SIAMESE NETWORKS FOR CHROMOSOME CLASSIFICATION

SWATI, GAURAV GUPTA, MOHIT YADAV, MONIKA SHARMA, LOVEKESH VIG

TCS RESEARCH, NEW DELHI, INDIA

ICCV 17
International Conference on Computer Vision 2017

INTRODUCTION

Karyotyping is the process of pairing and ordering 23 pairs of human chromosomes from cell images on the basis of size, centromere position, and banding pattern.
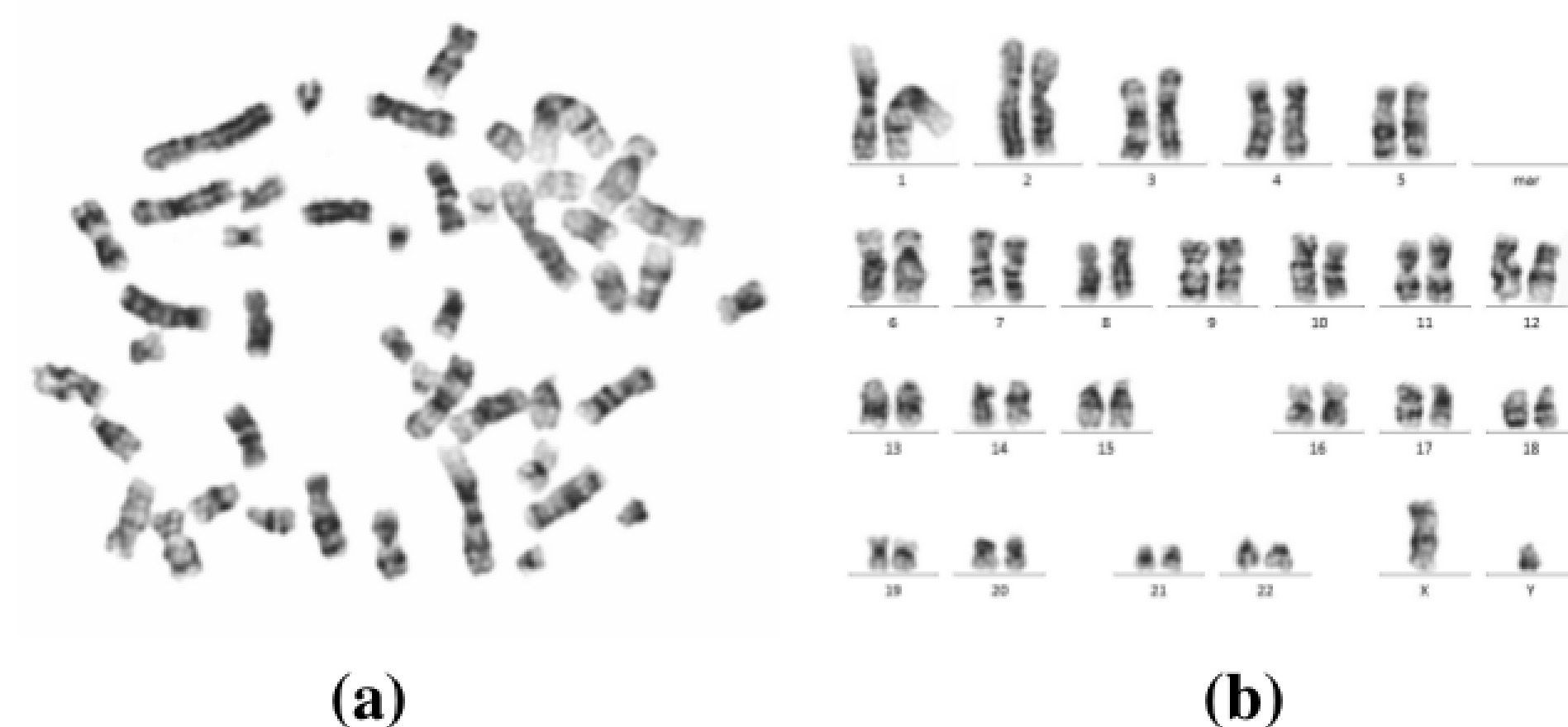

Figure 1: (a) Cell spread image during metaphase stage and (b) The standard and desired karyotyped image.

STRAIGHTENING METHODS

The following straightening methods are used for pre-processing of chromosome images:
1. SMAC: Straightening via Medial Axis extraction and Crowdsourcing [2]
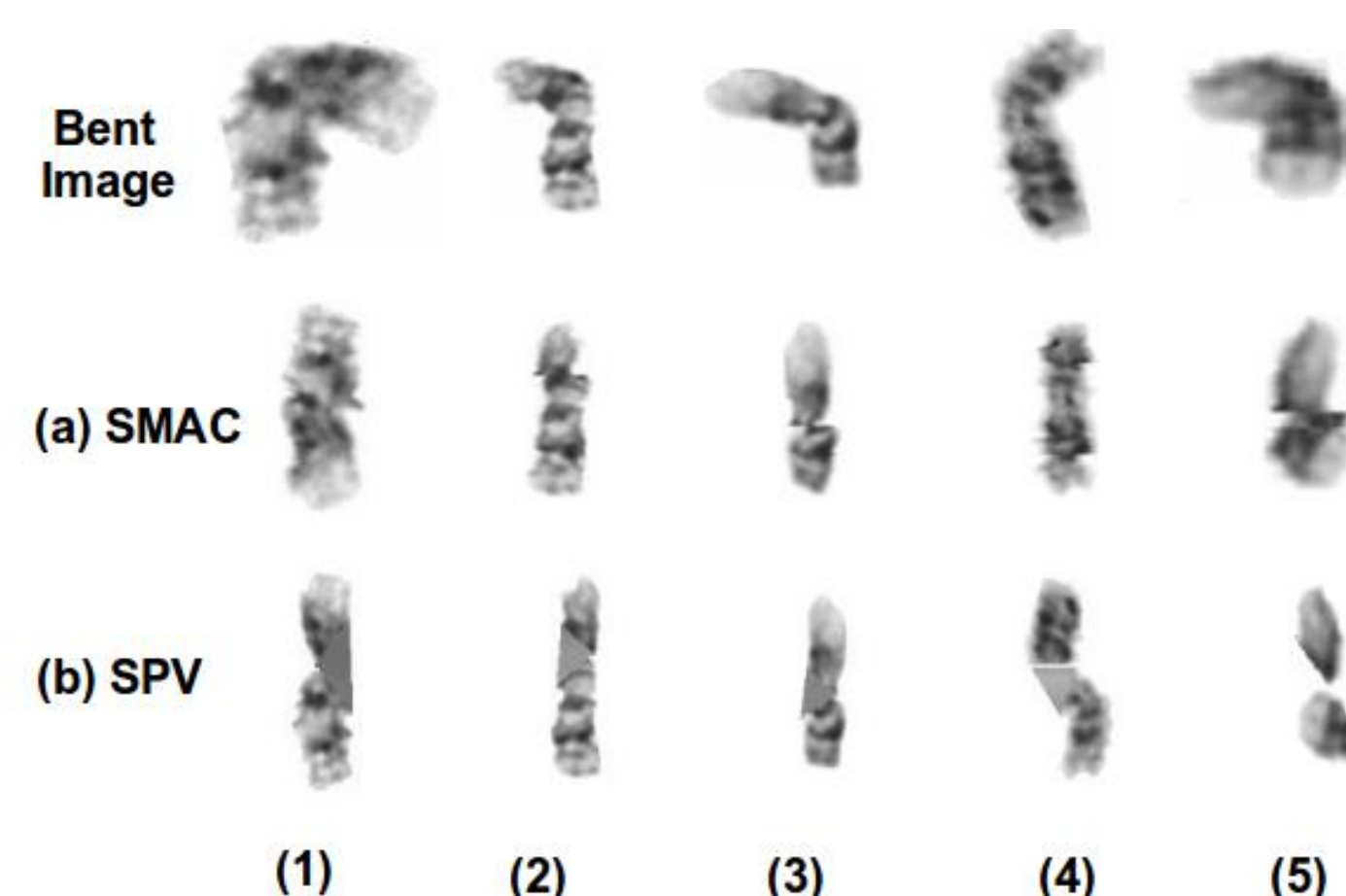2. SPV: Straightening via Projection Vectors [3]

Figure 3: Results of straightening algorithms

REFERENCES

[1] Monika Sharma, Oindrila Saha, Anand Sriraman, Lovekesh Vig, Ramya Hebbalaguppe, and Shirish Karande. Crowdsourcing for chromosome segmentation and deep classification. *CVPR*, 2017.

[2] Mehrsan Javan Roshtkhari and Seyed Kamaledin Setarehdan. A novel algorithm for straightening highly curved images of human chromosome. *Pattern Recognition Letters*, 2008.

[3] Sahar Jahani and S. Kamaledin Setarehdan. An automatic algorithm for identification and straightening images of curved human chromosomes. *Biomedical Engineering: Applications, Basis and Communications*, 2012.

[4] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, 2015.

CONTRIBUTIONS

We propose Siamese Networks for learning chromosome images similarity for the task of automatic classification with following variations :
- We employ a training method that advocates sampling from the dissimilar pairs in the Siamese Networks training set to account for the skewed distribution of similar and dissimilar pairs.
- A Multi-layer Perceptron (MLP) based feedforward network classifier is trained on the embeddings obtained from Siamese Networks.
- We explore two straightening based preprocessing methods and comparatively benchmark Siamese Networks against Deep Convolutional Neural Network [1] for the task of chromosome classification.

SIAMESE NETWORKS

Siamese networks [4] have twin neural networks (CNN in our case) with shared weights. It predicts the similarity or dissimilarity across images by computing distance between representations of the highest layer. The loss function used is contrastive loss.

Challenge I : Skewed ratio
Solution :
- Train multiple siamese models with random selection of dissimilar pairs and including all similar pairs.
- Train multiple siamese models with same number of dissimilar and similar pairs of images (balanced) from all the possible different class label pairs.

Challenge II : Prediction stage is memory expensive
Solution :
- Training Multi-layer perceptrons (MLPs) on top of all siamese base network embeddings.
- Training an MLP by concatenating the embedding of all siamese base networks.

CONCLUSION

- Vanilla siamese network surpasses the performance of Deep CNN but training time is too high.
- Proposed methods speed up both training and prediction by *83 and 3 folds*, respectively.
- Augmented Siamese Networks maintains the superior performance of vanilla Siamese Network under the low data domain.
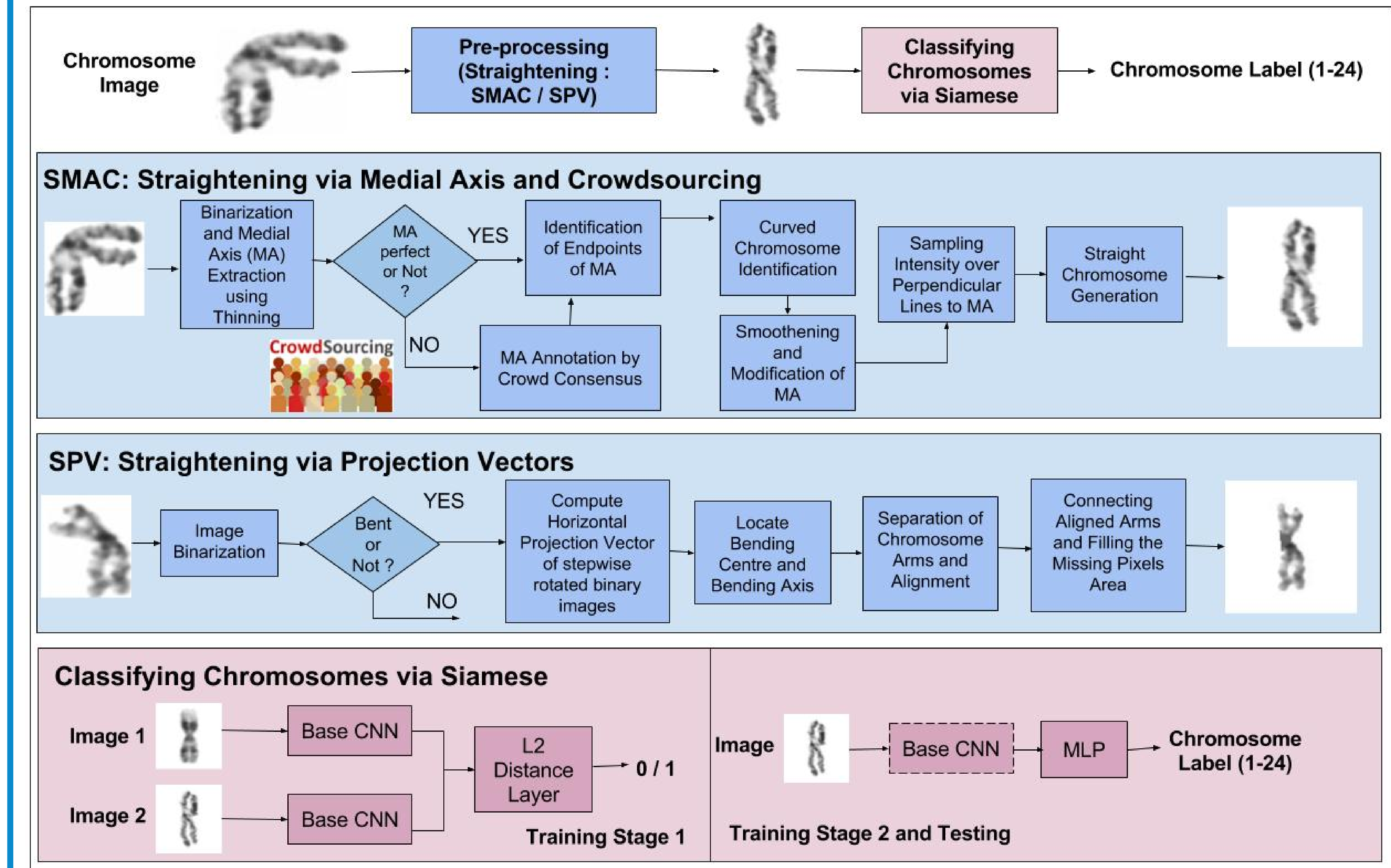
SYSTEM ARCHITECTURE

Figure 2: The flow-diagram of chromosome classification using Siamese Networks

RESULTS

Network Architecture	Similar:Dissimilar	Accuracy	Training time (h)	Pred Time (sec)
Two Layer MLP	——	59.7	0.16	4.088
Two Layer MLP+SMAC	——	67.9	0.16	4.088
Two Layer MLP+SPV	——	72.3	0.15	4.088
Deep CNN	——	68.5	5.35	12.085
Deep CNN+SMAC	——	78.4	2.09	12.085
Deep CNN+SPV	——	83.7	2.23	12.085
Siamese Network+SPV+Nearest	36.6K : 817K (R)	**85.6**	**124.17**	**15.088**
Siamese Network+MLP+SMAC(Avg.)	36.6K : 73.2K (R)	78.5	1.84	4.760
Siamese Network+MLP+SMAC(Max.)	36.6K : 73.2K (R)	79.4	1.92	4.760
Siamese Network+MLP+SPV(Avg.)	36.6K : 73.2K (R)	81.3	1.76	4.760
Siamese Network+MLP+SPV(Max.)	36.6K : 73.2K (R)	83.8	1.83	4.760
Siamese Network+MLP+SMAC(Avg.)	9K : 18K (B)	78.6	1.57	4.760
Siamese Network+MLP+SMAC(Max.)	9K : 18K (B)	80.4	1.58	4.760
Siamese Network+MLP+SPV(Avg.)	9K : 18K (B)	83.3	1.53	4.760
Siamese Network+MLP+SPV(Max.)	9K : 18K (B)	**85.2**	**1.50**	**4.760**
Con-Siamese Network+MLP+SMAC	——	79.8	1.53	4.760
Con-Siamese Network+MLP+SPV	——	**84.6**	1.52	4.760

Table 1: Results for the proposed Siamese Networks along with the multiple baselines created using a Deep CNN for the automatic chromosome classification. (R stands for *Random* and B stands for *Balanced*)