Glomerulus Classification via an Improved GoogLeNet

XUJING YAO, XINYUE WANG, YELİZ KARACA, JIAYONG XIE, SHUIHUA WANG

ABSTRACT Glomerulosclerosis is a pathomorphological feature of glomerular lesions. Early detection, accurate judgement and effective prevention of the glomeruli is crucial not only for people with kidney disease, but also for the general population. We proposed a method in combination of traditional image analysis with modern machine learning diagnosis system model based on GoogLeNet for recognizing and distinguishing different categories of glomerulus in order to efficiently capture the important structures as well as to minimize manual effort and supervision. We proposed a novel deep learning model based on GoogLeNet with added batch-normalization layers to extract useful features and subsequently entered the features into SoftMax for classification. We also incorporated Bayesian Optimization algorithm and k-fold cross validation in this system for achieving a more reliable result. Our method has eventually achieved an overall accuracy of 95.04±4.99%, and F1 score of 94.44±3.11% for no glomerulus category, 96.73±5.23% for normal glomerulus category and 93.66±7.82% for globally sclerosed glomerulus category, which means this method can accurately determine the degree of glomerulosclerosis with little supervision. The experimental result also shows that this method has better performance when compared with other state-of-art methods.

INDEX TERMS Bayesian optimization, glomerulus classification, glomerulosclerosis, GoogLeNet.

I. INTRODUCTION The incidence of chronic kidney disease (CKD) is gradually increasing worldwide and has become one of the most important diseases threatening human health. Glomerulosclerosis is a pathomorphological feature of glomerular lesions. This lesion is characterized by sclerosing changes in the glomerular capillary loops. It is clinically known as sclerosing glomerulonephritis, which is a glomerular end-stage disease caused by a variety of diseases. Glomerulosclerosis is an irreversible injury that should arouse wide public concern. According to relevant foreign research data on voluntary donors without history of kidney disease, 19% of the 18- to 29-year-old kidney donors, 47% of 40- to 49-year-old donors and 82% of renal donors aged 70 to 77 years had observed global glomerulosclerosis. Thus, early detection, accurate judgement and effective prevention of the glomeruli is crucial not only for people with kidney disease, but also for the general population. However, the time-consuming nature of the detection process as well as the non-specificity nature of signs and symptoms of glomerulosclerosis have constantly been a challenge for clinicians [1]. In addition to this, the supply of transplantable kidneys is very limited. Therefore, hospitals need to accurately assess the viability of kidneys before transplantation to determine which kidneys are suitable for transplantation and reduce organ discard. The
criteria for accepting or rejecting a donor kidney depend largely on whether the glomeruli are normal and sclerotic. The degree of glomerular sclerosis is a key indicator associated with transplant outcomes [2]. These reasons bring great significance for proposing a novel computer-aided diagnosis system model which could recognize and distinguish glomerulus categories as it can not only minimize the human effort and supervision, but also raise up the accuracy and stability of diagnosis.

In recent years, artificial intelligence (AI) has aroused widespread concern in the field of computer-vision, especially in computer-aided diagnosis expert system where many promising research findings are existent [3]–[12]. In 2017, a team from Stanford University compared the ability of identifying skin cancers between 21 dermatologists and the AI algorithm [13]. As reported one year later in the Nature journal, the AI algorithm was able to reach an agreement with all the results tested by experts in classifying skin cancers, which means it has the same ability as dermatologists. Studies show that using AI to make a primary diagnosis before a patient enters the emergency ward could create savings of $5 billion a year.

Nevertheless, with the coming of age of big data, traditional machine learning algorithms have suffered bottlenecks. The most obvious sign is the performance of traditional algorithms which does not increase with the enlargement of data volume, but gradually tends to be steady. In the meantime, deep learning shows obvious differences. It gets better as the data set gets bigger. Besides, theoretically, the deeper the network, the more expressive it is, and the more training data it can process [14].

Deep learning architectures, for instance, convolutional neural network (CNN) are capable of distinguishing the features of image with minimal human supervision [15]. CNN is comprised of layers of stacked neurons. Inside these neurons, input features are transformed into higher dimensional features through convolution layer. The convolution layer is usually followed with non-linear activation function such as ReLU. Besides, in the architecture, the aim of utilizing pooling layer is often for down-sampling. Eventually, soft-max classifier would receive the compressed high dimensional features and complete the classification task [16]. CNN has obtained satisfactory results in medical imaging analysis, such as skin cancer detection, pulmonary nodule classification and categorization of periodontal diseases from oral images [17]–[19].

However, traditional deep learning methods are isolated. In most cases, they train isolated models purely based on specific tasks and data sets without retaining knowledge. In this way, little knowledge or information can be migrated from one model to another. Transfer learning (TL) is an effective pattern recognition problem-solver [20]. In transfer learning, you can use previously trained models to gain knowledge such as features and weights to train new models, and even deal with new tasks with less data as shown in FIGURE 1. The different patterns in the light green circle just represent all the tasks are different.

In other words, when comparing with traditional methods, advantage of transfer learning is that it can do multi-task learning. Traditional models face different types of tasks and need to train multiple different models. While with transfer learning, we can first implement simple tasks and apply the knowledge obtained from simple tasks to deal with more difficult problems so as to solve problems faster and better as can be seen in FIGURE 2.

Therefore, the purpose of this paper is to propose a deep learning model for the rapid and accurate diagnosis of the degree of glomerulosclerosis based on the transfer learning method. Our paper is the first to report the use...
of improved GoogLeNet with the batch-normalization layer added in classifying glomerular sclerosis and we also incorporates Bayesian optimization in the system. The following chapter II, III, IV, V and VI will discuss the related work, dataset, methodology, experimental results and conclusion, respectively.

II. RELATED WORK
In previous studies, trichrome-stained image has been considered as reliable data set material and often used in glomerular related research. The team of Kannan et al. [21] presented a convolutional neural network (CNN) system based on Google’s Inception v3 CNN architecture to accurately discriminate the input trichrome-stained images from healthy and globally sclerosed (GS) images with a performance of 92.67% accuracy on test data. Ginley et al. [22], on the other hand, utilized CNN to detect glomerular boundaries and nuclei and an unsupervised technique to find other glomerular structures. A recurrent network architecture was also proposed by them to process the features they defined into the classification. Marsh et al. [2] described a method that utilized a deep learning model to identify and classify non-sclerosing and sclerosing glomeruli in a full-slide image of a frozen biopsy of a donor kidney. The performance of this model is very robust for sliding prepared workpiece related to frozen section preparation, and it overcomes the technical difficulty of applying the pre-trained CNN bottleneck model to the classification of the whole slide image and can successfully achieve the segmentation of the whole slide image. The model finally achieves 84.75% F1 score on non-sclerosing glomeruli category and 64.92% F1 score on sclerosing glomeruli category. In the paper of Altini et al. [23], the authors proposed a computer aided diagnosis (CAD) system based on a CNN to assess global glomerulosclerosis. In particular, the authors considered approaches based on semantic segmentation networks, such as SegNet and DeepLab V3+. At the target detection level, the optimal F1 score of the non-sclerosing glomeruli was 92.4%, and the optimal F1 score of the sclerosing glomeruli was 73%.

III. DATASET
Kidney biopsy procedures were performed on the selected patients treated at Xinghua People’s Hospital of Jiangsu Province between 2018.1 and 2019.12. A total of 183 images captured from kidney biopsy slides were available. These biopsy samples were taken from the adult patients who underwent a local biopsy, independent of the indication for the biopsy procedure. The inclusion criterion was the availability of pathology slides.

Biopsy samples were obtained in the form of individual trichrome-stained slides prepared from formalin-fixed, paraffin-embedded core-needle biopsy tissue. A selected core visible on each slide was imaged at 40 magnification (indicating a 4 objective and a 10 eyepiece) using a Nikon Eclipse TE-2000 microscope.

A sliding window operation was defined to crop each slide to smaller images. Each cropped image was then evaluated by at least 3 experts into 3 categories as displayed in FIGURE 3: (i) no glomerulus, (ii) normal glomerulus, and (iii) globally sclerosed (GS) glomerulus. ‘No glomerulus’ refers to the image that does not contain glomerulus. ‘Normal glomerulus’ means this image contains normal and healthy glomerulus which is circled by the black dotted line. ‘GS glomerulus’ means the glomerulus in this image is sclerotic. The sclerosing glomeruli is also circled with black dotted lines.

The dataset was processed with data augmentation for enrichment. And all the experiments on dataset were conducted on the laptop with GTX1060.

IV. METHODOLOGY
For the pre-processing of dataset, we utilized data augmentation, due to the reason that a successful neural network generally requires a large dataset, many of which are in the millions. However, the existing data in the actual situation are not as much as we imagine, and the acquisition of new data will bring about a huge amount of cost. Therefore, we choose to enhance the data, that is, conducting reflection, translation or rotation the existing data to create more data. In this way, the neural network will have a stronger generalization ability and be better adapted to application scenarios. We then combined traditional image analysis with modern machine learning diagnosis system model based on GoogLeNet for recognizing and distinguishing different categories of glomerulus in order to efficiently capture the important structures as well as to minimize manual effort and supervision. We added several batch-normalization layers into original GoogLeNet model and then utilized the new pre-trained model to extract useful features and finally entered the features into SoftMax for classification. K-fold cross validation was taken during the process to get a more reliable result. In addition, for fine turning the hyperparameters, we introduced Bayesian Optimization. A methods diagram is displayed in FIGURE 4.

A. TRANSFERLEARNING: IMPROVED GoogLeNet WITH BATCH-NORMALIZATION
GoogLeNet is a classic deep learning structure proposed by Szegedy et al. [24]. Unlike deeper the network to achieve better training performance, which will at the same time bring loads of negative effects, such as overfitting, gradient
disappearance, gradient explosion. GoogLeNet improves the training results by making more efficient use of computing resources, in other words, extracting more features for the same amount of computation [25], [26].

The basic structure is illustrated in FIGURE 5. The entire GoogLeNet structure is made up of several such inception modules in series. The inception structure has two main contributions: one is using $1 \times 1$ convolution to raise and lower the dimension, the other is carrying out convolution and reintegration on multiple scales at the same time.

On the basis of GoogLeNet, to further improve it, we chose to add some batch-normalization (BN) layers. Santurkar et al. [27] concluded in a research that BN prevents gradient explosions or dispersion, improves the robustness of the train-time model to different hyperparameters (learning rate, initialization), and keeps most of the activation functions away from its saturation area. Thus, all these natures of BN, can help us achieve a fast and robust training network. The relevant comparative experimental results will be presented in detail in the Chapter V.

**B. HYPERPARAMETER TUNING: BAYESIAN OPTIMIZATION INCORPORATED IN THE SYSTEM**

In machine learning, hyperparameter tuning is a tedious but crucial task because it affects the performance of the algorithm to a large extent. However, manual tuning is time-consuming, and when considering grid search and random search, although there is no need for human effort, these two searching methods require long running time and are easy to get the local optimal when coming across non-convex function problem. Compared with grid search, the advantages of Bayesian optimizers are that, on the one hand, the number of iterations is small, which saves considerable amount of time, on the other hand, they have robustness for non-convex problems [28]–[30].

The application of Bayesian Optimization to machine learning in adjusting hyperparameters was proposed by Snoek et al. [31] in 2012. The general concept is that given the optimized objective function (a generalized function that only specifies the input and output, without knowing the internal structure or mathematical properties), the posterior distribution of the objective function is updated by constantly adding sample points (a Gaussian process), until the posterior distribution basically fits the real distribution. To conclude, Bayesian Optimization adjusts the current parameter by considering the information of the previous parameter, in this way, it becomes possible to spare one from engaging in a lot of useless work. Using the Bayesian optimization to tune the hyperparameter can be divided into two parts: Gaussian process and Bayesian optimization.

Gaussian process is used to fit the optimization objective function by taking the previously explored points as shown in FIGURE 6.

Bayesian optimization, including ‘exploitation’ and ‘exploration’, is used to find the best value with the least cost. Exploitation refers to conduct sampling in the region where the global optimal solution is most possible to occur on the basis of posterior distribution. While exploration refers to obtain the sampling point in the unsampled area. For efficient sampling, we need the Acquisition Function, which is used to find the next Function of $x$. Among the mainstream Acquisition functions, this experiment selects ‘expected improvement per second plus’ and a flowchart of working scheme for Bayesian Optimization is displayed in FIGURE 7.

And to best utilize the power of Bayesian optimization, 30 objective function evaluations were performed [31]. A set of result graphs of Min objective vs. Number of function evaluations will be displayed in Chapter V.

**C. MEASURE: K-FOLD CROSS VALIDATION**

For the purpose of avoiding the problem of overfitting, k-fold cross validation technique is developed. In a certain sense, the original dataset was grouped, with one part as training set, one part as validation set and one part as testing set. K-fold cross validation will randomly divide the training data into k parts.
FIGURE 7. A flow chart for Bayesian optimization.

and do k times of training. The validation set was set to test the generalization error of the model [32].

The reason why k-fold is adopted in this experiment is that this method is very suitable for processing data sets with only a small amount of data. If the data volume is large, the training cost will also be increased by k times, which brings about substantial inconvenience. In machine learning modeling, data sets are usually divided into training sets and test sets. The testing set is data independent of the training and used only for the evaluation of the final model. One problem with this direct division is that the testing set does not participate in the training, which wastes this part of data and to some degree fails to optimize the model due to the reason that data determines the upper limit of program performance, which models and algorithms approach. Thus, it is significant to make good use of data sets, especially for this experiment. However, we do need the testing set as we need to verify the network generalization performance. K-fold cross validation successfully solve this problem as in the form of that, all data sets can be utilized, and finally, the model performance can be expressed reasonably by means of averaging.

In our experiment, we use 5-fold cross validation as it performs better after comparison. Images are split into 80% for training and validation, 20% for testing as displayed in FIGURE 8.

FIGURE 8. Five-fold cross validation.

All the images taken from a single patient were used either for training or the testing set to prevent over fitting due to patient-specific characteristics. In addition, the augmentation is respectively done to the sets, in this way, there will not be the case of the circumstance of having one original image in training set and the augmented image of that one in the testing set.

V. EXPERIMENTAL RESULTS AND DISCUSSION

This chapter will show statistical results of the comparison experiment results without adding BN layer, the comparison experiment results without Bayesian optimization and in detail display the final results of the improved model proposed by us. Since we are making predictions, we naturally hope that the predictions we make are as accurate as possible. Accuracy is the proportion of samples in which we make correct predictions. At the same time, in order to comprehensively evaluate the performance of the classifier, we used the harmonic average F-score, which considers the value of both precision and recall, as one of the evaluation criteria. Due to the reason that each category in our dataset has the same number of frames, we chose to use Macro-average method to calculate F1 score. In this method [33]–[36], the average of the evaluation indexes (Precision/ Recall/ F1 score) of different categories was obtained directly by adding them together, and the same weight was given to all categories. Supposing the number of samples classified correctly as ‘0’ is TP, the number of samples classified incorrectly into ‘0’ is FP, the number of samples in ‘0’ classified incorrectly into other categories is FN. The general expression of F1 score value for category ‘0’ is according to (1), (2), (3):

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{1}
\]
\[
\text{Recall} = \frac{TP}{TP + FN} \tag{2}
\]
\[
F1 \text{score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}
\]

TABLE 1. Data splits.

| Label | No. of training frames | No. of validation frames | No. of testing frames | Overall |
|-------|------------------------|--------------------------|-----------------------|---------|
| None  | 39                     | 10                       | 12                    | 61      |
| Normal| 39                     | 10                       | 12                    | 61      |
| GS    | 39                     | 10                       | 12                    | 61      |
| Overall| 117                    | 30                       | 36                    | 183     |

A. RESULT OF OUR MODEL WITHOUT BATCH-NORMALIZATION LAYER

In TABLE 2, our comparison experiment results show that if we do not add BN layers into GoogLeNet, the model will achieve an overall accuracy of 93.38±5.40%, and F1 score of 94.06±4.02% for no glomerulus category, 94.37±4.25%
TABLE 2. Statistical results on testing set.

| Set index | Overall Accuracy (%) | F1 score for Normal (%) | F1 score for GS (%) |
|-----------|----------------------|-------------------------|---------------------|
| 1         | 97.22                | 96.00                   | 95.65               |
| 2         | 86.11                | 92.31                   | 88.89               |
| 3         | 91.67                | 92.31                   | 91.67               |
| 4         | 91.89                | 89.66                   | 95.65               |
| 5         | 100.00               | 100.00                  | 100.00              |
| Average   | 93.38±4.40           | 94.06±4.02              | 94.37±2.25          | 91.10±10.75 |

for normal glomerulus category and 91.10±10.75% for GS glomerulus category.

B. RESULT OF OUR MODEL WITHOUT BAYESIAN OPTIMIZATION

In TABLE 3, our comparison experiment results show that if we do not utilize Bayesian optimization algorithm to adjust hyperparameters, the model will achieve an overall accuracy of 88.48±4.66%, and F1 score of 89.80±4.50% for no glomerulus category, 91.77±6.15% for normal glomerulus category and 82.07±13.70% for GS glomerulus category.

TABLE 3. Statistical results on testing set.

| Set index | Overall Accuracy (%) | F1 score for None (%) | F1 score for Normal (%) | F1 score for GS (%) |
|-----------|----------------------|-----------------------|-------------------------|---------------------|
| 1         | 91.67                | 92.31                 | 90.91                   | 91.67               |
| 2         | 89.56                | 92.31                 | 82.76                   | 58.82               |
| 3         | 88.89                | 91.67                 | 92.31                   | 81.82               |
| 4         | 89.19                | 81.82                 | 100.00                  | 85.71               |
| 5         | 92.11                | 90.91                 | 92.86                   | 92.31               |
| Average   | 88.48±4.66           | 89.80±4.50            | 91.77±6.15               | 82.07±13.70         |

C. RESULT OF OUR FINAL IMPROVED MODEL

1) CONFUSION MATRIX

In our experiment, class label ‘0’ represents ‘no glomerulus’, label ‘1’ represents ‘normal glomerulus’, and label ‘2’ refers to ‘globally sclerosed glomerulus’. We can see in FIGURE 9 that our model has good performance for classifying each category. During the first time of running, the classifier has correctly recognized 12 images of ‘no glomerulus’ as ‘no glomerulus’, 11 images of ‘normal glomerulus’ as ‘normal glomerulus’, 12 images of ‘globally sclerosed glomerulus’ as ‘globally sclerosed glomerulus’, but misidentified 1 image of ‘normal glomerulus’ as ‘no glomerulus’. In addition, the value 92.3% in confusion matrix means 92.3% of ‘sample’ predicted as ‘no glomerulus’ are classified right, and the value 91.7% means 91.7% of real ‘normal glomerulus’ sample are classified right.

2) STATISTICAL RESULTS

It is obviously displayed in TABLE 4 that our final model achieves an overall accuracy of 95.04±4.99%, and F1 score of 94.44±3.11% for no glomerulus category, 96.73±5.23% for normal glomerulus category and 93.66±7.82% for GS glomerulus category. These excellent results show that there does exist difference between these three categories of glomerulosclerosis and our model has the ability to find the difference and accurately classify them.

3) BAYESIAN OPTIMIZATION PROCESS

In FIGURE 10, an example graph of Min objective vs. Number of function evaluations achieved during the Bayesian optimization process is shown. The algorithm will train the system with different combination of hyperparameters such as section depth, initial learning rate and momentum.
TABLE 5. Statistical results on testing set.

| Author Model          | Class | Overall Accuracy | F1 score |
|-----------------------|-------|------------------|----------|
| Kamarudin [21] Inception v3 | Overall | 92.7             | /        |
| Marsh [22]: FCN + bio-detection | GS    | 64.9             | /        |
|                        | Normal | 84.8             |          |
| Altini [23]: DeepLab v3+ | GS    | /                | 92.4     |
|                        | Normal | /                | 63.6     |
| Ours: Without BN      | None              | 93.38±5.40       | 94.06±4.02 |
|                        | Normal            | 91.10±10.75      |          |
| Ours: Without Bayesian | None            | 89.80±4.50       |          |
|                        | Normal            | 91.77±6.15       |          |
|                        | GS                | 82.07±13.70      |          |
| Ours: GoogLeNet-BN-Bayesian | None | 95.04±4.99       | 94.44±3.11 |
|                        | Normal            | 96.73±5.23       |          |
|                        | GS                | 93.66±7.82       |          |

**FIGURE 11.** A summary of comparison between experimental results and state-of-art approaches.

**VI. CONCLUSION**

To conclude, our novel GoogLeNet-BN-Bayesian model can help to realize rapid, effective, stable and safe detection of degree of glomerulosclerosis, which is of great significance to clinical medicine and society. Our method eventually has achieved an overall accuracy of 95.04±4.99%, and F1 score of 94.44±3.11% for no glomerulus category, 96.73±5.23% for normal glomerulus category and 93.66±7.82% for GS glomerulus category. This proves that this method can accurately determine the degree of glomerulosclerosis with little supervision and it performs better than the previous state-of-art works. In the future, the system can be further optimized for internal network structure. We also plan to add segmentation techniques to calculate the area of glomerulosclerosis. In this way, patients can have a clearer understanding of the extent of their illness. We hope this computer diagnostic system can help more medical and computer related research.

**REFERENCES**

[1] M. C. Scott, “Defining and diagnosing sepsis,” *Emergency Med. Clin.*, vol. 35, no. 1, pp. 1–9, 2017.

[2] J. N. Marsh, M. K. Matlock, S. Kadose, T.-C. Liu, T. S. Stappenbeck, J. P. Gaut, and S. J. Swamidass, “Deep learning global glomerulosclerosis in transplant kidney frozen sections,” *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2718–2728, Dec. 2018.

[3] Y. Jiang, Y. Zhang, C. Lin, D. Wu, and C.-T. Lin, “EEG-based driver drowsiness estimation using an online multi-view and transfer TSK fuzzy system,” *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 27, 2020, doi: 10.1109/TITS.2020.2973673.

[4] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, S. Qiu, and L. Chin-Teng, “A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation,” *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Jan. 3, 2020, doi: 10.1109/TCBB.2019.2963873.

[5] Y. Jiang, A. Bi, K. Xia, J. Xue, and P. Qian, “Exemplar-based data stream clustering toward Internet of Things,” *J. Supercomput.*, vol. 76, pp. 1–29, Nov. 2019.

[6] Y. Jiang, J. Zhu, X. Gu, J. Xue, K. Zhao, T. Ni, and P. Qian, “Recognition of large-scale ncRNA data using a novel multitask cross-learning 0-order TSK fuzzy classifier,” *J. Med. Imag. Health Inform.*, vol. 10, no. 2, pp. 502–507, Feb. 2020.

[7] Y. Jiang, J. Xue, R. Wang, K. Xia, X. Gu, J. Zhu, L. Liu, and P. Qian, “Seizure recognition using a novel multitask radial basis function neural network,” *J. Med. Imag. Health Inform.*, vol. 9, no. 9, pp. 1865–1870, Dec. 2019.

[8] Y. Jiang, K. Zhao, K. Xia, J. Xue, L. Zhou, Y. Ding, and P. Qian, “A novel distributed multitask fuzzy clustering algorithm for automatic MR brain image segmentation,” *J. Med. Syst.*, vol. 43, no. 5, p. 118, May 2019.

[9] Y. Jiang, J. Zheng, X. Gu, J. Xue, and P. Qian, “A novel synthetic CT generation method using multitask maximum entropy clustering,” *IEEE Access*, vol. 7, pp. 119644–119653, 2019.

[10] Y. Jiang, D. Wu, Z. Deng, P. Qian, J. Wang, G. Wang, F.-L. Chung, K.-S. Choi, and S. Wang, “Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2270–2284, Dec. 2017.

[11] Y. Jiang, Z. Deng, F.-L. Chung, G. Wang, P. Qian, K.-S. Choi, and S. Wang, “Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system,” *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 3–20, Feb. 2017.

[12] S. Sun, X. W. Duan, X. Chen, R. T. Zhang, L. M. Chai, and L. L. Li, “Relationships among T-helper cell subtypes, cytokines, and traditional Chinese medicine syndromes in patients with psoriasis,” *J. Med. Imag. Health Inform.*, vol. 7, no. 1, pp. 41–46, 2017.

[13] B. Kalis, M. Collier, and R. Fu, “10 promising AI applications in health care,” *Harvard Bus. Rev.*, vol. 10, pp. 1–5, May 2018.

[14] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436, 2015.

[15] L. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[16] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolutional neural networks,” in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2016, pp. 2560–2567.

[17] Y. Yang, X. Feng, W. Chi, Z. Li, W. Duan, H. Liu, W. Liang, W. Wang, P. Chen, J. He, and B. Liu, “Deep learning aided decision support for pulmonary nodules diagnosing: A review,” *J. Thoracic Disease*, vol. 10, no. S7, pp. S867–S875, Apr. 2018.
