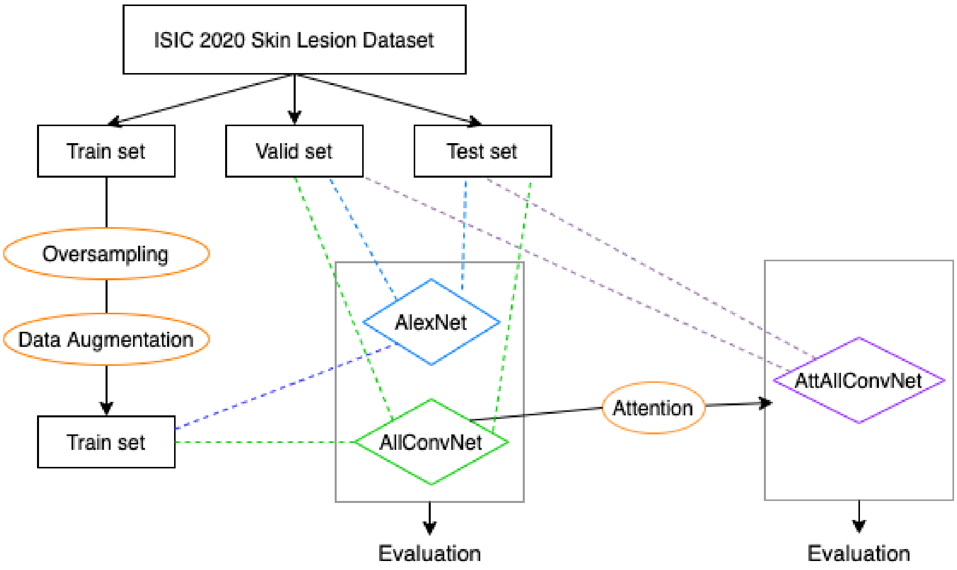
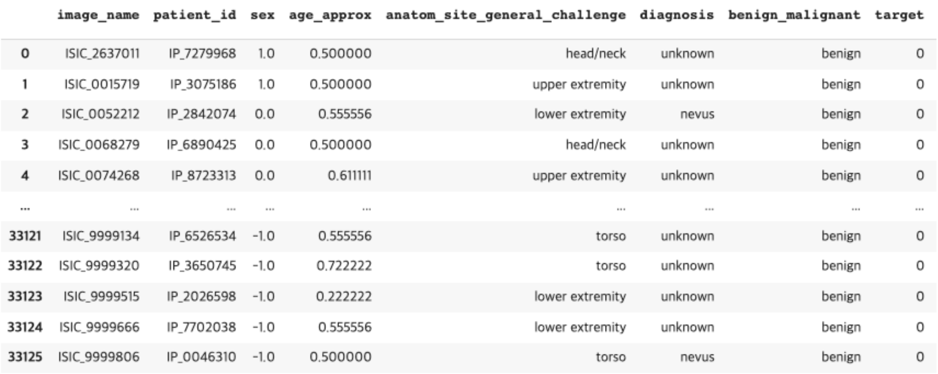


Figure 4.1: The procedure of research

4.1 Database

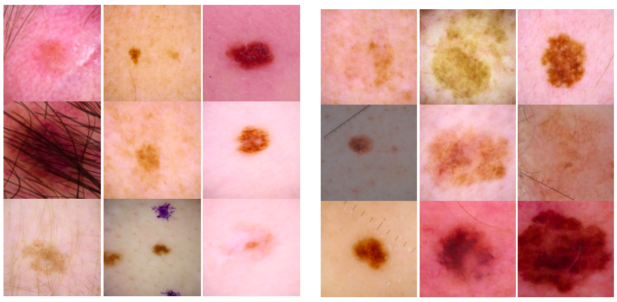
In this study, publicly available skin lesion image data set provided by ISIC in 2020 are used. The skin lesion images were cropped in a square shape based on the central lesion area and then adjusted to 128x128 size. In addition, the data includes not only the patient's skin lesion image, but also metadata such as the location of the mole and the patient's age and gender.

The most focused part of this study among them is the 'target' category Figure 4.2. It is represented as a binary number and is an entry indicating whether the patient is actually melanoma. Figure 4.3 shows images of a benign sample and a malignant sample classified according to the value of the target category.



	image_name	patient_id	sex	age_approx	anatom_site_general_challenge	diagnosis	benign_malignant	target
0	ISIC_2637011	IP_7279968	1.0	0.500000	head/neck	unknown	benign	0
1	ISIC_0015719	IP_3075186	1.0	0.500000	upper extremity	unknown	benign	0
2	ISIC_0052212	IP_2842074	0.0	0.555556	lower extremity	nevus	benign	0
3	ISIC_0068279	IP_6890425	0.0	0.500000	head/neck	unknown	benign	0
4	ISIC_0074268	IP_8723313	0.0	0.611111	upper extremity	unknown	benign	0
...	...	...	...	...	...	...	...	...
33121	ISIC_9999134	IP_6526534	-1.0	0.555556	torso	unknown	benign	0
33122	ISIC_9999320	IP_3650745	-1.0	0.722222	torso	unknown	benign	0
33123	ISIC_9999515	IP_2026598	-1.0	0.222222	lower extremity	unknown	benign	0
33124	ISIC_9999666	IP_7702038	-1.0	0.555556	lower extremity	unknown	benign	0
33125	ISIC_9999806	IP_0046310	-1.0	0.500000	torso	nevus	benign	0

Figure 4.2: ISIC 2020 skin lesion metadata



(a) Benign samples (b) Malignant samples

Figure 4.3: Benign and malignant samples in ISIC 2020 image data

4.1.1 Data Distribution

Before training, a total of 33126 data is split into train set : valid set : test set at a ratio of 80:10:10. In order to have the correct training performance, the ratio between benign and malignant samples should be kept the same during data set segmentation. See Figure 4.4.

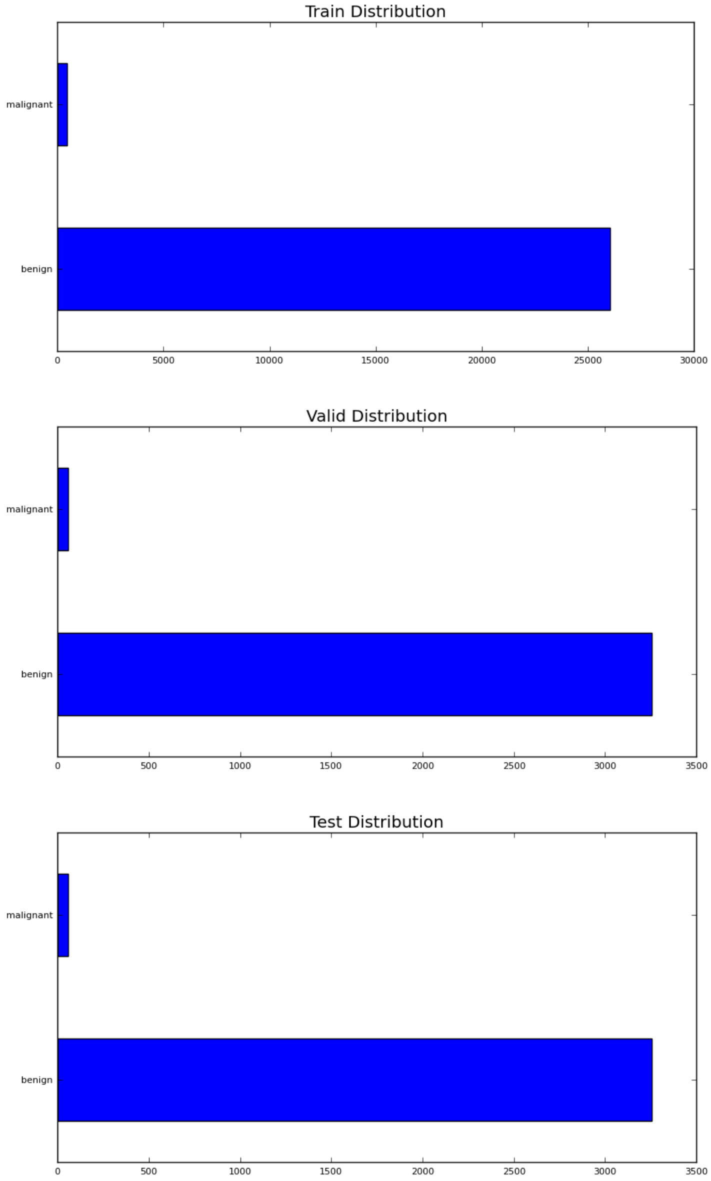


Figure 4.4: Split Data Distribution

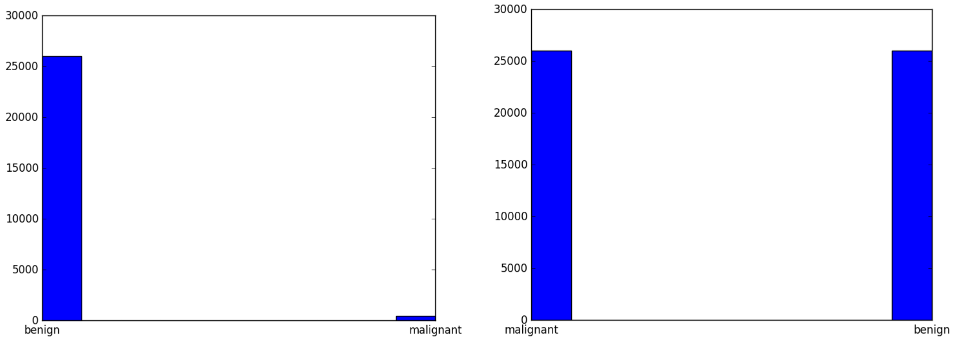
4.2 Preprocessing

4.2.1 Imbalanced Class Problem

It is well known what to do as part of the preprocessing process before we get into model training. It is to solve the problem of data imbalance between classes. Because most machine learning algorithms are designed to maximize accuracy and reduce errors, they work best when each class has approximately the same number of samples.

You can see that the difference in the number of samples between the two classes currently in the train set is too large. This often leads the class imbalance problem. As benign samples are about 55 times more present than malignant samples, this would cause to overfitting. Therefore, if the training proceeds as it is, there will be a problem of classifying all data only as positive.

As a solution to this, there are methods such as re-waiting, which imposes a weight on the loss function, and re-sampling, which adjusts the number of samples in the class itself. In this study, oversampling, a method of directly adjusting the number of samples by making a copy of the minority class in the data, was applied. Oversampling would be a good choice when you don't have much data to work with. I use the resampling module provided by the Scikit-Learn library in Python to randomly replicate samples of a minority class.



(a) Original distribution (b) Oversampled data distribution

Figure 4.5: Changes in sample distribution between classes depending on oversampling

It is important to note that before attempting oversampling techniques, always split into training and test sets. This is because oversampling the data before splitting it will likely result in exactly the same observations on both the training and test sets. This often causes the model to simply remember certain data points, leading to overfitting and poor generalization to the test data.

4.2.2 Data Augmentation

As another method to prevent overfitting in small and medium-sized data sets, a technique called image augmentation is very effective. In this study, I used the following functions from Almentations, one of the most popular and powerful Pytorch augmentation libraries: Transpose, Flip, Rotate, RandomBrightness, GaussNoise, OpticalDistortion, RandomContrast, MotionBlur, CLAHE, HueSaturationValue, MedianBlur, Gaus- sianBlur, GridDistortion, ShiftScaleRotate, ElasticTransform, Cutout [7][15].

4.3 Model

4.3.1 AlexNet

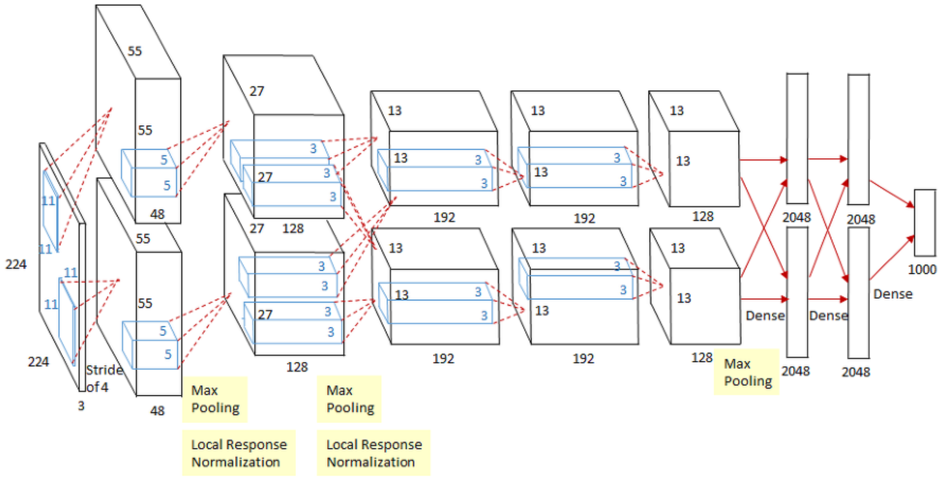


Figure 4.6: A structure of AlexNet

AlexNet is a convolutional neural network structure that won the ImageNet Large Scale Visual Recognition Challenge held in 2012. AlexNet consists of 8 layers. It consists of 5 convolutional layers and 3 fully-connected layers [1]. Figure 4.7 shows the structure of AlexNet. The second, fourth, and fifth convolutional layers are only connected to the feature maps of the same channel in the previous stage, while the third convolutional layer is connected to all the feature maps of the two channels in the previous stage. In this research, the number of last output dimension was changed from 1000 to 2 for binary classification.

4.3.2 AllConvNet

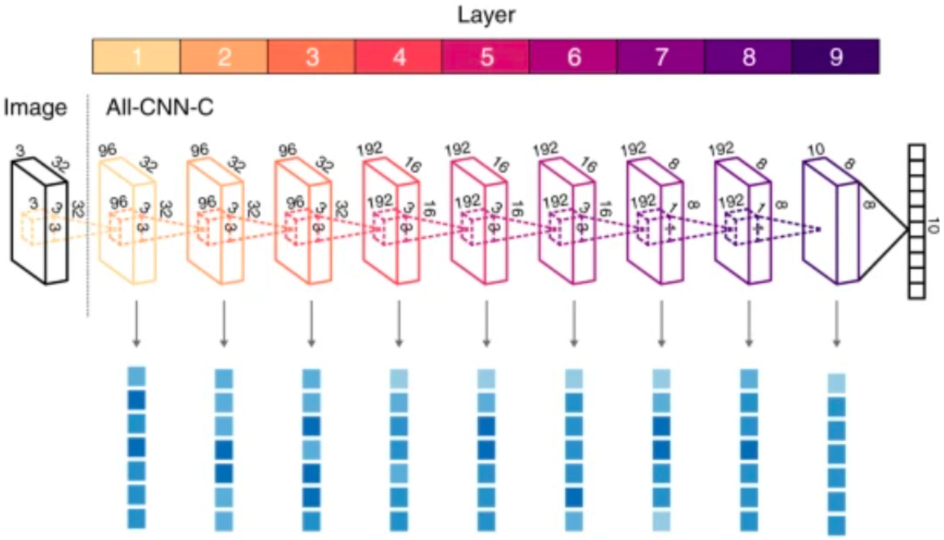


Figure 4.7: A structure of AllConvNet, reprinted from "Individual differences among deep neural network models, by Mehrer, J., Spoerer, C.J., Kriegeskorte, N. et al, Nat Commun 11, 5725 (2020)."

Most modern convolutional neural networks used for object recognition are built using the same principles. It is followed by alternating convolutional and max pooling layers and a few fully connected layers. AllConvNet questions the need for various components in pipelines with CNN. As a result, tasks such as Max-pooling could simply be replaced by convolution layers with increased stride length without loss of accuracy in pattern recognition [14]. The structure of AllConvNet constructed from this point of view is shown in Figure 4.8 above [13]. The number of last output dimension was changed from 10 to 2 for binary classification in this research.

4.4 Results

After running model training using the aforementioned two backbone models, the performance was compared using confusion matrix and ROC-curve.

4.4.1 Result of AlexNet

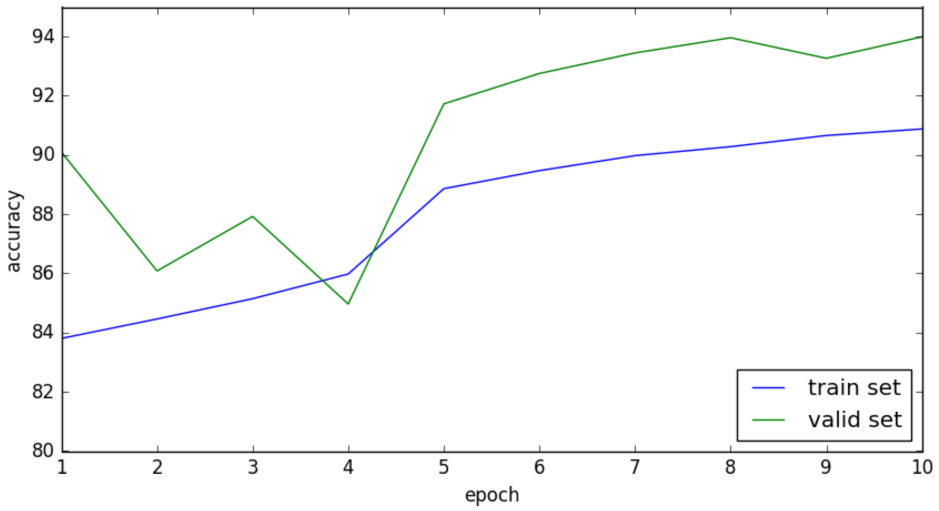
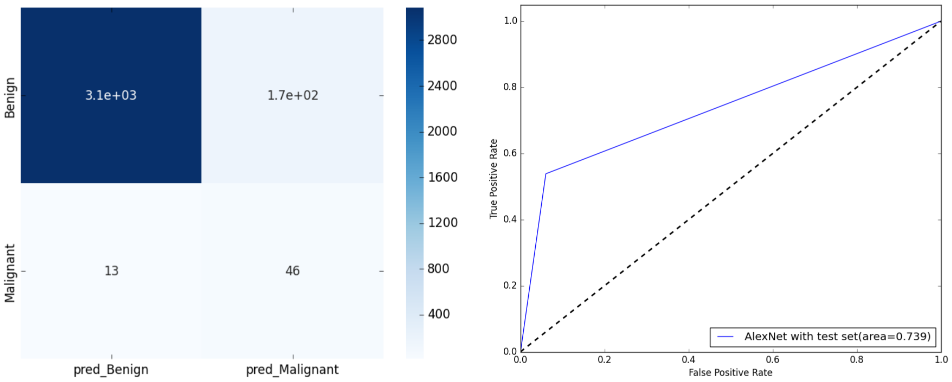


Figure 4.8: Accuracy on AlexNet model (#epochs=10)



(a) Confusion matrix

(b) ROC curve

Figure 4.9: The results of AlexNet model with test set

4.4.2 Result of AllConvNet

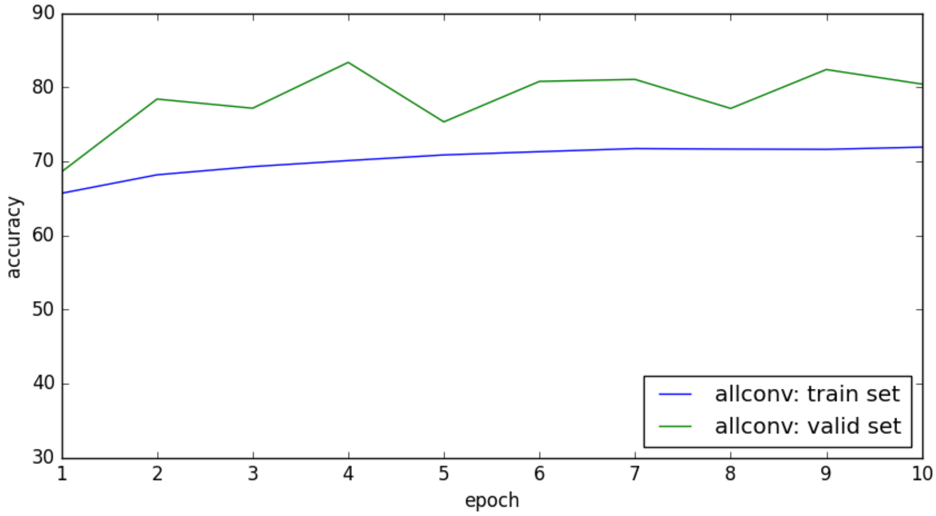
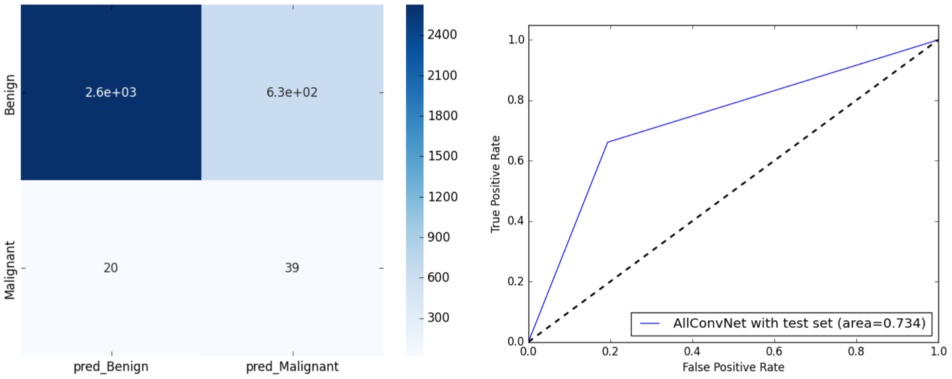


Figure 4.10: Accuracy on AllConvNet model (#epochs=10)



(a) Confusion matrix

(b) ROC curve

Figure 4.11: The results of AllConvNet model with test set

4.4.3 Evaluation

After running model training using the aforementioned two backbone models, the performance was compared using confusion matrix and ROC-curve. As a result of looking at the two models, we found that AlexNet is better in terms of accuracy and classification performance. However, due to the nature of the medical field, an accuracy of about 80% may be of some help, but it does not mean it is very useful. Therefore, we try to see if we can improve the existing model into a model with better performance by using a technique called attention mechanism to the relatively simple AllConvNet.

4.5 Attention Mechanism

4.5.1 Attention Mechanism

When training an image model, we want the model to be able to focus on important parts of the image. One way to achieve this is to use a trainable attention mechanism. It is not intended to change what the model would learn, but it is only used to provide insight into the model's decisions by using fixed weights [5].

In the context of attention mechanisms we refer to volitional cues as queries. Attention mechanisms, given all queries, bias selection over sensory input or intermediate features through attention pooling. These sensory inputs are called values. In general, all values are paired with a key, which are considered volitional cues of sensory input. Figure 4.13 shows the process of designing attention pooling so that a given query interacts with a key that guides a biased selection on a value. In other words, the attention mechanism is basically to find the similarity between the query and the key and reflect this weight to the value, so that the encoder element to pay attention to is reflected to the decoder [20] [19].

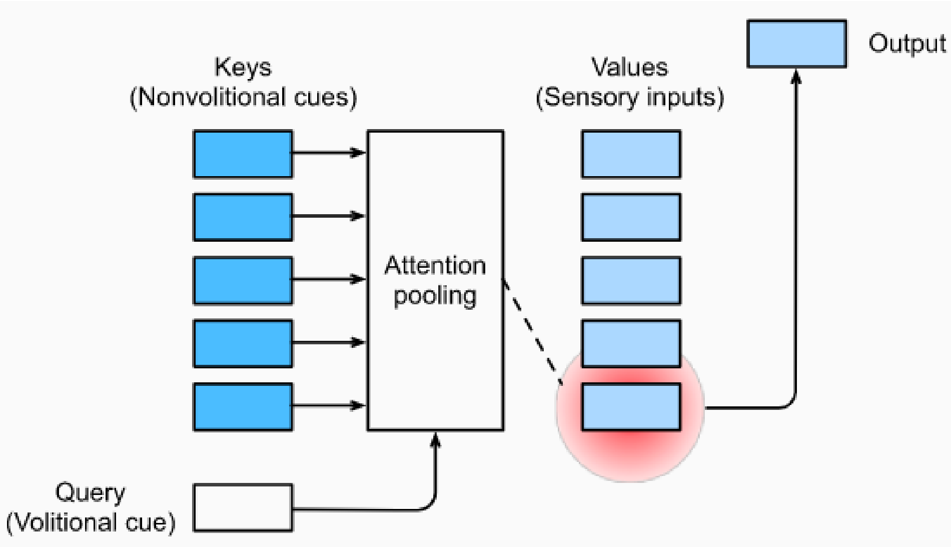


Figure 4.12: Using the volitional cue (want to read a book) that is task-dependent, attention is directed to the book under volitional control.

This is one of the newer mechanisms currently performing well in natural language processing. In this study, we construct a new attention layer for computer vision, not for natural language processing. The first convolution layer of backbone model is replaced with a newly constructed attentional convolution layer. In other words, the structure of the newly constructed model is exactly the same as the backbone model AllConvNet, except that only the first layer is replaced with an attention layer, and the structure, number of input / output channels, and kernel size are all exactly the same.

4.5.2 Results

Figure 4.15-17 show the results of training with 5 epochs and test values the valid dataset for the backbone model and the model to which the attention mechanism is applied.

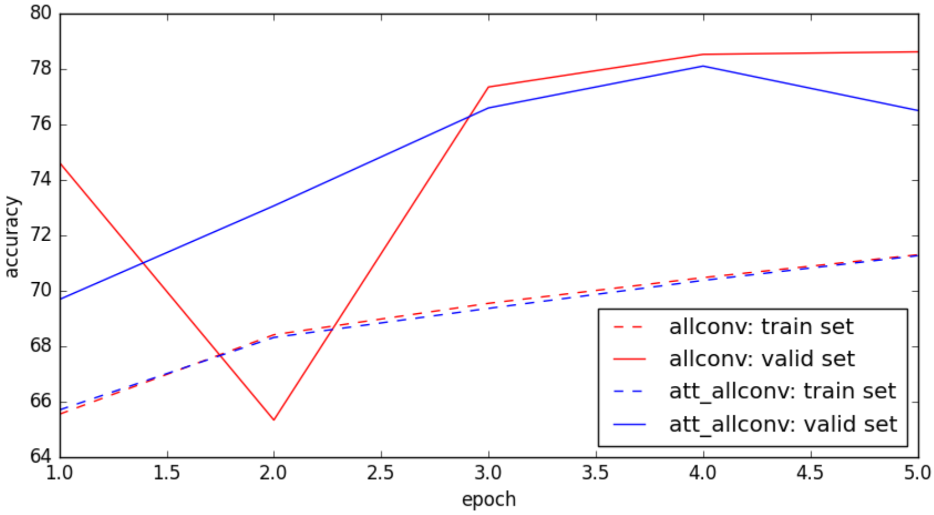


Figure 4.13: Accuracy per epoch (#epochs=5)

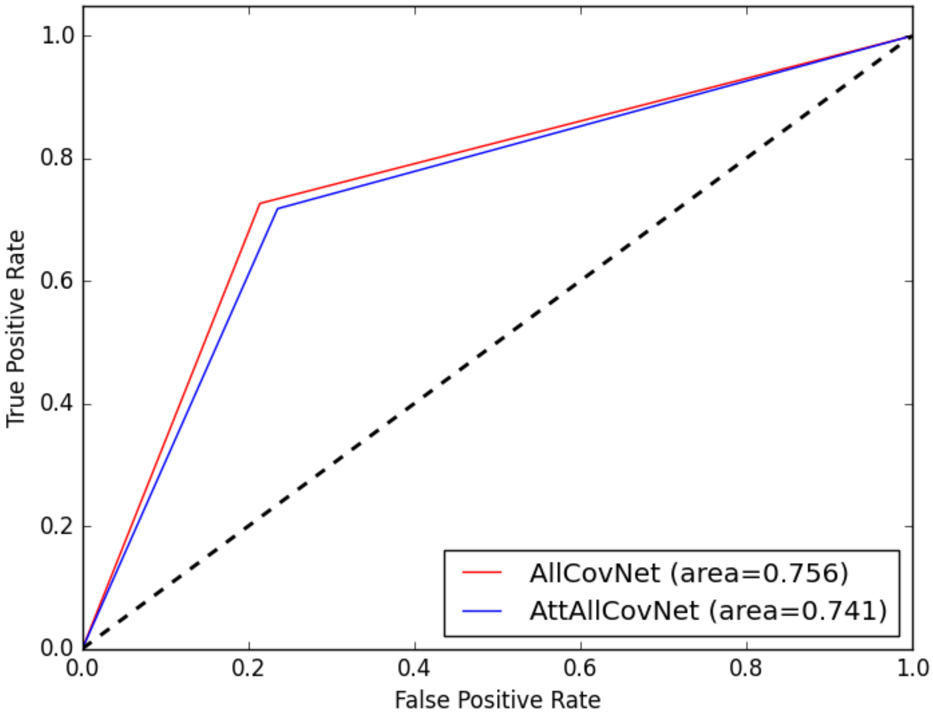
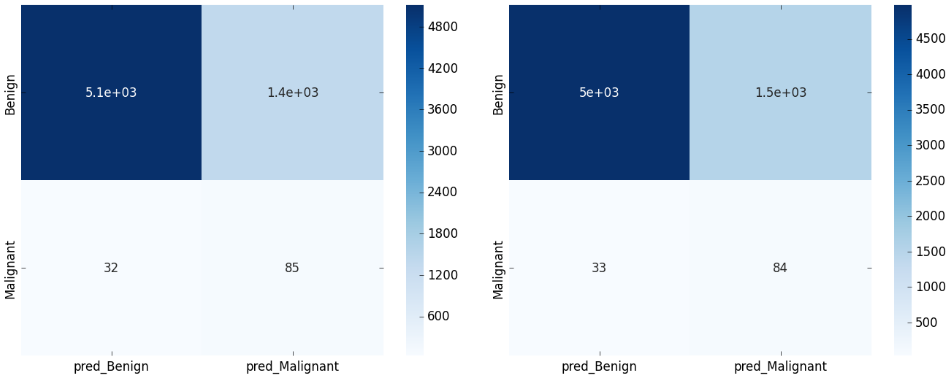


Figure 4.14: ROC curve



(a) Confusion matrix of AllConvNet (b) Confusion matrix of AttAllConvNet

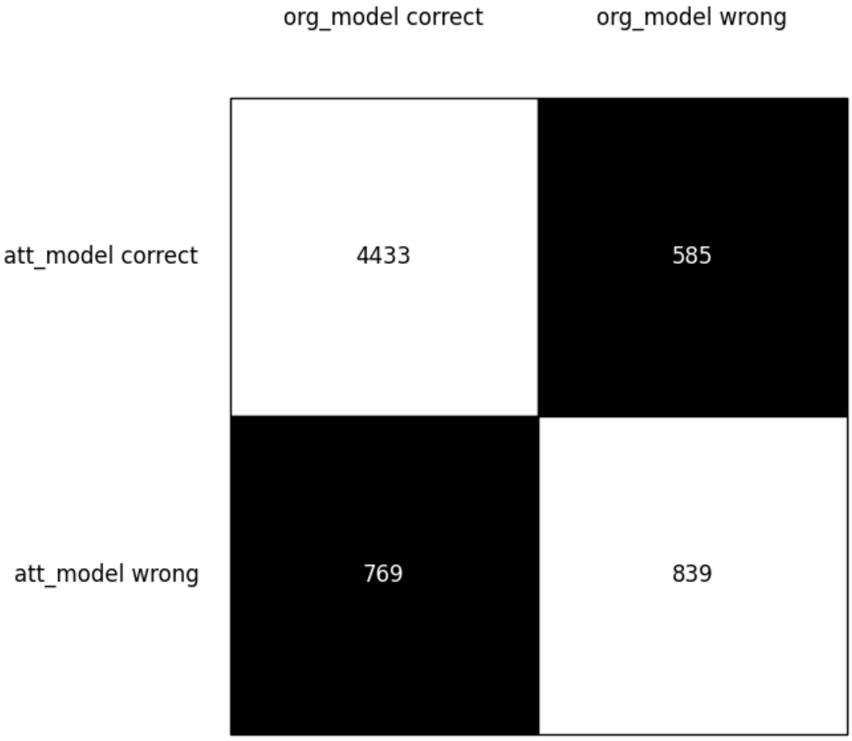
Figure 4.15: Confusion matrices of AllConvNet and AttAllConvNet

For the conventional backbone model, it correctly diagnoses 5117 out of a total of 6509 benign samples, and correctly diagnoses 85 out of a total of 117 malignant samples, showing about 79% and 73% accuracy for each class. This model has: precision = 0.7861, recall = 0.9937, and accuracy = 0.7851.

In the case of AttAllConvNet to which the attention mechanism is applied, 4978 out of a total of 6509 benign samples and 84 out of a total of 117 malignant samples are correctly diagnosed, with an accuracy of about 76% and 72%, respectively. This model has: precision = 0.7648 , recall = 0.9934, accuracy = 0.7639.

4.5.3 Evaluation

It should be noted that, contrary to our earlier expectations, when applying the attention mechanism, there is no obtained noticeable valid results. Although there is a very subtle difference in the classification results between the two models, this is too small to actually imply a clear performance difference between the models. Therefore, we use Mcnemar's test to decide the performance difference between the two classification models. The 2x2 continuity table used in Mcnemar's test is a useful tool for comparing two different models. Unlike a typical confusion matrix, this table compares two models to each other without displaying the false positives, true positives, false negatives, and true negatives of a single model prediction [16]. Figure 4.18 shows the contingency table comparing normal AllConvNet and Attentional AllConvNet.



	org_model correct	org_model wrong
att_model correct	4433	585
att_model wrong	769	839

Figure 4.16: Contingency table of two models

As a result of performing Mcnemar's test using this, it rejects the null hypothesis(H0) with the values of statistic=24.733 and p-value=0.000. In other words it does not assume that there is no difference between the two predictive models.

Chapter 5

Conclusions

5.1 Discussion

1. The number of malignant samples, which was too small compared to the benign samples, was very limited to training the model, despite using oversampling. Also, for some patients, their hairs are included in the mole image, which may affect the image learning performance of the model. If features such as hairs can be processed separately in the data augmentation process, there is a possibility that the performance of the model can be improved.

2. This study was about training an image model for a specific classification, and applying an attention mechanism to that model. In research process, there was not enough time to tune hyperparameters outside layer and model configuration. If you adjust the hyperparameters and choose a more complex and performance backbone architecture, you will get better results from that study.

3. In the process of applying the attention mechanism to the feedforward model, only the first convolution layer was replaced. This was an inevitable choice due to the limited GPU memory of the working environment. If we replace all convolution layers of the feedforward network with attentional layers given a faster and sufficiently large working environment, this can cause a noticeable difference in the comparison of diagnostic performance with the backbone model. However that does not mean it will have better performance.

4. Similarly, it seemed impossible to increase the number of training epochs to more than 5 per model because attention model training was time-consuming. However, according to the results of Mcnemar's test, it seems that there would be a difference between the attention model and the backbone model, so it is worth looking into in future

studies to check the results again by increasing the number of training epochs to 20 or more.

5. To replace the convolution layer of the existing network with an attentional layer, a relatively less complex model is chosen. However, this showed lower performance compared to more complex backbone models such as AlexNet. There are already more complex and high-accuracy models in the pytorch model library, such as Inception v3, MobileNet, and ResNet152 [4][11][9]. Therefore, this study only suggests a direction for how to apply the attention mechanism to a classification task, and it does not mean that you can make simple model work better than a good model that already exists even if putting more effort in a limited environment.

6. For future usability, this research requires seeking some opinion from an actual medical professional. In the medical field, no matter how high the accuracy is, it is not enough, and it is important to know how much value the model actually has clinically. One option is to present the images classified by the model directly to an expert for advice on whether they agree with these results.

5.2 Conclusion

In conclusion, this study investigates the ability of deep CNN in melanoma classification and suggests that how further research can occur in improving the performance of the model. Our results show that a deep learning architecture trained on dermatological image data using computer vision is actually better than the decision that dermatologists make by identifying with the naked eye. This study was conducted through pre-processing and transfer learning. However, judging from the results of additional studies applying the attention mechanism, it seems that there is no need to additionally replace the convolution layer of the existing network with an attention layer, because there are already sufficiently complex and high-performing networks. Nevertheless, the fact that computer vision technology using pattern recognition is helpful in making an initial diagnosis suggests a lot of potential for future applications in the medical field, such as the development of melanoma self-diagnosis applications or its use in areas such as developing countries, where there is a shortage of medical personnel. In addition to the image dataset used in the study, using metadata to extract information such as the site and size of frequent occurrence of melanoma can also help to increase the accuracy of the model.

Bibliography

[1] Ilya Sutskever Alex Krizhevsky and Geoffrey E. Hinton, *Imagenet classification with deep convolutional neural networks*, 2012, `https://papers.nips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf`.

[2] Ebtihal Almansour and M. Arfan Jaffar, *Classification of dermoscopic skin cancer images using color and hybrid texture features*, 2016, `http://paper.ijcsns.org/07_book/201604/20160421.pdf`.

[3] Roberto A Novoa Justin Ko Susan M Swetter Helen M Blau Andre Esteva, Brett Kuprel and Sebastian Thrun, *Dermatologist-level classification of skin cancer with deep neural networks*, 2017, `https://www.nature.com/articles/nature21056`.

[4] Suyanto Suyanto Arief Budhiman and Anditya Arifianto, *Melanoma cancer classification using resnet with data augmentation*, 2020, `https://ieeexplore.ieee.org/abstract/document/9034624`.

[5] Niki Parmar Jakob Uszkoreit Llion Jones Aidan N. Gomez Lukasz Kaiser and Illia Polosukhin Ashish Vaswani, Noam Shazeer, *Attention all you need*, 2017, `https://arxiv.org/abs/1706.03762`.

[6] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, *Neural machine translation by jointly learning to align and translate*, 2015, `https://arxiv.org/pdf/1409.0473.pdf`.

[7] E. Khvedchenya V. I. Iglovikov A. Buslaev, A. Parinov and A. A. Kalinin, *Albumentations: fast and flexible image augmentations*, 2018, `https://arxiv.org/abs/1809.06839`.

[8] Jacqueline Dinnes et al. and Cochrane Skin Cancer Diagnostic Test Accuracy Group, *Visual inspection for diagnosing cutaneous melanoma in adults*, 2018, `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6492463/`.

[9] Niharika Gouda1 and Amudha J, *Skin cancer classification using resnet*, 2020, `https://www.imedpub.com/articles/skin-cancer-classification-using-resnet.php?aid=28628`.

[10] Alex Graves, *Generating sequences with recurrent neural networks*, 2013, https://arxiv.org/abs/1308.0850.

[11] Cahyo Adhi Hartanto and Adi Wibowo, *Development of mobile skin cancer detection using faster r-cnn and mobilenet v2 model*, 2020, https://ieeexplore.ieee.org/abstract/document/9239197.

[12] Boni E. Elewski J. Daniel Jensen, *The abcdef rule: Combining the "abcde rule" and the "ugly duckling sign" in an effort to improve patient self-screening examinations*, 2015, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4345927/.

[13] Nikolaus Kriegeskorte Johannes Mehrer, Courtney J. Spoerer and Tim C. Kietzmann, *Individual differences among deep neural network models*, 2020, https://doi.org/10.1038/s41467-020-19632-w.

[14] Thomas Brox Jost Tobias Springenberg, Alexey Dosovitskiy and Martin Riedmiller, *Striving for simplicity: The all convolutional net*, 2014, https://arxiv.org/abs/1412.6806.

[15] Bo Liu Qishen Ha and Fuxu Liu, *Identifying melanoma images using efficientnet ensemble: Winning solution to the siim-isic melanoma classification challenge*, 2020, https://arxiv.org/abs/2010.05351.

[16] Matilda Q. R. Pembury Smith and Graeme D. Ruxton, *Effective use of the mcnemar test*, 2020, https://link.springer.com/article/10.1007/s00265-020-02916-y.

[17] Jochen Sven Utikal Niels Grabe Dirk Schadendorf Joachim Klode Carola Berking Theresa Steeb Alexander H Enk Titus Josef Brinker, Achim Hekler and Christof von Kalle, *Skin cancer classification using convolutional neural networks: Systematic review*, 2018, https://www.jmir.org/2018/10/e11936/.

[18] Jae-Young Choi Ulzii-Orshikh Dorj, Keun-Kwang Lee and Malrey Lee, *The skin cancer classification using deep convolutional neural network*, 2018, https://link.springer.com/article/10.1007/s11042-018-5714-1.

[19] Jeremy Kawahara Yiqi YanEmail and Ghassan Hamarneh, *Melanoma recognition via visual attention*, 2019, https://www2.cs.sfu.ca/~hamarneh/ecopy/ipmi2019.pdf.

[20] Cicero Nogueira dos Santos Mo Yu-Bing Xiang Bowen Zhou Yoshua Bengio Zhouhan Lin, Minwei Feng, *A structured self-attentive sentence embedding*, 2017, https://arxiv.org/abs/1703.03130.