Identification of Nail Lesions Based on Mask R-CNN Deep Learning Model

[1]Yu-Chieh Liu (劉宇捷), [1]Yong-Long Lin (林永隆), [1]Cheng-Ying Chou (周呈霙), [2]Chiou-Shann Fuh (傅楸善)

[1]Department of Biomechatronics Engineering,

National Taiwan University, Taipei, Taiwan,
steven880830@gmail.com queuzy68@gmail.com

[2]Department of Computer Science and Information Engineering,

National Taiwan University, Taipei, Taiwan,
fuh@csie.ntu.edu.tw

ABSTRACT

Nails are specialized skin tissue. It can reflect some health conditions by wrinkles and texture on the surface. The physicians mostly diagnosed through observing first, and confirmed with biopsy examination, which was time-consuming in clinical practice. Therefore, an instance segmentation algorithm Mask R-CNN deep learning model was utilized in this research to segment and classify 17 kinds of potential lesions within the nail images. In the results, the Mask R-CNN model could reach a precision rate of 87.69% in classification. For further comparison with the state-of-the-art DenseNet-201, whose precision rate was 84.37%, the Mask R-CNN model could outperform CNN-based algorithms. Hopefully, with the strength of the Mask R-CNN algorithm, it will access great diagnosis of nail diseases in clinical practice.

Keywords: Nail disease identification, instance segmentation, Mask R-CNN, classification.

1. INTRODUCTION

The composition of human nails is metabolized keratin, which continuously accumulates on the surface of skin. Since the diversity of age, metabolism and season, etc., there are different appearances of nails reflecting the body functionality, which can be taken as a sign of health. Normal nails have appearances that are brightly pink and without wrinkles and texture. To distinguish whether there is potential danger of health, we can examine if the surface of nails has vertical/horizontal lines, cracking or color difference, etc. However, it is difficult for people to self-detect the health condition of nails due to the lack of identification experience. Therefore, we decided to establish a deep learning algorithm Mask R-CNN for automatically detecting and pointing out the irregular region of nails. By the profound strengths of the deep learning model, we expected that it could provide effective nail health detection for people.

2. RELATED WORKS

In this section, two common identification methods in the other literature of nail lesions would be described, which were simple classification-based methods in Section 2.1 and segmentation-based methods in Section 2.2. Finally, a summary would be given in Section 2.3 based on the introduction of the above works and propose the sort of model that would be used in this research.

2.1. Classification of Nail Lesion

For the past research on imaging classification of nail diseases, there were 2 types that could be summarized roughly. First, there were researches which proposed the guidelines for identifying nail diseases. Mostly they were for physicians to identify lesions by observation through themselves or by nail dermoscopy in clinical practice [1-4]. However, it required the physicians who had been trained professionally and well-experienced with multiple patients. Otherwise, these types of guidelines were highly dependent on the visual features which tended to be defined by subjective observation [5-8]. In this way, it was not suitable for cases in the real world, because of the uncertainty under different sampling circumstances, i.e. lighting, imaging device format, or image resolution, etc. As a consequence, there existed noise in the sampling images commonly, which substantially affected the quality of analysis results, and it was not beneficial to establish golden standards in nail melanoma detection [9].

Therefore, there were other types of research dedicated to various algorithms that were based on computer vision techniques. In the past, it was common to use traditional machine learning methods for medical image identification such as Maniyan [10] et al., they used multiclass SVM (support vector machine) to assist in the early diagnosis of nail diseases on hand nail images. The model jointly predicted the classification of 23 nail diseases instead of predicting multiple classes separately, and achieved an average accuracy of about 90%. Nevertheless, since the rise and remarkable growth of deep learning in various fields, some scholars had gradually utilized deep learning methods in the auxiliary diagnosis of medical images. Winkler [11] et al. proposed to use convolutional neural network (CNN) to classify different types of nail melanoma into some categories. Their contribution was that their research used 6 dermoscopic image datasets. Hence, CNN could be more robust in the classification of nail lesions. Their CNN showed high-level performance in set-SSM, set-NM, and set-LMM dataset with almost 93% sensitivity and 65% specificity. However, the ability to classify the other three datasets is limited. In research of Nijhawan [12] et al., they made use of hybrid CNN for feature extraction to replace the use of traditional machine learning models such as Support Vector Machine (SVM), *K*-Nearest Neighbor (KNN), and Random Forest (RF). Their work was to classify a total of 11 types of nail lesions, and achieve an accuracy of 84.58%. In addition, there were also researches using CNN for single category classification, such as Mehra [13] et al. using a CNN with VGG-16 as the architecture for the classification of subungual melanoma. They presented good results on their in-door dataset. They also indicated that the GPU usage and the size of the dataset could be reduced effectively because of the use of transfer learning.

These methods based on computer vision provided an achievement with considerable accuracy in contrast to conventionally visual-based measuring methods. However, there were still some concerns about CNN-based methods. Due to the monotonicity of CNN classifying each image into a single category, it could not propose objective features to illustrate realistic information and measurements of the lesions, which were important for physicians to analyze in advance. It was not beneficial in clinical practice.

2.2. Segmentation of Nail Lesion

In the task of identifying the entire nail image of finger and toe, the machine learning and deep learning methods might not be able to focus on the texture and characteristics of the nail part to classify. The features of the background and even the skin around the nails in the image might affect the results of the classification. Therefore, models for segmenting objects were widely used, such as U-Net proposed by Ronneberger [14] et al. and Mask R-CNN proposed by He [15] et al., which could segment objects along the contour of objects. It could not only visualize which features were learned by the model, but also effectively segmented the lesion to facilitate clinical diagnosis. Hsieh [16] et al. took use of deep learning model Mask R-CNN for instance segmentation, and embedded a set of designed nail psoriasis image capturing and segmentation severity scoring system. This algorithm effectively segmented the 6 different appearances of nail psoriasis from the nail datasets, and analyzed the severity according to the Nail Psoriasis Severity Index (NAPSI) evaluation method, with an accuracy of 91.5%.

2.3. Summary

Since the great success reached by Hsieh [16] et al. in nail psoriasis, the ability of the segmentation-based algorithm to extract features and segment objects in nails was well recognized. Considering the importance of clinical diagnosis, we utilized the instance segmentation model Mask R-CNN to segment nail lesions and classify in this research.

3. METHODS AND MATERIALS

There were three parts in this section to describe the material and method. In Section 3.1, it showed the experimental environment setting of the computing server, including the hardware and framework used in this research. The dataset would be introduced in section 3.2, including the source of the dataset, the types of lesions, how to allocate the data set, etc. Section 3.3 described the data augmentation methods used in the research to enhance the generalization ability of the model, and a comparison of the augmented images with the original images would be presented. Section 3.4 introduced the image labeling software used in this research, and explained how the labeling technique of our work was performed. Finally, Section 3.5 described the network architecture of the deep learning model used in this research. The composition of each part of the deep learning model was explained as follows.

3.1. Experimental Environment

The experimental environment of this research was developed on the operating system Linux CentOS 7, equipped with two GeForce RTX 2080 ti graphics cards. The Mask R-CNN model was developed by Tensorflow 2.0 framework under CUDA 11.6 and cuDNN 8.4.0 environment.

3.2. Dataset

There were not many open-source nail datasets, so there was relatively little research in the field. Nijhawan [12] and Mehra [13] et al. conducted their research using indoor nail datasets. As for the dataset used in this research was the “nail dataset new” dataset published on Kaggle, which divided the image dataset into 17 classes according to the color, structure, position, and texture, etc. features of the nails. There were alopecia areata, Beau's lines, bluish nail, clubbing, dariers disease, eczema, koilonychia, leukonychia, Lindsay's nails, Muehrcke's lines, onycholycis, pale nail, red lunula, splinter hemmorrage, Terry's nail, white nail, yellow nails, respectively. Class names included naming after disease names as well as nail color. There was a total of 655 images in the dataset. In this research, the training, validation, and testing datasets were allocated in a ratio of 80%, 10%, and 10%, respectively. There were 524, 66 and 65 images within the training, validation and testing sets. Then, the number of images was increased after data augmentation, which was described in section 3.3.

Fig. 1. The 17 classes of images in the dataset.

3.3. Data Augmentation

In this research, pre-processing procedures such as reverse left and right, upside down, rotate -20 to +20 degrees, rotate -45 to +45, scaled 0.5 and 1.5 times were used for image augmentation to improve the generalization ability of the model on various types of imaging conditions. The results outputted from the augmentation were exhibited in Fig. 2.

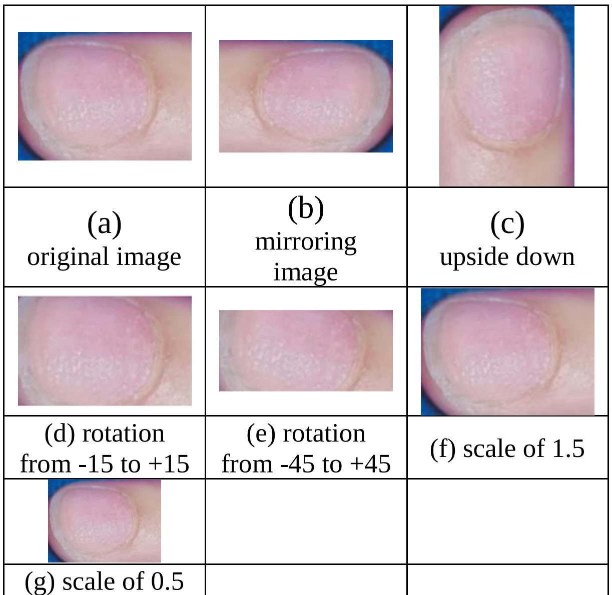


Fig. 2. Augmentation of training images.

3.4. Annotation

At the part of dataset annotation, VGG Image Annotator (VIA) [17] was utilized as the image annotation tool to extract ground truths of lesions in the nail images. The user interface of VIA was shown as Fig. 3. While annotating the regions of lesions, the type of polygons was used to mark up along with the contours of suspicious nail lesions. Then, there was an input text box to record the respective class name when the annotation of region was finished. At the stage of annotation, there was something significant to be aware of. First, the edge of the annotating region must match the contour of the lesion as closely as possible. If there existed margin against the ground truth contour, which implies the regions covering the non-lesion area or insufficient padding of lesions, the model would be confused by the inconsistency of annotation and fail to converge at the end. Therefore, it was essential to examine the annotations, thus ensuring the quality of training progress would not be affected by defects within the datasets.

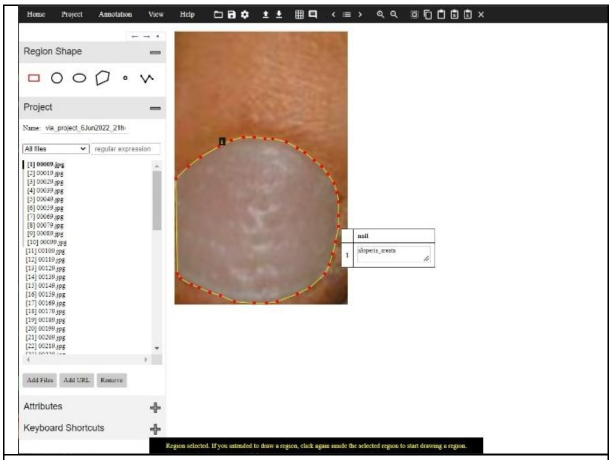


Fig. 3. User interface of VGG Image Annotator.

3.5. Network Architecture

In this research, the Mask R-CNN model was used to segment and classify the nail datasets, and the algorithm of this model was referenced from He [15] et al. Mask R-CNN model is an instance segmentation algorithm, which analyzes through the entire image and detects the target objects individually. The architecture of the Mask R-CNN was illuminated in Fig. 4. The overall progress of the algorithm could be divided into 3 parts: Firstly, the features of the lesions were extracted by the backbone network; Secondly, the potential objects were predicted by the Region Proposal Network (RPN); Finally, the three-branches network would classify each object into according class, predict confidence score, and output fine-grained mask. Differed from the research of Nijhawan [12] and others that performed the classification of the entire image through CNN, Mask R-CNN model paid more attention to the judgment of the boundaries of nail area, and meanwhile, identified each lesion with correct classification even if there had been overlapping conditions among the lesions. With the strength of this algorithm, the performance of the model could be validated more objectively through the range of predicting masks, not merely classified categories themselves.

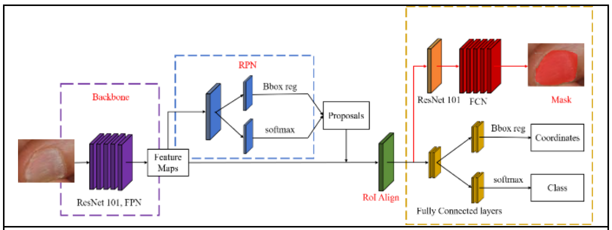


Fig. 4. Architecture of Mask R-CNN model.

To prove that the Mask R-CNN model achieved more success in advance, we introduced the state-of-the-art CNN model, DenseNet-201, to compare which type of algorithm would be more exceeding

4. EXPERIMENTAL RESULTS

In this section, there were 2 parts to exhibit the experimental results of 2 different algorithms in inferencing the nail diseases, including Mask R-CNN in Section 4.1 and DenseNet-201 in Section 4.2. Besides, there were several evaluation indices described in each section to validate the training performance of the model. In advance, there were some issues discussed after analyzing the results, to examine the validity of segmentation-based detection against the CNN-based algorithms.

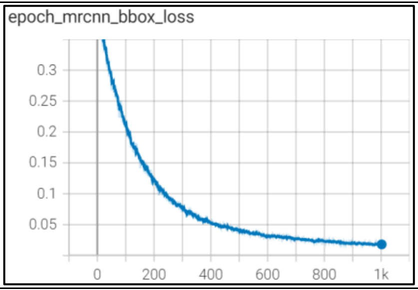
4.1. Mask R-CNN

At the part of backbone network selection, there were 5 different backbone networks, including Xception, DenseNet, ResNet-50, ResNet-101, and ResNet-152, which had been tested on extracting the lesions feature. The testing result was shown as Table 1. There were 4 types of indices, which were minimum loss, mean Average Precision (mAP), F1 score, and dice score, respectively, for comparison of the best performance on segmentation and classification tasks. Under the summary analysis of each network, it showed that the ResNet-101 network had the averagely outstanding scores of the 4 indices; therefore, ResNet-101 was then utilized as the best backbone network of Mask R-CNN model in this research.

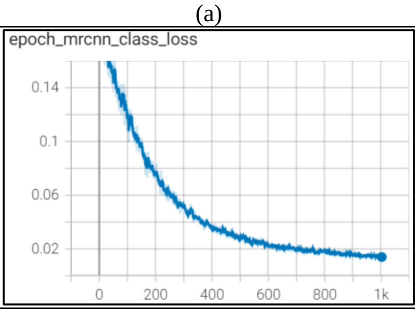
Table 1. Several backbone network comparisons.

Model Architechture	Backbone	**Minimum Loss**	**mAP***	**F1 Score***	**Dice Score***
Mask R-CNN	ResNet-50	0.432	0.48	0.51	0.39
	ResNet-101	0.412	0.48	0.51	0.39
	ResNet-152	0.478	0.49	0.52	0.38
	Xception	1.237	0.40	0.46	0.34
	DenseNet	1.034	0.41	0.47	0.38

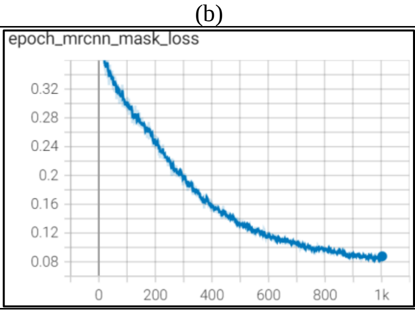
For evaluating the trained models, the loss curves reflected the condition of model convergence during training progress. According to various expected identification purposes, i.e., model classification accuracy, or object detection/segmentation precision, the loss function should be customized to achieve ideal outcome. In this research, several loss functions were utilized as the indices: Mask loss was used for calculating the residual of predicting masks against ground truths; Bounding box loss was for predicting bounding boxes against ground truths; Class loss was for predicting classes against ground truths. The loss curves of Mask R-CNN were shown as Fig. 5. Each curve had steadily declined to a lower loss value, which represented the model gradually converging to the ideal level, and there would be better expected predicting ability while inferencing. By examining the value of each loss curve, there was a minimum loss at 997 epochs, hence it was adopted as the representative weight on predicting tasks.



(a)



(b)



(c)

Fig. 5. Training and validation loss curves. (a) bounding box loss, (b) class loss, (c) mask loss.

In Fig. 6, it presented the confusion matrix of prediction results. The diagonal side of the matrix represented the true positive predictions among the testing samples, the other parts were the false positive ones that were mis-classified. Mis-classification was unavoidable since none of the models was perfect, but it could be identified the predicting accuracy of each class from the proportion of mis-classification. In the result confusion matrix of Mask R-CNN, there were scattered mis-classified samples among them, especially the class of red lunula and yellow nails that had the most false-positive rates of mis-classification. As for the reason, it was assumed that the 2 classes of samples shared similar appearance and patterns on other nail diseases. Therefore, the prediction of the 2 classes had higher false-positive rates.

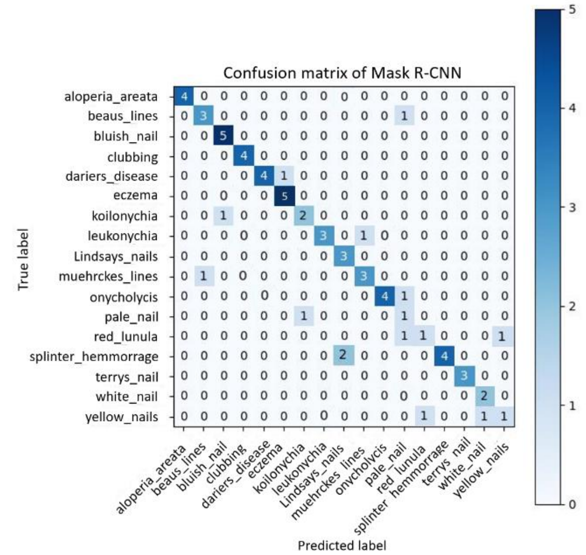
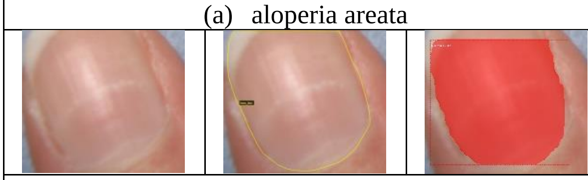


Fig. 6. Multi-class confusion matrix of Mask R-CNN.

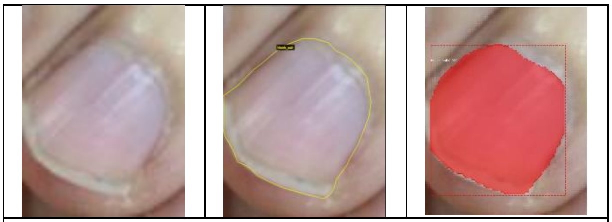
In Fig. 7, it showed the actual visualized results produced by the Mask R-CNN model, the region of lesions would be labeled with contour lines which were referred as masks. There was also other supplemental information provided on the masks, including the bounding boxes, predicting class names and confidence scores. In result figures, it could be observed that the segmentation fineness could suit the edge of the nails. Compared with the ground truth classification label, the result could also achieve great accuracy with the precision rate of 87.69%.

original image	ground truths	predicted image

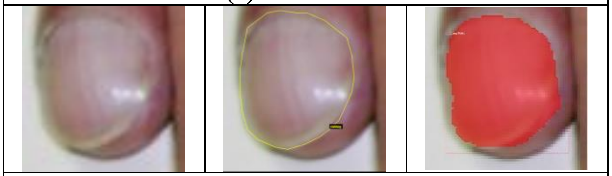
(a) aloperia areata



(b) beau's lines

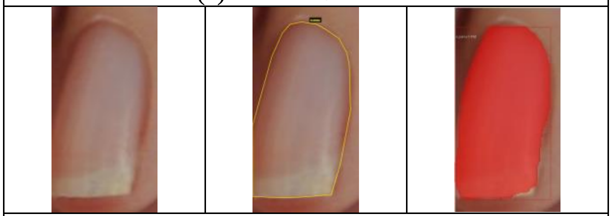


(c) bluish nail

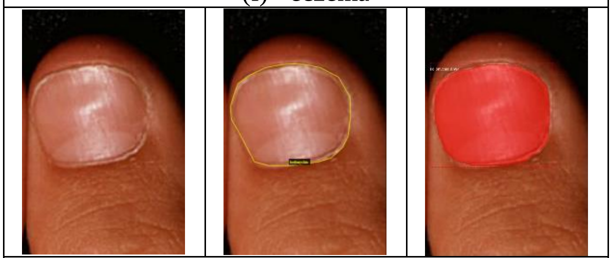


(d) clubbing

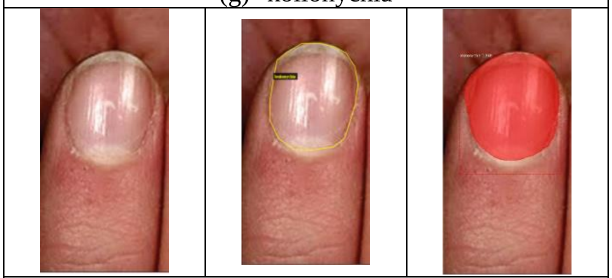
(e) Darier's disease



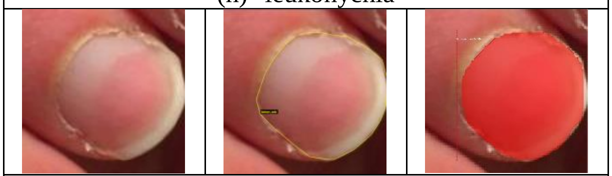
(f) eczema



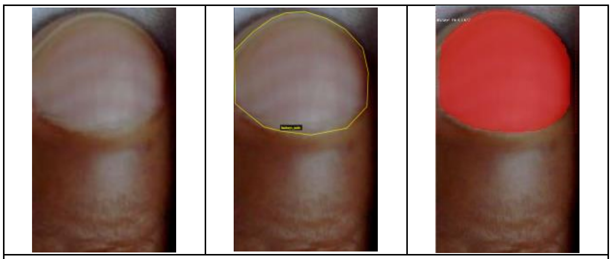
(g) koilonychia



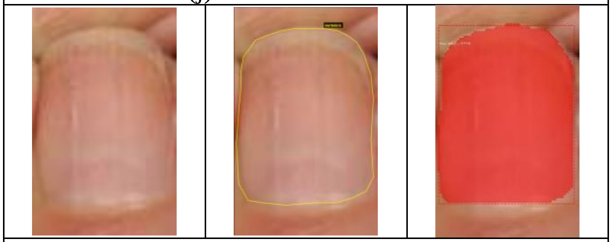
(h) leukonychia



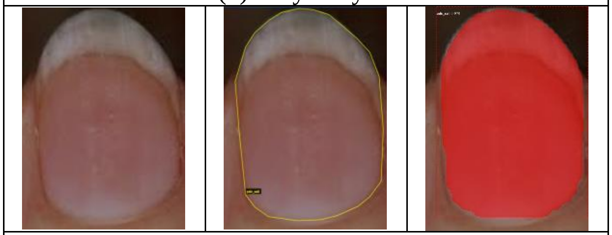
(i) Lindsay's nails



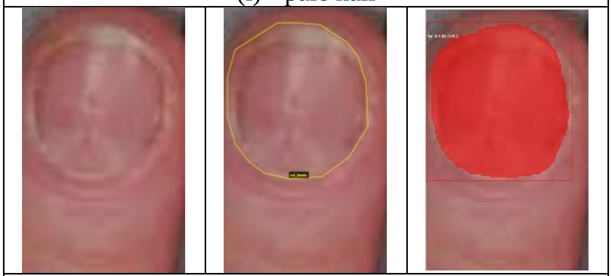
(j) Muehrck-e's lines



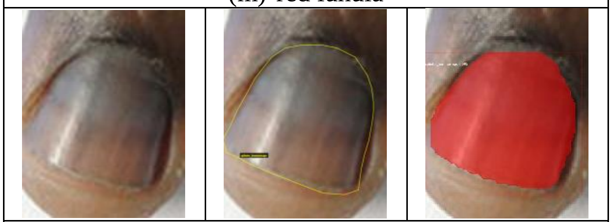
(k) onycholycis



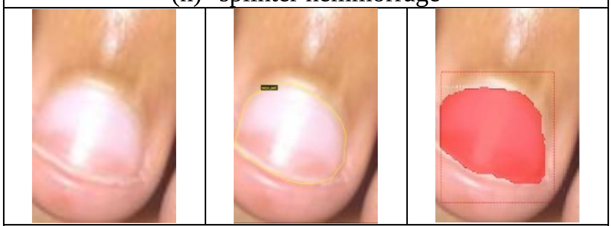
(l) pale nail



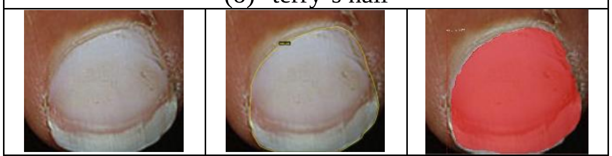
(m) red lunula



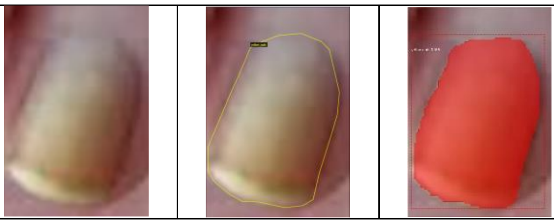
(n) splinter hemmorrage



(o) terry's nail



(p) white nail

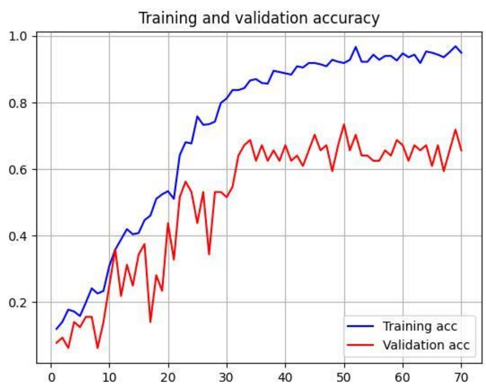


(q) yellow nails

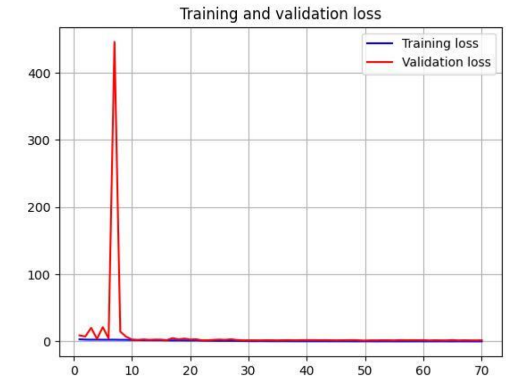
Fig. 7. Each nail lesion class with their original images, ground truths, and predicted images.

4.2. DenseNet-201

In comparison with the instance segmentation-based algorithm Mask R-CNN, here introduced the state-of-the-art CNN model, DenseNet-201, as the confrontation of our work to examine whether additional procedure of instance segmentation would improve the prediction accuracy. The DenseNet-201 model was not like the Mask R-CNN model; it received the entire image to predict class solely, which is commonly doubtful that the model would learn from non-targeting regions, and further lowered the quality of prediction. In Fig. 8, it showed the overall training accuracy and loss curves with a total of 70 epochs. From the trend of the loss curves, the model had converged at 50 epochs, with the highest validation accuracy of 73.44% and least validation loss of 1.4068. In Fig. 9, it showed the confusion matrix of predicting results. There were more scattered mis-predictions in comparison with the confusion matrix of Mask R-CNN in Fig. 6. It implied that the CNN-based algorithm DenseNet-201 had less robustness in predicting the nail diseases.



(a) Accuracy curve.



(b) Loss curve.

Fig. 8. Training accuracy and loss curves of DensNet-201.

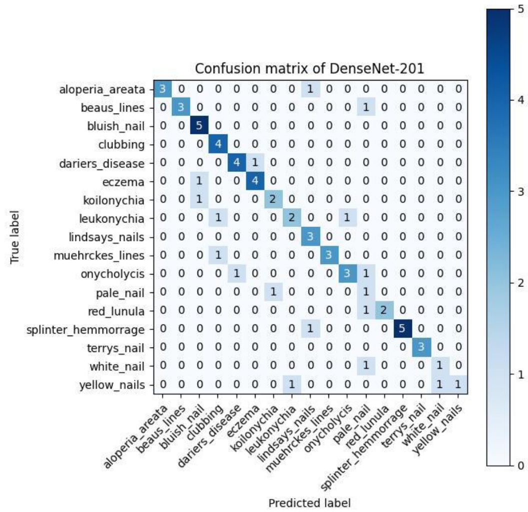


Fig. 9. Confusion matrix of DenseNet-201.

5. CONCLUSION

Most people did not have the clinical experience and domain knowledge of dermatologists, and thus even when there were obvious nail diseases in the appearance or texture of the nails on the hands and feet, it was difficult to take professional treatment immediately. Therefore, in this study, the deep learning algorithm Mask R-CNN was used to segment the nail disease images and classify up to 17 classes of lesions, which is not only quite challenging, but also credible with open datasets instead of indoor data. Model could overcome the various imaging conditions of different equipment and shooting angles, and achieve a precision rate of 87.69% in classification. In the future, it is expected that it can be further developed as a portable mobile APP, so as to realize self-detection of nail condition in no time.

REFERENCES

[1] B. Richert, P. Lecerf, M. Caucanas, and J. André, “Nail tumors,” Clinics in Dermatology, Vol. 31, No. 5, pp. 602-617, 2013.

[2] V. J. James, “Fibre diffraction of skin and nails provides an accurate diagnosis of malignancies,” Int J Cancer, Vol. 125, No. 1, pp. 133-145, 2009.

[3] G. Hinds, and V. D. Thomas, “Malignancy and Cancer Treatment-Related Hair and Nail Changes,” Dermatologic Clinics, Vol. 26, No. 1, pp. 59-68, 2008.

[4] B. M. Piraccini, F. Bruni, and M. Starace, "Dermoscopy of non-skin cancer nail disorders," Dermatologic Therapy, Vol. 25, No. 6, pp. 594-602, 2012.

[5] D. L. Perruchoud, C. Varonier, E. Haneke, R. E. Hunger, H. Beltraminelli, L. Borradori, and A. Ehnis Pérez, “Bowen disease of the nail unit: a retrospective study of 12 cases and their association with human papillomaviruses,” Dermatology and Venereology, Vol. 30, No. 9, pp. 1503-1506, 2016.

[6] K. A. Mannava, S. Mannava, L. A. Koman, L. Robinson-Bostom, and N. Jellinek, “Longitudinal melanonychia: detection and management of nail melanoma,” Hand Surgery, Vol. 18, No. 1, pp. 133-139, 2013.

[7] I. R. Bristow, and D. A. de Berker, “Development of a practical guide for the early recognition for malignant melanoma of the foot and nail unit,” Journal of Foot and Ankle Research, Vol. 3, No. 1, 2010.

[8] H. Koga, T. Saida, and H. Uhara, “Key point in dermoscopic differentiation between early nail apparatus melanoma and benign longitudinal melanonychia,” The Journal of Dermatology, Vol. 38, No. 1, pp. 45-52, 2011.

[9] T. S. Indi, and Y. A. Gunge, "Early Stage Disease Diagnosis System Using Human Nail Image Processing," International Journal of Information Technology and Computer Science (IJITCS), Vol. 7, pp. 30-35, 2016.

[10] P. Maniyan, and B. L. Shivakumar, "Detection of Diseases using Nail Image Processing Based on Multiclass SVM Classifier Method," International Journal of Engineering Science and Computing (IJESC), Vol. 8, No. 5, pp. 17382-17390, 2018.

[11] J. K. Winkler, K. Sies, C. Fink, F. Toberer, A., Enk, and T. Deinlein, et. al., “Melanoma recognition by a deep learning convolutional neural network—Performance in different melanoma subtypes and localisations,” European Journal of Cancer, Vol. 127, pp. 21–29, 2020.

[12] R. Nijhawan, R. Verma, Ayushi, S. Bhushan, R. Dua and A. Mittal, "An Integrated Deep Learning Framework Approach for Nail Disease Identification," 2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 197-202, 2017.

[13] M. Mehra, S. D'Costa, R. D'Mello, J. George and D. R. Kalbande, "Leveraging Deep Learning for Nail Disease Diagnostic," 2021 4th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), pp. 1-5, 2021.

[14] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234-241, 2015.

[15] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-cnn,” arXiv preprint arXiv:1703.06870, 2017.

[16] K. Y. Hsieh, H. Y. Chen, S. C. Kim, H. Y. Chiu, and G. Y. Chen, "A Mask R-CNN Based Automatic Assessment System for Nail Psoriasis Severity," Computers in Biology and Medicine, Vol. 143, 2022.

[17] A. Dutta, A. Zisserman, “The VIA annotation software for images, audio and video,” Proceedings of the 27th ACM International Conference on Multimedia, pp. 2276–2279, 2019.