

RESEARCH

Open Access



Bi-stream CNN Down Syndrome screening model based on genotyping array

Bing Feng^{1,2}, William Hoskins², Yan Zhang^{2,5}, Zibo Meng², David C. Samuels⁴, Jiandong Wang², Ruofan Xia², Chao Liu¹, Jijun Tang^{1,2,5*} and Yan Guo^{3*}

From Selected articles from the IEEE BIBM International Conference on Bioinformatics & Biomedicine (BIBM) 2017: medical genomics
Kansas City, MO, USA. 13-16 November 2017

Abstract

Background: Human Down syndrome (DS) is usually caused by genomic micro-duplications and dosage imbalances of human chromosome 21. It is associated with many genomic and phenotype abnormalities. Even though human DS occurs about 1 per 1,000 births worldwide, which is a very high rate, researchers haven't found any effective method to cure DS. Currently, the most efficient ways of human DS prevention are screening and early detection.

Methods: In this study, we used deep learning techniques and analyzed a set of Illumina genotyping array data. We built a bi-stream convolutional neural networks model to screen/predict the occurrence of DS. Firstly, we built image input data by converting the intensities of each SNP site into chromosome SNP maps. Next, we proposed a bi-stream convolutional neural network (CNN) architecture with nine layers and two branch models. We further merged two CNN branch models into one model in the fourth convolutional layer, and output the prediction in the last layer.

Results: Our bi-stream CNN model achieved 99.3% average accuracies, and very low false-positive and false-negative rates, which was necessary for further applications in disease prediction and medical practice. We further visualized the feature maps and learned filters from intermediate convolutional layers, which showed the genomic patterns and correlated SNPs variations in human DS genomes. We also compared our methods with other CNN and traditional machine learning models. We further analyzed and discussed the characteristics and strengths of our bi-stream CNN model.

Conclusions: Our bi-stream model used two branch CNN models to learn the local genome features and regional patterns among adjacent genes and SNP sites from two chromosomes simultaneously. It achieved the best performance in all evaluating metrics when compared with two single-stream CNN models and three traditional machine-learning algorithms. The visualized feature maps also provided opportunities to study the genomic markers and pathway components associated with Human DS, which provided insights for gene therapy and genomic medicine developments.

Keywords: Deep learning, Convolutional neural networks, Human down syndrome, Genotyping

*Correspondence: jtang@cse.sc.edu; yaguo@salud.unm.edu

¹ College of Education, Zhejiang University, 310058 Hangzhou, Zhejiang, People's Republic of China

² Department of Computer Science and Engineering, University of South Carolina, 29208 Columbia, SC, USA

Full list of author information is available at the end of the article



Background

Human Down syndrome (DS) is usually caused by genomic micro-duplications and dosage imbalances of human chromosome 21 [1]. It is associated with many genomic and phenotype abnormalities [2, 3]. Currently, human DS occurs at a very high rate, which is about 1 per 1000 births worldwide [4]. Human DS is also associated with a group of serious diseases, including congenital heart defects, intellectual disability, leukemia, Alzheimer's disease, Hirschsprung disease, early aging, physical abnormalities, and other abnormalities [1, 5–7]. Current treatments of human DS mainly concentrate on physical therapy [8, 9], emotional and behavioral therapies [10, 11], educational therapy, and early intervention [10, 12]. However, these therapies only have some limited effects that cannot cure DS fundamentally.

DS screening has been studied for more than 50 years. Currently, widely used approaches include combined genomic test [13], blood test [14], sequencing test [15], and ultrasound measurement of nuchal translucency [16]. However, 1/16 of positive screening women may still suffer from further invasive diagnostic procedures, which might result in fetal loss [15, 17]. Therefore, an accurate and error-less DS screening method could significantly reduce the risk of human DS screening procedures.

Recent genome-wide association studies (GWAS) and single nucleotide polymorphisms (SNPs) studies have proved strong correlations between genomic abnormalities and occurrences of different kinds of diseases [18–21]. DS related GWAS studies also showed that SNP variations, gene copy number variations (SNVs), and lots of unidentified genomic variations were associated with the complex genomic disorders and abnormalities of Human DS [22, 23]. However, only a few biomarkers have been discovered to associate with Human DS, such as chorion gonadotropin, unconjugated estriol, and alpha-fetoprotein [24, 25]. Human chromosome 21 (Hsa21) encodes more than 500 genes [26, 27] and have various functions, including RNA splicing protein modifiers, cell surface receptors, transcription factors, adhesion molecules, and biochemical pathway components [27, 28]. Currently, 160 of Hsa21 genes have already been annotated as protein-coding genes by SwissProt. Five of them are microRNAs. Most of them have unknown functions [29]. The over-expression of Hsa21 genes results in complex genomic disorders and perturbations of biological processes and pathways [28]. Illumina has introduced a new exome genotyping array technique to identify rare single-nucleotide polymorphisms, which is an alternative technique of high-throughput sequencing. The Vanderbilt University Medical Center and Center for Quantitative Sciences developed an exome chip-processing protocols for this technique [23].

Machine learning has already been applied to human diseases and genomic pattern predictions [30–32]. Based on our knowledge, only limited types of traditional machine learning techniques have been used in human DS studies [27, 33]. Most of them are performed on mice DS models [23, 27, 34]. Zhao et al. used hierarchical constrained regional model and independent component analysis to detect Human Down syndrome of pediatric patients [35]. Cao et al. used a Naive Bayes model to predict locomotor activities in mice models Ts65Dn and Ts1Cje under the treatments of N-methyl-D-aspartate receptor [34]. Clara et al. designed an unsupervised self-organizing map model to identify biological differences in mice model Ts65Dn [27]. Recently, deep neural networks, especially convolutional and recurrent neural networks, have achieved impressive performances in disease screening, predictions and diagnosis studies [30, 36–38].

In this study, we used convolutional neural networks to construct human Down Syndrome screening/prediction models based on Illumina genotyping array data. Firstly, we built image input data by converting the intensities of SNP sites into chromosome SNP maps. Then we proposed a bi-stream convolutional neural network architecture with nine layers and two branch CNN models, which took two input chromosome SNP maps simultaneously. We also constructed another two single-stream CNN models, which took one chromosome SNP map as input image using the same dataset. Next, we used three traditional machine learning algorithms Random Forest, SVM, and Decision Trees to construct DS screening/prediction models with the same dataset. We evaluated, compared, and analyzed the performance metrics for all models mentioned above. We concluded that our bi-stream CNN model had best performances in all evaluation metrics when compared with other models. At last, we visualized feature maps and learned filters from intermediate layers to study the genomic patterns and correlated gene and SNP variations. We also analyzed and discussed the characteristics and strengths of the bi-stream CNN model.

Result

Building human chromosome SNP maps

The genotyping dataset used in this study was Illumina exome genotyping array data, which targeted rare single-nucleotide polymorphisms. The dataset contained 378 samples, including 63 DS samples and 315 control samples. Each sample contained the intensity information of 5458 SNPs sites from 321 Hsa21 coding genes. The SNP intensities were normalized to the interval [0,1]. As shown in Fig. 1, we built two chromosome SNP maps to represent the intensities of all SNP site for two Hsa21 chromosomes. Each column of chromosome SNP map represented one single gene. Each row represented adjacent SNP sites within the same gene. Therefore, each

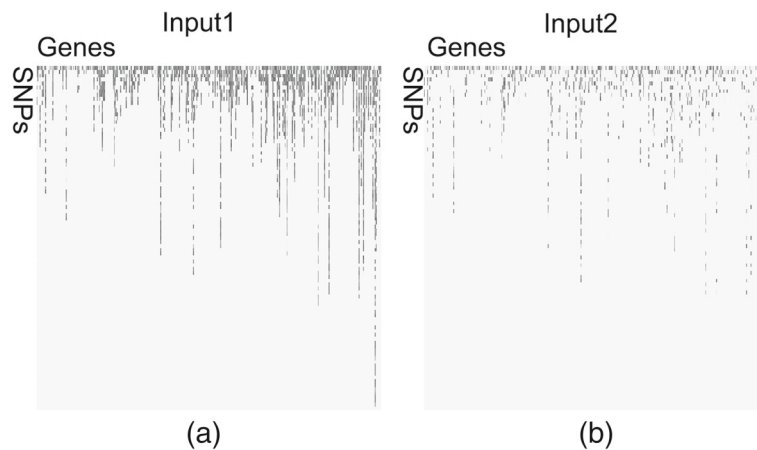


Fig. 1 Chromosome SNP maps to represent the intensities of all SNP site on HSA21. Each column represents the information of one single gene located on the chromosome. Each row represents adjacent SNP sites within the same gene. Therefore, each pixel of the chromosome SNP map is used to represent the intensity of each SNP site of genes

pixel could be used to represent the intensity of each SNP site. In this study, we used chromosome SNP maps as input images to construct and evaluated CNN models. For traditional machine learning algorithms, we used original Illumina genotyping array dataset to construct and evaluate the models. For each model construction and evaluation, we did ten parallel experiments with ten sample datasets by randomly sampling the original dataset ten times. Each sample dataset randomly selected 75% data for training and the rest 25% for testing. We calculated average evaluating metrics to provide reliable evaluations.

Bi-stream convolutional neural network architecture

Figure 2 showed the architecture of the bi-stream CNN model used in this study, which was merged from two branch CNN models. Each branch model had one input layer, three convolutional layers, and one max-pooling layer. Therefore, our bi-stream CNN model could take two chromosome SNP maps as input images simultaneously. We merged two branch CNN models into one CNN model in the fourth convolutional layer, which was followed by a max-pooling layer. Next, we had another three fully connected layers and one output layer. We added

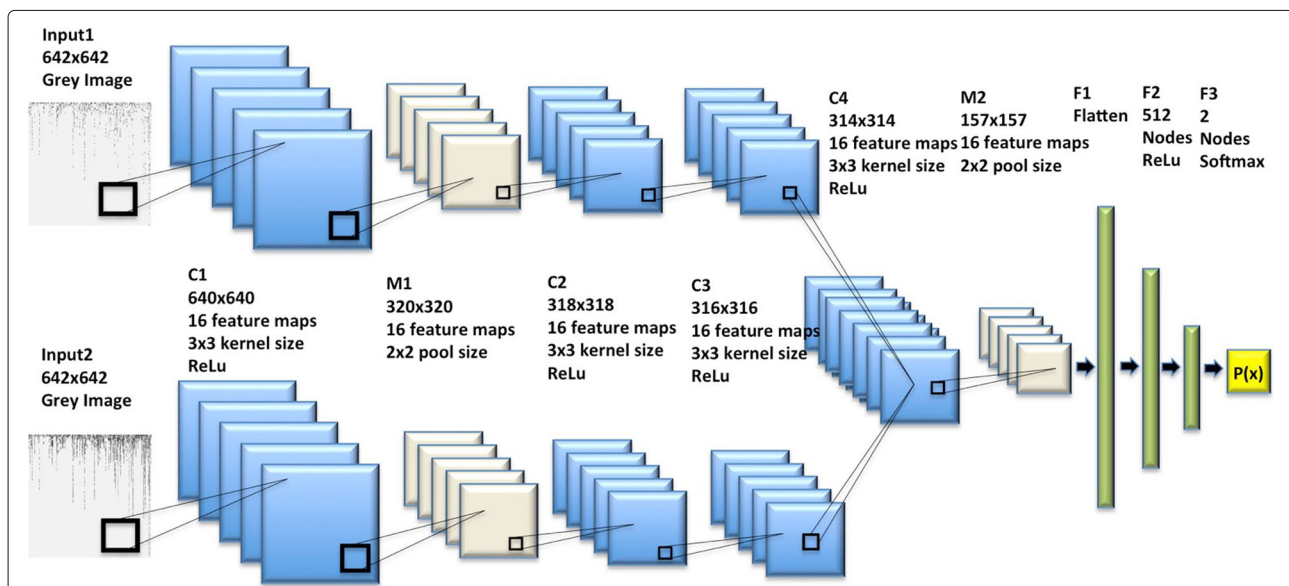


Fig. 2 Bi-stream CNN architecture taking two chromosome SNP maps as inputs The upper CNN branch model and the lower CNN branch model both take one chromosome SNP map as input image. We merged two branch CNN models into one CNN model in the fourth convolutional layer C4, which was also followed by a max-pooling layer. Detailed CNN architecture construction and configurations are available in the [Method](#) section

dropouts for each hidden layer for reducing over-fitting. Detailed CNN architecture and configurations were available in the [Method](#) section.

Bi-stream CNN DS screening/prediction model

We first constructed human DS screening/prediction model using bi-stream CNN architecture proposed in the last section. To provide reliable and confidence evaluation, we ran ten parallel experiments on ten randomly sampled dataset and calculated average performance metrics. As shown in [Table 1](#), our bi-stream CNN model achieved 99.3% average accuracy in ten parallel experiments. The average precision, recall, and F-score were 99.2, 98.4, and 99.3%. It was worth to notice that the bi-stream CNN model had very low false-positive and false-negative rates, which were 0.6 and 1.1%. We only mis-predicted five non-DS samples and two DS samples in all ten experiments. Our results showed that the bi-stream CNN architecture could construct very accurate and robust human DS screening/prediction models.

Comparing with traditional machine learning DS screening/screening models

We further applied three different traditional supervised learning algorithms to construct human DS prediction models using the original Illumina genotyping array data with total 5458 SNP features. We also ran ten parallel experiments and further compared the performances with our bi-stream CNN model. [Table 1](#) showed that Random Forest, SVM, and Decision Tree models could achieve very high average accuracies, which were all above 96%. The model built from Random Forest achieved the best performance in all evaluation metrics among all three traditional learning algorithms. Nevertheless, [Table 1](#) also showed that the bi-stream CNN model produced higher accuracy, precisions, recalls, and F-scores when compared with traditional machine learning algorithms. Furthermore, the false negative rates of Random Forest, SVM, and Decision Tree models were very high, which were 8.1, 5.3, and 8.0% respectively. Models with such high false-negative rate were impractical to be applied in real-life clinical prediction and medical practice. However, the bi-stream CNN models achieved significantly better performances in false-positive and false-negative rates, which

were only 0.6 and 1.1%. The result above demonstrated that the bi-stream CNN model achieved better performances when compared with the traditional machine learning algorithms. It was more suitable for human DS screening.

Comparing with single-stream CNN model

Here we built two new single-stream CNN models using the same configurations and datasets with our bi-stream CNN model proposed above. The only difference between bi-stream and single-stream CNN models was that single stream model only had one CNN branch and took one chromosome SNP map as the input image. We further compared and evaluated the performances of two single-stream CNN models. As [Table 2](#) shown, our bi-stream CNN model achieved the best performance over all three models in all evolutionary metrics. The other two single-stream CNN model also achieved over 96% accuracies. However, the recall of the first single-stream model and the precision of the second single-stream model were very low, which were 84.0% and 88.7% respectively. Furthermore, the false positive and false negative rate of the single-stream CNN models were significantly higher than the bi-stream CNN model. In general, our bi-stream CNN model had significantly better performances than the single-stream CNN models. The single-stream models could only extract the genome features from one single chromosome, which completely neglected the genomic patterns from the other one. Therefore, they were not as accurate as the bi-stream CNN model. The bi-stream CNN model was more comprehensive, accurate, and reliable when compared with the single-stream DS prediction models.

Visualization of feature maps and trained filters of bi-stream model

In this section, we visualized the trained filters and feature maps from intermediate convolutional hidden layers of our trained bi-stream CNN model. The bi-stream CNN model had a few advantages when compared with traditional machine learning algorithms. First of all, we used chromosome maps to represent the genotyping array information, which converted one-dimensional genome data to images. Secondly, We used 16 convolutional 3x3

Table 1 Evaluation metrics of bi-stream CNN and conventional machine learning models

Models	Evaluation metrics of different models					
	Accuracy	Precision	Recall	F-score	False-positive rate	False-negative rate
Decision tree	96.9(+1.0)%	94.1%	95.4%	94.6%	2.2%	8.0%
Random forest	97.1(+0.7)%	94.4%	94.9%	94.7%	1.9%	8.1%
SVM	96.7(+0.9)%	92.7%	95.9%	94.2%	2.9%	5.3%
Bi-Stream CNN	99.3(+0.4)%	99.2%	98.4%	99.3%	0.6%	1.1%

Table 2 Evaluation metrics of different CNN models

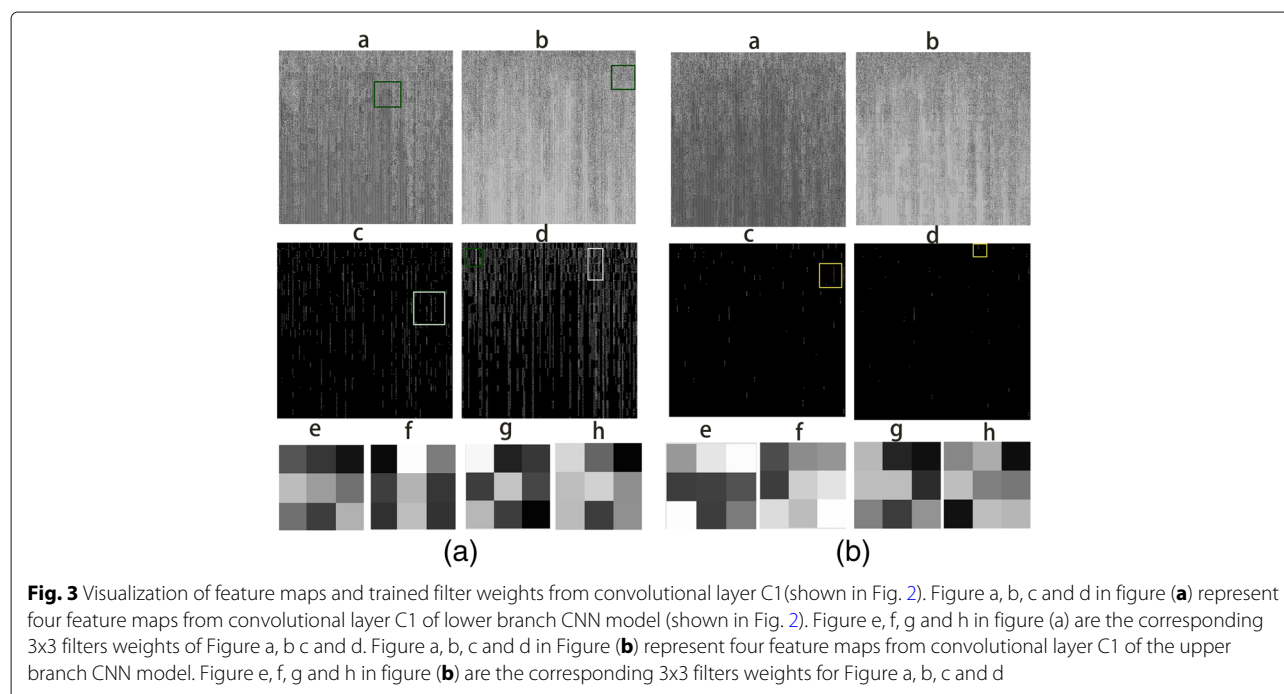
Models	Evaluation metrics of different CNN models					
	Accuracy	Precision	Recall	F-score	False-positive rate	False-negative rate
Bi-Stream CNN	99.3(+0.4)%	99.2%	98.4%	99.3%	0.6%	1.1%
Single-stream CNN (ChrA)	96.4(+0.5)%	94.7%	84.0%	96.4%	5.2%	3.2%
Single-stream CNN (ChrB)	96.6(+0.6)%	88.7%	92.9%	96.6%	11.2%	4.3%

size kernels to capture local genomic features and detect patterns from adjacent genes and SNP sites from two chromosome SNP maps. Thirdly, two branch CNN model could capture the genomic features from two chromosomes at the same time. Figure 3a and b showed the output feature maps and their corresponding trained filters from convolutional layer C1 of each branch CNN models. Some trained filters could highlight the most important and informative SNP sites from the chromosome SNP maps, and neglect less informative ones (marked as yellow squares). The green rectangles showed that our trained filters could sharpen input images and capture local motifs, which represented the correlated variations in genome regions. The bi-stream model could also detect continuous gene and SNP intensity variations by capturing adjacent variation patterns in line (marked as white rectangles). Our bi-stream CNN model could detect the simultaneous or causal SNPs variations in human genomes. These genome characterizations and extracted genomic patterns provided signals to classify DS and normal samples. However, traditional machine learning algorithms tended to build models with a global view

from all available features and treated each feature independently. Therefore they were hard to extract signals from regional genomic patterns and correlations between adjacent genes and SNPs sites.

Discussion

Previous studies illustrated that gene expressions and SNP variations were highly correlated within local genome regions [39–41]. Genome-wide association studies also demonstrated that human DS was usually associated with many gene copy number and SNPs variations, and many unidentified genomic abnormalities [23, 42, 43]. In this study, our bi-stream CNN model could learn the genomic features and associated variations among adjacent genes and SNP sites from chromosome SNP maps. Currently, human DS treatments only have limited effects and can not cure DS fundamentally. There isn't any clear effect or benefit on human DS treatments using traditional drugs either [44–47]. The feature maps and extracted genome features could identify DS related markers and pathway components. These genome features explained the genomic characteristics and pathological mechanisms



of human DS, which could be further be applied in gene therapy and genetic medicine developments.

An accurate non-invasive DS screening method offers a low-risk way to screen human DS. It helps low-risk patients avoid taking further invasive diagnostic procedures, which might result in fetal loss. Nowadays, genotyping array analyses on fetal genomes could be performed on the trophoblast cells with non-invasive procedures after the fifth week gestation [42, 43]. In this study, we developed a novel method to construct accurate DS screening model by using bi-stream CNN and genotyping array data. The results showed that our bi-stream CNN model had the best performance in every evaluation metric when compared with two single-stream CNN models and three traditional machine learning models. The CNN model achieved over 99.3% accuracies, as well as very low false positive and false negative rates. It was very important to disease prediction and medical practice. Even though traditional machine learning algorithms obtained over 96% accuracies, their high false-negative rates are not suitable for clinical screening tests. Traditional machine learning algorithms treated each SNP sites as single feature independently. They were hard to extract signals from regional genomic patterns and variation correlations between adjacent genes and SNPs sites. Although the single-stream models could extract features and patterns from local genome features and adjacent SNP sites, they could only learn these features from one single chromosome, which completely neglected the genomic patterns of the other one. In deep learning studies, large datasets were great obstacles in the model construction and optimization. We used each pixel to represent the intensities of SNP site, and used chromosome SNP maps to represent the genome information, which significantly reduced data and model complexity. Furthermore, our bi-stream CNN architecture could learn local genomic patterns and extracted regional features, which could also be applied to building prediction models from genotyping array data for more diseases.

Method

Data

In this study, the rare single-nucleotide polymorphisms were measured by newly introduced Illumina exome genotyping array technique. Illumina exome genotyping array could identify rare single-nucleotide polymorphisms, which was an alternative technique of high-throughput sequencing. The Vanderbilt University Medical Center and Center for Quantitative Sciences had developed an exome chip-processing protocols for this technique [23], and provided us the experiment data. The dataset contained the intensity information of total 5458 SNP sites from 321 coding genes on Hsa21 [48]. There

were total of 378 samples, including 63 DS samples and 315 control samples.

Bi-stream CNN architecture

Our bi-stream CNN model was merged from two branch CNN models. Each branch CNN model had one input layer, three convolutional hidden layers, and one max pooling layer. We fed two input chromosome SNP maps to the two branch CNN models at the same time. Two branch CNN models were further merged into one CNN model in the sixth layer, which was also a convolutional hidden layer. Figure 4 showed the detailed deep neural network structure and configurations for each layer. Detailed information and configurations were shown as below:

Each branch model contained five layers: Layer 1, the input layer took one size 642×642 grey chromosome SNP map image as input. Layer 2, one convolutional layer with $16 \ 3 \times 3$ filters and ReLU activation. Layer 3, one max pooling layer with 2×2 pool size to down-sample the data, followed by a dropout (0.25) to reduce over-fitting. Layer 4, one convolutional layer with $16 \ 3 \times 3$ filters and ReLU activations, followed by dropout (0.25). Layer 5, one convolutional layer with $16 \ 3 \times 3$ filters and ReLU activations, followed by dropout (0.25). Next, in layer 6, we merged two branch CNN models into one convolutional layer with $16 \ 3 \times 3$ filters and ReLU activations. Layer 7 was another max pooling layer with 2×2 pool scale, followed by dropout (0.25). Layer 8 was a fully connected layer to flatten all features into one-dimension. Layer 9 was a fully connected layer with 512 nodes and ReLU activation. Layer 10 was the output layer with two nodes and Softmax activation. We used stochastic gradient descent optimizer (SGD) and binary cross-entropy as loss function, with a learning rate of 0.01, $1e-6$ decay and 0.9 nesterov momentum. We used Tensor-flow and Keras construct all CNN models that used in this study. We use a NVIDIA GeForce GTX TITAN X GPU to build our model on a Ubuntu 14.04.5 LTS machine[48].

Conventional machine-learning algorithms

In this study, we used Python and Scikit Learn package [49] to implement and construct models for traditional machine learning algorithms SVM, Random Forest, and Decision Tree[49]. SVM was implemented using C-Support Vector Classification algorithm, which used "one-vs-one" scheme. Random Forest and Decision Tree used entropy and Gini impurity to measure features' splitting qualities. There was no maximum depth limit for Random Forest sub-trees, unless there were less than two samples or all leaves were pruned. We used Classification and Regression Trees algorithm to implement the Decision Tree models. We constructed binary tree with the largest information gains on each splitting node, which

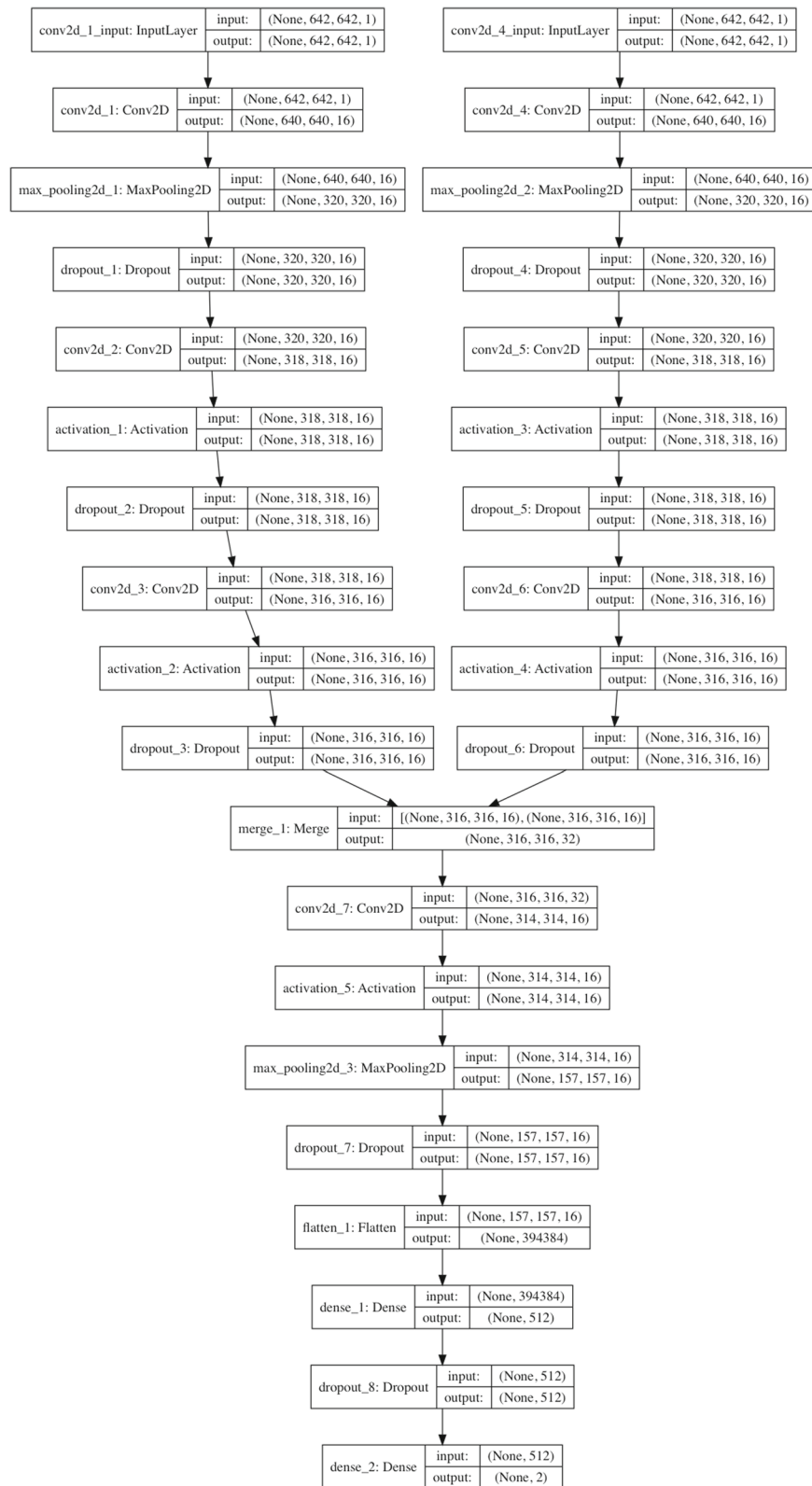


Fig. 4 Detailed configurations and structures for each layer of the bi-stream CNN DS prediction/screening model

was very similar to C4.5 decision tree algorithm. No depth limit was preset before training decision tree models. The maximum features used in model building was set to the total number of features [48].

Conclusion

In this study, we proposed bi-stream convolutional neural network architecture to construct accurate and robust human Down Syndrome screening and prediction model using Illumina genotyping array data. Our bi-stream CNN model was merged from two branch CNN models, which used two chromosome SNP maps as input images simultaneously. Two branch CNN models were further merged into one CNN model in a deeper convolutional layer. The comparison results showed that the bi-stream CNN model achieved the best performances in all evaluation metrics when compared with other three traditional machine learning algorithms and two single-stream CNN models. The CNN model could achieve 99.3% accuracies with very low false-positive and false-negative rates. Even though the conventional learning algorithms also obtained over 96% accuracies, their high false negative-rates made them hard to be applied in real life clinical screening test. Our bi-stream model used two branch CNN models to learn the local genomic pattern and regional correlations of the adjacent genes and SNPs from two chromosomes simultaneously. However, the single-stream CNN models only learn genomic features from one single chromosome, which completely neglected the genomic patterns of the other chromosome. The genomic patterns, correlated genes and SNPs variation identified by our CNN model provided opportunities to study the genomic markers and pathway components associated with human DS, which could be further applied in gene therapy and genomic medicine developments. Therefore, our method could learn local genomic patterns and extracted regional features from chromosome SNP maps, which could be applied to building prediction models from genotyping array data for more diseases.

Acknowledgments

We want to thank the National Key R&D Program of China 2017YFC0908400. We thank National Science Foundation of China (NSFC61772362). We are grateful* to the NVIDIA Corporation for the TITAN X GPU through a NVIDIA Hardware Grant.

Funding

JT and BF was supported by the US National Science Foundation 1161586, National Key R&D Program of China, grant number 2017YFC0908400, and National Natural Science Foundation of China, grant number 61702456. YG was supported by the National Institute of Cancer center supporting grant P30CA118100. Publication costs were funded by National Key R&D Program of China, grant number 2017YFC0908400.

Availability of data and materials

Please access the data and model by this link: goo.gl/2HLEFM.

About this supplement

This article has been published as part of *BMC Medical Genomics Volume 11 Supplement 5, 2018: Selected articles from the IEEE BiBM International Conference on Bioinformatics & Biomedicine (BiBM) 2017: medical genomics*. The full contents of the supplement are available online at <https://bmcmedgenomics.biomedcentral.com/articles/supplements/volume-11-supplement-5>.

Authors' contributions

JT, YG, and BF conceived and designed the project. BF, YG and JT designed and performed the experiments. All authors analyzed experiments results of this project. BF wrote the manuscript. All authors reviewed the manuscript. All authors read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹College of Education, Zhejiang University, 310058 Hangzhou, Zhejiang, People's Republic of China. ²Department of Computer Science and Engineering, University of South Carolina, 29208 Columbia, SC, USA. ³School of Medicine, The University of New Mexico, 87131 Albuquerque, NM, USA. ⁴Vanderbilt University School of Medicine, Vanderbilt University, 37232 Nashville, TN, USA. ⁵School of Computer Science and Technology, Tianjin University, 300072 Tianjin 300072, People's Republic of China.

Published: 20 November 2018

References

1. Antonarakis SE. Down syndrome and the complexity of genome dosage imbalance. *Nat Rev Genet.* 2016.
2. Gardiner KJ. Molecular basis of pharmacotherapies for cognition in down syndrome. *Trends Pharmacol Sci.* 2010;31(2):66–73.
3. Prandini P, Deutsch S, Lyle R, Gagnebin M, Vivier CD, Delorenzi M, Gehrig C, Descombes P, Sherman S, Bricarelli FD, et al. Natural gene-expression variation in down syndrome modulates the outcome of gene-dosage imbalance. *Am J Hum Genet.* 2007;81(2):252–63.
4. Weijerman ME, De Winter JP. Clinical practice. *Eur J Pediatr.* 2010;169(12):1445–52.
5. Patterson D. Molecular genetic analysis of down syndrome. *Hum Genet.* 2009;126(1):195–214.
6. Wiseman FK, Alford KA, Tybulewicz VL, Fisher EM. Down syndrome—recent progress and future prospects. *Hum Mol Genet.* 2009;18(R1):75–83.
7. Asim A, Kumar A, Muthuswamy S, Jain S, Agarwal S. Down syndrome: an insight of the disease. *J Biomed Sci.* 2015;22(1):41.
8. Chavez MC. Hippotherapy versus aquatic therapy use in early intervention physical therapy in children with down syndrome; 2016. PhD thesis, Division of Physical Therapy, School of Medicine, University of New Mexico.
9. Wentz EE. Importance of initiating a “tummy time” intervention early in infants with down syndrome. *Pediatr Phys Ther.* 2017;29(1):68–75.
10. Wuang Y-P, Chiang C-S, Su C-Y, Wang C-C. Effectiveness of virtual reality using wii gaming technology in children with down syndrome. *Res Dev Disabil.* 2011;32(1):312–21.
11. Greenspan SI, Wieder S, Simons R. *The Child with Special Needs: Encouraging Intellectual and Emotional Growth*. Boston: Addison-Wesley/Addison Wesley Longman; 1998.
12. Guralnick MJ. Early intervention approaches to enhance the peer-related social competence of young children with developmental delays: A historical perspective. *Infants Young Child.* 2010;23(2):73.

13. Driscoll DA, Gross S. Prenatal screening for aneuploidy. *N Engl J Med*. 2009;360(24):2556–62.
14. Ehrich M, Deciu C, Zwiefelhofer T, Tynan JA, Cagasan L, Tim R, Lu V, McCullough R, McCarthy E, Nygren AO, et al. Noninvasive detection of fetal trisomy 21 by sequencing of dna in maternal blood: a study in a clinical setting. *Am J Obstet Gynecol*. 2011;204(3):205–1.
15. Palomaki GE, Kloza EM, Lambert-Messerlian GM, Haddow JE, Neveux LM, Ehrich M, van den Boom D, Bombard AT, Deciu C, Grody WW, et al. Dna sequencing of maternal plasma to detect down syndrome: an international clinical validation study. *Genet Med*. 2011;13(11):913–20.
16. Spencer K, Souter V, Tul N, Snijders R, Nicolaidis K. A screening program for trisomy 21 at 10–14 weeks using fetal nuchal translucency, maternal serum free β -human chorionic gonadotropin and pregnancy-associated plasma protein-a. *Ultrasound Obstet Gynecol*. 1999;13(4):231–7.
17. of Obstetricians AC, Gynecologists, et al. Acog practice bulletin no. 88, december 2007. invasive prenatal testing for aneuploidy. *Obstet Gynecol*. 2007;110(6):1459.
18. Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, et al. Genome-wide association study identifies five novel susceptibility loci for crohn's disease and implicates a role for autophagy in disease pathogenesis. *Nat Genet*. 2007;39(5):596.
19. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK, et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring il2 and il21. *Nat Genet*. 2007;39(7):827.
20. Corradin O, Cohen AJ, Luppino JM, Bayles IM, Schumacher FR, Scacheri PC. Modeling disease risk through analysis of physical interactions between genetic variants within chromatin regulatory circuitry. *Nat Genet*. 2016;48(11):1313.
21. Warren CR, O'Sullivan JF, Friesen M, Becker CE, Zhang X, Liu P, Wakabayashi Y, Morningstar JE, Shi X, Choi J, et al. Induced pluripotent stem cell differentiation enables functional validation of gwas variants in metabolic disease. *Cell Stem Cell*. 2017;20(4):547–57.
22. Ramachandran D, Zeng Z, Locke AE, Muller JG, Bean LJ, Rosser TC, Dooley KJ, Cua CL, Capone GT, Reeves RH, et al. Genome-wide association study of down syndrome-associated atrioventricular septal defects. *G3: Genes, Genomes, Genet*. 2015;5(10):1961–71.
23. Sailani MR, Makrythanasis P, Valsesia A, Santoni FA, Deutsch S, Popadin K, Borel C, Migliavacca E, Sharp AJ, Sail GD, et al. The complex snp and cnv genetic architecture of the increased risk of congenital heart defects in down syndrome. *Genome Res*. 2013;23(9):1410–21.
24. Brock DJ, Sutcliffe RG. Alpha-fetoprotein in the antenatal diagnosis of anencephaly and spina bifida. *Lancet*. 1972;300(7770):197–9.
25. Wald NJ, Cuckle HS, Densm JW, Nanchahal K, Royston P, Chard T, Haddow JE, Knight GJ, Palomaki GE, Canick JA. Maternal serum screening for down's syndrome in early pregnancy. *Bmj*. 1988;297(6653):883–7.
26. Sturgeon X, Gardiner KJ. Transcript catalogs of human chromosome 21 and orthologous chimpanzee and mouse regions. *Mamm Genome*. 2011;22(5–6):261–71.
27. Higuera C, Gardiner KJ, Cios KJ. Self-organizing feature maps identify proteins critical to learning in a mouse model of down syndrome. *PLoS ONE*. 2015;10(6):0129126.
28. Dierssen M, de la Torre R. Pathways to cognitive deficits in down syndrome. *Down Syndr: Underst Neurobiol Ther*. 2012;197:73.
29. Gardiner K, Herval Y, Lott IT, Antonarakis SE, Reeves RH, Dierssen M. Down syndrome: from understanding the neurobiology to therapy. *J Neurosci*. 2010;30(45):14943–5.
30. Roth HR, Lu L, Liu J, Yao J, Seff A, Cherry K, Kim L, Summers RM. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Trans Med Imaging*. 2016;35(5):1170–81.
31. Park Y, Kellis M. Deep learning for regulatory genomics. *Nat Biotechnol*. 2015;33(8):825–6.
32. Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D, Initiative ADN, et al. Random forest-based similarity measures for multi-modal classification of alzheimer's disease. *NeuroImage*. 2013;65:167–75.
33. Feng B, Hoskins W, Zhou J, Xu X, Tang J. Using supervised machine learning algorithms to screen down syndrome and identify the critical protein factors. In: International Conference on Intelligent and Interactive Systems and Applications. New York: Springer; 2017. p. 302–8.
34. Nguyen CD, Costa AC, Cios KJ, Gardiner KJ. Machine learning methods predict locomotor response to mk-801 in mouse models of down syndrome. *J Neurogenet*. 2011;25(1–2):40–51.
35. Zhao Q, Okada K, Rosenbaum K, Zand DJ, Sze R, Summar M, Linguraru MG. Hierarchical constrained local model using ica and its application to down syndrome detection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. New York: Springer; 2013. p. 222–9.
36. Esteve A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115–8.
37. Sun W, Tseng T-LB, Zhang J, Qian W. Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. *Comput Med Imaging Graph*. 2017;57:4–9.
38. Faust O, Acharya UR, Sudarshan VK, San Tan R, Yeong CH, Molinari F, Ng KH. Computer aided diagnosis of coronary artery disease, myocardial infarction and carotid atherosclerosis using ultrasound images: A review. *Physica Medica*. 2016.
39. Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, et al. A genome-wide association study identifies il23r as an inflammatory bowel disease gene. *Science*. 2006;314(5804):1461–3.
40. Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoshani N, Whitton H, Ryan RJ, Shishkin AA, et al. Genetic and epigenetic fine-mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337.
41. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, Duncan L, Perry JR, Patterson N, Robinson EB, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236–41.
42. Jain CV, Kadam L, van Dijk M, Kohan-Ghadr H-R, Kilburn BA, Hartman C, Mazzorana V, Visser A, Hertz M, Bolnick AD, et al. Fetal genome profiling at 5 weeks of gestation after noninvasive isolation of trophoblast cells from the endocervical canal. *Sci Transl Med*. 2016;8(363):363–43634.
43. Petry C, Mooslehner K, Prentice P, Hayes M, Nodzinski M, Scholtens D, Hughes L, Acerini C, Ong K, Lowe W, et al. Associations between a fetal imprinted gene allele score and late pregnancy maternal glucose concentrations. In: *Diabetes & Metabolism*; 2017.
44. Mohan M, Bennett C, Carpenter PK. Memantine for dementia in people with down syndrome. In: *The Cochrane Library*; 2009.
45. Kishnani PS, Heller JH, Spiridigliozzi GA, Lott I, Escobar L, Richardson S, Zhang R, McRae T. Donepezil for treatment of cognitive dysfunction in children with down syndrome aged 10–17. *Am J Med Genet A*. 2010;152(12):3028–35.
46. Lott IT, Doran E, Nguyen VQ, Tournay A, Head E, Gillen DL. Down syndrome and dementia: a randomized, controlled trial of antioxidant supplementation. *Am J Med Genet A*. 2011;155(8):1939–48.
47. Gardiner KJ. Pharmacological approaches to improving cognitive function in down syndrome: current status and considerations. *Drug Des Devel Ther*. 2015;9:103–25.
48. Feng B, Samuels DC, Hoskins W, Guo Y, Zhang Y, Tang J, Meng Z. Down syndrome prediction/screening model based on deep learning and illumina genotyping array. In: *Bioinformatics and Biomedicine (BIBM), 2017 IEEE International Conference On*. New York: IEEE; 2017. p. 347–52.
49. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res*. 2011;12(Oct):2825–30.