Skin Cancer Detection Based on Hybrid Model by Means of Inception V3 and ResNet 50

Yuyu Zeng[1], †, Xingsheng Zhu[2], *, †

[1]School of Science, Harbin University of Science and Technology, Harbin, China
[2]Software College, Northeastern University, Shenyang, China
**Corresponding author. Email: 20195650@stu.neu.edu.cn*
†These authors contributed equally.

Abstract

Due to the rapid growth of skin cancer patients, being diagnosed and treated at an early stage has become more and more necessary. However, only mature, and experienced doctors are capable for the precise detection of skin cancer by carrying out some expensive examination. To solve this issue, several computer-based, deep-learning detection methods have been proposed over recent years. However, most of current methods only focus on algorithms and do not provide with industrial applications. In addition, the current accuracy based on these algorithms can be further improved. This paper presents an improved deep Convolutional Neural Network (CNN) model of feature fusion by merging InceptionV3 and ResNet50, to classify images of skin cancer into eight types. By concatenating, the proposed model has not only the Inception Modules, but also the Residual Blocks, which obtains the advantages of reducing the number of parameters and mitigating the gradient problem. The experiment has been carried out on the augmented HAM10000 dataset, which contains quite several images of different types of skin cancer. The experimental results show that the proposed model has the best training performance with a validation accuracy at 87.11% among InceptionV3, ResNet50, VGG19 and itself. Besides, during prediction, the proposed model also holds the best achievement with accuracy at 0.822, precision at 0.824, recall at 0.822, and F1-score at 0.817 among the four models, which has a rivalrous performance than VGG19. Furthermore, an iOS App was also developed to provide a better interactive experience for users to acquire a diagnosis through one single image of skin lesion and communicate about their results conveniently.

Keywords: *Skin Cancer Classification, Deep Convolution Neural Network, InceptionV3, ResNet50, Feature Fusion, NLP, iOS App*

1 INTRODUCTION

Skin cancer is the most common cancerous carcinoma among the Caucasians in the United States [11]. One-sixth of American citizens developed skin cancer on some occasions. There are seven main types of skin cancer, e.g. basal cell carcinoma (bcc), melanocytic nevi (nv). Around 75% of all deaths of skin cancer infect with malignant melanoma [7] [11]. Fortunately, nearly each type of skin cancer can be cured if diagnosed and treated at an early stage. Thus, early detection of skin cancer has become a crucial problem for the patients. Biopsy methods and dermocopy analysis are the main medical ways in skin cancer detection. Biopsy is a sample of tissue taken from the body to examine it more clearly. Evidently, the biopsy methods are not only painful but take a long time to get the result. Dermocopy is a non-invasive diagnostic technique for the invivo observation of pigmented skin lesions (PSLs) [4]. However, it has analytical limitation in the diagnosis of skin cancer in a very early stage which lacks main features [14]. Therefore, with the significant progress of artificial intelligence in multiple fields, sorts of computer-based techniques are considered for solving this medical problem, which can provide a relatively accurate and objective detection of skin cancer.

Throughout two decades of relevant research, it is found that machine learning (ML) and deep learning (DL) are widely used to complete the task of skin cancer detection in a form of image classification [2] [3] [11]. The previous researchers provide a great number of images and feed them through the structure of ML or DL algorithms to get expected results of classification. In an early study, using Super Vector Machine (SVM) and k-

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 275-282, 2023.
https://doi.org/10.2991/978-94-6463-040-4_42

Nearest Neighbour (k-NN) classifiers for the classification of skin cancer acquires an accurate diagnosis rate of 61% on certain dataset [12]. In another study, Convolutional Neural Network (CNN) is used for the classification of four different skin lesions and an accurate diagnosis rate of 77% is obtained [2]. More recently, Li et al. [9] builds a Soft-Attention Model with InceptionResnetV2 for the skin cancer classification task, acquiring an accuracy of 93%, being better than some previous methods in literatures. As time goes by, deep CNNs are more and more usually applied to medical field to detect skin cancer because of their advantages of excelling in image processing, which have already achieved notable results [3]. However, most of current methods neglect to test whether a skin lesion is skin cancer or not while detecting the types of it. In addition, many algorithms in current research ignore the possibility of merging some deep CNNs to achieve a better performance. Some studies only focus on algorithms and do not provide with industrial applications for diagnosis feedback and convenient communication.

InceptionV3 (Szegedy 2015) and ResNet50 (He 2016) are two typical deep CNNs for image classification, each of which has its unique advantages and are proved in many studies [10] [18]. In this paper, by merging InceptionV3 and ResNet50, an improved deep CNN model of feature fusion was proposed to classify images of skin cancer, whose performance on the augmented HAM10000 dataset is better than each of the two models (i.e. InceptionV3, Resnet50) after the same number of epochs. It is known that the higher the accuracy, the more difficult it is to improve. Even so, the proposed model has a high validation accuracy that achieves 87.11%. This study also compares the proposed model with some other typical deep CNN models, like VGG19 [12] [13], to discover whether the proposed model can perform better than them. Then, pre-trained model was used to implement an iOS App embedded with a chatbot for users to detect skin cancer by a picture and gain more information of their results through communication.

2 METHOD

2.1 Dataset and Pre-processing

2.1.1 Dataset in the study

Training process of most Convolutional Neural Networks for automatic diagnosis of pigmented skin lesions (PSLs) is circumscribed by the lack of diversity and sufficiency of available dataset of dermatoscopic images. In our work, we used HAM10000 (Human Against Machine with 10,000 images) dataset [17], which is provided by Harvard University. The dataset contains 10,015 RGB dermatoscopic images with a resolution of 600×450 pixels, including 7 types of skin cancer, which are respectively:

- Actinic keratoses and intraepithelial carci-noma / Bowen’s disease (akiec);
- basal cell carcinoma (bcc);
- benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl)
- melanoma (mel)
- dermatofibroma (df)
- melanocytic nevi (nv)
- vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc)

Moreover, a further dataset of healthy skin with 105 images, was also added to the final dataset, named as ‘nod’, which stands for no disease. Thus, our final integrated dataset consists of 8 classes, each of whose samples are displayed in Figure 1.

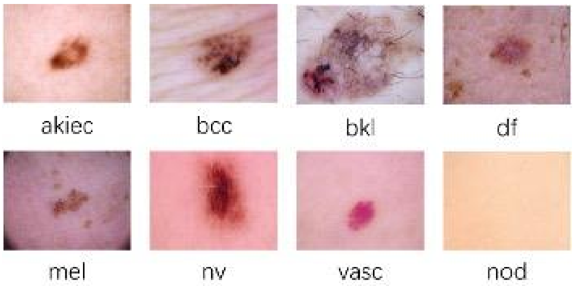


Figure 1: Dataset examples.

2.1.2 Pre-processing

The integrated dataset was separated into three sets with a split ratio of 6:2:2, a training set, a validation set, and a testing set. After the division, the training set has 6,072 images, the validation set and the testing set each have 2024 images. The training set was then augmented by cropping, flipping and rotating specific images. What’s more, the augmentation needs to meet a new proportional criterion, which is, the ratio of images in the training, validation and testing sets should turn to 86:7:7. In order to get better convergence in back-propagation and reduce the relative loss, we transformed the images to the required format. All images in the three sets were normalized and resized to 224×224 pixels through the ‘ImageDataGenerator’ from Keras [8].

2.2 Proposed Model

2.2.1 Background

A CNN is a special kind of neural network that significantly reduces the number of parameters in a deep neural network with many units without losing too much in the quality of the model. Its two indispensable and critical characteristics are convolution and pooling. With convolutional layer, local features can be fully extracted, which mitigates overfitting. With pooling layer, parameters are greatly reduced, same to the computation volume. Pooling also brings translation invariance to the neural network.

Up to now, many deep CNNs have been proposed and found applications in image classification where they beat many previously established benchmarks. InceptionV3 and ResNet50 are two typical ones among them.

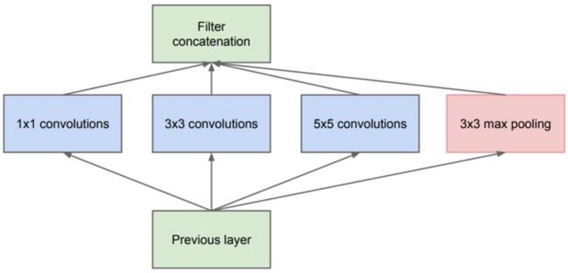


Figure 2: Structure of InceptionV3.

InceptionV3 was proposed by Google in 2015, which is composed of multi-overlaid improved Inception Module. The basic structure of Inception Module has four components: 1×1 convolution, 3×3 convolution, 5×5 convolution, and 3×3 max pooling, as shown in Figure 2 [16]. Finally, the results of the four component operations are combined on the channel. This is the core idea of Inception Module. By extracting information from different scales of the image through multiple convolution kernels and finally fusing them, a better representation of the image can be obtained. InceptionV3 also contains the idea of splitting a multi-dimensional convolution into two smaller convolutions, whose advantage is reducing the number of parameters. By this asymmetric convolutional splitting, it is better than symmetric splitting into several identical convolutions and can handle more, richer spatial features.

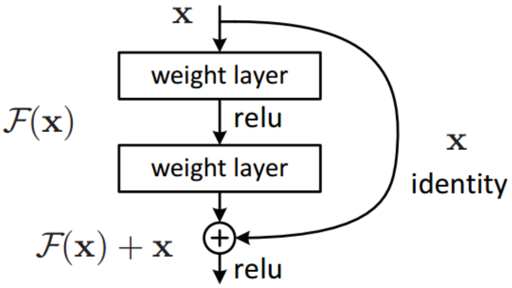

x
weight layer
relu
weight layer
$\mathcal{F}(\mathbf{x})$
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 3: Structure of ResNet50.

ResNet50, stands for Residual Networks 50, contains 49 convolutional layers and one fully connected layer. The method of enhancing the learning ability of the network by increasing the number of layers of the network may lead to the problem of gradient disappearance, which leads to a decrease in the accuracy of the network. The key of ResNet lies in the residual blocks in its structure that can solve the gradient problem, and the increase in the number of layers of the network also makes it express better features and correspondingly better performance in detection or classification. As shown in Figure 3 [6], the residual block contains cross-layer connections, which allow the input to be passed directly across layers, mapped equally, and later added to the result of the convolution operation.

Due to the unique advantages of both InceptionV3 and ResNet50, we hope to build an ensemble model that keeps the two deep CNNs' advantages, to take the best of both and get a better performance in image classification.

2.2.2 Ensemble Model Based on InceptionV3 and ResNet50

With the background knowledge described above, we then make feature fusion for InceptionV3 and ResNet50. Firstly, let a pooling layer by global average, a dropout layer, and a fully connected layer with 512 units follow behind the two models. Secondly, ensemble the two former part by concatenating. Thirdly, link a dropout layer and a fully connected layer with 512 units behind, and then come the output layer with 8 units, which represents the 8 types of diagnostic results of skin cancer. This far, we obtain our ensemble model, as shown in Figure 4.

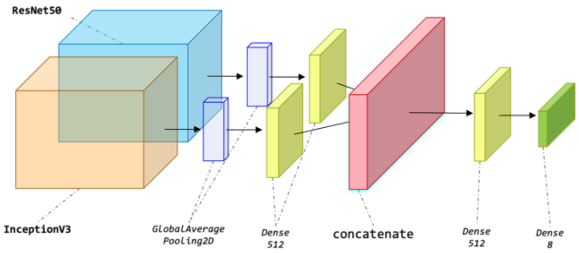


Figure 4: Structure of the ensemble model.

2.3 Implementation

2.3.1 Building

The proposed ensemble model was built by the framework of TensorFlow [1] in Python. In the model building step, we coded our model based on the structure defined above. We designed (224,224,3) as the input shape and placed 8 nodes in output layer corresponding to 8 types. Besides, a method called transfer learning was implemented as well. The motivation of transfer learning is that after transferring another image classification model to the custom model, the proposed model can be trained in a comparatively short time and performs well.

2.3.2 Training

In the model training process, training set and validation set were used to train the model. An Adam optimizer and a learning rate of 0.0001 are used to train the Neural Network [5]. We used 'ReduceLROnPlateau' to realize a dynamic decrease in learning rate based on certain measurements during training [5].

$$CategoricalCrossEntropy = -\sum_{i=1}^{output\ size} y_i * log\, \hat{y}_i \quad (1)$$

Categorical cross entropy was used as loss function, which calculates the loss of an example by computing the equation (1), where $\hat{y}_i$ is the i-th scalar value in the output, $\hat{y}_i$ is the corresponding target value, and output size is the number of scalar values in the output.

It is worth mentioning that during the training process, we made use of the 'class_weight' parameter to adjust the loss function. This parameter allows the loss function to pay more attention to data with insufficient sample size when dealing with unbalanced training data.

Besides, we set epoch to 50 and used 'Modelcheckpoint' to save weights after each epoch [5].

2.3.3 Evaluating and Predicting

In the evaluating part, we took accuracy, loss function into consideration, to check the performance of our proposed model by drawing line graphs. The accuracy is defined as equation (2), where TP represents true positive, TN represents true negative, FP represents false positive, FN represents false negative.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

In addition, a confusion matrix was used to summarize the performance of our classification model. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class.

In the predicting part, by feeding the test set to our pre-trained model, calculate recall, precision and F1-score to reveal our model's performance, whose equations are as follows as equation (3-5):

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$F1 = \frac{2 \times pre \times rec}{pre+rec} \quad (5)$$

2.4 Application

To provide a platform for diagnosis feedback and convenient communication, we developed an iOS app embedded with a chatbot for users to detect skin cancer by a picture and gain more information of their diagnosis results through communication. This industrial application has two main parts: chatbot by NLP and user interface.

2.4.1 Chatbot Based on NLP

We implemented a simple NLP chatbot with Natural Language Toolkit (NLTK) for diagnosis communication. This chatbot can answer limited questions from users. There are two steps of it: question classification, and response generation.

In the step of question classification, we divided user's questions into two categories: generic and medical. We used Naive Bayes Classifier to separate the questions into two types, generic questions and medical questions, and then saved the classifier for chatbot to classify. In this step, we gave some generic and medical questions as a training set.

In the step of response generation, we defined two functions to generate medical and generic responses for corresponding questions. For medical responses, we firstly traversed medical question-answer pairs in a dictionary found online [15]. Then, select the corresponding medical response based on the overlap between question and answer in the dictionary, in order, first 100%, then 80%, last 70%. For generic responses, we defined a greeting function to answer greeting inputs like 'Hello' and 'Hey'. Then, we considered the input question's content. If the input is in the 'GREETING INPUTS' set, choosing a random answer in 'GREETING RESPONSES' set. Or traverse generic question-answer pairs in dictionaries found online, select corresponding responses based on the overlap between question and answer in the dictionaries in order, just like the way we resolved medical responses [15]. After all this, we need to consider the cases where none of the above is satisfied, specifically.

2.4.2 User Interface on iOS

As for UI, we chose to go with a rather trendy approach, which is iOS app. We used Xcode to develop front interface in SwiftUI and Python Flask framework to construct our back end. Front-end and back-end interaction is implemented via HTTP requests of 'GET' and 'POST'. We built three pages, which are homepage, image upload page and chat page, respectively has their own functions.

3 RESULT AND DISCUSSION

3.1 *Classification Performances*

To compare the performance of proposed model with InceptionV3, ResNet50 and VGG19, we carried out four sets of experiments with a same epoch at 50 and batch size at 32. As shown in Table 1, Figure 5, and Figure 6, it is shown that in this classification task, our proposed model has the best training performance with a validation accuracy at 87.11% among InceptionV3, ResNet50, VGG19 and itself. In prediction part, our proposed model holds the best achievement with accuracy at 0.822, precision at 0.824, recall at 0.822, and F1-score at 0.817 among the four models, as shown in Table 2. Through these evaluation metrics, we can easily tell that the higher the accuracy before assembling, the more difficult it is to improve. Therefore, a 1% to 2% improvement of the accuracy of our proposed model from pre-synthesis models is relatively a big promotion.

Table 1: Training performance of different models.

Models	Performance			
	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
Proposed Model (epoch=50, batchsize=32)	98.41%	0.0383	87.11%	0.5178
InceptionV3 (epoch=50, batchsize=32)	98.24%	0.0443	86.33%	0.5226
ResNet50 (epoch=50, batchsize=32)	98.17%	0.0430	85.55%	0.5396
VGG19 (epoch=50, batchsize=32)	89.91%	0.2761	83.59%	0.4438

Table 2: Testing performance of different models.

Models	Performance			
	Accuracy	Precision	Recall	F1-score
Proposed Model	0.8224	0.8241	0.8224	0.8168
InceptionV3	0.8051	0.8184	0.8051	0.8014
ResNet50	0.8115	0.8061	0.8115	0.7924
VGG19	0.7783	0.7800	0.7783	0.7373

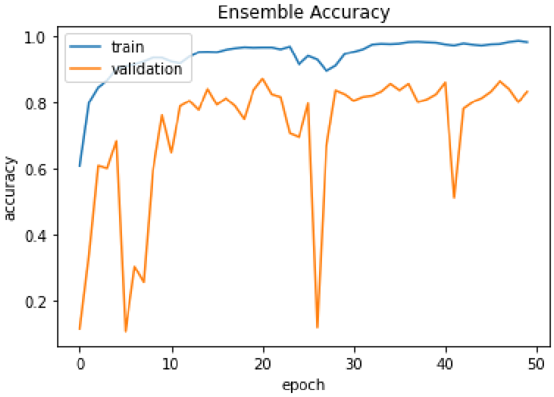


Figure 5: Accuracy of the proposed model.

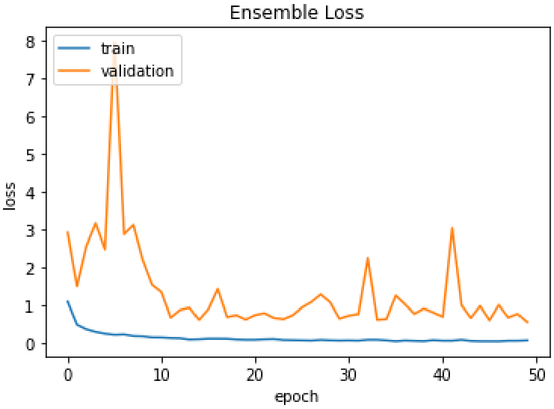


Figure 6: Loss of the proposed model.

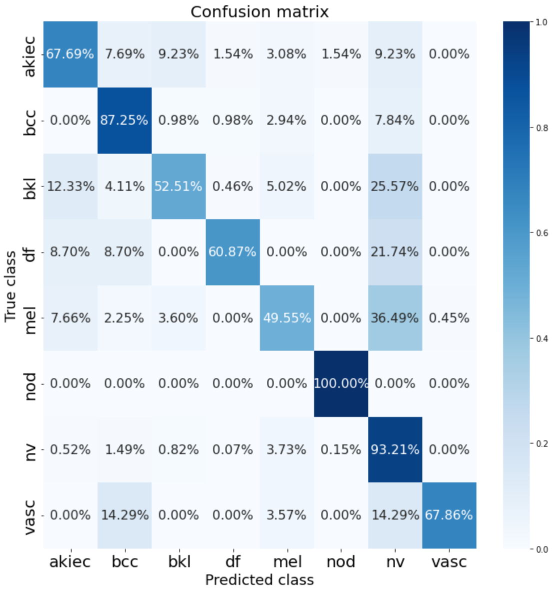


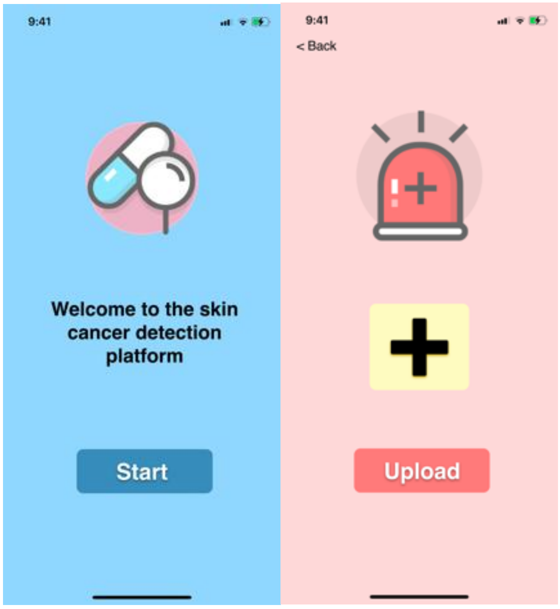
Figure 7: Confusion matrix of proposed model.

According to our proposed model's confusion matrix of prediction in Figure 7, we can find that the prediction of the eight types of skin cancer is relatively accurate. However, it's easily to be found that almost all eight types have a high probability of being misclassified as 'nv', especially for the 'mel' type, whose prediction accuracy only achieves at 49.55%. After analysis we found that in our validation set and test set, the number of images of type 'nv' is several times higher than other types, which causes imbalance. Beyond that, the features extracted by the proposed model maybe too similar between the 'mel' and 'nv' types, which causes the lowest prediction accuracy among the eight types. Therefore, it may need a more balanced dataset and a more detailed data processing to gain a better classification performance of our proposed model.

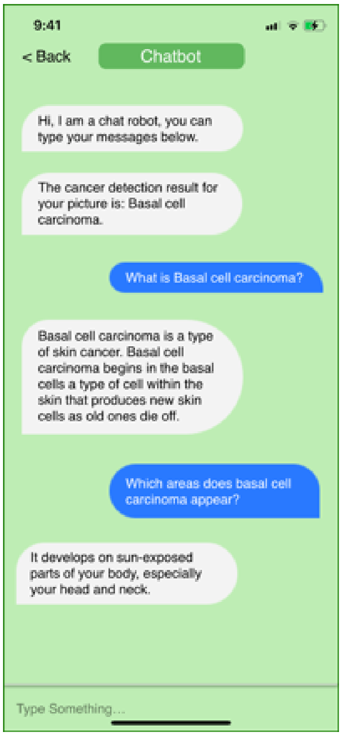
3.2 *App Interaction*

We have developed iOS app embedded with a chatbot for users to detect skin cancer by a picture and gain more information of their diagnosis results through communication. We build front-end in Swift UI and back-end in Python Flask. Operation steps of this app are shown in Figure 8.

As shown in Figure 8(a), the home page of our application has a start button. Click the start button to go to the next page, which is upload page, shown in Figure 8(b). Then click the plus symbol to choose a picture existing in phone or take a photo immediately and click upload button to post the image to our pre-trained model. After processing, we can get a classification result from pre-trained model, then come to the chat page, shown in Figure 8(c). In this page, we can ask some questions about the disease, and we can get corresponding results.



(a) (b)



Hi, I am a chat robot, you can type your messages below.

The cancer detection result for your picture is: Basal cell carcinoma.

What is Basal cell carcinoma?

Basal cell carcinoma is a type of skin cancer. Basal cell carcinoma begins in the basal cells a type of cell within the skin that produces new skin cells as old ones die off.

Which areas does basal cell carcinoma appear?

It develops on sun-exposed parts of your body, especially your head and neck.

(c)

Figure 8: UI of the designed application.

4 CONCLUSION

In this paper, an ensemble model was proposed by merging InceptionV3 and ResNet50, this model was trained and tested by a multi-class dataset called HAM10000 dataset released by Harvard, to classify types of skin cancer from images, and applied it into an industrial iOS application providing diagnosis results and convenient communication for users. Several experiments were designed to evaluate the proposed model. The result shows that the proposed model has a high validation accuracy that achieves 87.11%, which is commensurate or even has a better performance than InceptionV3, ResNet50 and VGG 19. In the future, further study plans to improve the structure of the model in terms of feature fusion by considering a more efficient way of integrating Inception Module and Residual Block rationally and resolve the problem of data imbalance by using Generative Adversarial Networks (GAN) to generate more images for training. In addition, more attempts will be made to apply this model into more practical applications.

REFERENCES

[1] Abadi M, Barham P, Chen J, et al. (2016). {TensorFlow}: A System for {Large-Scale} Machine Learning[C]//12th USENIX symposium on operating systems design and implementation (OSDI 16). 265-283.

[2] Albahar, M. A. (2019). Skin Lesion Classification Using Convolutional Neural Network With Novel Regularizer. IEEE Access 7: 38306-38313.

[3] Ali, M., et al. (2021). An Enhanced Technique of Skin Cancer Classification Using Deep Convolutional Neural Network with Transfer Learning Models. Machine Learning with Applications, vol. 5, p. 100036.

[4] Baldi A, Quartulli M, Murace R, et al. (2010). Automated dermoscopy image analysis of pigmented skin lesions. Cancers (Basel). 2(2):262-273. Published 2010 Mar 26. doi:10.3390/cancers2020262

[5] Gulli A, Pal S. (2017). Deep learning with Keras. Packt Publishing Ltd.

[6] He K, Zhang X, Ren S, et al. (2016). Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 770-778.

[7] Jerant A F, Johnson J T, Sheridan C D, et al. (2000). Early detection and treatment of skin cancer. American family physician, 62(2): 357-368.

[8] Kingma D P, Ba J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv: 1412.6980.

[9] Li Y, Wang D, Xu Z, et al. (2021). Intelligent Skin Cancer Detection System Based on Convolutional Neural Networks[C]//Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences. 188-198.

[10] Qiu, Y., Chang, C. S., Yan, J. L., Ko, L., & Chang, T. S. (2019). Semantic segmentation of intracranial hemorrhages in head CT scans. In 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) (pp. 112-115). IEEE.

[11] Scotto J, Fears T R, Fraumeni J F. (1983). Incidence of nonmelanoma skin cancer in the United States.

[12] Sheha M A, Mabrouk M S, Sharawy A. (2012). Automatic detection of melanoma skin cancer using texture analysis. International Journal of Computer Applications, 42(20): 22-26.

[13] Simonyan K, Zisserman A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[14] Skvara H, Teban L, Fiebiger M, Binder M, Kittler H. (2005). Limitations of Dermoscopy in the Recognition of Melanoma. Arch Dermatol. 141(2): 155-160. doi:10.1001/ archderm.141.2.155.

[15] Suraksha RV. (2021). "Suraksharv/MedBot: Chatbot for Skin Cancer Diagnosis." GitHub, https://github.com/SurakshaRV/MedBot.

[16] Szegedy C, Liu W, Jia Y, et al. (2015). Going deeper with convolutions//Proceedings of the IEEE conference on computer vision and pattern recognition, 1-9.

[17] Tschandl, P, et al. (2018). "The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions." Scientific Data, vol. 5, no. 1, 2018, https://doi.org/10.1038/sdata, 161.

[18] Wen, L., Li, X., & Gao, L. (2020). A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Computing and Applications, 32(10), 6111-6124.