Chromosome Image Classification Techniques for Down Syndrome Detection: A Review of AI-Based Genetic Diagnostic Approaches

Dr. Khalid Al-Mutairi[1]*, Dr. Nora Al-Zahrani[1]
[1]King Saud University College of Medicine, Department of Medical Genetics and Clinical Cytogenetics, Riyadh, Saudi Arabia

ABSTRACT
Down syndrome (DS) is one of the numerical abnormality which is characterized by a change in the chromosome number. Down syndrome is a genetic disorder with genome dosage imbalances and micro-duplication of human chromosome 21. It is usually associated with a group of serious diseases, including intellectual disabilities, cardiac diseases, physical abnormalities, and other abnormalities. In the existing methods we use various classifiers such as *k*-means, SVM and ANN. In our proposed method, the down syndrome is detected from the G-banded metaphase chromosomes. The abnormal chromosomes are classified using CNN (Convolution Neural Network).

KEYWORDS: Down syndrome, Numerical abnormality, G-banded metaphase chromosomes, Connected component labeling.

1. INTRODUCTION
Down syndrome (DS) is a genetic disorder with genome dosage imbalances and micro-duplication of human chromosome 21 (HAS21) ie., DS occurs when an extra copy of chromosome gets added to the chromosome number 21. It is usually associated with intellectual disabilities, congenital heart defects, childhood Leukemia, Alzheimer's disease, early aging, physical abnormalities, and other abnormalities. Even though DS occurs in a high rate worldwide. It has been well studied ,researchers haven't found any effective cure method. Currently Human DS therapy are mainly focusing on early intervention, educational therapy, physical therapy as well as behavioral therapies. These therapies only have limited effects. . Here we are handling with G-banded metaphase chromosomes. G-banding is a technique which will produce bands on the chromosome. Hence the identification became easier and more accurate. G-banding is the oldest and widely used in chromosome analysis technique used to produce a unique alternate pattern in chromosome. The existing method for classification are ANN (Artificial Neural Network) which is trained using back propagation network, another classifiers are K-means and SVM (Support Vector Machine). In this paper, CNN (Convolutional Neural Network) is used for classification purpose. The chromosome count from the images are obtained through connected component labeling.

2. HUMAN CHROMOSOME 21
Human chromosome 21(HSA21) is the smallest human chromosome. The length of the long arm (21q) is 33.5 Mb and the short arm (21p) is 5-15 Mb. There are approximately 300-400 genes on this chromosome. Many of these genes are important both for the formation of body organs and for maintaining numerous functions of the organism. It is spanning about 48 million base pairs (the building blocks of DNA) and representing 1.5 to 2 percent of the total DNA in cells. Down Syndrome (DS) or otherwise called Trisomy is a case occurred in chromosome 21. DS occurs when an extra copy of chromosome gets added to the chromosome number 21.

3. CAUSES OF DOWN SYNDROME
Down syndrome is usually caused by errors in cell division. These are non-disjunction and Robertsonian translocation. The former is a failure of the pair of chromosome to separate during meiosis. It is the process by which egg and sperm cells replicate themselves and divide. Non-disjunction results in both 21^{st} chromosomes being carried to one cell and none to the other. The 80% of children born with down syndrome are born to women under 35 years of age. The latter accounts for only 3 to 4% of cases of DS. In this, part of chromosome 21 breaks off during cell division and attaches to another chromosome, while the total number of chromosomes in the cell remain 46. Unlike non-disjunction, maternal age is not linked to the risk of translocation. In the one-third of translocation incidents, one parent is carrier of a translocated chromosome.

4. MATERIALS AND METHODS
4.1 *k*-means : *k*-means clustering is a method of ,vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. *k*-means clustering aims to partition n observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. *k*-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes. The most common algorithm uses an iterative refinement technique, it is often called the *k*-means algorithm. The algorithm has converged when the assignments no

longer change. The algorithm does not guarantee to find the optimum. The algorithm is often presented as assigning objects to the nearest cluster by distance. Using a different distance function other than (squared) Euclidean distance may stop the algorithm from converging.

The applications of the *k*-means clustering are that it is rather easy to apply to even large data sets. It has been successfully used in market segmentation, computer vision, and astronomy among many other domains. It often is used as a preprocessing step for other algorithms, for example to find a starting configuration.

4.2 SVM (Support-Vector Machines): In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. An SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification implicitly mapping their inputs into high-dimensional feature spaces.

A support-vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection. a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin, the lower the generalization error of the classifier. SVMs can be used to solve various real-world problems:

- SVMs are helpful in text and hypertext categorization, as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.
- Classification of images can also be performed using SVMs. This significantly achieve higher accuracy.
- Hand-written characters can be recognized using SVM.
- The SVM algorithm has been widely applied in the biological and other sciences. They have been used to classify proteins with up to 90% of the compounds classified correctly.

4.3 ANN (Artificial Neural Network) : Automated chromosome classification has been an important pattern recognition problem. Artificial neural network (ANN) is ideal for this task. Artificial neural network is a machine learning technique used for classification problems. It allows application of expert knowledge and experience through network training. A large number of different based ANNs have been tested and evaluated for chromosome classification. An artificial neural network is a network of simple elements called artificial neurons, which receive input, change their internal state (activation) according to that input, and produce output depending on the input and activation. On the basis of connection ANN can be classified into two categories: feed-forward network and recurrent network. Feed forward neural network is the network in which connections between units do not form cycle whereas in recurrent neural network connection form cycle.

Because of their ability to reproduce and model nonlinear processes, Artificial neural networks have found many applications in a wide range of disciplines. Application areas include :

- System identification and control (vehicle control, natural resource management),
- Quantum chemistry and general game playing,
- Pattern recognition (radar systems, face identification, signal classification, 3D reconstruction, object recognition and more),
- Sequence recognition (gesture, speech, handwritten and printed text recognition),
- Medical diagnosis,
- Finance(e.g. automated trading systems),
- Data mining, visualization, machine translation, social network filtering and e-mail spam filtering.

5 CONCLUSION

In this paper, we have provided various classifiers along with their applications on various fields. These classifiers are following various algorithms. *k*-means clustering is a vector quantization method. It tends to find clusters of comparable spatial extent. The algorithm has converged when the assignments no longer change. The algorithm does not guarantee to find the optimum. The algorithm is often presented as assigning objects to the nearest cluster by distance. Support-vector machines (SVM) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. SVMs can efficiently perform a non-linear classification implicitly mapping their inputs into high-dimensional feature spaces. An artificial neural network is a network of simple elements called artificial neurons, which receive input, change their internal state (activation) according to that input, and produce output depending on the input and activation. It allows application of expert knowledge and experience through network training. A large number of different based ANNs have been tested and evaluated for chromosome classification

REFERENCES

[1] A. T. Natarajan, "Chromosome aberrations: past, present and future," *Mutation Research/Fundamental and Molecular Mechanisms of Muta- genesis*, vol. 504, no. 1, pp. 3–16, 2002.

[2] A. Theisen and L. G. Shaffer, "Disorders caused by chromosome abnormalities." *The application of clinical genetics*, vol. 3, pp. 159–174, 2010.

[3] J. Piper, "Automated cytogenetics in the study of mutagenesis and cancer," in *Advances in Mutagenesis Research*. Springer, 1990, pp. 127–153.

[4] B. Lerner, H. Guterman, I. Dinstein, and Y. Romem, "Medial axis transform-based features and a neural network for human chromosome classification," *Pattern Recognition*, vol. 28, no. 11, pp. 1673–1683, 1995.

[5] D. Ming and J. Tian, "Automatic pattern extraction and classification for chromosome images," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 31, no. 7, pp. 866–877, 2010.

[6] C. Markou, C. Maramis, A. Delopoulos, C. Daiou, and A. Lam- bropoulos, "Automatic chromosome classification using support vector machines," *Google Scholar*, pp. 1–24, 2012.

[7] N. Madian and K. Jayanthi, "Analysis of human chromosome classifi- cation using centromere position," *Measurement*, vol. 47, pp. 287–295, 2014.

[8] P. Biyani, X. Wu, and A. Sinha, "Joint classification and pairing of human chromosomes," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2, no. 2, pp. 102–109, April 2005.

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

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

[11] G. Gupta, M. Yadav, M. Sharma, L. Vig *et al.*, "Siamese networks for chromosome classification," in *Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on*. IEEE, 2017, pp. 72–81.

[12] Y. Wu, Y. Yue, X. Tan, W. Wang, and T. Lu, "End-to-end chromosome karyotyping with data augmentation using gan," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 2456–2460.

[13] R. J. Stanley, J. Keller, C. W. Caldwell, and P. Gader, "Centromere attribute integration based chromosome polarity assignment." in *Pro- ceedings of the AMIA Annual Fall Symposium*. American Medical Informatics Association, 1996, p. 284.

[14] X. Wang, B. Zheng, S. Li, J. J. Mulvihill, and H. Liu, "A rule-based computer scheme for centromere identification and polarity assign- ment of metaphase chromosomes," *computer methods and programs in biomedicine*, vol. 89, no. 1, pp. 33–42, 2008.

[15] A. S. Arachchige, J. Samarabandu, J. H. Knoll, and P. K. Rogan, "In- tensity integrated laplacian-based thickness measurement for detecting human metaphase chromosome centromere location," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 2005–2013, 2013.

[16] E. Loganathan, M. Anuja, and N. Madian, "Analysis of human chromo- some images for the identification of centromere position and length," in *Point-of-Care Healthcare Technologies (PHT), 2013 IEEE*. IEEE, 2013, pp. 314–317.

[17] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural infor- mation processing systems*, 2012, pp. 1097–1105.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition,"

arXiv preprint arXiv:1409.1556, 2014.
[21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks.” in *CVPR*, vol. 1, no. 2, 2017, p. 3.
[23] T.-Y. Lin, A. RoyChowdhury, and S. Maji, “Bilinear cnn models for fine- grained visual recognition,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1449–1457.
[24] J. Fu, H. Zheng, and T. Mei, “Look closer to see better: Recurrent atten- tion convolutional neural network for fine-grained image recognition,” in *CVPR*, vol. 2, 2017, p. 3.
[25] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, “Multi-scale convo- lutional neural networks for lung nodule classification,” in *International Conference on Information Processing in Medical Imaging*. Springer, 2015, pp. 588–599.
[26] L. Zeng, X. Xu, B. Cai, S. Qiu, and T. Zhang, “Multi-scale convolutional neural networks for crowd counting,” in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 465–469.
[27] W. Lotter, G. Sorensen, and D. Cox, “A multi-scale cnn and curriculum learning strategy for mammogram classification,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2017, pp. 169–177.
[28] S. Zagoruyko and N. Komodakis, “Wide residual networks,” *arXiv preprint arXiv:1605.07146*, 2016.
[29] R. Caruana, “Multitask learning,” *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.
[30] R. Girshick, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.
[31] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems*, 2015, pp. 91–99.
[32] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.
[33] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
[34] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” in *NIPS-W*, 2017
[35] L. v. d. Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.