

Received: 19 October 2016 | Revised: 17 April 2017 | Accepted: 14 May 2017

DOI: 10.1002/humu.23272

SPECIAL ARTICLE



DeepBipolar: Identifying genomic mutations for bipolar disorder via deep learning

Sundaram Laksshman* | Rajendra Rana Bhat* | Vivek Viswanath | Xiaolin Li 

National Science Foundation Center for Big Learning, University of Florida, Gainesville, Florida

Correspondence

Xiaolin Li, National Science Foundation Center for Big Learning, University of Florida, Gainesville, Florida

Email: andyli@ece.ufl.edu

* Sundaram Laksshman and Rajendra Rana Bhat contributed equally to this work.

Contract grant sponsor: NIH (U41HG007446, R13HG006650, and R01GM110240).

For the CAGI Special Issue

Abstract

Bipolar disorder, also known as manic depression, is a brain disorder that affects the brain structure of a patient. It results in extreme mood swings, severe states of depression, and overexcitement simultaneously. It is estimated that roughly 3% of the population of the United States (about 5.3 million adults) suffers from bipolar disorder. Recent research efforts like the Twin studies have demonstrated a high heritability factor for the disorder, making genomics a viable alternative for detecting and treating bipolar disorder, in addition to the conventional lengthy and costly post-symptom clinical diagnosis. Motivated by this study, leveraging several emerging deep learning algorithms, we design an end-to-end deep learning architecture (called DeepBipolar) to predict bipolar disorder based on limited genomic data. DeepBipolar adopts the Deep Convolutional Neural Network (DCNN) architecture that automatically extracts features from genotype information to predict the bipolar phenotype. We participated in the Critical Assessment of Genome Interpretation (CAGI) bipolar disorder challenge and DeepBipolar was considered the most successful by the independent assessor. In this work, we thoroughly evaluate the performance of DeepBipolar and analyze the type of signals we believe could have affected the classifier in distinguishing the case samples from the control set.

KEYWORDS

bipolar disorder, convolutional neural network, deep learning, exome single-nucleotide polymorphisms analysis

1 | INTRODUCTION

Bipolar disorder, also known as manic-depressive illness, is a complex genetic disorder, resulting in hysteria or severe states of depression. Medical research reports that bipolar disorder affects close to 5.7 million adults in the United States aged 18 or above ("Bipolar disorder statistics," 2014). Suicides alone kill 15%–17% of the bipolar disorder patients. Around 51% of the patients are estimated to go untreated due to lack of efficient and effective detection and treatment solutions ("Bipolar disorder statistics," 2014). Twin studies of bipolar disorder have yielded estimates of heritability up to 90% (Craddock, & Sklar, 2013), the highest among all mental disorders (Shih, Belmonte, & Zandi, 2004; Sullivan, Daly, & O'donovan, 2012). The Psychiatric Genomics Consortium (PGC) (Psychiatric GWAS Consortium Bipolar Disorder Working Group, 2011) identified 19 genome-wide significant loci for this disorder based on genome-wide association studies (GWAS). This study was the first significant result identifying genetic bipolar disorder factors. Motivated by these recent discoveries, Critical Assessment of Genome Interpretation (CAGI) announced the bipolar disorder exomes challenge, in search of predictor methods for

bipolar phenotypes from specific genotypes. As a part of the challenge, we proposed a deep learning-based framework (called DeepBipolar) to predict bipolar disorder phenotypes from genotypes.

Deep learning combines lower-level representations to yield higher-level representations of the input datasets. Unlike conventional/shallow machine learning techniques that require feature engineering, deep learning models learn the features automatically through multiple layers (tens or hundreds of layers) of hierarchical representations, without handcrafted features (LeCun, Bengio, & Hinton, 2015). Deep learning achieved state-of-the-art results in several domains, such as computer vision (Krizhevsky, Sutskever, & Hinton, 2012), natural language processing (Mikolov, Karafiat, Burget, Cernock, & Khudanpur, 2010), and speech recognition (Hinton et al., 2012), with applications in fields such as genomics (Zhou & Troyanskaya, 2015) and astronomy (Huertas-Company et al., 2015) as well. Some commonly used deep learning models include deep autoencoders (Hinton & Salakhutdinov, 2006), Convolutional Neural Networks (LeCun, Bottou, Bengio, & Haffner, 1998), and long short-term memory (Hochreiter & Schmidhuber, 1997). DeepBipolar leverages the design of convolutional models and performs convolutions

with trainable filters/kernels on the input genomic sequences. The convolving kernels move around in a window length across the input sequences, forming feature maps of the input sequences. Additionally, adding further layers of convolution causes the extraction of richer intricate features within the data, resulting in feature maps. Pooling is a common operation applied on these feature maps and is based on spatial invariance in the input dataset. The idea is that key features are more important than their surroundings. Two popular pooling techniques are max pooling and mean pooling.

DeepBipolar combines the state-of-the-art deep convolutional neural networks with biological indicators to predict these phenotypes. We show that our technique achieved the best accuracy in predicting the diseased group from the control group, while substantially outperforming other conventional machine learning techniques. After elaborating on the framework, we focus on explaining the features learnt by the model to decipher the biological indicators of the disease. Then, through an understanding of the features learnt by the model, we carry out post analysis studies to identify the key contributors to the disorder. We found that the data could have issues related to sequencing. Identifying these as practices prevalent in the exomes sequencing community, we acknowledge and highlight the drawbacks of such a data preparation. Machine learning algorithms are suitable for identifying features that differentiate between various data points. Traditionally, experiments are designed to minimize any random noise in the data preparation phase to ensure the employed machine learning or statistical methods work on the actual biological indicators of the disorder and less on the artifacts. Therefore, the scope for systematic or batch artifacts is high in genome sequencing studies, based on such approaches that are used to sequence the case and control sets. Furthermore, some cost-effective measures used in genomic communities, say, using a control set of another sequencing study, can introduce artifacts in the datasets that are picked up by the machine learning model. After introducing our framework, we identify the features learnt and design experiments to measure the usability of the exome sequencing SNP for identifying bipolar disorder phenotypes.

2 | RELATED WORK

Bipolar disorder studies are based on the structure of the human brain and its activities using functional magnetic resonance imaging and positron emission tomography instruments ("Bipolar disorder," 2016). These studies highlight the smaller size of the prefrontal cortex of the brain in bipolar patients, with the brain activity in such patients being similar to patients suffering from other brain disorders like schizophrenia ("Bipolar disorder," 2016).

Twin studies of bipolar disorder show heritability as the most responsible factor causing the disorder (Craddock & Sklar, 2013). This factor is close to 90% and is among the highest across all other mental disorders (Shih et al., 2004; Sullivan et al., 2012). There have been previous studies analyzing the lifetime prevalence of the disorder and the heritability associated with it (Sullivan et al., 2012), affirming the link between nongenetic factors and bipolar disorder (Alloy et al., 2005), therefore leading to GWAS. Studies conducted by the PGC concluded

that there is less than 25% total variance in liability of bipolar disorder that could be explained by all SNPs (Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013). This shows that variants that were poorly tagged and missed by GWAS contribute to the risk of bipolar disorder (Bodmer & Bonilla, 2008; Cirulli & Goldstein, 2010; Manolio et al., 2009). These studies prompted bipolar disorder study using exome sequencing with the bipolar exome challenge, released by CAGI, being one among them.

Various machine learning algorithms have been employed for predicting phenotype from genotype, but there are few works in the literature that leverage the remarkable capability of deep neural networks for such tasks. Deep learning works on spatial invariance between input features, especially convolutional neural networks that are used extensively in computational biology. This model utilizes the sliding window technique to understand the interactions between the SNPs and their context. For various challenging problems like variant interpretation, deep learning algorithms have yielded state-of-the-art results with genome sequences data (Alipanahi, Delong, Weirauch, & Frey, 2015; Xiong et al., 2015). During sequencing analysis, a large number of variants poses a greater challenge and carrying out studies on such a large pool of the human population is tedious and expensive. This necessitates higher and superior dimensionality reduction algorithms for machine learning models to work efficiently, and the promise shown by convolutional autoencoders to extract and compose robust features is effective, as shown by previous research (Vincent, Larochelle, Bengio, Manzagol, 2008).

3 | DEEPBIPOLAR

This section introduces the dataset we worked on the challenge related to the prediction of bipolar phenotype. We also elaborate on our post result analysis the insight we gained to better understand the data and through our DeepBipolar framework.

3.1 | Datasets

The dataset was provided by The Regents of the University of California under the challenge "Bipolar Exomes" in the CAGI-4 experiment. The dataset consists of 1,000 samples, whose exome sequences are available, and 500 samples of the dataset are labeled to indicate whether the sample belongs to the diseased or the control group. Within the labeled dataset, there is a 50-50 split between the test and the control group. These data samples were randomly sampled from the pool of 1,000 samples. The challenge is to predict the labels of the next 500 samples to indicate whether they belong to the diseased or the control group. For target capture and sampling, 60 genes of bipolar disorder were used. Intronic information of 1,422 synaptic genes and NimbleGen SeqCap EZ v2.0 Exome arrays were also utilized (CAGI-4, 2016). Additionally, variants with more than 10% missing data or a Hardy-Weinberg disequilibrium value of P less than 10^{-6} with the genotype of read depth less than 10, or genotype quality less than 20 were excluded (CAGI-4, 2016).

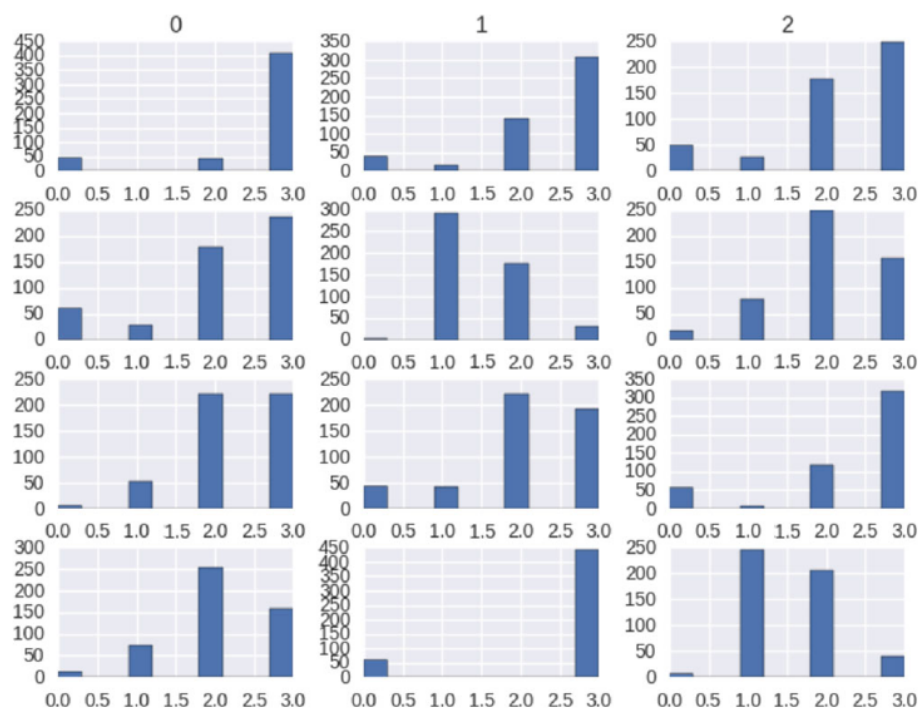


FIGURE 1 Histogram plot of bipolar datasets

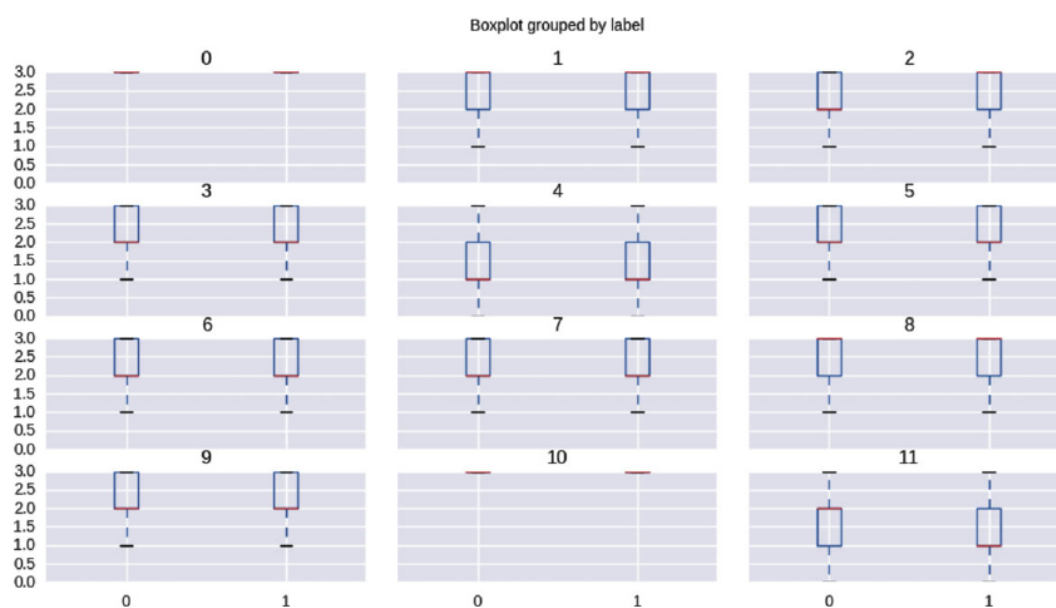


FIGURE 2 Boxplot of bipolar datasets

3.2 | Data visualization

Histograms and box plots are plotted to visualize the distribution of our datasets. It is easier to plot the distribution as all the genotypes in our dataset are unphased and we encode them with the numerical value, as explained in the data preprocessing section. As the number of variants for all the chromosomes is greater than $5e5$, we use feature selection to reduce feature dimension. This is achieved through L1-based

feature selection with regularization penalty parameter, $C = 0.85$. This reduces to less than 1% of the total features and we use the top 12 variants for visualization purposes. Figures 1 and 2, are the histogram and box plot, respectively, for our bipolar datasets. Similarly, Figure 3 summarizes the mapping quality, quality depth, and genotype quality for our data sets, without employing any threshold on low-quality variants.

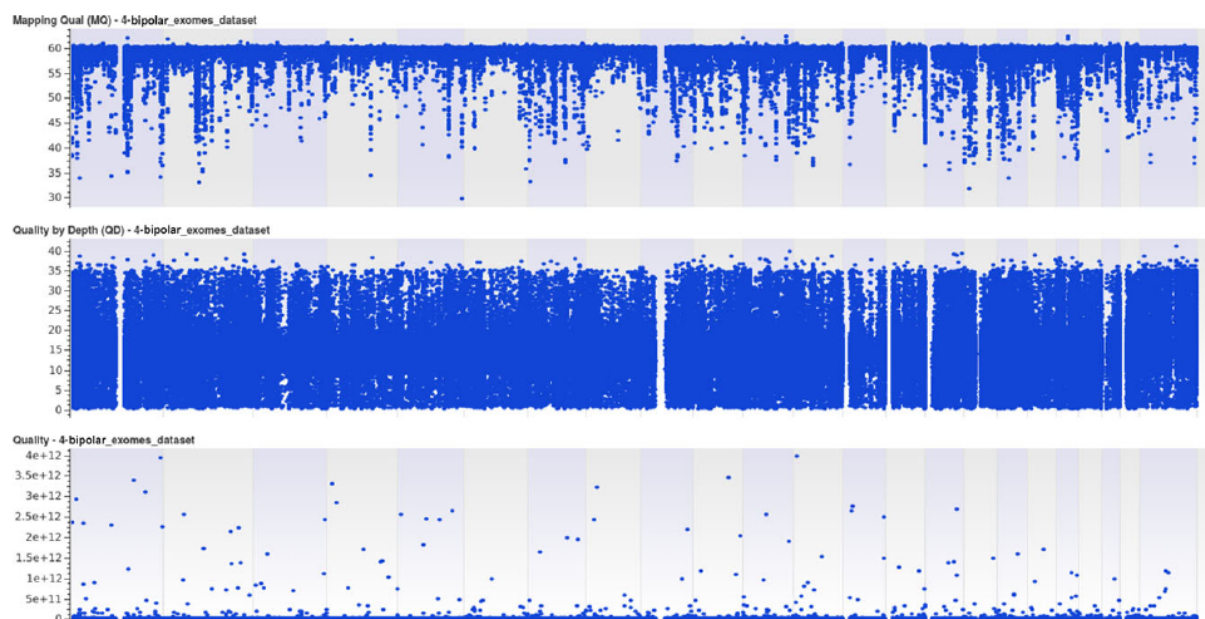


FIGURE 3 Mapping quality, quality depth, and genotype quality of bipolar datasets

3.3 | Data preprocessing

The data have the genotype information along with allele depth, read depth, and genotype quality. To make the data consumable for deep learning models, we extract the genotype information for each variant that has a genotype quality greater than 50. As convolution works on spatial invariance and there is no ordering between variants across chromosomes, we use convolution within variants belonging to the same chromosome. For all the variants, we had their unphased genotype information, for example, “reference allele”/“alternate allele” 1, “reference allele”/“alternate allele” 2, and so on. To ensure the deep learning models are unbiased and do not favor one genotype over the other, we create one-hot-encoded vectors for all the types of genotypes available. 0/0 is made as 001, 0/1 is made as 010, and so on. This encoding ensures that all the categories are equidistant from each other.

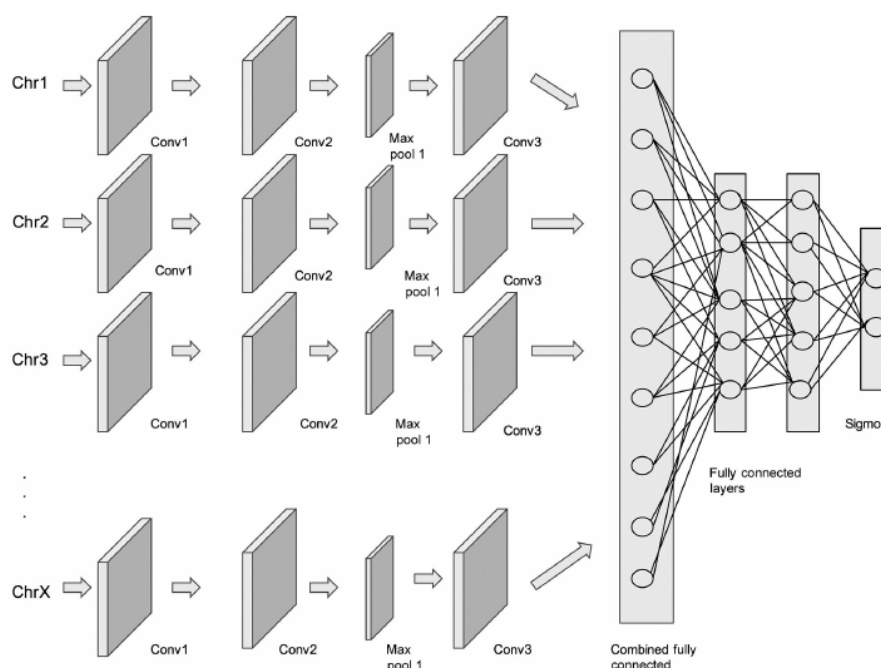
We also separate out variants from each chromosome separately because deep learning networks work on spatial invariance. There is an inherent ordering of variants among each chromosome but there is no such ordering between chromosomes. For this reason, convolution could not be applied over all the chromosomes, and is hence separated. Next, for each of the samples, we feed to our deep learning model all its variants across each chromosome and the target label of 1 or 0, with 1 being the diseased group. For validation purposes, we randomly sample 400 from both the case and the control groups for training and remaining 100 with labels for validation, and later predict the remaining 500 unlabeled test sets using this trained model. In typical optimization algorithms, it is important for the loss function to be computed on both the classes because in an unbalanced dataset, machine learning models tend to favor the bigger class as it reduces the loss. Here, we ensure that it is less likely for our models to learn one-class features better as we have evenly split the training dataset.

3.4 | DeepBipolar architecture

Inspired by state-of-the-art convolutional architectures, the architecture of DeepBipolar takes in a 23-channel input, with each input being the variants of a chromosome starting from 1 to 22 and chromosome X.

The richer and finer interactions between the variants of each chromosome are captured by the 1D convolution and are applied across each channel of input. Figure 4 illustrates our DeepBipolar framework. Specific details of the number of layers of convolution with respect to each channel is obtained from Table 1. We have a block of two convolutional layers, followed by a max-pooling layer resulting in sparser feature maps containing only the most important features. In each block, the max-pooling layer is followed by a batch normalization layer that helps in keeping the activation of the neurons in nonsaturation regions. This further avoids overfitting the data by regularizing the activations of each max-pooling layer. Next, we combine the feature maps from all the input channels by using a fully connected layer. This fully connected layer is then connected to one other fully connected layer with the final layer being the sigmoidal layer. This sigmoidal layer produces probabilities of output from 0 to 1, with control set belonging to class 0 and diseased group belonging to class 1.

Each neuron consists of ReLU activation to introduce nonlinearity in the network. This helps the model get deeper by eliminating the vanishing gradient problem observed with tanh or sigmoid nonlinear activations of neurons. The deep neural network is then trained using gradient descent by standard backpropagation algorithm (LeCun et al., 2015). We use a much quicker version of gradient descent, that is, stochastic gradient descent with a batch size of 32, that is, the gradient descent happens for all the 32 samples together. This makes the order of input data important and, hence, for this reason, we reshuffle the input training dataset after every epoch. At the end of one epoch

**FIGURE 4** DeepBipolar architecture**TABLE 1** Bipolar CNN architecture

Chromosome	Layer	Number of kernel	Kernel size	Pool size
Chromosome 1–chromosome X	Convolution 1	64	60	–
	Convolution 2	64	60	–
	Max pool 1	–	–	2
	Convolution 4	128	30	–
	Convolution 5	128	30	–
	Max pool 2	–	–	2
	Fully connected 1	64	–	–
Combined layers	Combined fully connected 1 (merge mode: sum)	128	–	–
	Combined fully connected 2	64	–	–
Output layer	Sigmoid	2	–	–

over all the batches, we obtained two sets of vectors. The first set of vectors is the value of the predicted classes for all the training samples, and the second set of vectors is the true class of the samples in the training set. Taking these two vector, we calculate the Bernoulli distance by computing the KL divergence. This step provides us with the loss function, representing a distance measure between the predicted and true labels. We then run the next epoch, initializing weights to reduce the loss function obtained in epoch 1. Each epoch runs until the model converges with a patience of five epochs, that is, the training was stopped when the validation accuracy did not improve for more than five epochs, usually implying model overfitting (on the training data). Empirically, it takes around 100 epochs to converge with a learning rate of 10^{-4} using the Adam optimizer. Training the model end-to-end takes close to 6 hr on Nvidia Tesla M40 servers. In Figure 5, the whole process of training is illustrated by taking chromosome 1 as an example. The sample process went into all the chromosome channels.

3.5 | Comparison with conventional machine learning techniques

Apart from convolutional neural networks, we also employ decision trees and random forests techniques for obtaining the classification results. We train 400 samples for training and 100 for validation with the rest of the 500 samples for testing. Accuracy of 0.536 in the case of decision trees and 0.548 for random forests with 10 estimators (default) is recorded. Classification score summary for both the approaches is given in Table 2. AUC scores recorded were 0.53 and 0.55, respectively.

ROC curve for the decision trees and random forests is shown in Figure 6. We applied a threshold of 0.5. Both of these classifiers are able to classify the bipolar disorder patient almost equally well as shown in Figure 6. Compared with traditional machine learning techniques, the winning performance results were obtained by using a deep

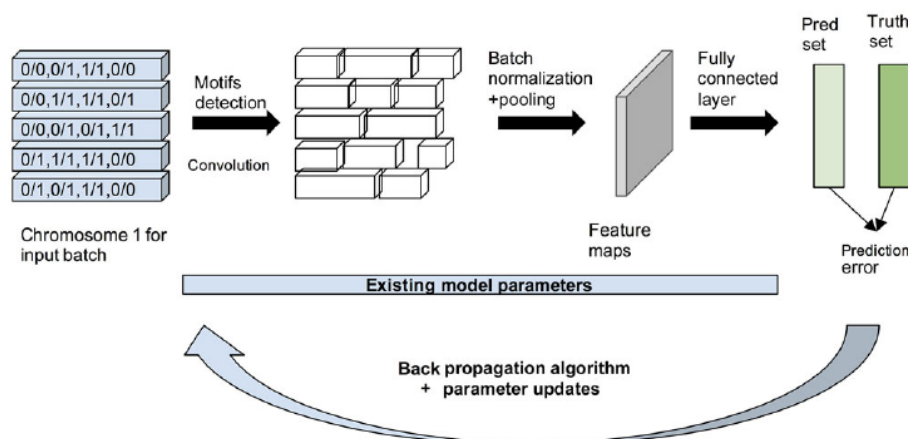


FIGURE 5 Model training workflow of DeepBipolar

TABLE 2 Classification score summary for decision tree/random forest/convolutional neural network

	Precision	Recall	F1 score	Support
Unaffected-0	0.53/0.53/0.63	0.53/0.73/0.68	0.53/0.62/0.65	249
Bipolar disorder-1	0.54/0.58/0.66	0.54/0.37/0.61	0.54/0.45/0.63	251
Average/total	0.54/0.56/0.64	0.54/0.55/0.64	0.54/0.53/0.64	500

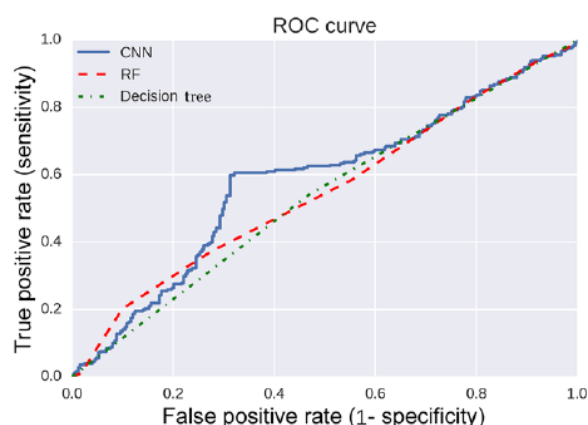


FIGURE 6 ROC curve of DeepBipolar versus shallow machine learning

TABLE 3 Prediction score for deepbipolar

Class	Predict unaffected	Predict bipolar disorder
Unaffected-0	169	80
Bipolar disorder-1	99	152

convolutional neural network architecture. The accuracy of the deep convolutional neural network model in predicting both the classes is about 65%. The typically hard nature of the problem might explain such a modest accuracy metric winning the prediction challenge. Table 3 and Figure 6 illustrate the performance metrics of the convolutional design. Tables 2 and 3 describe the accuracy with precision, recall parameters, and Figure 6 is the ROC curve for this approach. We applied Scikit-learn (v0.181) and Keras (v1.2.2) Python software packages for these analyses.

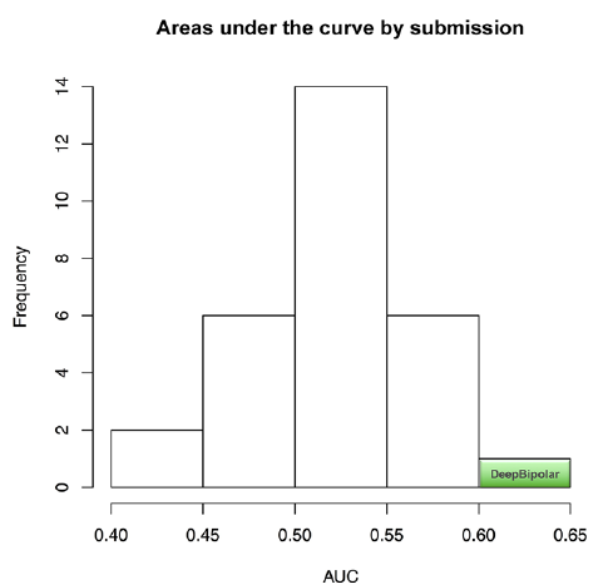


FIGURE 7 Performance of DeepBipolar against other competitors

3.6 | Performance of other competitors

There were 29 submissions overall for the challenge and the AUC for these submissions is shown in Figure 7 (Morgan, 2016). The prediction methods had an AUC range from 0.40 to 0.65 with most predictions ranging between 0.5 and 0.55. Our method is in the far right of the AUC curve, having a winning AUC score of 0.65; it achieved the best prediction metric among the fellow competitors. This demonstrates the ability of deep convolution-based neural networks to understand genomic datasets without handcrafted features. Having introduced the framework and proving its superiority over conventional machine learning

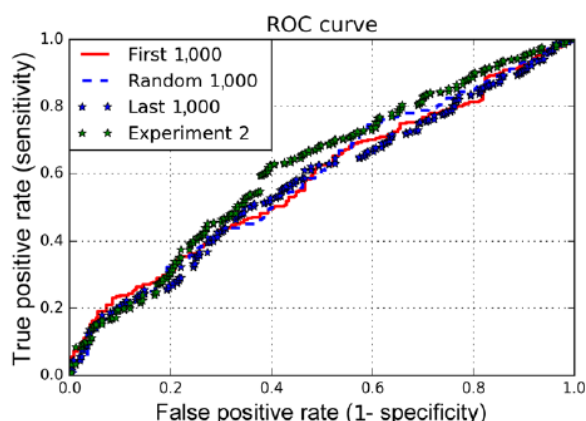


FIGURE 8 ROC curves for experiment 1 and 2

model, we proceed to understand the features learnt and the usability of the data in the coming sections.

3.7 | Post result analysis

As a post result analysis, we want to analyze the features learnt by the DeepBipolar framework and the suitability of the exome sequencing SNPs for the prediction of bipolar disorder. The fact that such expensive studies of bipolar disorder yielded biologists with few credible biological indicators responsible for the disorder made us question whether the features present in the dataset made the machine learning effective. The key question was: "Are there some artifacts that were unknowingly introduced into the dataset during the data preparation process?" An example of an artifact could be sequencing of the disease group together and using the control set that was sequenced for another study. As mentioned below, we ran two experiments where we restricted the information available to the model by reducing the variants and observed how it classified the samples. By changing the variants as seen by the model, we expected a drop in the classification performance and the experiments below were designed to study it.

3.7.1 | Experiment 1: Controlling variants in each chromosome

The hypothesis we are trying to answer is, "If we reduce the number of variants in each chromosome channel, does the performance of the machine learning models drop and do these models fail to accurately classify the two classes of sample set?" Essentially, if no changes in the classification ability of the machine learning models are observed, then the machine learning models are most likely picking up some artifacts that differentiate between these classes.

1. We pick only the first 1,000 variants of each chromosome.
2. We picked at random 1,000 variants of each chromosome.
3. We pick only the last 1,000 variants of each chromosome.

ROC curve for all three cases, along with that of experiment 2 is shown in Figure 8. Prediction score for our three experiments are shown in Table 4, where first, second, and third values correspond to

TABLE 4 Prediction score for experiment 1

Class	Predict unaffected	Predict bipolar disorder
Unaffected-0	137/131/130	112/118/119
Bipolar disorder- 1	114/106/98	137/145/153

first 1,000, random 1,000, and last 1,000 variants of each chromosome, each separated by "/." The results of all the three techniques of data preprocessing with our machine learning model are identical. The ROC curves indicate the model's predictions being significantly affected by the absence of the variants across these chromosomes. Therefore, the model needs all the variants to replicate the winning performance at the CAGI challenge. Had there been artifacts present in the dataset, the models might have performed without a drop in the performance and that would have been a consistent signal of case and control sets across the variants and chromosomes.

3.7.2 | Experiment 2: Training with chromosome 1 and 2/testing with chromosome 5, 6, 7, and 8

In this experiment, we train the deep learning models with variants only from chromosomes 1 and 2. During the testing time, we show the variants only from chromosomes 5, 6, 7, and 8. Given that an artifact signal would be consistent, even without showing the model all the chromosomes, the model should be able to replicate the winning performance, that is, a model that trains with all variants. As we see, there is a huge drop in the performance. Any differences that might arise due to sequencing should be uniform across chromosomes and the models would have expected to work similarly well in training on one set of chromosomes and testing on another set. Given the significant drop in performances, the likelihood of the presence of such signals being picked up by the neural network models for classification is quite remote as the performance takes a considerable hit on these experiments.

4 | OBSERVATIONS

We initially suspected the neural networks of picking up signals arising out of different sequencing technologies, but the experimental results suggest otherwise. The experiments were designed to show the presence of signal in the datasets to distinguish the case from the control case. The results from the experiment sufficiently prove the existence of such a signal and have shown that the neural networks are picking up signals that are not differences in sequencing technologies. Given the lack of interpretability with neural networks, we had to rely on the input signal and the changes it effects on the output performance. Given that convolutional neural networks excel at picking interlinked features, the models might be picking up complex patterns across the samples to classify the diseases and rely less on noise. Recently, there have been significant advances in interpreting neural networks and, hence, as a future work, we plan to analyze and score each of the variants with respect to its impact on classifying the variants as benign or pathogenic.

5 | CONCLUSION

We propose a deep learning-based framework called DeepBipolar and analyze the genotype information to predict the bipolar phenotype using convolutional neural networks. The technique described in this paper achieved the winning results in the CAGI-4 bipolar prediction challenge. Given the lack of need for hand-engineering the feature set for classifying the disease samples, such a technique might be a useful tool in computational biology.

ACKNOWLEDGMENT

We would like to thank the organizers of the Bipolar Disorder challenge, the Critical Assessment of Genome Interpretation (CAGI) experiment, and the data providers for the challenge. The CAGI experiment coordination is supported by NIH U41 HG007446 and the CAGI conference is supported by NIH R13 HG006650. The work presented in this paper is supported in part by NIH R01GM110240.

REFERENCES

- Alipanahi, B., Delong, A., Weirauch, M. T., & Frey, B. J. (2015). Predicting the sequence specificities of DNA and RNA-binding proteins by deep learning. *Nature Biotechnology*, 33, 831–838.
- Alloy, L. B., Abramson, L. Y., Urosevic, S., Walshaw, P. D., Nusslock, R., & Neeren, A. M. (2005). The psychosocial context of bipolar disorder: Environmental, cognitive, and developmental risk factors. *Clinical Psychology Review*, 25(8), 1043–1075.
- Bipolar disorder. (2016, April). National Institute of Mental Health and Education. Retrieved from <https://www.nimh.nih.gov/health/topics/bipolar-disorder/index.shtml>
- Bipolar disorder statistics. (2014, August 1). *Biopolarlives*. Retrieved from <http://www.bipolar-lives.com/bipolar-disorder-statistics.html>
- Bodmer, W., & Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics*, 40, 695–701.
- CAGI-4. (2016). Distinguish between exomes of bipolar disorder patients and unaffected individuals, San Francisco, CA, USA. Retrieved from <https://genomeinterpretation.org/content/4-bipolar-exomes>.
- Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, 11, 415–425.
- Craddock, N., & Sklar, P. (2013). Genetics of bipolar disorder. *The Lancet*, 381(9878), 1654–1662.
- Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., ... International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, 45, 984–994.
- Disorders share risk gene pathways for immune, epigenetic regulation, genome-wide findings add to evidence blurring traditional psychiatric categories. (2015, January 29). *National Institute of Mental Health*. Retrieved from <https://www.nimh.nih.gov/news/science-news/2015/disorders-share-risk-gene-pathways-for-immune-epigenetic-regulation.shtml>
- Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-R., Jaitly, N., ... Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, 29, 82–97.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313, 504–507.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Huertas-Company, M., Gravet, R., Cabrera-Vives, G., Pérez-González, P. G., Kartaltepe, J. S., Barro, G., ... McIntosh, D. H. (2015). A catalog of visual-like morphologies in the 5 candels fields using deep learning. *The Astrophysical Journal Supplement Series*, 221, 1–23.
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *Proceedings of the conference on Advances in Neural Information Processing Systems Foundation*, Lake Tahoe, NV, USA, 1106–1114.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE. USA: IEEE*, 2278–2324.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., ... Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461, 747–753.
- Mikolov, T., Karafiat, M., Burget, L., Cernock, J. H., & Khudanpur, S. (2010). Recurrent neural network based language model. *Interspeech*, 2, 1045–1048.
- Morgan, A. (2016). Bipolar assessors presentation. CAGI conference. San Francisco, CA, USA.
- Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*, 43, 977–983.
- Shih, R. A., Belmonte, P. L., & Zandi, P. P. (2004). A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *International Review of Psychiatry*, 16(4), 260–283.
- Sullivan, P. F., Daly, M. J., & O'donovan, M. (2012). Genetic architectures of psychiatric disorders: The emerging picture and its implications. *Nature Reviews Genetics*, 13(8), 537–551.
- Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on machine learning (ICML-08)*. *The International Machine Learning Society*, Helsinki, Finland. 1096–1103.
- Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K. C., ... Frey, B. J. (2015). The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 347(6218), p. 1254806 (1–9)
- Zhou, J., & Troyanskaya, O. G. (2015). Predicting the effects of noncoding variants with deep learning-based sequence model. *Nature Methods*, 12, 931–934.

How to cite this article: Lakshman S, Bhat RR, Viswanath V, Li X. DeepBipolar: Identifying genomic mutations for bipolar disorder via deep learning. *Human Mutation*. 2017;38,1217–1224. <https://doi.org/10.1002/humu.23272>