Chromosome Classification with Convolutional Neural Network based Deep Learning

Wenbo Zhang* Sifan Song* Tianming Bai* Yanxin Zhao* Fei Ma*§ Jionglong Su*‡§ Limin Yu†§
*Mathematical Sciences Department, Xi'an Jiaotong-Liverpool University, Suzhou, China
†Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China
‡Neusoft Corporation, Shenyang, China
§Research Center for Precision Medicine, HT-URC, Xi'an Jiaotong-Liverpool University
Suzhou, China, Tel: (86) 512 8816 1633, Email: limin.yu@xjtlu.edu.cn, jionglong.su@xjtlu.edu.cn

***Abstract*—Karyotyping plays a crucial role in genetic disorder diagnosis. Currently Karyotyping requires considerable manual efforts, domain expertise and experience, and is very time consuming. Automating the karyotyping process has been an important and popular task. This study focuses on classification of chromosomes into 23 types, a step towards fully automatic karyotyping. This study proposes a convolutional neural network (CNN) based deep learning network to automatically classify chromosomes. The proposed method was trained and tested on a dataset containing 10304 chromosome images, and was further tested on a dataset containing 4830 chromosomes. The proposed method achieved an accuracy of 92.5%, outperforming three other methods appeared in the literature. To investigate how applicable the proposed method is to the doctors, a metric named proportion of well classified karyotype was also designed. An result of 91.3% was achieved on this metric, indicating that the proposed classification method could be used to aid doctors in genetic disorder diagnosis.**

I. Introduction

The human cell normally contains 23 pairs of chromosomes, including 22 pairs of autosomes (the ones exist in both males and females) as well as sex chromosomes X and Y. Females have double X chromosomes as one pair of sex chromosomes, while males have both X and Y. Chromosome abnormality, namely aneuploidy (having abnormal number of chromosomes in a cell) and structural abnormalities (including deletions, duplication, translocation, inversion, insertions rings, and isochromosome) may cause genetic disorder such as Down's syndrome. It is important to inspect the cells of a patient and identify any irregular, extra or missing parts for diagnostic purposes. Karyotyping, the process of separating and classifying human chromosomes from a cell image, plays a crucial role in this diagnosis process [1].

However, accomplishing this work efficiently not only requires considerable manual efforts, domain expertise and experience, but also consumes a lot of time. Since 1980, with the motivation of lightening the load of cytogeneticists, automatic diagnosis systems for chromosome karyotyping and analysis have become a popular and important task.

J. M. Cho chose the two-layer artificial neural network with the error backpropagation training as chromosome classifier, which resulted an overall classification error rate of 6.52% in the 460 chromosomes images [2]. To overcome the higher classification error, J. Cho et. al. [3] proposed a hierarchical multi-layer network as chromosome classifier and an error back-propagation training algorithm. The overall result of classification error in this method was 5.9% which was based on the 7 experiments.

S. Delshadpour [4] reduced the complexity of an ANN in order to increase the performance of ANN and combined an improved multi-layer perceptron neural network for automated classification of chromosomes. The overall accuracy of classification was increased to 88.3% on 304 chromosomes. However, their classification results on 24 classes vary and on many classes the results are not very accurate. To overcome the problems, B. C. Oskouei and J. Shanbehzadeh [5] proposed a classifier based on the wavelet neural network. They obtained an accuracy rate of 93.35%. S. Gagulapalalic and M. Can [6] proposed a novel method based on the Competitive Neural Network Teams (CNNTs) to distinguish 22 types of the autosomes. Their method achieved better classifying results, i.e. approximately 96.64%, on 150 chromosome images with each type of autosomes.

M. J. Roshtkhari and S. K. Setarehdan [7] presented a wavelet transform based linear discriminant analysis to classify normal and automatically straightened chromosomes, and a three layers feed-forward perceptron neural network which was trained using the backpropagation algorithm. The overall outcome of correct classification was 99.3% after 303 highly curved chromosomes were straightened. A subspace-based approach was proposed by Q. Wu et. al., which synthesized the prototype chromosome images and utilized transformation coefficients as the feature measurements [8]. The result shows that this method could synthesize highly visual prototype chromosome images which were previously unseen in

978-1-5386-7604-2/18/$31.00 ©2018 IEEE

chromosome classification. Nevertheless, most of the proposed systems are partially automated and still need much manual assistance, since defining and extracting features of images with overlapping or even touching chromosomes are still difficult steps [6].

With the development of CNNs, they have been utilized in medical image sector to deal with complicated features. CNN based models have been explored to analyze chromosome images. In 2016, Qiu et al. proposed a 8-layer convolutional neural network and its lowest testing error achieved 13.3%, which is the first research proposing a CNN model to classify metaphase chromosome images [9]. In 2017, Sharma et al. utilized a 4 deep learning blocks to construct a model and achieved an accuracy of 68.5% (without straightening) and 86.7% (with straightening) [10]. Swati et al. proposed a model based on the Siamese networks which is composed by a twin neural network and the best model yields 85.2% classification accuracy [1].

Although above methods are based on convolutional networks, their models do not utilize dropout layers and norm regularization methods (such as L1 and L2) to avoid overfitting, and the setting of parameters may not be optimal. In this study, we consider to construct a new CNN based model to address the above issues. The rest of the paper is organized as follows. In Section II, we outline the datasets used in the experiment and pre-processing approaches. The structure of our CNNs model is shown in Section III. In Section IV, we utilized three metrics to evaluate our model and show the results. Also, we compare our result with other methods. Finally, conclusion and future work are presented in Section V.

II. Data set and Summary Statistics

A. Data preprocessing

The raw data were collected from a local company. The data set includes both cell images and their corresponding keryotypings. An example of a cell image and its karyotype is shown in Fig.1.

For image preprocessing, we first use an image binarization operator on each karyotype to turn it into binary images. We then applied an area filter to remove noises such as the labels [11]. After that, the regionprop function of Matlab can easily generate bounding boxes to contain individual chromosomes [12]. On each karyotype, 46 single chromosome can be identified in sequence. An example is shown in Figure 1. Each single chromosome were centrally placed in a 142x282 picture with black background, as shown in Figure 2. In the end, we can obtain 23 sets of chromosomes for each karyotype.

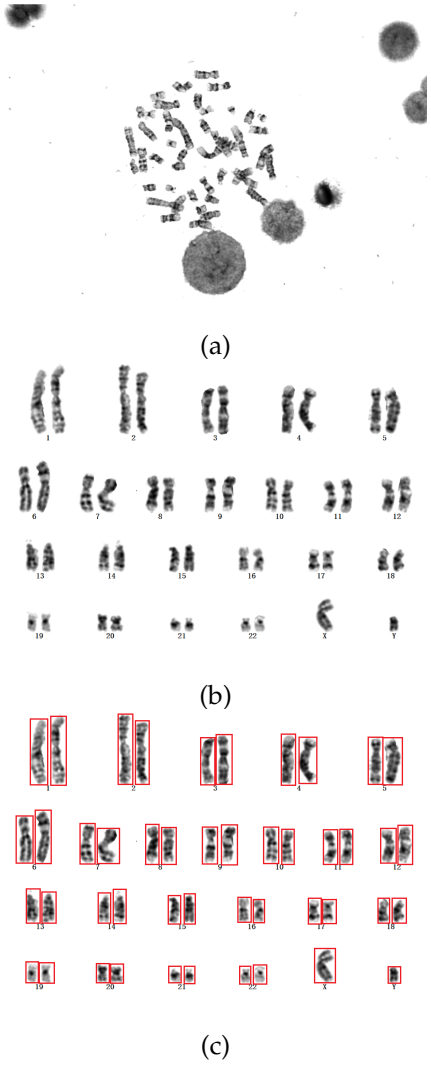

Fig. 1. (a) A raw cell image with its 46 chromosomes; (b) karyotype of image (a); (c) individual chromosomes with bounding boxes.

Fig. 2. A standardized individual chromosome

TABLE I
The datasets used in the experiment

Dataset	training	test	total
1	8243	2061	10304
2 (Test Set)	0	4830	4830

B. training and test dataset

After data pre-processing, we obtained images with each of which contains only one individual chromosome. The images were adjusted to the size of 120×40. In this study, we utilized two datasets from these processed data and they are shown in Table I. The size of two datasets were 10304 and 4830 respectively. For dataset

1, we randomly divided it into a training set (8243) and a test set (2061). The labels of training set range from 1 to 24, corresponding to chromosome classes. Dataset 2 is used only for test. It consists 105 chromosome images, each with 46 chromosomes labeled with the same image number. The dataset will be used with a new evaluation metric which will be introduced in Section V.

III. Classification model based on CNNs

Convolutional neural networks (CNNs) belong to artificial neural networks which have been widely used in computer vision, natural language processing and other fields. VGG-16 [12], ResNet50 [13], Inception network [14] have achieved excellent efficiencies in areas such as image classification and understanding. Five main operations in these methods are convolutional kernel, nonlinear transformations (activation functions), pooling method, fully connected layers and loss function (classification). The emerging of the convolutional kernel is illumined by the unique structure of cerebral cortex neurons of cats to learn and integrate deep features of image [15]. The activation functions and pooling approaches assist convolutional layers to subsequently extract and filter more useful signals [16]. As classification engines, final fully connected layers and loss function can screen and supervise deeply learned features to their specific labels [17].

In this study we construct a new deep network structure based on CNN methods. There are five types of layers used in our model, including convolution layer, pooling layer, dropout layer, flatten layer and dense layers. The main structure of the network is shown in Fig. 3. In our architecture, the number of convolution layers is four and each convolution layer use filters of size 3*3 except the first one (5*5). This is because small size filters can decrease the amount of computation and hence to train the model faster. Also, the activation function of each convolution layer is rectified linear units (ReLU) rather than sigmoid:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases} \tag{1}$$

The reason is that compared with sigmoid function, ReLU could efficiently solve gradient vanishing problem in the backpropagation process of updating parameters and reduce computation [16].

At the beginning of our model, we pile two convolution layers together and the two layers have 256 and 128 filters respectively. The reason is that it could enhance the ability of model in learning features. For example, compared with one layer, this structure incorporates two non-linear relu activation functions rather than one single function, which makes the decision function more discriminative [12].

After a couple of convolution layers, a pooling layer is used, which could decrease the amount of parameters and accelerate the computation. Then a dropout layer with parameter 0.5 connects to the end of pooling layer. Dropout layer is utilized to randomly set some dimensions of input vector to be zero with certain probability (according to the parameter we set). Therefore, it does not have any trainable parameters, meaning that there is no updating during the process of training. This kind of layer can mitigate overfitting to a large extent [18]. After that, there are the third convolution layer, consisting of 256 filters, and a pooling layer. The final convolution layer with 128 filters take the input from the last layer. In terms of flatten layers, since the shape of data flow is image (the format is array) which does not match the desired type (vector), flatten layer is utilized to transform the data flow. The final fully connected dense layer with 120 neurons followed by one output layer produces a distribution over the 24 output labels.

To cater to multi-class classification, the activation function chosen in the output layer is softmax:

$$y(z) = \frac{e_i^z}{\sum_j e_j^z} \tag{2}$$

Where z is the input of the softmax layer, i represents ith class.

The optimization algorithm used here is Adam [19] with learning rate 0.0001. Since our task in this study is multi-classification, the loss function used here is cross-entropy objective function:

$$L(\{x,y\}_1^N) = -\sum_n \sum_i y_i^{(n)} * \log(\hat{y}_i^{(n)}) \tag{3}$$

where $\{x,y\}_1^N$ are the training data with corresponding label, $y_i^{(n)}$ represents whether nth sample belongs to class i, and the value is 1 if it is true, and 0 otherwise. $\hat{y}_i^{(n)}$ is the probability that the network assigns the nth sample to ith class.

IV. Experiments and evaluation metrics

In the experiments, the pre-processed dataset 1 and 2 were used. Training set of dataset 1 was firstly fed into our model and the batch size was set at 128. After being trained through 100 epoches, this mode was used to predict chromosome types on test set of dataset 1 and dataset 2. In this study, the ground truth of karyotype provided by doctors are used to evaluate the results of the proposed model and then the performance is captured by two metrics, accuracy and proportion of well classified karyotype (PWCK).

A. Accuracy

Accuracy refers to the proportion of correctly classified chromosomes in all chromosomes. This metric is widely used in classification tasks since it can evaluate the performance efficiently and gives clear evaluations. As mentioned before, we evaluated 2061 chromosome images of testing set of dataset 1, by using this metric.

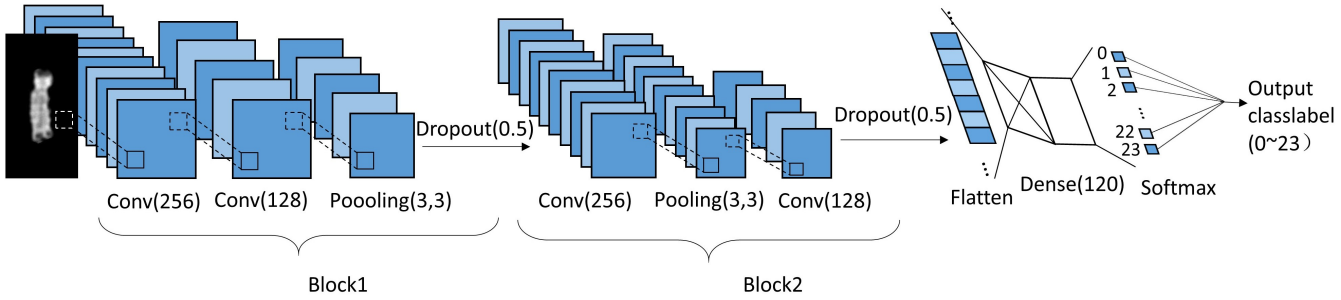

Fig. 3. CNN architecture used for chromosome classification

On this testing set, we obtained an accuracy of 92.5%. Table II shows this result.

TABLE II
PERFORMANCE OF THE PROPOSED METHOD AND OTHER THREE METHODS FROM THE LITERATURE

Author	Method	Accuracy
S. Delshadpour [4]	Multi-layer Perceptron	88.3%
Swati [1]	Siamese Network	85.6%
M. Sharma [10]	deep CNN	86.7%
Proposed	based on LeNet	92.5%

To further evaluate the performance of our proposed method, We compared our result with one previous chromosome classifying algorithms of M. Sharma [10] by using the same evaluation metrics. Since the parameters of Sharma's model might not be suitable for our data sets, we optimized the settings of data for Sharma's model. From the results of all methods shown in Tabel II, our algorithm outperformed Sharma's approaches in terms of accuracy. S. Delshadpour's method [4] and Swati's [1] approach were also listed in table II. However, their results come directly from their papers, which were obtained in their studies with their own testing datasets different to the ones used in this study.

B. Proportion of well classified karyotype (PWCK)

Doctors usually check a patient's chromosomes by taking a karyotype image as a unit. They often focus on whether the chromosomes of one karyotype image are well classified. If the accuracy achieves a certain threshold, doctors would regard the karyotype image as a qualified karyotype image, otherwise the image will be discarded. Proportion of well classified karyotype (PWCK) evaluates the proportion of acceptable karyotype classified by the proposed method. The definition of PWCK is as follows:

$$PWCK = \frac{\sum_i I(\text{Accuracy}(i) > 80\%)}{N} \tag{4}$$

where I is the indicator function,

$$I(x) = \begin{cases} 0 & x \text{ is false} \\ 1 & x \text{ is true} \end{cases} \tag{5}$$

N is the number of karyotype image, Accuracy(i) is the classification accuracy of 46 chromosomes in ith karyotyping image and its definition is as follows:

$$\text{Accuracy(i)} = \frac{\text{correct classified chromosomes}}{46} \times 100\% \tag{6}$$

The threshold of 80% in (4) was identified by two doctors in the research group who commented that the classification results on a karyotype is acceptable if over 80% of chromosomes in the karyotype were correctly classified. Here we use PWCK to evaluate dataset 2 with 105 karyotype images (4830 chromosomes). We achieved a result of 91.3%, which shows that our method is applicable in real life tasks.

V. DISCUSSION AND CONCLUSION

In this study, an automatic-classification method based on CNNs was proposed. The model extracts chromosome images from karyotype and output their classes. Compared with three other methods and deep learning algorithms, our method achieved a better accuracy. Our experiment also shows that our method is applicable in real life tasks. The results of this study showed that CNNs is useful in extracting features in terms of preprocessed medical images. To investigate if our proposed method is actually applicable and acceptable to doctors, we proposed a new metric PWCK, which is closely related to the medical application, since doctors focus on the accuracy of individual karyotype images. The results suggest that this proposed automatic classification method can be utilized in chromosome classification, and can potentially help doctors to save a lot of time.

Furthermore, in the process of obtaining PWCK, we found the average accuracy was not as good enough as we expected, because there was a small discrepancy between this value and the overall accuracy (92.5%). This difference might be caused by the different construction of datasets, since dataset 1 and dataset 2 we used in this task contained different human chromosomes. Since the model was trained by part of dataset 1, the model was familiar with chromosome images from the people who provided training data. Therefore, our model need to consider the inner variance of chromosomes of different

people. A data set containing chromosomes from more people would improve the ability of generalization of our model.

In this study, we only focus on classifying vertical chromosomes. In other situations, some chromosomes have different orientations, which should be considered in our future work.

ACKNOWLEDGMENT

This study is supported by the National Natural Science Foundation of China (Grant No. 61501380), Key Program Special Fund in XJTLU (KSF), Sando Medical Laboratories, Inc and the Open Program of Neusoft Corporation, Item number SKLSAOP1702.

REFERENCES

[1] Swati, G. Gupta, M. Yadav, M. Sharma, and L. Vig, "Siamese networks for chromosome classification," in *IEEE International Conference on Computer Vision Workshop*, 2018, pp. 72–81.

[2] J. M. Cho, "Chromosome classification using backpropagation neural networks," *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, no. 1, pp. 28–33, 2000.

[3] J. Cho, S. Y. Ryu, and S. H. Woo, "A study for the hierarchical artificial neural network model for giemsa-stained human chromosome classification," in *International Conference of the IEEE Engineering in Medicine & Biology Society*, vol. 2, 2004, pp. 4588–4591.

[4] S. Delshadpour, "Reduced size multi layer perceptron neural network for human chromosome classification," in *International Conference of the IEEE Engineering in Medicine & Biology Society*, vol. 3, 2003, pp. 2249–2252.

[5] B. C. Oskouei and J. Shanbehzadeh, "Chromosome classification based on wavelet neural network," in *International Conference on Digital Image Computing: Techniques and Applications*, 2010, pp. 605–610.

[6] S. Gagulapalalic and M. Can, "Human chromosome classification using competitive neural network teams (cnnt) and nearest neighbor," in *Ieee-Embs International Conference on Biomedical and Health Informatics*, 2014, pp. 626–629.

[7] M. J. Roshtkhari and S. K. Setarehdan, "Linear discriminant analysis of the wavelet domain features for automatic classification of human chromosomes," in *2008 9th International Conference on Signal Processing*, Oct 2008, pp. 849–852.

[8] Q. Wu, Z. Liu, T. Chen, Z. Xiong, and K. R. Castleman, "Subspace-based prototyping and classification of chromosome images," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1277–1287, Sept 2005.

[9] Y. Qiu, X. Lu, S. Yan, M. Tan, S. Cheng, S. Li, H. Liu, and B. Zheng, "Applying deep learning technology to automatically identify metaphase chromosomes using scanning microscopic images: an initial investigation," *Proceedings of SPIE*, vol. 9709, 2016.

[10] M. Sharma, O. Saha, A. Sriraman, R. Hebbalaguppe, L. Vig, and S. Karande, "Crowdsourcing for chromosome segmentation and deep classification," in *Computer Vision and Pattern Recognition Workshops*, 2017, pp. 786–793.

[11] J. Blahuta, T. Soukup, and P. ermk, "Image processing of medical diagnostic neurosonographical images in matlab," in *Wseas International Conference on Computers*, 2011, pp. 85–90.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International conference on learning representations*, 2015.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Computer vision and pattern recognition*, pp. 770–778, 2016.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Computer vision and pattern recognition*, pp. 1–9, 2015.

[15] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: a convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997.

[16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[17] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[18] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout:a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[19] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computer Science*, 2014.