IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 February 2023, accepted 6 March 2023, date of publication 14 March 2023, date of current version 3 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3257045

RESEARCH ARTICLE

Feature Fusion Classifier With Dynamic Weights for Abnormality Detection of Amniotic Fluid Cell Chromosome

WENBIN HE[1], ZEYU SHI[1], YINXIA LIU[2], TING LIU[1], JINGUANG DU[1], JUN MA[1], YANG CAO[1], AND WUYI MING[3]

[1]Henan Key Laboratory of Intelligent Manufacturing of Mechanical Equipment, Henan Engineering Technology Research Center for Green Manufactory and Precision Measurement, Zhengzhou University of Light Industry, Zhengzhou 450002, China
[2]Laboratory Medicine of Dongguan Kanghua Hospital, Dongguan 523808, China
[3]Guangdong Provincial Key Laboratory of Manufacturing Equipment Digitization, Guangdong HUST Industrial Technology Research Institute, Dongguan 523808, China

Corresponding authors: Yinxia Liu (liuyinxia-511@163.com) and Wuyi Ming (mingwuyi@gmail.com)

This work was supported in part by the Key Research and Development Projects of Henan Province in 2022 under Grant 221111240200, in part by the Dongguan Key Project of Social Development Technology in 2022 under Grant 20221800905062, in part by the Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program under Grant 2017BT01G167, and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515140066.

ABSTRACT Chromosomal karyotype is important to determine whether a newborn has a genetic disorder. There are two main categories of chromosomal abnormalities: structural abnormalities, in which the chromosome structure is altered, and chromosome number abnormalities. Manual karyotyping is complex and takes a lot of time because it requires a high degree of domain expertise. Based on this investigation, we propose a new method of chromosome defect detection based on deep learning with 20,299 chromosome images from Dongguan Kanghua Hospital as data that integrates the diversity of chromosome features and trains a classifier model based on feature fusion for chromosome abnormality detection. We put forward a feature fusion classifier with dynamic weights (FFCDW) for chromosomal abnormality detection, after data augmentation with three deep learning networks, ResNet, SENet, and VGG19, the three trained models are combined using a dynamic weighting approach. Experiments prove the FFCDW method outperforms these mainstream models of ResNet, SENet, and VGG19. The proposed method based on FFCDW achieves a precision of 0.8902 and an F1-score of 0.8805 with a small standard deviation (0.00903 and 0.00892, respectively). In addition, the algorithm can automatically assign weights based on the results of a single model, and the strategy with dynamic weights outperforms the strategy with fixed weights in the proposed feature fusion classifier.

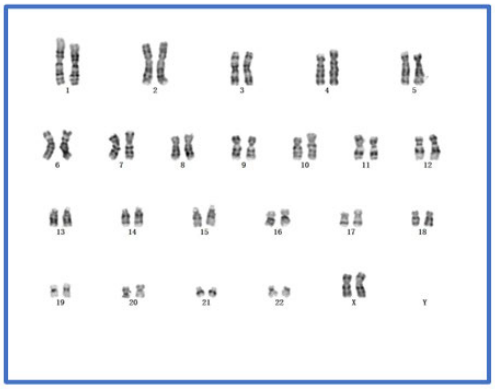
INDEX TERMS Chromosome karyotype analysis, deep learning, machine vision, dynamic weighting, feature fusion classifier, model fusion.

I. INTRODUCTION

Chromosomes are the carriers of human genetic material (DNA), and there are 23 pairs (46 chromosomes) in human somatic cells, 22 pairs of which are autosomes unrelated to sex, and the remaining two (1 pair) are sex chromosomes, of which the female sex chromosomes appear in pairs as XX and the male as XY [1]. Chromosomal abnormalities include numerical and structural abnormalities; numerical abnormalities are an increase or decrease in the number of chromosomes, which is the main form of chromosomal abnormalities [2], and structural abnormalities are due to the breakage and reunion of chromosome fragments. Karyotyping is an important screening and diagnostic procedure for detecting several genetic disorders or chromosomal abnormalities (e.g., Edwards syndrome, Turner syndrome, and Down syndrome) [3], [4] and other genetic disorders as an important clinical procedure [5]. It is obtained from

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

(a) Original image



(b) Image after processing

FIGURE 1. Chromosome slide diagram and images of well-sorted chromosomes.

stained intermediate chromosomes by using staining techniques [6]. The karyotype analysis method analyzes microscopic images of chromosomes in intermediate divisions by banding techniques and then diagnoses diseases based on changes in the structure and number of chromosomes [7]. Typical micrographs of chromosomes are shown in Fig. 1a. Correspondingly, Fig. 1b shows the karyotype map obtained by extraction and classification, which can be used as a basis for disease diagnosis.

A spectral karyotyping method is a new approach for the easy identification of different chromosomes, but it also has the disadvantages of high costs and complex experiments. Most automated karyotyping systems currently in use provide a graphical environment with basic segmentation operations and manual chromosome classification [8], [9]. Interactive systems are time-consuming, laborious, and dependent on specialized technicians. In the case of hematological tumors, for example, to ensure diagnostic accuracy and comprehensiveness, technicians need to select at least 30 mid-term images suitable for analysis and then segment the chromosomes in each image. Training professionals who can independently and effectively perform karyotyping analysis takes more than two years. This leads to the fact that in smaller hospitals, karyotyping cannot be performed effectively due to a lack of professionals, while in specialty hospitals, the pressure to analyze is increasing as more patients go there for an accurate diagnosis.

To alleviate the burden of karyotype analysis, many automated classification methods have been developed for the analysis of intermediate chromosomes [10], [11], [12]. With the development of research in recent years, deep learning has shown great research potential in medical image processing because of its ability to extract and process complex features in these images [11], [13]. Recent studies have replaced the traditional feature extraction approach with hand-crafted features and classification, with encouraging results. A convolutional neural network (CNN) is one of the network structures for deep learning. It is an important tool for image classification [14] and object detection tasks [15]. It is just like the human brain perceives the world visually. It consists of neurons that are basic computational units and are activated by specific signals. Layers are stacked into neurons, and a series of these layers form the CNN [16], [17], [18]. The CNN is mainly composed of the following types of layers: convolutional layers are responsible for feature extraction; activation layers are functions that determine whether a neuron will output; the fully connected layer integrates the features extracted from the convolutional layer, which connects each neuron in the current layer to each neuron in the next layer. Finally, the classification layer selects the most likely classes [19], [20].

Detecting chromosomal abnormalities is an important part of determining whether this study can be used for practical detection. Various methods have been tried for structural and numerical abnormalities of chromosomes. Wang et al. [21] combined chromosome size and other information to identify 22 pairs of chromosomes, and then applied the template matching method for normal/abnormal classification. However, the data just include 30 positive and 30 negative cases of bone cancer, which has poor generalization ability, and the template matching method has poor robustness; Legeand et al. [22] combined different karyotypes to specify reference chromosome density profiles and then used dynamic time warping (DTW) to identify chromosome density profiles for translocation and recombination sites. He et al. [23] used a machine learning-based random forest algorithm to predict trisomy 21 syndrome and achieved an accuracy of 85.2%; Mona et al. [24] used YOLOv2 to detect individual chromosomes on mid-term karyotype images and adapted the VGG19 network using two different methods. The final detection of common chromosome number abnormalities (13, 18, 21, X) obtained an accuracy of 96.67%, but it cannot detect chromosomal structural abnormalities. Yang et al. [25] utilized the deep convolutional neural network (DCNN) model to classify 2,424 normal chromosomes and 544 abnormal chromosome, including 24 normal chromosomes (autosomal 1-22 and sex X or Y) and 8 abnormal chromosomes, with a classification accuracy of 87.76%. Nimitha et al. [26] used VGG16 for transfer learning, and the classification accuracy of chromosome number abnormalities was 95.5%. Ezhumalai et al. [27] used the

DCNN model to distinguish five chromosome number abnormalities: trisomy 13 syndrome, trisomy 18 syndrome, trisomy 21 syndrome, trisomy XXY syndrome, and X chromosome. The precision and F1-score of the model are 98.65% and 98.86%, respectively.

However, there are still many problems with the current method of chromosome abnormality detection. Firstly, most of the current studies on chromosome image processing focus on the chromosome segmentation part, and there are fewer studies on chromosome abnormalities, and the existing research mainly focuses on the identification of chromosome number abnormalities, while the research on chromosome structure abnormalities is less, especially the detection research based on deep learning. On the other hand, abnormal chromosome discrimination needs to identify chromosomal multi-morphological information. However, the learning ability of a single model is limited, and the actual effect is poor. Based on this, this paper tries to put forward a feature fusion classifier with dynamic weights based on the multi-model fusion method (FFCDW) to obtain better results through multi-level feature learning.

In this study, the main structure of the paper is the following: In Section I, we briefly introduce the significance and status of research on chromosomal defect detection. In Section II, the experimental data are presented in detail, and the principle of methods for defect detection of chromosomes is given. The equipment for defect detection at the chromosomal level and the evaluation criteria for this problem are described in Section III. As shown in Section IV, single model training optimization and performance comparisons of combinatorial classifiers are given. The discussion and conclusions are given in Sections IV and V, respectively.

II. METHOD AND DATASET

A. DATASET

The data in this investigation came from Dongguan Kanghua Hospital, and the images are all accurately classified autosomal 1-22, X, and Y sex chromosomes. All data are from samples taken by a microscope, and chromosome types were confirmed by professional doctors. Due to the confidentiality agreement related to the data used in this paper, the detailed information cannot be made public for the time being.

Figs. 2a-2d depict images of abnormal chromosomes in structure and abnormal chromosomes in number, respectively. The dataset is the chromosome images completed by professional doctors, including 18,088 normal chromosome samples, 1,740 abnormal chromosomes in structure samples, and 471 abnormal chromosomes in number samples, totaling 20,299, with details in Table 1. The dataset shows serious imbalance, so we used horizontal flipping, rotation, adding noise, etc. to augment the data for structural abnormality and numerical abnormality chromosomes, and the composition of the dataset is shown in Fig. 3.

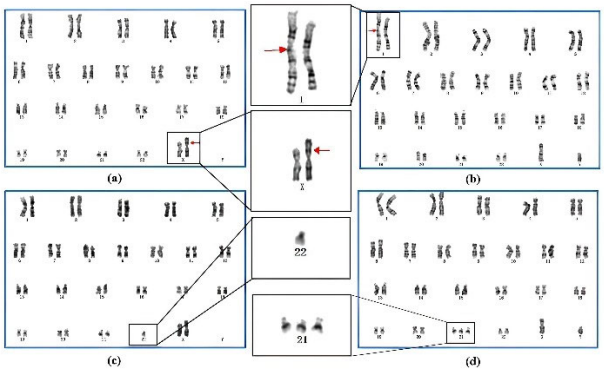


FIGURE 2. Chromosome defect detection data set; (a) Abnormal structure of chromosome X; (b) Abnormal structure of chromosome; (c) Abnormal number of chromosome 22; (d) Abnormal number of chromosome 21.

TABLE 1. Raw data statistics of chromosome database.

Category	Number
Normal	18,088
Structural Abnormality	1,740
Numerical Abnormality	471

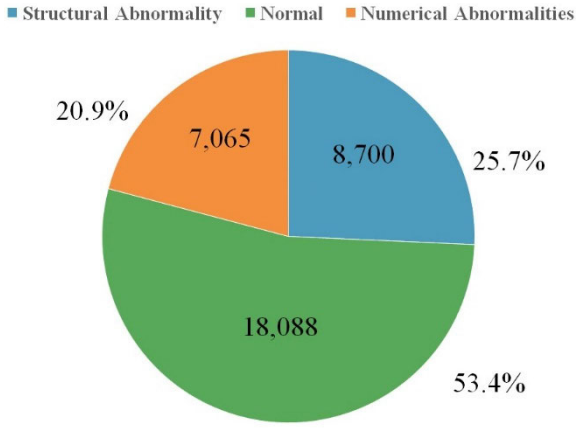


FIGURE 3. Composition of the augmented dataset of chromosome database.

B. CLASSIFIER

Fig. 4 depicts the architecture of the proposed feature fusion classifier for chromosome abnormality detection. First, due to the unbalanced dataset, data augmentation is used to make the number of images in each category roughly equal to narrow the gap. Then, the data-augmented dataset is imported into three different deep learning models to train three different classifiers, ResNet50, SENet, and VGG19, respectively. The trained three different deep learning classifiers are then combined into the FFCDW using a dynamic weights algorithm, which is fully described in Section II-B2, and finally the FFCDW is obtained and tested for the desired accuracy.

1) DEEP LEARNING CLASSIFIER

Unlike traditional image processing methods, the neural network method does not require manual feature extraction

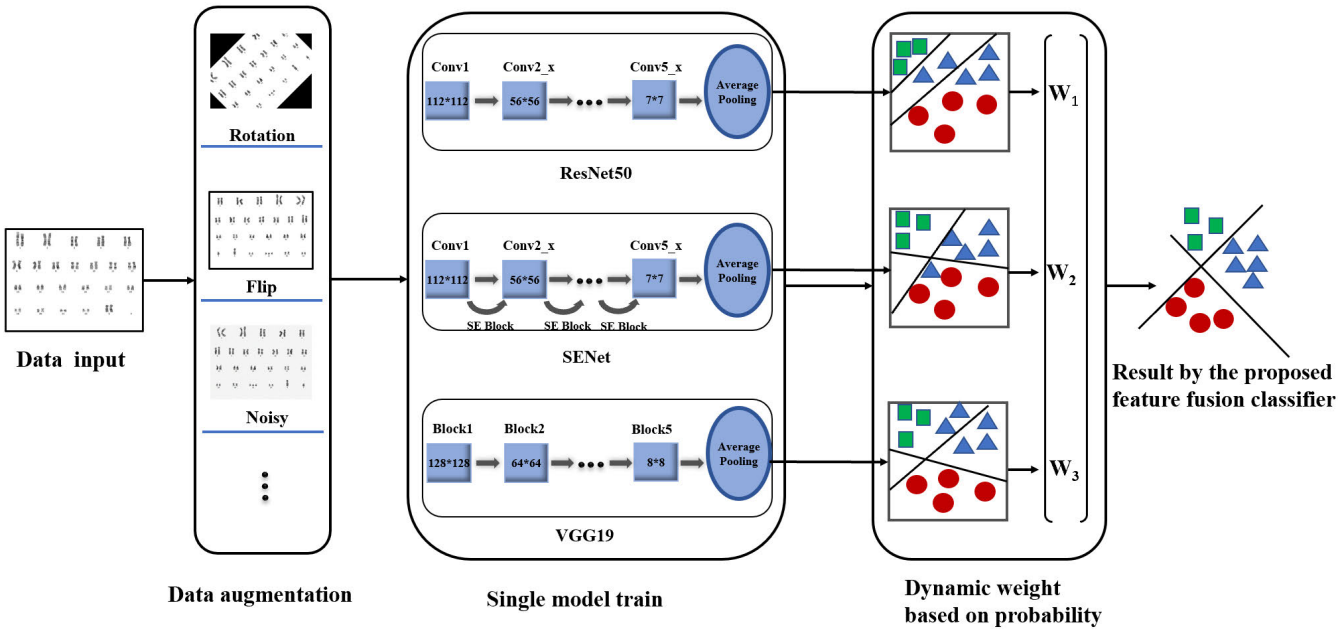


FIGURE 4. Architecture of FFCDW for chromosome abnormality detection.

of images but can automatically perform feature extraction, which is robust and applicable to more fields. In addition, through the training of many images of the same type, deep learning classifiers can better identify minute features that cannot be handled by traditional image processing methods, so deep learning methods are more suitable for medical image processing. Since the successful application of neural networks to handwriting recognition [28], they, especially CNN, have been widely used in the field of image recognition [15]. The three different CNN models chosen in this paper (ResNet, SENet, and VGG) are all widely used classification networks with their advantages and disadvantages.

The ResNet network model [29], which borrows the idea of highway networks controlling information flow through gating, changes the flow of feature information, and the output of the L layer no longer affects the output of the $L+1$ layer singularly but also affects the output of the $L+2$ layer. Thereafter, every two layers can form a residual learning block, and the block changes the learning goal in disguise. The entire ResNet network consists of multiple stacks of residual blocks interspersed with pooling layers. The training process only needs to learn the difference between input and output, which protects information integrity and simplifies the learning goal and difficulty. The ResNet network solves the gradient vanishing problem encountered by previous deep learning networks that simply deepened the network structure. However, most of its network structures prevent model degradation, which leads to errors. The mathematical expression of ResNet is as follows: where $F(x)$ is the network mapping before summation and $H(x)$ is the network mapping from input to summation.

$$H(x) = F(x) + x \tag{1}$$

The core idea of SENet [30] is to train the model to achieve better results using a network that learns feature weights according to the loss so that effective feature maps are weighted heavily and ineffective feature maps are weighted less. SENet is very simple to construct and can be easily deployed without introducing new functions or layers. It mainly consists of two operations: squeeze and excitation. Squeeze is a global average pooling, and excitation is shown as follow: This operation is designed to fully capture channel dependencies, z is the result of squeeze, $W_1 z$ is a fully connected operation, same as W_2. Finally, s is obtained by signature.

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)) \tag{2}$$

The VGG network [31] is a deeper and wider evolution of the AlexNet network, which locally replaces a large convolutional kernel with a tandem stack of more small convolutional kernels. The VGG network demonstrates that a deeper network can extract richer feature information to obtain better results. Its advantage is that the deeper layers make the feature map wider and more suitable for large datasets. Its deeper network structure and smaller convolutional kernel can both ensure the perceptual field of view and reduce the parameters of convolutional layers, which are more capable of feature learning. As a classical network, the VGG network still maintains excellent accuracy and is widely used today.

The three classifiers are selected in this investigation not only because they have the most stable performance among the current classification models, but also because they can learn features from different perspectives. The VGG network can learn tiny features by increasing the learning depth, while the ResNet network can constantly correct the learning errors

of the deep network by adding the residual mechanism so that it can learn more global features. Compared with ResNet, SENet adds an attention mechanism that can pay attention to local features of important parts while learning the global features. These three models can learn different levels of features, which can increase the diversity of the combinatorial classifier and improve its performance.

2) FEATURE FUSION CLASSIFIER

Deep learning methods are trained using a single model for tuning, which often leads to a single extraction capability and does not extract features at different levels of the image at multiple scales. The weight of each classifier is dynamically modifiable thanks to the dynamic weight technique used in this work. On the adjustment foundation, the weight is specifically distributed equally by the accuracy of various classifiers. Previous studies by Shipp et al. [32] and Zhou [33] have also shown that the same sample has different classification results for different classifiers, especially those that are prone to problems because the performance of each classifier is different. This phenomenon suggests that different classifiers have complementary information about each other, and combining them often gives better results [34], [35].

The three models, ResNet, SENet, and VGG, were mainly selected in this study for two reasons. First, the three models show the best accuracy performance, and the accuracy of other classifiers, such as DenseNet and EfficientNet, is 0.6–0.7 in the single model training, which fails to meet the requirements. The second point is the consideration of practical application. The purpose of this study is to explore the effectiveness of the feature fusion method, and to reduce the calculation time and training cost, these three models are finally selected as the basic models of feature fusion.

Dynamic weight refers to weighting the model based on the actual impact of the prediction result of each model on the final output result during the training process to improve the robustness of the model. It must be pointed out that the dynamic weight proposed in this study is not the weight transmitted in each layer connection of CNN, but the weight of the whole model in the classifier. Assuming that the ensemble-based classifier contains T basic h_1, h_2, … h_T, the $h_i(x)$ represents the output of base classifier h_i. The strategy of FFCDW used in this study is dynamic weights, as shown in the following equation:

$$H(x) = \sum_{i=1}^{T} W_i h_i(x) \tag{3}$$

In the above equation 3, W_i is the weight of $h_i(x)$, and $H(x)$ is the output of the FFCDW. Usually, W_i is greater than zero, and it meets the below equation.

$$\sum_{i=1}^{T} W_i = 1 \tag{4}$$

III. EXPERIMENTS

A. TOOLS

Our implementation code is built with the Google Open Deep Learning Cloud Server colab, developed using Python and PyTorch framework. The selection of initial parameters is crucial to obtaining the best classifier performance. Table 2 below lists the initial learning parameters of three deep learning models, and the main parameters in deep learning are learning rate and number of epochs. Dynamic weights are directly implemented by a soft voting algorithm, in which three separate deep learning models are trained, and the three models are directly weighted according to the comparison of their training accuracy to generate weights to achieve dynamic weights.

B. EVALUATION STANDARDS

Accuracy, recall, and F1-Score are usually used as evaluation metrics for machine learning algorithms. The higher the prediction rate and F1-score, the better the classifier. There are four metrics, namely true positive (TP), false positive (FP), true negative (TN), and false negative (FN). In this study, TP (FP) denotes the number of sub-images classified as abnormal and matching truth labels (and not matching truth labels); TN (FN) denotes the number of sub-images classified as normal and matching truth labels (and not matching truth labels). Thus, the performance evaluation criteria of the classifier are listed as follows:

$$precision = \frac{TP}{TP + FP} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} = \frac{2TP}{2TP + FP + FN} \tag{7}$$

IV. RESULTS AND ANALYSIS

A. SINGLE MODEL TRAINING

To achieve the best performance of the feature fusion classifier, it is important to train a single model. It is necessary to set the model parameters. Set input image size, batch size, momentum, and weight decay to 224 × 224 pixels, 64, 0.9, and 0.0001, respectively. To improve the accuracy and training speed of a single classifier for chromosomal abnormality detection, a combination of migration learning, and distribution training is used to train the single classifier.

Firstly, we load the initial weights set by each model, which are the initial weights retained by the model when training the ImageNet image set. Secondly, use the transfer learning method to quickly obtain the basic recognition ability of the model. Finally, use the distributed training method to conduct multiple training sessions based on the pre-training. It can be drawn from Fig. 5a that all three models can achieve a relatively good accuracy after 10 epochs of learning by migration, and the models converge to a relatively stable result between 30 and 40 epochs. This confirms that 50 epochs of learning are enough for these three models. The final single model accuracy of these three models is shown in Fig. 5b. After eight iterations of stepwise learning, the accuracy of the three

TABLE 2. Main parameters for single model training.

No.	Classifier	Initial value of parameters
1	ResNet50	
2	SENet	$Lr = 0.001$; epoch = 50
3	VGG19	

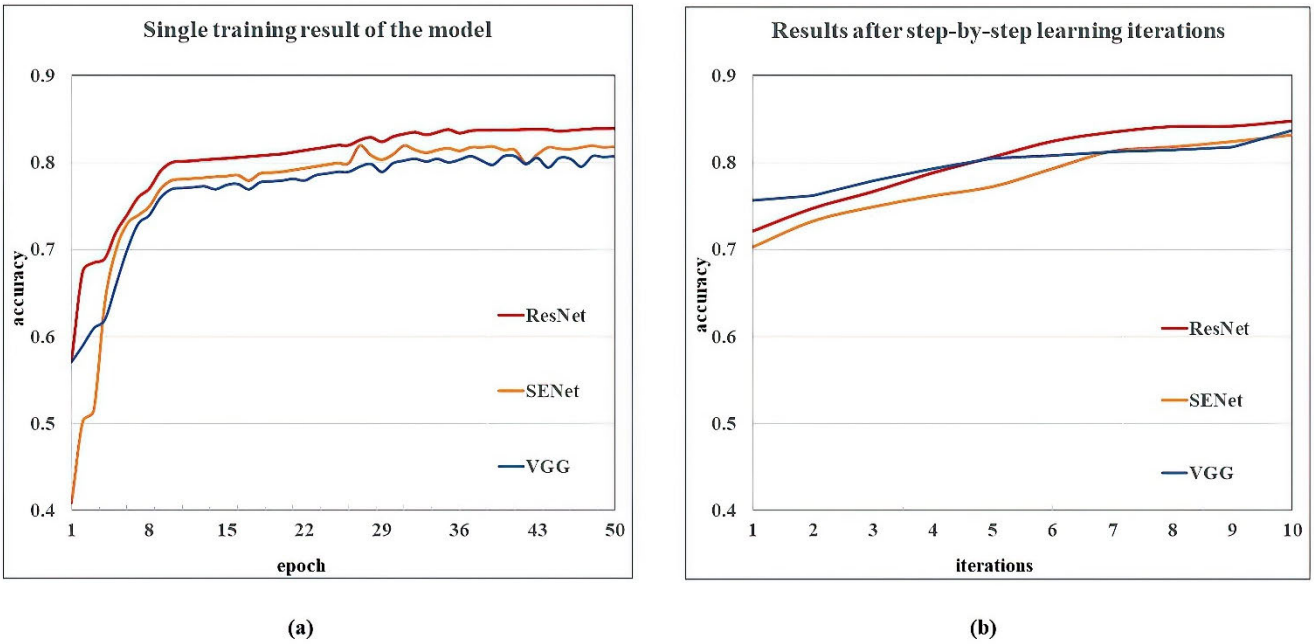


FIGURE 5. Single model training results; (a) Single training result of the model; (b) Results after step-by-step learning iterations.

models has stabilized, and we finally selected the result of the tenth iteration as the final model accuracy, with the final mean model accuracy of 0.8370 for ResNet, 0.8086 for VGG, and 0.818 for SENet. The parameter weights trained from the final accuracy were retained as the model weights for the feature fusion classifier.

B. IMPACT OF SINGLE MODEL PRECISION IMPROVEMENT

When the model continues to learn, using the deep learning model to learn will provide its model recognition ability. The current widely used learning method is to have deep learning load the pre-training weights of the model first and then improve the classification accuracy of the target data set after the pre-training model learns for the target data set. Moreover, because the deep learning algorithm uses dynamic weight decay to ensure the stability of the model and prevent overfitting, the learning accuracy of a single model will continue to improve with relearning.

To verify the effect of single model accuracy on the accuracy of the proposed feature fusion classifier, three different training weights were selected in the process of optimizing the single model, and they were arbitrarily combined to form the classifier. From Table 3, the single model accuracy affects the performance of the feature fusion classifier, and when the ResNet model accuracy is improved from 0.8012 to 0.8591, the accuracy of feature fusion classifier is also improved to varying degrees. And it can be found that the improvement of the feature fusion classifier is influenced to the greatest extent by the single model with the highest accuracy. When the ResNet model improves the accuracy, the combined most obvious change in the accuracy of the classifier is found in the ResNet model, while the improvement of the feature fusion classifier is not obvious in the SENet and VGG models when the accuracy is improved. This result may be because more accurate models have higher weights.

C. THE PERFORMANCE OF THE FFCDW

The discriminant model is tested on the test sets of ResNet50, SENet, and VGG19, and the fusion model based on the soft voting method, respectively. To evaluate the performance of FFCDW, comparative experiments using different datasets are required. In this study, the entire dataset is divided into a training set, a validation set, and a test set in the ratio of 8:1:1,

TABLE 3. Effect of different accuracy of a single model on the accuracy of the feature fusion classifier.

Method	Different combinations of single-model precision								
	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9
ResNet	0.8012	0.8326	0.8591	0.8012	0.8326	0.8591	0.8591	0.8326	0.8591
SENet	0.7744	0.7744	0.7744	0.7994	0.7994	0.7994	0.8127	0.8127	0.8127
VGG	0.7561	0.7561	0.7561	0.7786	0.7786	0.7786	0.7988	0.7998	0.7988
FFCDW	0.8212	0.8458	0.8723	0.8374	0.8682	0.8726	0.8792	0.8853	0.8913

TABLE 4. Comparison prediction accuracy of classifier performances for detection of chromosomal abnormality detection.

Method	Experiment No.1			Experiment No.2			Experiment No.3			Mean	Standard deviation
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd		
ResNet	0.8326	0.8568	0.8154	0.8418	0.8219	0.8315	0.8591	0.8413	0.8329	0.8370	0.01367
SENet	0.8118	0.8250	0.8132	0.8325	0.8315	0.8100	0.8010	0.8130	0.8240	0.8180	0.01009
VGG	0.8063	0.8045	0.8049	0.8127	0.8375	0.8150	0.7926	0.8023	0.8015	0.8086	0.01190
FFCFW	0.8539	0.8698	0.8317	0.8618	0.8496	0.8524	0.8719	0.8615	0.8532	0.8562	0.01137
FFCDW	0.8894	0.8917	0.8703	0.9012	0.8831	0.8914	0.9027	0.8924	0.8894	0.8902	0.00903

TABLE 5. Comparison F1-score of classifier performances for detection of chromosomal abnormality detection.

Method	Experiment No.1			Experiment No.2			Experiment No.3			Mean	Standard deviation
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd		
ResNet	0.8228	0.8459	0.8093	0.8317	0.8143	0.8209	0.8453	0.8307	0.8214	0.8269	0.01197
SENet	0.8095	0.8030	0.8102	0.8232	0.8152	0.7998	0.7983	0.8028	0.8168	0.8088	0.00799
VGG	0.8052	0.7924	0.8010	0.8091	0.8287	0.8102	0.7902	0.7983	0.7981	0.8037	0.01094
FFCFW	0.8451	0.8607	0.8214	0.8508	0.8354	0.8419	0.8591	0.8488	0.8397	0.8448	0.01145
FFCDW	0.8803	0.8854	0.8612	0.8946	0.8741	0.8807	0.8893	0.8791	0.8801	0.8805	0.00892

and it is divided into different datasets using a random division method, with the number of different datasets being 3. To obtain the statistical performance of these classifiers, the experiments are repeated three times on each different dataset. To verify the performance advantage of FFCDW, a feature fusion classifier with fixed weights (FFCFW) was set up, and the weights of the three single models were fixed uniformly at 1/3.

As can be seen in Tables 4 and 5, the test means of classifiers with fixed weights and those with dynamic weights significantly outperformed the prediction results of the single model. The mean prediction accuracy of the three single models ranged from 0.80-0.84, with the worst, mean value being the VGG classifier with a score of 0.8086 and the best mean value being the ResNet classifier with a score of 0.8370. The mean F1-score of the three deep learning models ranged from 0.80-0.83, with the worst being the VGG classifier with a score of 0.8037. Therefore, the overall effect of the VGG model is unsatisfactory, and it has the lowest weight in the feature fusion classifier. In addition, the performance comparison of the classifiers on three different datasets is shown in Fig. 6. As can be seen in Figs. 6a-6c, similar trends can be derived from these three different datasets. In addition, the standard deviations of the fusion model test results are all around 0.01, which proves the robustness of the proposed method.

To balance the performance of classifiers in terms of mean and standard deviation, two feature fusion classifiers (fixed weights and dynamic weights) are proposed in this paper. As shown in Fig. 6d, the mean prediction accuracy rate of the feature fusion classifier is significantly better than the single deep learning model effect. Although the standard deviation of the feature fusion classifier is slightly higher than that of the single classifier, the standard deviations are small, both being around 0.01. Evaluating the performance of the F1-score, a similar trend is derived from Fig. 6d. In addition,

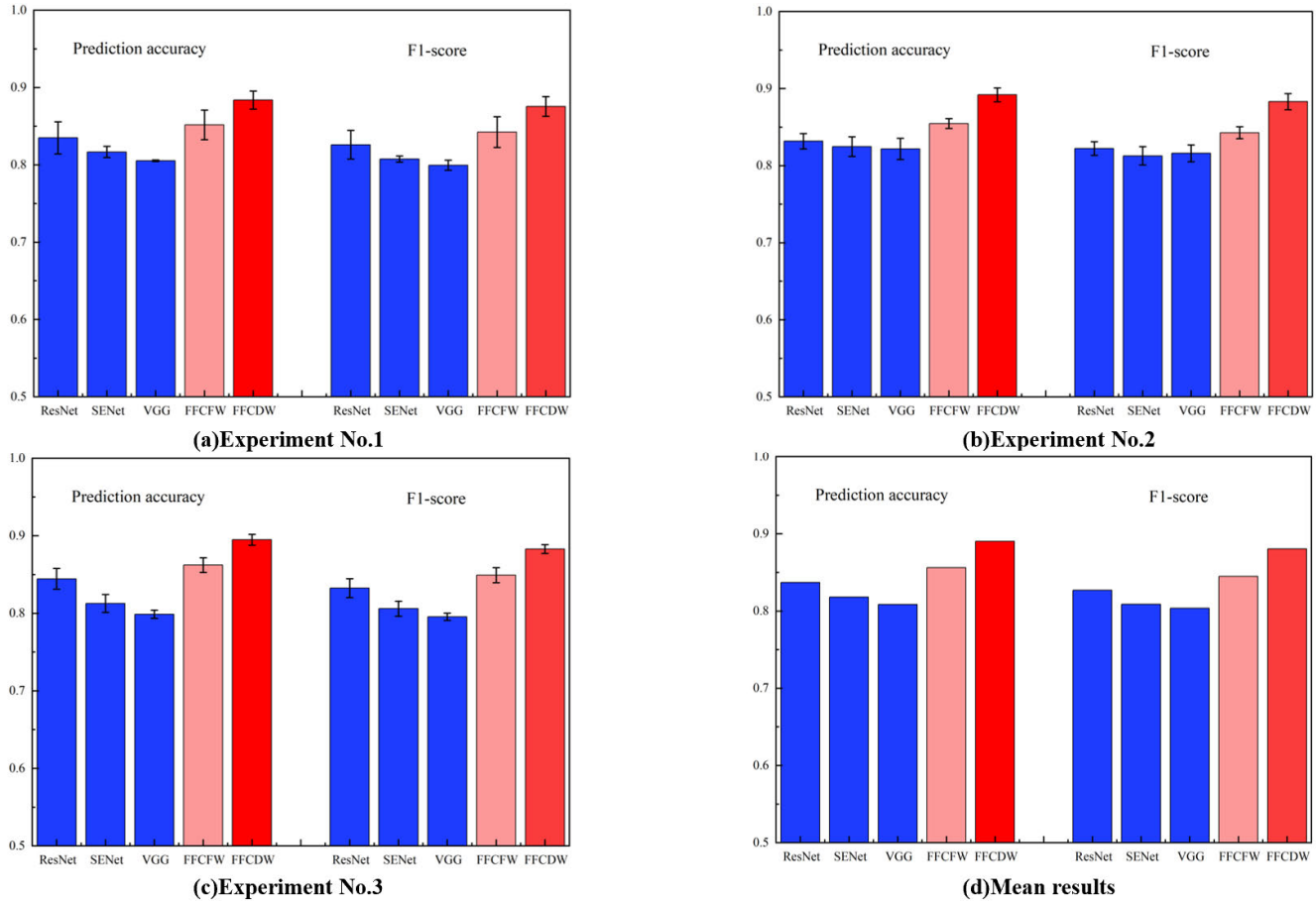


FIGURE 6. Comparison of classifier performances for detection of chromosomal abnormality detection; (a) Comparison prediction accuracy and F1-score of experiment No.1; (b) Comparison prediction accuracy and F1-score of experiment No.2; (c) Comparison prediction accuracy and F1-score of experiment No.3; (d) Comparison prediction accuracy and F1-score of mean results.

the performance of the FFCFW was slightly lower than that of the FFCDW. In all nine tests, the mean values of the prediction rate and F1-score of FFCDW are no less than 0.8703 and 0.8612, respectively. This indicates that our proposed method has good resistance to performance fluctuations for defect classification.

D. EFFECT OF SAMPLE SIZE ON FFCDW

To further test the stability of the proposed combinatorial classifier, the performance of the classifier was tested under different sample sizes using multiple experiments. Firstly, a single model combination was retrained based on the three different randomly divided datasets in Experiment 4.3 and then randomly selected into four collections, for 1/4 of the data volume, 1/2 of the data volume, 3/4 of the data volume, and the entire data volume, respectively. Each set of data was repeated three times and eventually averaged. Fig. 7 shows a scatter plot of the experimental results, from which in some cases the results are higher for the small sample than for the full data set. However, this is because there is a small gap in the characteristics of some data, so the training results of this part will be better.

From Table 6, it can be seen that the proposed FFCDW still has stability in the combination experiments with different data sizes. No matter what data size, the accuracy of the model can always remain around 0.89, fluctuating from 0.8868 to 0.8902 with insignificant fluctuation, which can prove that the proposed FFCDW could maintain relatively stable results under the change of data sample. This can prove that the proposed FFCDW can maintain relatively stable results under the change of data samples and has robustness.

V. DISCUSSION

A. THE EFFECT OF FEATURE FUSION CLASSIFIER WITH DYNAMIC WEIGHTS

In this study, we propose a feature fusion classifier for chromosomal abnormality detection. To verify the effectiveness of the model fusion method, two methods, FFCFW and FFCDW, are used to detect and classify chromosomal abnormalities to verify the effectiveness of the dynamic weight rule. As described in the previous section, FFCFW is a feature fusion classifier with fixed weights, and FFCDW is a feature fusion classifier with dynamic weights, and we can find that the model performance of the approach with dynamic weights

TABLE 6. Precision of FFCDW with different data volume.

Data	Experiment No.1			Experiment No.2			Experiment No.3			Mean	Standard deviation
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd		
1/4 data	0.8804	0.8913	0.8659	0.8907	0.8818	0.8913	0.8915	0.8958	0.8925	0.8868	0.00878
2/4 data	0.8814	0.8893	0.8724	0.8914	0.8972	0.8924	0.8859	0.8898	0.8853	0.8872	0.00678
3/4 data	0.8898	0.8951	0.8758	0.9063	0.8843	0.8839	0.8913	0.8892	0.8869	0.8892	0.00795
all data	0.8894	0.8917	0.8703	0.9012	0.8831	0.8914	0.9027	0.8924	0.8894	0.8902	0.00903

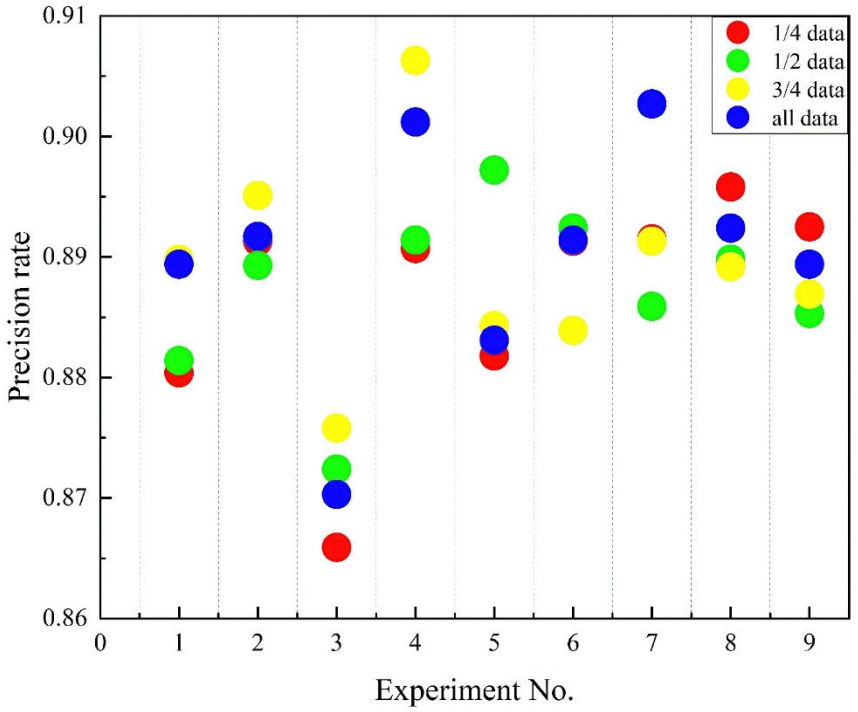


FIGURE 7. Effect of different data volumes on the performance of FFCDW.

is significantly better than that with fixed weights. When comparing our results to those of previous studies, it must be pointed out that, compared with Wang et al. [21], who used template matching to detect chromosome abnormalities, although the accuracy rate reached 93.3%, this result only compared 30 negative samples and was not robust. Compared with He et al. [23], which detected trisomy 21 with an accuracy of only 85.2%, and Al-Kharraz et al. [24], which only detected chromosome number abnormalities, the detection scope of this study is wider, the results are stable, and it is not limited to the detection of single chromosome abnormalities or only the detection of chromosome number abnormalities. Nimitha et al. [26] and Nimitha et al. [27] similarly, only abnormal chromosome number was detected, but abnormal chromosome deconstruction was not detected. Yang [25] classified eight different abnormal structural chromosomes and normal chromosomes, achieving a precision of 87.76%. Both the detection type and the precision should be significantly improved in this paper.

According to the results, it can be found that the accuracy of FFCDW is significantly better than the performance of a single model. This result ties well with previous studies by Shipp and Kuncheva [32] and Zhou [33]. The main reason is that the features learned by a deep learning model are homogeneous under the same model, and although continuous parameter optimization will improve the performance of the model, such improvement cannot change the limitations of the model architecture itself. Despite this, the integration of different methods will increase the calculation cost and processing time. The average reasoning time of a single image is 0.21s when the dynamic weights strategy is adopted, while the average reasoning time of a single image is 0.15s when the fixed weights strategy is adopted. Although the dynamic weights strategy has a disadvantage in terms of

processing time and computing cost, the improvement effect is still obvious. Its processing time is also within acceptable limits. In addition, the performance of the overall model can be improved by combining multiple models. This can prove the stability of the model itself when the size of the data set changes.

Compared with FFCFW, FFCDW has a better effect, which is mainly due to the difference in model accuracy. To test the impact of single model accuracy on the feature fusion classifier, it can be found that the model with good performance in the single model has a significantly higher impact on the final accuracy of the feature fusion classifier than the model with poor performance. It can be seen from Table 4 that when the accuracy of the ResNet model is increased by 3.89%, the accuracy of the feature fusion classifier is increased by nearly 3%, while when the accuracy of the SENet and VGG19 models is increased by 3.26% and 2.98%, the accuracy of the feature fusion classifier is only increased by 0.96%. This indicates that the accuracy of the feature fusion classifier model will be most affected by the model with the highest accuracy in the single model, and the proportion of impact itself has exceeded the difference in weight between the models. From the above discussion, it can be concluded that different precision models have different influences on the formation of the final feature fusion classifier. If FFCFW is adopted, the advantages of the high-precision model cannot be fully transferred to the final feature fusion classifier. Therefore, the dynamic weight method can better play the advantages of the high-precision model in the feature fusion classifier and finally get a better result.

B. LIMITATION

Although this study reveals important findings, there are limitations. First, this study does not use public datasets for experiments, and its results are not directly comparable with those of other studies. Second, only deep learning models were selected for experiments in this study, and no other machine learning methods or traditional image processing methods were selected for combinations. Finally, only three models were selected for combinations in this paper. Even if it is preliminary, this study can clearly demonstrate the effectiveness of the feature fusion classifier. Although this study did not use a public dataset, the dataset used in this study was not only judged and labeled by professional doctors, but was also large enough to ensure the credibility of the experimental results. This is enough to prove that the method of this study is credible.

C. APPLICATION

Based on the previous summary of the problem, an important future direction for chromosomal abnormality detection research is the feature fusion classifier, which can be further investigated by adding other machine learning models. Such as support vector machines (SVM) or template matching and other traditional image processing methods for experiments, or adding classification models such as Vision Transformer, which may further improve the feature fusion classifier's effect. The accuracy of the feature fusion classifier may also be improved when classifier models with better robustness and higher accuracy are proposed. This study is expected to provide useful feedback for chromosomal abnormality detection efforts and contribute to alleviating the pressure on medical resources.

This study is a research project in cooperation with the hospital, and its main purpose is to assist doctors in the detection of chromosome abnormalities. At present, the accuracy of this study is not enough to completely replace doctors in the detection process, but it can effectively assist doctors in the process. In a practical application, the chromosome karyotype image can be judged by the algorithm first, and then reviewed by professional doctors, which can save doctors' actual working time and improve efficiency. It is also expected that in the future, it will completely replace doctors for testing and diagnosis.

VI. CONCLUSION

In this paper, a novel computer vision system using a dynamic weight feature fusion classifier is developed for chromosome abnormality detection, and the experimental results show that the method has good performance for chromosome abnormality recognition. The main contributions and innovations of this study are as follows:

(1) Three different deep learning models are used to improve the diversity of the feature fusion classifier, with SENet containing SE blocks and ResNet containing residual blocks. In addition, this study integrates three deep models into the proposed FFCDW to improve the generalization ability of the feature fusion classifier.

(2) The proposed FFCDW outperforms the mainstream deep learning classifiers and obtains an average prediction accuracy of 89.02% and an F1-score of 0.8805 with a small standard deviation, which demonstrates its good ability to effectively identify chromosomal abnormalities.

(3) Among the proposed combinatorial classifiers, the strategy of applying dynamic weights is better than that of applying the fixed weight method, and the soft voting algorithm can automatically assign dynamic weights directly according to the accuracy of the deep learning model itself. And the model performance is very stable regardless of the sample size, which proves that the method is robust.

In general, this paper proves the feasibility of dynamic weights strategy. Compared with the single model method, the combined model has better results, and the dynamic weights strategy has more advantages than the fixed weights strategy. Despite this, this paper still has a lot of flaws. The following are potential directions for further research.

(1) It is possible to research the dynamic weight adjustment method. This work proposes a dynamic weight technique that distributes weights according to precision. Future study may focus on this area if the weight difference widens and the weight ratio of the model with the favorable effect increases.

(2) This work only examines the combination impact of three classifiers with strong performance due to computational time and cost constraints, and it validates the viability of the dynamic weight feature fusion classifier approach. To verify the effectiveness of multi-classifier models, future research can include more diverse classifiers, such as SVM and other machine learning or deep learning algorithms.

REFERENCES

[1] T. Arora and R. Dhir, "Geometric feature-based classification of segmented human chromosomes," *Int. J. Image Graph.*, vol. 20, no. 1, Jan. 2020, Art. no. 2050006.

[2] L. Shaffer and Theisen, "Disorders caused by chromosome abnormalities," *Appl. Clin. Genet.*, vol. 3, p. 159, Dec. 2010.

[3] G. Stetten, "The principles of clinical cytogenetics," *Amer. J. Hum. Genet.*, vol. 76, no. 3, p. 543, 2005.

[4] D. Patterson, "Molecular genetic analysis of down syndrome," *Hum. Genet.*, vol. 126, no. 1, pp. 195–214, Jul. 2009.

[5] P. Devika, B. B. Rao, and A. Saralaya, "A survey on automation of chromosome based genetic diagnosis using machine learning," in *Proc. IEEE 6th Int. Conf. Comput., Commun. Autom. (ICCCA)*, Dec. 2021, pp. 166–170.

[6] J. Zhang, W. Hu, S. Li, Y. Wen, Y. Bao, H. Huang, C. Xu, and D. Qian, "Chromosome classification and straightening based on an interleaved and multi-task network," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 3240–3251, Aug. 2021.

[7] L. Uzolas, J. Rico, P. Coupe, J. C. SanMiguel, and G. Cserey, "Deep anomaly generation: An image translation approach of synthesizing abnormal banded chromosome images," *IEEE Access*, vol. 10, pp. 59090–59098, 2022.

[8] D. Huber, L. V. Von Voithenberg, and G. V. Kaigala, "Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH?" *Micro Nano Eng.*, vol. 1, pp. 15–24, Nov. 2018.

[9] D. Ming and J. Tian, "Automatic pattern extraction and classification for chromosome images," *J. Infr., Millim., THz Waves*, vol. 31, no. 7, pp. 866–877, 2010.

[10] N. Madian and K. B. Jayanthi, "Analysis of human chromosome classification using centromere position," *Measurement*, vol. 47, pp. 287–295, Jan. 2014.

[11] F. Abid and L. Hamami, "A survey of neural network based automated systems for human chromosome classification," *Artif. Intell. Rev.*, vol. 49, no. 1, pp. 41–56, 2018.

[12] Y. Wu, Y. Yue, X. Tan, W. Wang, and T. Lu, "End-to-end chromosome karyotyping with data augmentation using GAN," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2456–2460.

[13] Y. Qin, J. Wen, H. Zheng, X. Huang, J. Yang, N. Song, Y.-M. Zhu, L. Wu, and G.-Z. Yang, "Varifocal-Net: A chromosome classification approach using deep convolutional networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 11, pp. 2569–2581, Nov. 2019.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.

[15] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3642–3649.

[16] M. Sharma and L. Vig, "Automatic chromosome classification using deep attention based sequence learning of chromosome bands," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.

[17] P. N. Druzhkov and V. D. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," *Pattern Recognit. Image Anal.*, vol. 26, no. 1, pp. 9–15, 2016.

[18] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[19] A. Lumini, L. Nanni, and G. Maguolo, "Deep learning for plankton and coral classification," *Appl. Comput. Inform.*, 2020.

[20] S. Hijazi, R. Kumar, and C. Rowen, *Using Convolutional Neural Networks for Image Recognition*, vol. 9. San Jose CA USA: Cadence Design Systems, 2015.

[21] X. Wang, "Automated identification of abnormal metaphase chromosome cells for the detection of chronic myeloid leukemia using microscopic images," *J. Biomed. Opt.*, vol. 15, no. 4, Jul. 2010, Art. no. 046026.

[22] B. Legrand, C. S. Chang, S. H. Ong, S. Y. Neo, and N. Palanisamy, "Automated identification of chromosome segments involved in translocations by combining spectral karyotyping and banding analysis," *IEEE Trans. Syst., Man, A, Syst. Humans*, vol. 38, no. 6, pp. 1374–1384, Nov. 2008.

[23] F. He, B. Lin, K. Mou, L. Jin, and J. Liu, "A machine learning model for the prediction of down syndrome in second trimester antenatal screening," *Clinica Chim. Acta*, vol. 521, pp. 206–211, Oct. 2021.

[24] M. S. Al-Kharraz, L. A. Elrefaei, and M. A. Fadel, "Automated system for chromosome karyotyping to recognize the most common numerical abnormalities using deep learning," *IEEE Access*, vol. 8, pp. 157727–157747, 2020.

[25] C. Yang, T. Li, Q. Dong, and Y. Zhao, "Chromosome classification via deep learning and its application to patients with structural abnormalities of chromosomes," *Available SSRN 4260020*, to be published.

[26] N. Nimitha, S. Abbiraamavallee, E. Elakiya, J. Harini, and Y. Kotishree, "Supervised chromosomal anomaly detection using VGG-16 CNN model," *AIP Conf. Proc.*, vol. 2405, no. 1, 2022, Art. no. 030010.

[27] N. Nimitha, P. Ezhumalai, and A. Chokkalingam, "An improved deep convolutional neural network architecture for chromosome abnormality detection using hybrid optimization model," *Microsc. Res. Technique*, vol. 85, no. 9, pp. 3115–3129, Sep. 2022.

[28] Y. LeCun, "Handwritten digit recognition with a back-propagation network," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2, 1989, pp. 1–9.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[32] C. A. Shipp and L. I. Kuncheva, "Relationships between combination methods and measures of diversity in combining classifiers," *Inf. Fusion*, vol. 3, no. 2, pp. 135–148, 2002.

[33] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.

[34] W. Ming, F. Shen, H. Zhang, X. Li, J. Ma, J. Du, and Y. Lu, "Defect detection of LGP based on combined classifier with dynamic weights," *Measurement*, vol. 143, pp. 211–225, Sep. 2019.

[35] W. Ming, J. Du, D. Shen, Z. Zhang, X. Li, J. R. Ma, F. Wang, and J. Ma, "Visual detection of sprouting in potatoes using ensemble-based classifier," *J. Food Process Eng.*, vol. 41, no. 3, May 2018, Art. no. e12667.

WENBIN HE was born in Henan, China, in 1970. He received the bachelor's degree from Henan Polytechnic University, in 1991, and the master's and Ph.D. degrees from Tsinghua University, in 2002 and 2014, respectively, all in engineering.

He is currently the Vice Dean of the School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry. He has successively published more than 40 academic articles in important domestic and foreign journals, such as *Journal of Mechanical Engineering*, *Forging Technology*, *International Journal of Advanced Manufacturing Technology*, and *Ceramics International*. His main research interests include digital design and manufacturing, structural analysis and optimization of heavy equipment, and artificial intelligence.

ZEYU SHI was born in Henan, China, in 1998. He is currently pursuing the master's degree with the College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry. His research interests include machine learning and deep learning.

YINXIA LIU was born in Guangdong, China, in 1984. She is the Deputy Chief Doctor. She is major in chromosome karyotype analysis, quality control process establishment, and the application testing of informationized and intelligent prenatal diagnostic management systems. She has published more than ten articles and presided over one municipal key project. She holds one authorized patent. Her research interest includes prenatal diagnostic genetics.

TING LIU was born in Henan, China, in 1999. She is currently pursuing the master's degree with the College of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry. She has published a paper "A review: The detection of cancer cells in histopathology based on machine vision" in computers in biology and medicine. Her research interests include machine learning and deep learning.

JINGUANG DU was born in China, in 1984. He received the Ph.D. degree in engineering from Harbin Institute of Technology. His main research interests include machining process and control and machine learning and technology.

JUN MA was born in China. He received the bachelor's and master's degrees from Taiyuan University of Science and Technology, in 2004, and the Ph.D. degree from Zhejiang University, in 2008, all in engineering. Currently, he is the Director of the Department of Mechanical Manufacturing and Automation, Zhengzhou University of Light Industry and has been engaged in the research of mass customization design and collaborative manufacturing of complex products for a long time.

YANG CAO was born in China. She received the degree in mechanical design and theory from Harbin Engineering University, in June 2015. She was a Teaching Model with the School of Mechanical and Electrical Engineering, Zhengzhou University of Light Industry, Henan. Her main research interests include advanced material and structure design and analysis.

WUYI MING was born in Hubei, China, in 1981. He received the bachelor's degree from Huazhong Agricultural University, in 2002, the master's degree from Zhengzhou University, in 2010, and the Ph.D. degree from Huazhong University of Science and Technology, in 2014, all in engineering. He has done the theoretical and applied research in EDM and deep learning. He is currently works with the Guangdong Provincial Key Laboratory of Manufacturing Equipment Digitization. He has published nearly 90 SCI indexed articles (more than 1000 citations, H21) in journals, such as *Journal of Cleaner Production*, *Applied Surface Science*, *Journal of Materials Processing Technology*, *Measurement*, and IEEE Access. His main research interests include digital technology and equipment for difficult-to-process materials. He won the Second Prize of the Henan (China) Science and Technology Progress Award, in 2018.