Chatbot Combined with Deep Convolutional Neural Network for Skin Cancer Detection

Qianfei Hu[1*†], Haochong Xia[2†] and Tianrui Zhang[3†]

[1]College of Letters and Science, University of California - Berkeley, Berkeley, California, U.S.A.
[2]College of Liberal Arts and Sciences, University of Illinois Urbana - Champaign, Urbana, Illinois, U.S.A.
[3]Institute for Interdisciplinary Information Science, Tsinghua University, Beijing, China
[*]Corresponding author: qianfei_hu@berkeley.edu, hxia5@Illinois.edu, zhangtr21@mails.tsinghua.edu.cn
[†]These authors contributed equally.

Abstract

For the sake of raising people's serious awareness of skin cancer and providing a convenient way to diagnose skin cancer, this study designs an application to help users figure out the true situation of their skin. In this field, many current studies still cannot achieve high accuracy for skin cancer detection. To increase the accuracy of the previous models from others, firstly, this study trains the proposed model by numbers converted by RGB using a 10k jpg file. Then, the deep convolutional neural network was applied with three units: a 2D convolutional unit to convolute the kernel with all the data in its receptive field, a pooling unit to decrease the dimension of the input, and a dense unit to generalize traits from the data. In addition, since it is noticed that the application that gives feedback on the diagnosis is missing, this paper set up a chatbot based on natural language processing and decides what functions shall the proposed bot provide. The developed application can diagnose seven kinds of skin symptoms related to skin cancer and give out some basic information and advice based on the result. Finally, the results indicated that the model achieves higher accuracy and makes the interaction with the user becomes true.

Keywords: *Convolutional Neural Network, Chatbot, Skin Cancer Detection, NLP*

1 INTRODUCTION

Nowadays, more and more people begin to enjoy sunbathing, which can be beneficial for the human body. Nevertheless, it also can cause skin cancer if people are exposed to the sun for a long time. According to the information provided by the skin cancer organization in 2022, due to the mutations induced by broken DNA, cells located at the surface layer of the skin become growing irregularly and may finally turn into malignant tumors that could be diagnosed as skin cancer [13]. In addition, since it is a little difficult to discover, many people easily ignore the importance of spotting it earlier. Besides, it is also inconvenient to get to the hospital frequently, especially in the context of the rapid spread of the COVID-19 Therefore, designing a diagnosis application to help everyone to figure out the true situation of one's skin is necessary. In general, the procedure for diagnosing skin cancer is called dermoscopy. A dermatologist can use special tools such as a microscope and magnifying lens to view suspicious areas in more detail. Biopsy and imaging, which requires professional knowledge and tools, are two ways that have been mostly used in diagnosing skin cancer. It's worth noting that the diagnosis of skin cancer usually begins with a visual inspection. If there is any suspicion, the doctor will first check the area for size, shape, color, texture, bleeding, or peeling. Doctors can also check surrounding lymph nodes to make sure they are enlarged.

Although computer science has been used to aid skin cancer diagnosis, there are two main flaws in existing algorithms. The first one lies in the choice of models. For example, in this study [9], the authors use the artificial neural network (ANN), fuzzy rule-based system, or adaptive fuzzy inference neural network (AFINN) as their classifiers. Note that AFINN would adjust weights using the Backpropagation (BP) algorithm. Similarly, in [3], the authors utilize BP in addition to ANN as their classifiers. However, these traditional neutral network technologies have their limitations. For ANN, it requires 2-dimensional image inputs to be converted into 1-dimensional vectors. This process not only leads to partial losses of spatial information but also results in an exponentially increasing number of trainable parameters

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 35-41, 2023.
https://doi.org/10.2991/978-94-6463-040-4_7

as the image size increases. If authors decide to employ BP to update the parameters' weights, they will potentially face a vanishing gradient problem (VGP) during the training time, which is common for multilayer neural networks using BP [11]. Regarding the fuzzy inference system, it requires prior knowledge from human experts to make decisions, while issues such as "knowledge representation, approximate reasoning, and reasoning under uncertainty" are present when adapting fuzzy logic to artificial intelligence [4].

The other issue with the existing algorithm relates to the outdated datasets used to train the models. In this study [18], the authors use 5 skin lesion datasets in the process of developing their detection system. Sets 3, 4, and 5 each consist of 1000 – 1500 skin lesion images, while set 1 and 2 each only contain around 100 images. Moreover, these datasets only have two types of skin classifications, benign lesions, and melanoma, without further details or other classes. In contrast, this modern dataset from ISIC 2018 [2] [16] has more than 10,000 images for training and another separate set for validation. The dataset only includes 7 disease categories, such as melanoma, melanocytic nevus, and basal cell carcinoma, to provide more details to the patient. When designing a medical application that emphasizes cancer detection, employing an updated and advanced dataset to improve classification accuracy is always desirable.

Concerning these problems in the existing algorithm and the dataset, it seems that a better structure is needed. And as convolutional neural network (CNN) behaved well in many aspects, like medical purposes--recognizing lung cancer from CT images [12] [17], etc, or industrial usage like automatic driving [1] [10], CNN was chosen in this study for further image processing.

The proposed model seems to overcome the mentioned studies by adopting a better network structure--the convolutional neural network, which can significantly decrease the number of parameters to be trained, since this structure uses convolutional layers which have fewer parameters than the dense layers, to get the trait of the input. Consequently, the exponential explosion of the number of parameters will be settled. For the VGP, this study applied the batch normalization method to avoid this as soon as possible. In this study [15], the author put forward this method as a way to avoid the VGP and reasoned the underlying principle behind the algorithm. Also, this study applied the method introduced in the thesis publisheb by Srivastava in 2014, called dropout to decrease the dependency between parameters and avoid overfitting from happening, which is a common problem for the deep neural network. In this study [14], the original author introduced the dropout method and explained in detail the reason why dropout can reduce the probability of overfitting. Also, for CNN, there's no prior need, all it needs is the input with the correct label, it can gain the trait of different labels by itself.

Also, to settle the problem of the too small and too concentrated dataset, a new dataset of skin cancer, named ISIC-2018, was used in this study, which consists of images and corresponding labels of different kinds of skin cancer, including melanoma, melanocytic nevus, etc., and some common non-cancer skin symptoms.

2 METHODS

2.1 Data preparation

Both the training dataset and the testing dataset of this paper were extracted from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets [2] [16]. There are 10, 015 sample points in the training dataset, and each image consists of 2, 352 RGB pixels used for its feature space. Every sample point is classified as one of the following seven categories of skin symptoms: benign keratosis-like lesions (bkl), melanocytic nevi (nv), dermatofibroma (df), melanoma (mel), pyogenic granulomas and hemorrhage (vasc), basal cell carcinoma (bcc), and actinic keratoses and intraepithelial carcinoma (akiec). The number distribution of these seven types in the training dataset is displayed in Figure 1. In addition, some of the sample images are shown in Figure 2.

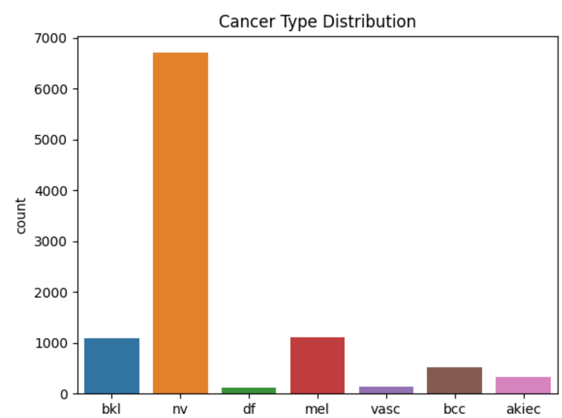
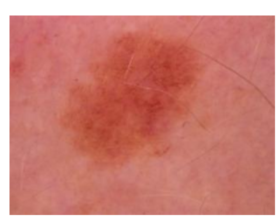


Figure 1. Distribution of Sample Points in the Training Set



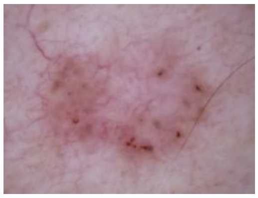


Figure 2. Some Images from Our Dataset

Throughout the training process, 20% of the sample points were used for validation. The testing dataset consists of 200 separate images, which is approximately 2% of the training dataset's size.

Batch normalization was employed to better support the training process. As normal distribution with mean 0 and variation 1 indicates a smooth distribution of random noise without shift, it is desirable that the numerical input of each image could follow the Normal(0, 1) distribution. However, with the deep neural network, due to the application of nonlinear functions, such as Sigmoid and ReLU, as the activation function, the distribution of the data would deviate further from the Normal(0, 1) distribution as the model becomes deeper with more hidden layers. Therefore, BatchNormalizatin was utilized to counter this trend. The BatchNormalization algorithm would normalize each feature selected in the mini-batch and then output the scaled and shifted version of each normalized feature.

2.2 Convolutional neural networks

The structure for this study is the CNN model, which is a widely used model for image processing. As it says in [8], this model is composed of 3 kinds of units, which are named as "convolutional unit", "pooling unit", and the usual fully connected layer--"dense unit". The structure of the model starts with 1 layer of convolutional units, then a layer of pooling units. This structure occurs in the whole model repeatedly until the model is ended by a layer of fully connected units.

For the convolutional unit, as in this study, the input is a 3D vector, since it is composed of the coordinate of the pixel and the RGB information about the pixel, the 2D convolutional unit is applied. The input is a tensor of shape (number of inputs) × (input height) × (input width) × (input channels). For the convolutional layer, it has a convolutional kernel, and it will convolute with the data in its receptive field.

$$\mathrm{O}_{\mathrm{X,Y}} = \mathrm{F}_{\mathrm{X,Y}} \cdot \mathrm{I}_{\mathrm{X,Y}} \quad (1)$$

Where O denotes the output of the convolution, I denote the input of the convolution, and F denotes the kernel; X, Y is the coordinate of the corresponding data.

And after convoluting the input, the output of the convolutional layer will be the input of the pooling layer. There exist several pooling strategies. In this study, max pooling is applied. In the max-pooling progress, the pooling layer chooses the maximal element in a block, which is produced by the kernel in a single step, to represent this block. And in this progress, we decrease the dimension of the input. Consider a kernel (also named a filter) of size 2×2, the max-pooling strategy gives out that:(S is a block; X, Y is the coordinate of the block)

$$f_{X,Y}(S) = max_{x,y\in\{0,1\}}(S_{2X+a,2Y+b}) \quad (2)$$

And in the final stage of this model, there come the fully connected units. In this layer, activation functions are applied to generalize traits from the data which already has been processed by the convolutional layer and pooling layer.

Figure 3 shows the structure of the neural network in this study, where the input is of shape (600, 450, 3), indicating that the input picture is consisting of 600 x 450 pixels and each pixel has a 3-tuple to denote its color. For the choice of the size of the kernel, we apply the suggested size in the article [5].

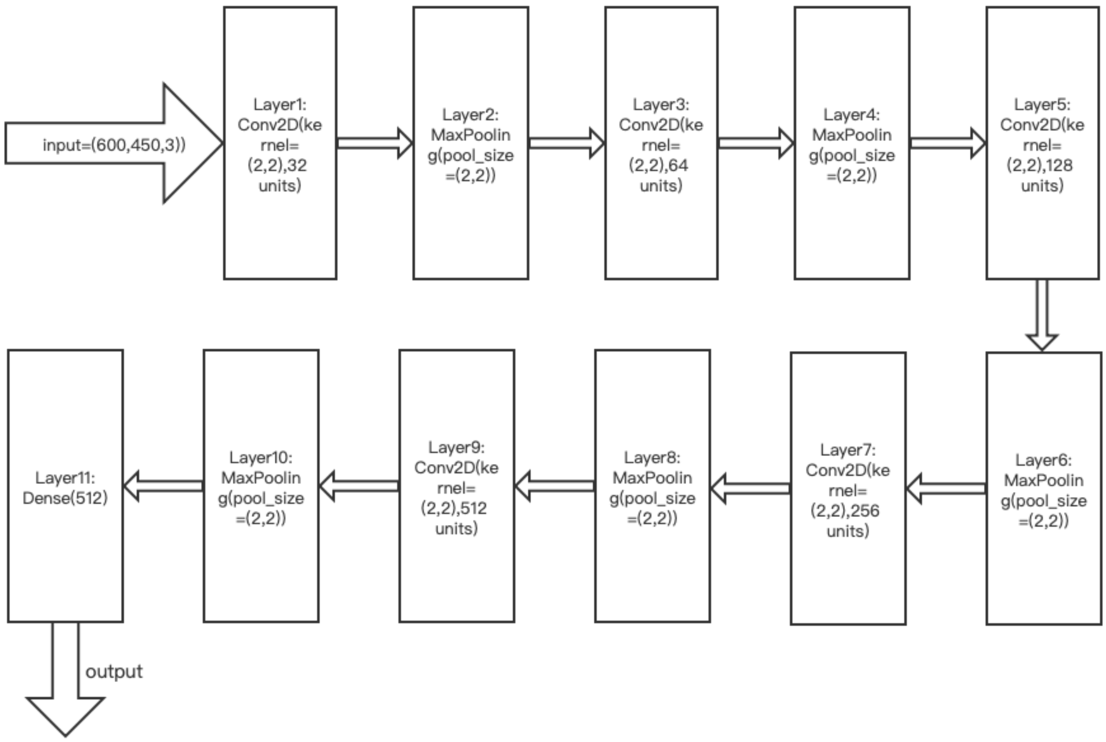


Figure 3. Model Training Procedure

2.3 Implementation details

To better utilize both the first and the second moments of the gradient, this paper's model used the Adam optimizer. The Adam optimizer uses stochastic optimization and a constant learning rate. Moreover, the suggested values of $\alpha = 0.001$, $\beta 1 = 0.9$, and $\beta 2 = 0.999$ by Kingma and Ba were used to achieve a faster convergence [6].

Throughout the training process, the sparse categorical cross-entropy function was used as the loss function. This cross-entropy loss function is a widely used loss function for classification. It comes from the binary cross-entropy, where the lost L is calculated as in equation (3).

$$L = -\sum_x [yln(p) + (1 - y)\ln(1 - p)] \quad (3)$$

In this formula, x is a sample point in the dataset, y is the label value, and p is the predicted probability of sample point x in class y. To expand this function to cover multi-class classification, the lost L is now calculated as in equation (4).

$$L = -\sum_c \sum_{\{x:\, y=c\}} yln(p) \quad (4)$$

In this updated formula, c represents all classes in the dataset, and p is the probability of the corresponding sample point in class c, while the meanings of the rest notations remain the same.

In addition, the training process set the epoch number to be 40 and the batch size of each epoch to be 256.

2.4 Chatbot

Our chatbot (Doctor G4) provides a basic chatting function and can respond to the diagnosis based on the image uploaded by the user. We first use some methods to clean up the data. In the beginning part, we use the "sent_tokenize" and "word_tokenize" functions from the nltk package to tokenize the words and sentences. After that, we clean up the tokens by Lemmatizing words by their root form, removing the stopwords the most frequent words used in Natural Language that are not useful for the text classifier, and removing the punctuations. Then, we define a method called "LemNormalize" to use the "LemTokens" function and the "remove_punct_dict" dictionary that we defined for lemmatizing to clean and tokenize the text. After all these preparations we can proceed to the actual processing. We write a function called "chat" to packet all chatting functions. It can call "greeting" to reply with greeting words politely including "hello, hi, greetings, sup, what's up, hey, ai, next".

Our "is_path" function will find the image through the path provided by the user, If the image cannot be found, Doctor G4 will reply with a message "Sorry, cannot find the picture through your input path. Please try again. "If the image is found, Doctor G4 can give out 7 kinds of cancers, including melanocytic nevi, melanoma, benign keratosis-like lesions, basal cell carcinoma, pyogenic granulomas, and hemorrhage, Actinic keratoses and intraepithelial carcinoma, a dermatofibroma. Besides, it will give out basic information and advice regarding what kind of skin cancer is discovered as well. In the end, it will also reply with some polite words with the "response" function and end the chat. A diagram of our chatbot is shown in Figure 4.

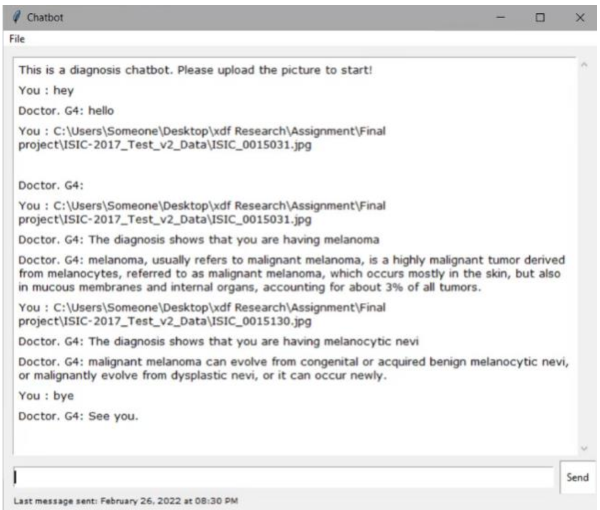


Figure 4. Medical Chatbot Demonstration

3 Result and discussion

The experiments showed that after 40 epochs, the model was able to achieve an accuracy rate of 96% on the training dataset and 93% on the validation dataset. The two graphs in Figure 5 describe the accuracy versus epoch number and the loss versus epoch number.

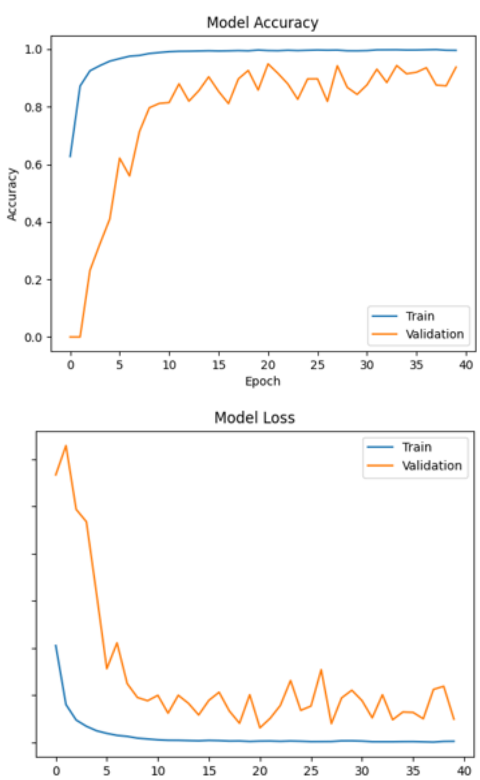


Figure 5. Accuracy vs. Epoch and Loss vs. Epoch

Also, the model achieved an accuracy rate of 82% on the testing dataset.

From the plots, we can see that a deep convolutional neural network performed relatively well in detecting skin cancer/disease. When training the model, we found that the model would suffer from overfitting if the epoch number was set to 60 or higher. Consequently, we decided to choose 40 as our final epoch number to avoid excessive overfitting while maintaining the accuracy of our model. In future research, one could employ ResNet to increase the training epoch number while better handling the issue of overfitting. Meanwhile, it is important to point out that a carefully designed traditional neural network could still achieve high accuracy in skin disease diagnosis. For example, when detecting solely melanoma, the multilayer perceptron (MLP) neural network [7] was able to obtain accuracies in the range of 80% to 88% depending on the hyperparameters. However, these models would be less likely to perform as accurately as CNN when the number of categories increases.

Based on the results, improvements in the dataset could be done to further improve the model. As shown in figure 1, the types of skin cancer in the dataset were not evenly distributed. Most of the sample points in the training set were melanocytic nevi (nv), while only a few were dermatofibroma (df) or pyogenic granulomas and hemorrhage (vasc). This unequal distribution could lead to underfitting for underrepresented classes in the dataset and undermine the final accuracy of the model. Therefore, by collecting more data from underrepresented classes, one could ameliorate the prediction of these classes as well.

4 CONCLUSIONS

The research implemented an application to diagnose various kinds of skin cancers and give simple suggestions on it. This helps the user to discover skin cancers at an early stage and have a better chance of recovering. For the final result of the CNN model, the model receives an accuracy rate of 96% on the training set, 93% on the validation set, and 82% on the test set. For the NLP, the chatbot can handle basic conversations, including simple greetings, asking for the input image, and giving basic suggestions based on the result of the prediction. A better data pre-processing may improve the result of the research. Thus, more pre-processing methods for images could be needed, including segmentation or boundary extraction. Also, the overfitting phenomenon is a problem that affects our results greatly. Merely applying the dropout method or adjusting the epoch size cannot solve this problem completely. Thus, more in-depth research is needed. Also, the NLP chatbot is not intelligent enough, and one could build another neural network for a more intelligent NLP bot.

REFERENCES

[1] Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., ... & Zieba, K. (2016). End to end learning for self-driving cars. arXiv preprint arXiv:1604.07316.

[2] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., ... & Halpern, A. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368.

[3] Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin Cancer Detection: A Review Using Deep Learning Techniques. International journal of environmental research and public health, 18(10), 5479. https://doi.org/10.3390/ijerph18105479

[4] Guillaume, S., & Charnomordic, B. (2012). Fuzzy inference systems: An integrated modeling environment for collaboration between expert knowledge and data using FisPro. Expert Systems with Applications, 39(10), 8744-8755.

[5] Hasan, M., Barman, S. D., Islam, S., & Reza, A. W. (2019, April). Skin cancer detection using convolutional neural network. In Proceedings of the 2019 5th international conference on computing and artificial intelligence (pp. 254-258).

[6] Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, 448–456. Lille, France: JMLR.org.

[7] Kingma, D.P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.

[8] Lecun, Y. , & Bottou, L. . (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[9] Mehta, Palak & Shah, Prof. (2016). Review on Techniques and Steps of Computer Aided Skin Cancer Diagnosis. Procedia Computer Science. 85. 309-316. 10.1016/j.procs.2016.05.238.

[10] Qiu, Y., Yang, Y., Lin, Z., Chen, P., Luo, Y., & Huang, W. (2020). Improved denoising autoencoder for maritime image denoising and semantic segmentation of USV. China Communications, 17(3), 46-57.

[11] Roodschild, M., Sardiñas, J.G., & Will, A. (2020). A new approach for the vanishing gradient problem on sigmoid activation. Prog. Artif. Intell., 9, 351-360.

[12] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging, 35(5), 1285-1298.

[13] Skin cancer information. The Skin Cancer Foundation. (2022, January 20). Retrieved March 4, 2022, from https://www.skincancer.org/skin-cancer-information/

[14] 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.

[15] Toussaint, M (2009) Robot trajectory optimization using approximate inference. In: Danyluk, AP, Bottou, L, Littman, ML (eds.) Proceedings of the 26th Annual International Conference on Machine Learning (ICML 2009), Montreal, Quebec, Canada, 14–18 June 2009 (ACM International Conference Proceeding Series, Vol. 382). New York: ACM Press, pp. 1049–1056.

[16] Tschandl, P., Rosendahl, C. & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi:10.1038/sdata.2018.161.

[17] Xu, S., Lu, H., Ye, M., Yan, K., & Jin, Q. (2020). Modified Hybrid Task Cascade for Lung Nodules Segmentation in CT Images with Guided Anchoring. ICMLC 2020: 2020 12th International Conference on Machine Learning and Computing.

[18] Yuan, X., Situ, N., Yuan, X., & Zouridakis, G. (2013). Computer-Aided Cancer Detection and Diagnosis: Recent Advances - Automated Melanoma Screening and Early Detection. doi: 10.1117/3.1002311.ch7