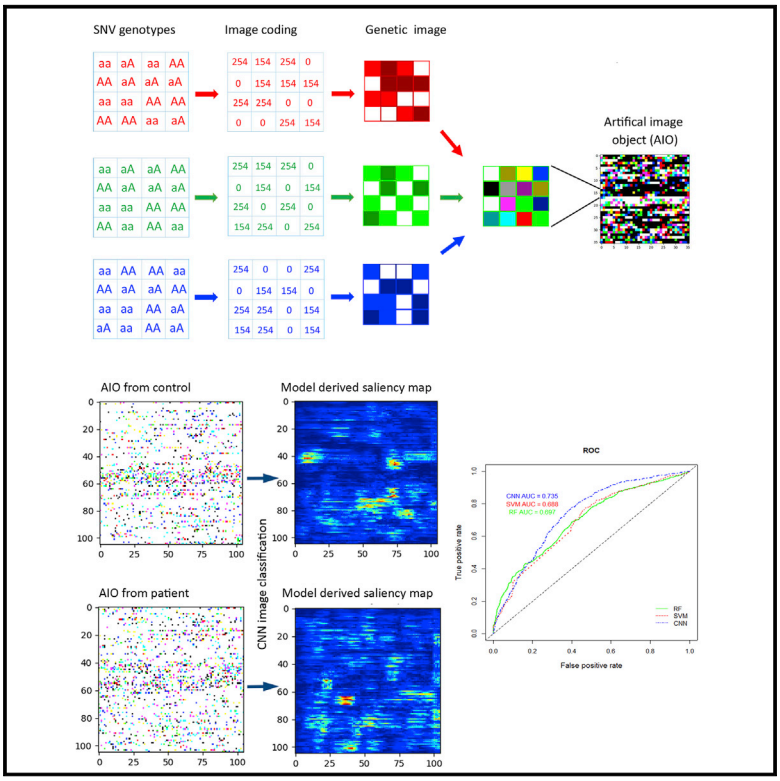
Patterns

Artificial image objects for classification of schizophrenia with GWAS-selected SNVs and convolutional neural network

Graphical abstract



Highlights

- Introduce a technique to transform genomics data into AIOs
- Apply CNN algorithms to classify genomics derived AIOs
- Showcase the technique with GWAS-selected SNVs to classify schizophrenia diagnosis

Authors

Xiangning Chen, Daniel G. Chen, Zhongming Zhao, Justin Zhan, Changrong Ji, Jingchun Chen

Correspondence

va.samchen@gmail.com (X.C.), jingchun.chen@unlv.edu (J.C.)

In brief

GWAS has discovered genetic risk variants for many diseases. It is a challenge to use these variants to promote effective prevention, diagnosis, and treatment. In this report we introduce a new technique to transform genomic data into artificial image objects and apply CNN algorithms to classify these image objects. Using GWAS-selected SNVs, we can predict schizophrenia diagnosis with a reasonable accuracy. This technique can be applied to other omics data and combine them together to model disease risks.

Chen et al., 2021, Patterns 2, 100303
August 13, 2021
https://doi.org/10.1016/j.patter.2021.100303

CellPress

Patterns

CellPress
OPEN ACCESS

Article

Artificial image objects for classification of schizophrenia with GWAS-selected SNVs and convolutional neural network

Xiangning Chen,[1,2,7,*] Daniel G. Chen,[1] Zhongming Zhao,[3,4] Justin Zhan,[5] Changrong Ji,[2] and Jingchun Chen[6,*]
[1]410 AI, LLC, 10 Plummer Ct, Germantown, MD 20876, USA
[2]A3.AI INC., 10530 Stevenson Road, Stevenson, MD 21153, USA
[3]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[4]Department of Psychiatry and Behavioral Sciences, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[5]Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, AR 72701, USA
[6]Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA
[7]Lead contact
*Correspondence: va.samchen@gmail.com (X.C.), jingchun.chen@unlv.edu (J.C.)
https://doi.org/10.1016/j.patter.2021.100303

THE BIGGER PICTURE Genome-wide association studies have discovered many genetic variants that contribute to human diseases. However, it remains a challenge to effectively utilize these variants to facilitate early and accurate diagnosis and treatment. In this report, we propose a new approach that transforms genetic data into AIOs so that they can be classified by advanced artificial intelligence and machine learning algorithms. Using schizophrenia as a case study, we demonstrate that genetic variants can be transformed into AIOs and that the AIOs can be classified by CNN algorithms consistently. Our approach can be applied to other omics data and combine them to jointly model disease risks and treatment responses.

Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

In this article, we propose a new approach to analyze large genomics data. We considered individual genetic variants as pixels in an image and transformed a collection of variants into an artificial image object (AIO), which could be classified as a regular image by CNN algorithms. Using schizophrenia as a case study, we demonstrate the principles and their applications with 3 datasets. With 4,096 SNVs, the CNN models achieved an accuracy of 0.678 ± 0.007 and an AUC of 0.738 ± 0.008 for the diagnosis phenotype. With 44,100 SNVs, the models achieved class-specific accuracies of 0.806 ± 0.032 and 0.820 ± 0.049, and AUCs of 0.930 ± 0.017 and 0.867 ± 0.040 for the bottom and top classes stratified by the patient's polygenic risk scores. These results suggest that, once transformed to images, large genomics data can be analyzed effectively with image classification algorithms.

INTRODUCTION

In the last decade, large-scale genome-wide association studies (GWASs) have identified risk variants for many human diseases and other complex traits,[1] including schizophrenia.[2] This provides a great opportunity to utilize these variants to promote personalized medicine through early diagnosis, prevention, and optimized treatment. However, common diseases have complex genetic architecture, thousands of genetic variants, if not more, contribute to these diseases.[3,4] The effects of the individual variants are very small. These make it a challenge to develop an effective strategy to utilize these GWAS-identified variants for early diagnosis and better treatment. In the literature, the use of polygenic risk score (PRS) is popular and it has been demonstrated that PRS has high predictive power for many diseases.[5–8] There are also reports that a panel of selected risk variants could predict clinical outcomes. While useful and promising, these approaches have some limitations. For the PRS

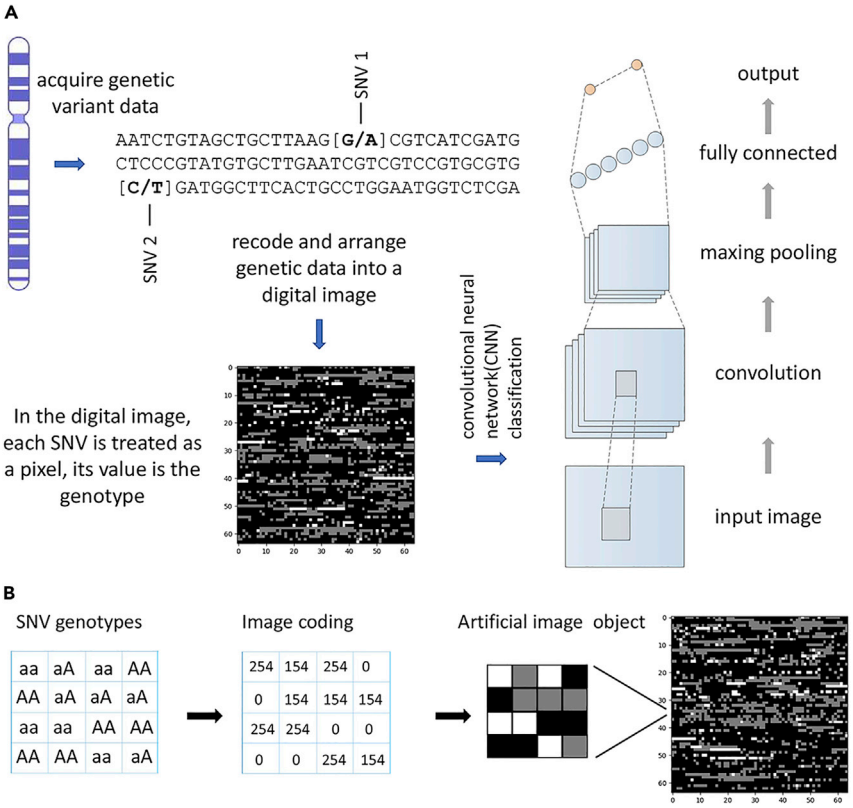


Figure 1. Schematic drawing showing the overall design of the study

(A) Study design.

(B) Recoding of SNV genotypes into an artificial image object (AIO). A set of SNVs can be selected from GWAS data and recoded and rearranged into an AIO. In the process, each SNV is treated as a pixel, and its genotype, AA, aA, or aa, is assigned a unique, arbitrary value. In the figure, 0, 154, and 254 are used for the 3 genotypes. The AIOs made from the genetic markers of a group of individuals can be classified as digital images by advanced AI algorithms, such as a convolutional neural network (CNN).

approach, the effects of individual variants could not be followed, because the effects of individual variants are aggregated across the genome. For the approach that uses selected variants to form a panel, its predictive power is generally low, and most models do not consider the interactions among these selected markers. Can we develop a new approach that preserves the effects of individual markers and at the same time harnesses the power of hundreds of thousands of markers collectively?

In recent years, computer science has seen rapid progress in image classification analysis using artificial intelligence (AI) and machine learning (ML) algorithms, exemplified by the highly accurate identification of objects in image files.[9,10] Many of these advanced techniques have been implemented in software frameworks, such as TensorFlow[11] and Keras (https://keras.io/about/). One approach, the convolutional neural network or CNN, has been widely used in image classification and prediction, and has achieved unprecedented accuracy. Impressed by the performance and simplicity of the implementation, we have thought of whether we can adapt these algorithms to the analyses of genetic and genomic data.

Imaging analysis relies on the detection and extraction of spatial patterns of image elements, such as pixel intensity and color. This process is commonly referred to as feature extraction and is accomplished by the CNN algorithm. In the process, images are analyzed pixel by pixel. In an image, a pixel can carry a series of values and colors. In contrast, genetic markers, for a majority of them, have limited variations. For example, for many single-nucleotide variations (SNVs), the most common form of genetic variation, there are only two alleles. For a diploid organism, such as humans, there are only three possible genotypes for a given locus. If we consider each genetic marker as a pixel in a digital image, we can arrange a collection of genetic markers from an individual to form an artificial image object (AIO), then we can use the advanced image analysis algorithms to train and classify these AIOs as regular images. Based on this rationale, we obtained genotype data from the Molecular Genetics of Schizophrenia (MGS),[12] the Swedish Case-Control Study of Schizophrenia (SCCSS),[13] and the Clinical Antipsychotic Trials for Intervention Effectiveness (CATIE),[14,15] and made AIOs for all individuals in these datasets. Then we performed image classification using the TensorFlow/Keras packages and CNN architecture. The overall design of this study is outlined in Figure 1A. Here, we use schizophrenia as a case study to evaluate this new approach and report our findings from analyses of the AIOs derived from schizophrenia patients and normal controls.

RESULTS

Set I SNV classification of normal controls and diagnosed patients

Using the 4,096 set I SNVs selected by GWAS p value and linkage disequilibrium (LD) pruning, we made 64 × 64 AIOs for the subjects in the MGS, SCCSS, and CATIE datasets. We combined the MGS and SCCSS datasets as training data and used the CATIE data as testing data. Table 1 summarizes the results. The models achieved an accuracy of 0.678 ± 0.007 (mean ± SD) and an area under the curve (AUC) of 0.738 ± 0.008. The precision, recall, and f1-score were 0.677 ± 0.013, 0.675 ± 0.029, and 0.675 ± 0.014, respectively. To evaluate the impact of confounding factors on the model performance, we used the PLINK program to calculate 20 principal components from imputed genotypes for the 3 datasets and made an ancillary AIO with sex and the 20 principal components. The ancillary AIO was integrated into the main model as a separate layer. We found that the impact was minimal, and the confounders did not change the overall performance. For example, in a typical run with the ancillary AIO included, the classification accuracy was 0.670,

Table 1. Classification of schizophrenia diagnosis and PRS class by set I SNPs

Phenotype	Group/Class	Accuracy[a]	AUC[b]	Precision	Recall	f1-score
Diagnosis	CTRL	0.666 ± 0.014		0.667 ± 0.015	0.711 ± 0.025	0.688 ± 0.011
	SCZ	0.687 ± 0.012		0.688 ± 0.011	0.638 ± 0.033	0.662 ± 0.016
	average	0.678 ± 0.007	0.738 ± 0.008	0.677 ± 0.013	0.675 ± 0.029	0.675 ± 0.014
PRS class	class I	0.747 ± 0.065	0.937 ± 0.006	0.749 ± 0.065	0.866 ± 0.035	0.801 ± 0.024
	class II	0.364 ± 0.023	0.747 ± 0.006	0.363 ± 0.024	0.523 ± 0.073	0.427 ± 0.037
	class III	0.839 ± 0.021	0.873 ± 0.012	0.840 ± 0.020	0.563 ± 0.069	0.673 ± 0.048
	average	0.650 ± 0.036	0.852 ± 0.008	0.650 ± 0.036	0.650 ± 0.059	0.634 ± 0.036

[a]For PRS classes, categorical accuracy is reported.
[b]For PRS classes, class-specific AUC is reported.

the AUC was 0.726, and the precision, recall, and f1-score were 0.670, 0.675, and 0.667, respectively, which were not significantly different from the results listed in Table 1.

To visualize the patterns of the correctly classified controls and cases, we compared the original AIOs along with the saliency maps derived from the model using the saliency map visualization method,[16] a technique that visualizes which pixels of the image contribute the most to the prediction. We selected correctly predicted individuals and calculated the saliency maps for these individuals. We found that the correctly classified AIOs did have some distinct features (see Figures 2A and 2B) between individuals. We confirmed the observation after visually inspecting hundreds of correctly classified cases and controls. To view the patterns between the control and case groups, we extracted all correctly classified controls and cases and calculated the mean values for each pixel of the saliency maps for each group. Then we plotted the mean values as an image for each group (Figure 2C, CTRL and SCZ) and the difference between the means of the two groups (Figure 2C, Diff btwn CTRL & SCZ). From these images, we could see that, at group level, the differences between the groups were almost all quantitative. The distinct differences observed between individual patients and controls were not observed at group level.

Set I SNV classification of PRS-stratified classes

Using the same set I SNVs, we set up models to classify PRS-stratified classes. In these analyses, the AIOs were the same for the subjects but the phenotype was changed from diagnosis to PRS classes. We used similar model architecture and strategy to classify these AIOs. Table 1 summarized the report matrices. From this table, it could be seen that, while the overall average of the accuracy was similar to that of the diagnostic model (0.678 ± 0.007 versus 0.650 ± 0.036), the categorical accuracy and class-specific AUC for classes I and III were significantly improved. For example, for class III, the categorical accuracy and class-specific AUC reached 0.839 ± 0.021 and 0.873 ± 0.012, respectively (Table 1), much better than the accuracy and AUC of the diagnostic model. We noticed that class II, which had a PRS score within 0.25 SD of the mean of diagnosed subjects, the performance was much worse than classes I and III. This was somewhat expected because class II had a PRS score between classes I and III, it was a mixed group with members from the high and low end of classes I and III, respectively. As shown in Figure S1, the PRSs from the controls and cases overlapped significantly.

Similar to the model that classified diagnosed patients and controls, we also compared the original AIO images with the model-generated gradient saliency maps for the three classes (Figures 3A–3C). Like what we observed for the diagnostic phenotype, there were distinct patterns between individual patients and controls. When we took the average pixel intensity for each class and plotted them as an image (Figure 3D), the patterns between the classes became more similar. To quantify the differences between class I and class III, which corresponded to the controls and cases in the diagnostic model, we took the difference of pixel intensity between class I and class III, and plotted them as an image. This image showed where the differences between the two classes were located. From the image, it was clear that many pixels, i.e., SNVs, were required to form the pattern that separated the two classes.

Set II SNV classification of normal controls and diagnosed patients

Set II SNVs were selected by GWAS p value $\leq 5 \times 10^{-5}$ and minor allele frequency ≥ 0.05. The 44,100 SNVs were arranged as a 210 × 210 AIO and classified using the CNN algorithm. The average accuracy for the models was 0.685 ± 0.021 and the AUC was 0.708 ± 0.030 (Table 2). When the performance of the two SNV sets were compared, the results were very close across all measured matrices (see Tables 1 and 2), suggesting that more SNVs did not improve the performance when the SNVs were selected from the same GWAS p value threshold. Similar to the set I SNVs, we also evaluated the impact of confounders on model performance. Using a similar approach to integrate the confounders into the main model, we found similar results as the set I SNVs that these confounders did not change the overall performance.

When we applied the saliency map visualization technique to examine and compare the model extracted feature maps (Figures 4A and 4B), the distinction between individual patients and controls was clear. However, at group level (Figure 4C, CTRL and SCZ), the differences became very difficult for humans to see but could be quantified (see the image between CTRL and SCZ).

We explored the effect of AIO configuration on classification performance using the set II SNVs, here the same 44,100 SNVs were arranged as a 105 × 105 × 4 AIO, the third dimension could be treated technically as different color channels as in regular digital images. We compared the performance of the

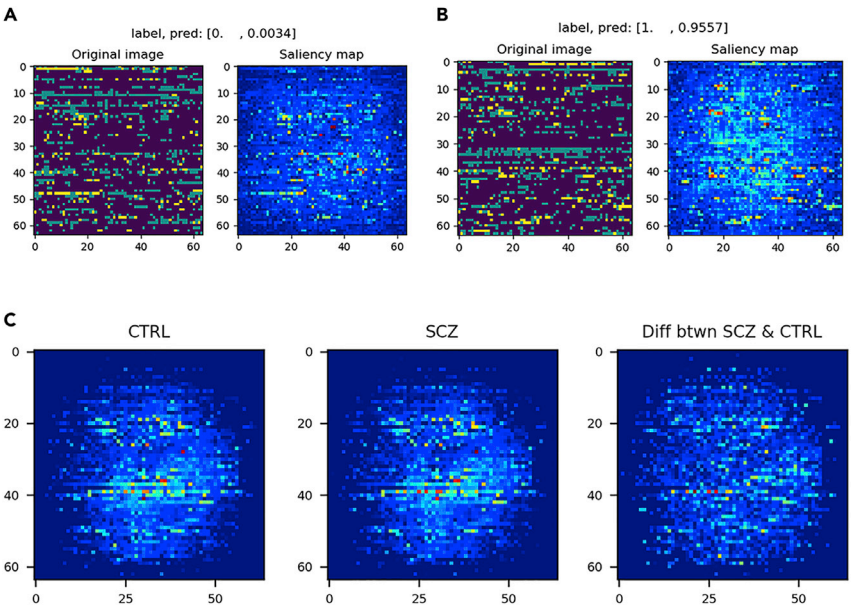


Figure 2. Classification of diagnosed patients and normal controls by set I SNVs

(A and B) Comparisons of original AIOs and model-predicted saliency maps for the normal control (CTRL) and diagnosed (SCZ) subjects. The images were plotted with the jet color map series (pseudo color) as defined in the pyplot library of the Matplotlib package. For the jet color map, the color changed from blue to green to red, with the red as the highest value. The label and pred in the panels were labels for the subjects (0 for normal controls and 1 for diagnosed schizophrenia patients) and predicted probabilities from the model.

(C) The saliency maps of the CTRL and SCZ groups and the difference (Diff btwn SCZ & CTRL) between the two groups.

210 × 210 and 105 × 105 × 4 configurations and found that they performed similarly (see Tables 2 and S1), the accuracies were 0.685 ± 0.021 and 0.657 ± 0.014 and the AUCs were 0.708 ± 0.030 and 0.708 ± 0.047, respectively, for the two configurations. When the saliency maps were compared, the colored AIOs, i.e., 105 × 105 × 4 AIOs, were more simplistic and distinctive (compare Figures 4 and S2). For the colored AIOs, some subjects had only a few clustered pixels that were critical to the correct classifications while other subjects had slightly more clustered pixel patterns. We also noticed that small changes in training hyperparameters could lead to quite different saliency maps, but the overall performance matrices did not change much (data not shown).

Set II SNV classification of PRS-stratified classes

We applied the set II SNVs to classify the PRS phenotype using the 210 × 210 AIO configuration. The performance of the model was summarized in Table 2. Similar to the results of set I SNVs, the performances for classes I and III (accuracy: 0.806 ± 0.032 and 0.820 ± 0.049; AUC: 0.930 ± 0.017 and 0.867 ± 0.040, respectively) were significantly better when compared with the results using the case control design (i.e., binary classification of normal controls and diagnosed patients) (Table 2). But the performances between the two SNV sets were grossly similar (see Tables 1 and 2) for all measured matrices.

We generated the saliency maps for some individuals (Figures 5A–5C for each of the three classes of the PRS-stratified phenotype) as we did for the set I SNV models. As seen previously, individually these subjects looked distinctive, but since the number of markers in set II were 10 times more than that in set I, the patterns seemed more complex in human eyes. Visually, it was difficult to see the differences between the classes (Figure 5D, class I, class II, and class III). To reveal the difference between class I and class III, we took the differences in pixel intensity between the two classes and plotted them as an image (Figure 5D, Diff btwn class I & class III). Although the feature differences between the classes looked alike by human eye, the classification model could distinguish the two classes with high discriminative power (class-specific AUC for classes I and III were 0.930 ± 0.017 and 0.867 ± 0.040) and reasonable sensitivity (class-specific sensitivity or recall for classes I and III were 0.815 ± 0.035 and 0.605 ± 0.115) (see Table 2).

Classification of diagnosis by PRS with RF and SVM algorithms

To compare the performance of our CNN model, we conducted classification analyses with the random forest (RF) and support vector machine (SVM) algorithms using the PRS calculated at $p = 5 \times 10^{-5}$, the same threshold we used to select SNVs for our AIOs, along with sex and 20 principal components. We did a grid search to find the optimal parameters for both the RF and SVM algorithms and used the selected parameters to build RF and SVM models separately. For the RF model, the classification accuracy for the training samples (MGS and SCCSS combined) was 0.675. With this model, we obtained an accuracy of 0.637 ± 0.003 and AUC of 0.699 ± 0.002 for the external CATIE samples. For the SVM model, the classification accuracy for the training samples was 0.678. Applying this SVM model to the CATIE samples, the accuracy and AUC were 0.647 ± 0.001 and 0.687 ± 0.001, respectively. For direct comparison, we plotted the receiver operating characteristics (ROCs) for the CNN, RF, and SVM models together (Figure 6). From the plot, it was clear that, while the performances of the RF and SVM models were comparable, our AIO-based CNN model was about 4% better.

DISCUSSION

In this study, we developed a method to transform genetic data into AIOs and then used advanced image classification algorithms to classify the AIOs. Our goal was to evaluate whether genetic data, once transformed into an image, could be used effectively to distinguish diagnosed patients from normal controls using schizophrenia as a case study.

In this study, we selected two sets of SNVs from GWAS data to create AIOs and applied CNN image classification algorithms to

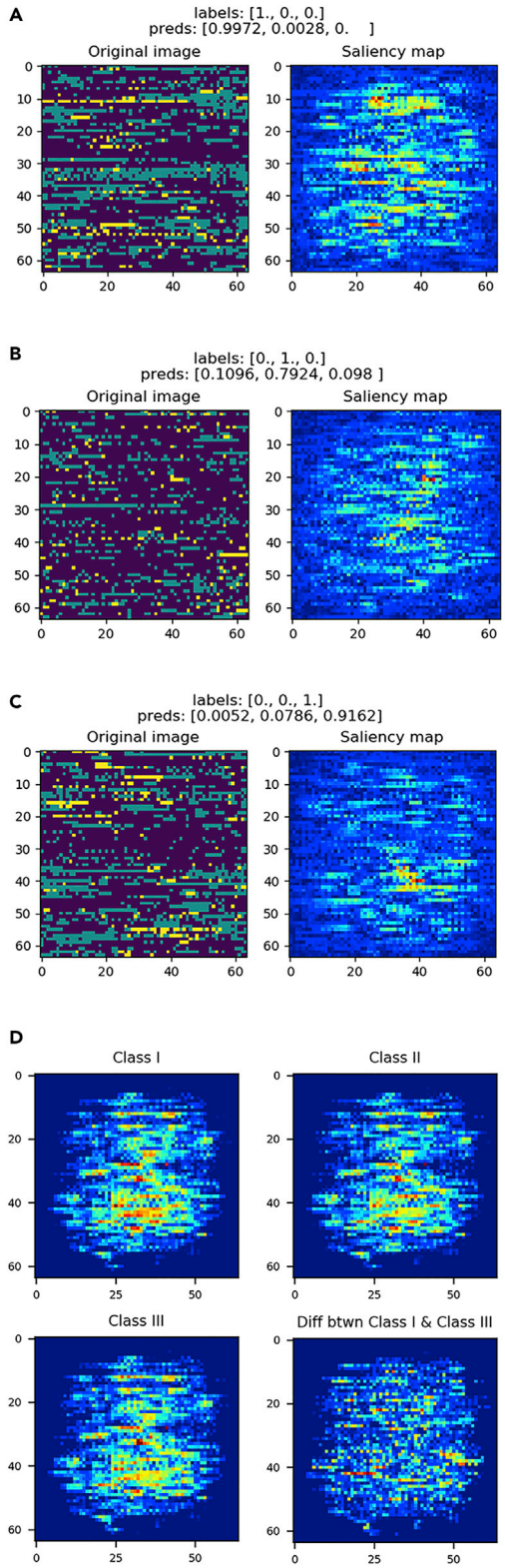


Figure 3. Classification of PRS-stratified phenotypes by set I SNVs

(A–C) Comparisons between original AIOs and saliency maps derived from the classification model for selected individuals from the three classes, respectively. The labels and preds in these panels listed the labels and prediction probabilities for the corresponding classes.

(D) The saliency maps for the three PRS classes by averaging the saliency maps of correctly classified subjects for each class (class I, class II, and class III). The fourth image in the panel was the difference in pixel intensity between class I and class III.

classify these AIOs. We used two phenotypes, one was the disease diagnosis status, for which we used the case control status for the three datasets directly. The other one was a PRS-stratified class, where we used the schizophrenia PRS calculated at $p = 5 \times 10^{-5}$ to group the subjects into three classes. Class I were those subjects whose PRSs were smaller than or equal to the mean of PRS of the diagnosed subjects minus 0.25 SD; class III were those subjects whose PRSs were greater than or equal to the mean of PRS plus 0.25 SD; and class II were those subjects whose PRSs were between classes I and III. Using the 4,096 set I SNVs, which were selected based on GWAS p value and LD pruning, the classification accuracy was 0.678 ± 0.007 and the AUC was 0.738 ± 0.008 for the diagnostic phenotype. For the PRS-stratified phenotype, while the overall average accuracy and AUC were similar to the diagnostic phenotype, the categorical accuracy and class-specific AUC for classes I and III were significant improved, reaching 0.747 ± 0.065 and 0.937 ± 0.006; and 0.839 ± 0.021 and 0.873 ± 0.012, respectively (Table 1). Using the 44,100 set II SNVs, which were selected based on GWAS p value and minor allele frequency, the classification results were similar, accuracy was 0.685 ± 0.021 and AUC was 0.708 ± 0.030 for the diagnostic phenotype. For the PRS-stratified phenotype, the categorical accuracy and class-specific AUC for classes I and III were 0.806 ± 0.032 and 0.930 ± 0.017, and 0.820 ± 0.049 and 0.867 ± 0.040, respectively (Table 2). Compared with the RF and SVM models using PRS to predict schizophrenia, the AIO-based classification model performed about 4% better. In the literature, there were multiple studies that used genetic markers to predict schizophrenia diagnosis,[17] including a study from our group.[18] Most of these studies used an SVM algorithm, and the variables were mostly individual SNVs. Only two studies, one of which was our previous study, used external validation samples. The AUCs from these studies varied substantially, from 0.54 to 0.95. The AIO results reported here fared better than most the studies reviewed, including our previous study,[18] which used not only the PRSs of schizophrenia but also PRSs from other comorbid traits. These results indicated that genetic data, once transformed into AIOs, could be used effectively to classify the diagnosed patients and unaffected controls. We evaluated the effect of sex and potential population stratification on model performance. We found that the models with the inclusion of sex and 20 principal components had similar performance as the models without these confounders. One possible reason might be that the influence of confounders on SNVs had changed after the SNVs were transformed into pixels.

The performance of the PRS-stratified classes was interesting. For both classes I and III, the categorical accuracies were better than 0.80 and class-specific AUC values were better than 0.85 when set II SNVs were used (Table 2). For set I SNVs, except for the categorical accuracy for class I, all other measures were similar to the set II performance. This raised an interesting question. If we could define schizophrenia with some objective

Table 2. Classification of schizophrenia diagnosis and PRS class by set II SNPs

Phenotype	Group/Class	Accuracy[a]	AUC[b]	Precision	Recall	f1-score
Diagnosis	CTRL	0.687 ± 0.027		0.690 ± 0.032	0.688 ± 0.016	0.690 ± 0.013
	SCZ	0.685 ± 0.014		0.680 ± 0.010	0.680 ± 0.054	0.683 ± 0.034
	average	0.685 ± 0.021	0.708 ± 0.030	0.685 ± 0.021	0.684 ± 0.035	0.686 ± 0.024
PRS class	class I	0.806 ± 0.032	0.930 ± 0.017	0.808 ± 0.030	0.815 ± 0.035	0.810 ± 0.006
	class II	0.376 ± 0.026	0.730 ± 0.052	0.373 ± 0.026	0.570 ± 0.114	0.450 ± 0.017
	class III	0.820 ± 0.049	0.867 ± 0.040	0.823 ± 0.035	0.605 ± 0.115	0.688 ± 0.065
	average	0.667 ± 0.036	0.842 ± 0.037	0.668 ± 0.031	0.663 ± 0.088	0.649 ± 0.029

[a]For PRS classes, categorical accuracy is reported.
[b]For PRS classes, class-specific AUC is reported.

measures, such as PRS, we would be able to predict the disorder with good confidence.

When the predicted feature patterns, the saliency maps, were examined, the distinctions between individual subjects were clear, and this was true for both the diagnostic and PRS class phenotypes (see Figures 2, 3, 4, 5, and S2). However, when we tried to find the common feature patterns for a group, the group-wise or class-wise feature patterns became more complex (see the difference panels in Figures 2, 3, 4, and 5). Group-wise distinctions seemed unclear; instead, the differences between groups were mostly seen as change of signal intensities for many pixels. It was possible, maybe likely, that there were multiple patterns for each group; when we took an average of multiple subjects, this would make the distinction between groups less clear. This was consistent with the observation in image classification, where the same object label, such as a dog or a cat, could present multiple feature patterns and these patterns did not have to have common features. Genetically, the same diagnostic group or class could have multiple variant profiles. In fact, there was ample evidence that schizophrenia is genetically heterogeneous; therefore, it should be no surprise that schizophrenia patients as a group did not show different predicted feature maps or saliency maps as observed at individual level. In the fields of image classification and computer vision, model performance was measured by accuracy, AUC, precision and recall, and model generalization, we believed that it should be appropriate to apply the same measurements when we adopted the AI algorithm to AIO classification. It might not be necessary to discover the common feature maps for a group or class.

From our study, we concluded that, by transforming genetic data into AIOs, we could apply advanced image classification algorithms to analyze these data. This approach had several advantages over regression and other ML models. First, the AIO approach could use a large number of SNVs to predict/classify a trait of interest. In this study, we used two sets of SNVs, set I had 4,096 SNVs and set II had 44,100 SNVs. For most regression or RF and SVM models, it would not be easy to incorporate so many SNVs into a single model. An AIO of 1,000 × 1,000 pixels would use one million SNVs and it could be easily analyzed with a desktop computer with a GPU. The ability to analyze a large number of variables efficiently was advantageous to other approaches. Second, there was no need to conduct variable selection because, in image analysis, pixel correlation had no impact on the results of classification. In this study, we did use GWAS p value and LD pruning to screen SNVs, the purpose was to demonstrate the principle of the AIO technique with a reasonable number of SNVs without exhausting our computational resources. Our selection of the SNVs was categorical, not on individual SNVs, which was a prerequisite for most prediction/classification models because variable correlation or collinearity had significant impacts on the outcomes. The set I SNVs were the first 4,096 SNVs out of the 4,109 SNVs sorted by chromosome number and position without consideration of their effects to the disease. Third, the association found with the AIO approach would be a two- or three-dimensional multi-pixel pattern association (i.e., the saliency map) where each pixel represented a single SNV. The pixels that made the unique association pattern could be traced back to the original pixel map of the AIO, identifying those SNVs that were necessary and sufficient to classify the subject. Biologically, this would be equivalent to the identification of multi-factor interaction and association. This opened a new window for multi-factor gene-gene interaction analysis.

We noticed that the classification results from set I and set II SNVs were similar (accuracy 0.678 ± 0.007 and 0.685 ± 0.021), but the number of SNVs in set II was 10 times more than that in set I. This was somewhat surprising and interesting. Set I SNVs were selected based on GWAS $p \leq 5 \times 10^{-5}$ and pairwise LD pruning ($r^2 \leq 0.5$), and set II SNVs were selected based on the same GWAS p value and minor allele filtering (minor allele frequency ≥ 0.05). Based on these conditions, most SNVs in set I should be included in set II. Our initial plan was to test whether additional SNVs with high LD would affect model performance. The results indicated that these additional SNVs did not impact the overall performance, and neither improved nor deteriorated the performance (the difference in performance was not statistically significant). The results also suggested that, once transformed into pixel signals, the SNVs lost their property as individual markers, the high correlation among the SNVs, a collinearity issue that confounded regression analysis, would not impact image classification analysis. Instead, the SNVs were treated as structural elements that contributed to a signature pattern or image object for which the labels were trained for and classified by. The SNV to pixel transformation, therefore, made it easier to pick the SNVs to form the AIO, because the correlation among the individual SNVs could be ignored.

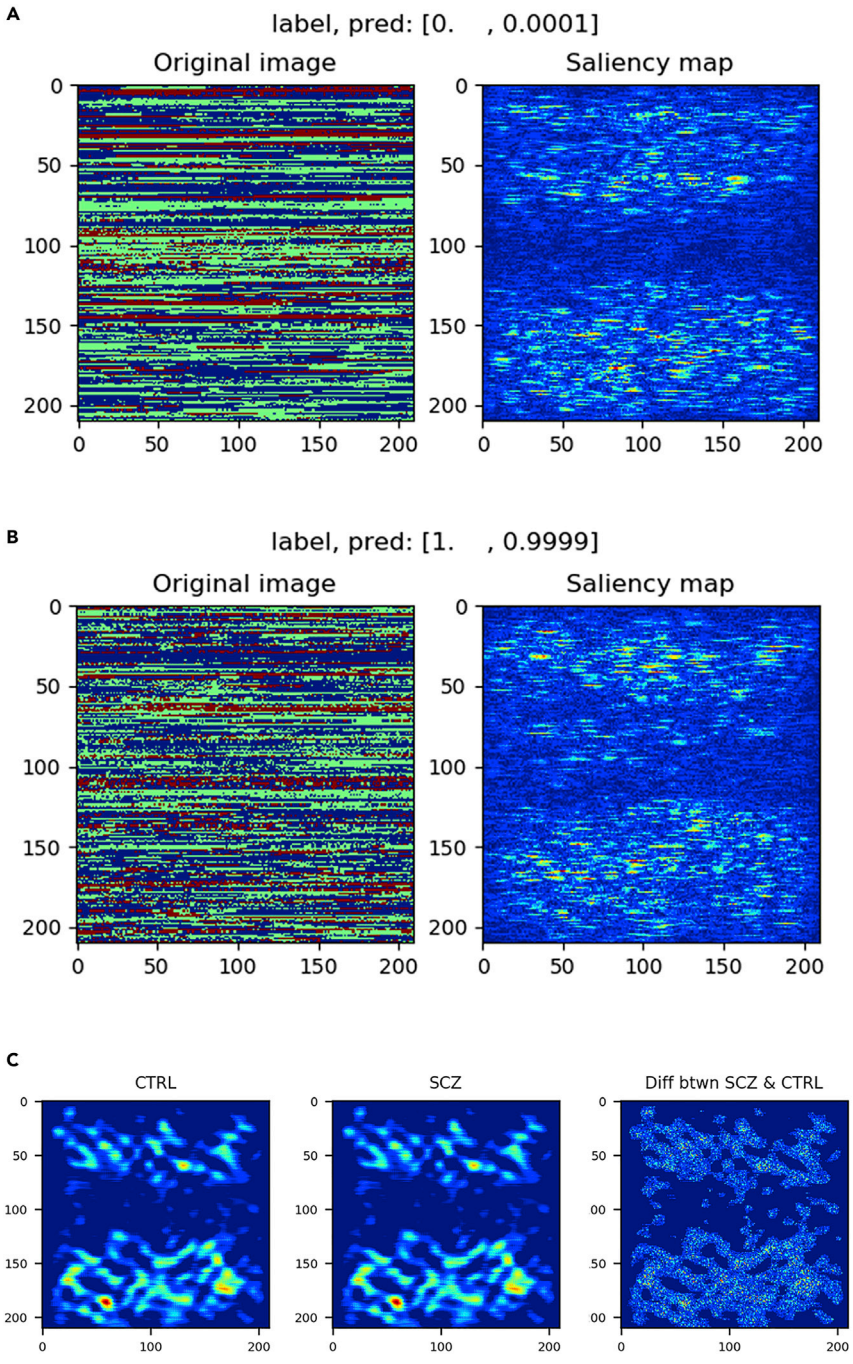


Figure 4. Classification of normal controls and diagnosed schizophrenia patients by set II SNVs

(A and B) Comparisons between the original AIOs and model-derived saliency maps for a control and a patient. The label and pred in (A and B) listed the labels (0 for control, and 1 for schizophrenia patient) and prediction probabilities for the individuals.

(C) The group-wise saliency maps from averaging saliency maps of all correctly predicted subjects for each group (CTRL and SCZ). The pixel intensity differences between the two groups were shown in the Diff btwn SCZ & CTRL image.

However, the categorical selection of SNVs merited some discussion. In this study, we used a p value threshold of 5×10^{-5} to select SNVs. Some researchers would use a more stringent p value threshold, such as GWAS significance level, and others might use a more lenient threshold. In the literature, there were many examples that used a variety of thresholds to select SNVs for modeling disease risks and classification.[19–21] We believed that the selection criterion should depend on the objectives of the study and the nature of the problem intended to solve. In our case, we intended to demonstrate the analytic capacity of the AIO technique in classification models, and the threshold we used gave us tens of thousands SNVs to work with. Should we use the GWAS significance threshold, there were not enough SNVs for us to work with (just about 100 SNVs reaching $p = 5 \times 10^{-8}$ for the GWAS used in this study). Furthermore, the inherent feature extraction function of the CNN algorithm provided us the luxury to include more SNVs to maximize our model performance. This was because that GWAS tested SNVs individually and did not consider multi-locus interactions. In contrast, CNN algorithms worked by learning the spatial patterns that were multi-locus by nature. Some SNVs, they might not show significant association with schizophrenia diagnosis individually, but when they were placed in a network of higher-order interactions, such as the multi-locus spatial patterns learned by the CNN model, they could make significant contribution to these signature patterns, leading to better model performance.

In summary, in this study we demonstrated that a set of SNVs selected from GWAS could be used to form an AIO and the AIO could be classified by advanced AI algorithms, such as CNN image classification. With two sets of SNVs selected by GWAS $p = \leq 5 \times 10^{-5}$, we obtained similar classification accuracies (0.678 ± 0.007; 0.685 ± 0.021) and AUCs (0.738 ± 0.008; 0.708 ± 0.030). Our approach, the AIO technique that transformed individual genetic markers into image pixel signals, allowed us to analyze a significant number of SNVs jointly to evaluate their utility in diagnosis classification. Compared with the RF and SVM models that used PRS calculated at the same GWAS p value threshold, our AIO approach had about 4% better performance. We believed that other types of genomic data, such as gene expression and methylation, could be transformed into AIOs and classified by AI algorithms as well. The AIO method, therefore, had the potential to build classification models for clinical applications.

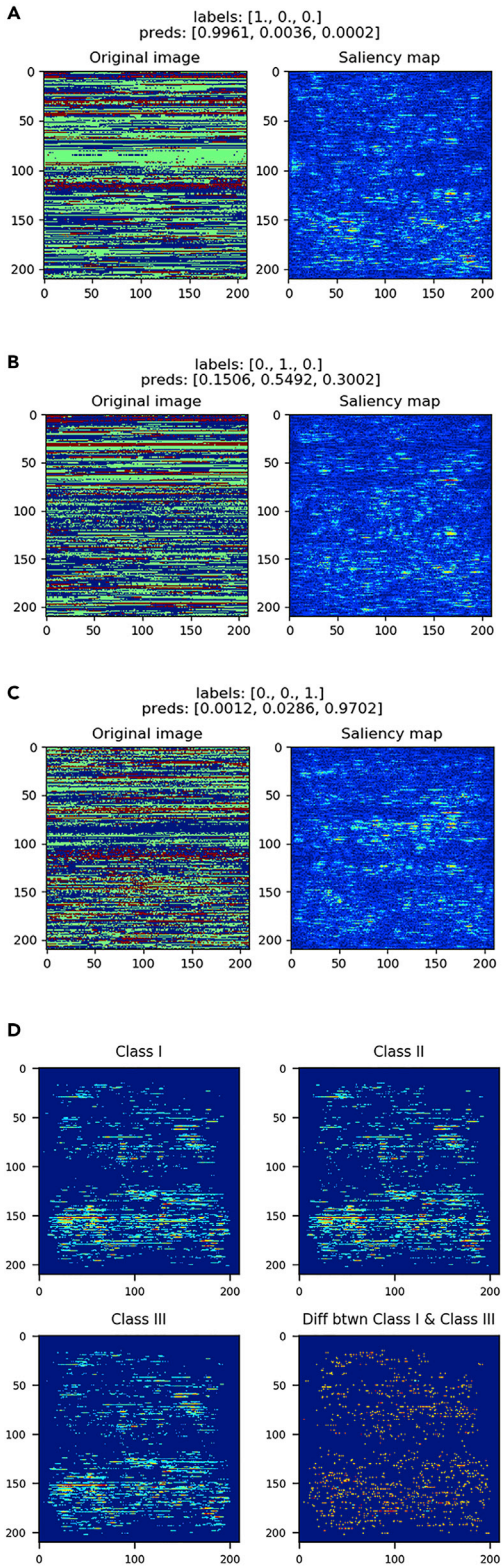


Figure 5. Classification of PRS-stratified classes by set II SNVs

(A–C) Comparisons for the three classes between the original AIOs and model-derived saliency maps for the three classes, respectively. The labels and preds listed the labels and prediction probabilities for each class.

(D) Saliency maps for the three classes by averaging the saliency maps of correctly predicted subjects for each class (class I, class II, and class III). The fourth image in the panel was a plot showing the intensity difference between class I and class III pixel by pixel.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Request for information and resources used in this article should be addressed to Dr. Xiangning Chen at this email: va.samchen@gmail.com.

Materials availability

There were no physical materials associated with this study.

Data and code availability

The genetic and clinical data used in this study were from the Genetics Repository of the National Institute of Mental Health (https://www.nimhgenetics.org/). This is a controlled access data available for qualified researchers. Please refer to the website for the policies for data access application. The codes used in the paper are available at the Github website: https://github.com/mdsamchen/AIO_scripts.

Genetic data

We applied for and obtained the genotype and clinical data for the MGS,[12] SCCSS,[13] and CATIE[14,15] datasets from the NIMH Genetics Repository (https://www.nimhgenetics.org/). The MGS dataset had 2,681 affected subjects and 2,653 controls. The SCCSS had 2,895 affected subjects and 3,836 controls. The CATIE had 741 affected subjects and 751 controls. All subjects used in the 3 datasets were of European ancestry. The MGS and CATIE datasets were genotyped with Affymetrix 6.0 microarray with 906,600 SNVs. The SCCSS was typed with Illumina OmniExpression array with 713,599 SNVs. To have the same markers across the MGS, SCCSS, and CATIE datasets, we used the IMPUTE2[22] to impute the missing genotypes using the 1000 Genome haplotypes as reference. Markers with the INFO value <0.4 were filtered out. Details of imputation were described previously.[23] In this study, we combined the MGS and SCCSS data as training data for model training. The CATIE was used to validate the models trained with the MGS and SCCSS data.

The rationale for AIO design

For most SNVs, there are two alleles, A and a. Since humans have two chromosome copies, for a given SNV, therefore, there are three possible genotypes, AA, Aa, and aa. In this study, we propose a new approach to improve the efficiency for SNV analysis by adapting image classification algorithms. We consider an SNV as a pixel in an image, and its value takes the genotype for the given individual at the specified SNV locus. This would allow us to arrange a collection of SNVs from an individual to create an AIO (Figure 1B). In the AIO, the physical distance and relationship of SNVs on chromosomes could be indexed because each SNV occupies a specific address on the AIO just like each pixel occupies a specific address in an image, the spatial relationship between any two pixels, therefore, is clearly defined. Since the same SNV from different individuals has the same address on the AIOs, the relationship among the SNVs would be preserved across individuals. This arrangement would allow us to not only analyze the relationship between a single SNV and the trait of interest (analogous to traditional single-point association analysis), but also identifying the complex relationship between a specific pattern made of multiple SNVs and the trait (multipoint interaction and association). This multi-SNV, spatial pattern or signature is the target that image classification algorithms are designed to discover and used for object classification. With this signature, we would be able to classify and distinguish people with different disease status if the disease is indeed caused by these genetic variations. In addition, other variables such as sex, age, family history, ethnicity, and computed principal components, can be recoded and organized as an ancillary AIO, and integrated into the classification models.

Selection of SNVs to make AIOs

First, we downloaded the summary statistics of the schizophrenia GWAS[2] from the Psychiatric Genomics Consortium (PGC) website (www.med.unc.

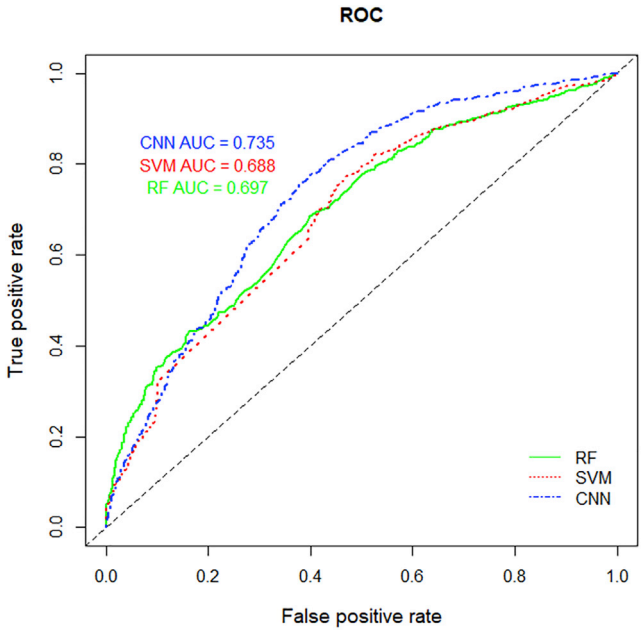


Figure 6. Performance comparison of the CNN, RF, and SVM models
RF, random forest; SVM, support vector machine; CNN, convolutional neural network. Among the three models, the CNN model had the best performance.

edu/pgc/results-and-downloads). SNVs with association $p \leq 5 \times 10^{-5}$ were selected. The main reason for using this p value threshold was to select a reasonable number of SNVs that could demonstrate the analytic capacity of our AIO approach. Based on the GWAS summary statistics, there were 51,855 SNVs with $p \leq 5 \times 10^{-5}$. Since many of these SNVs are in high LD with each other, we further pruned the SNVs to exclude SNVs with pairwise LD (r^2) $\geq$ 0.5 using the PLINK program[24,25] (plink command –indep 100 10 2). The reason that we used LD pruning was that we wanted a list of SNVs well spaced (by LD) across the chromosomes. For a linkage region, there might be more than one SNV selected, which was different from the LD clumping technique where only the best SNV would be selected from an LD region. The pruning produced 4,109 SNVs with $p \leq 5 \times 10^{-5}$. For the 4,109 SNVs with $p \leq 5 \times 10^{-5}$, we selected the first 4,096 SNVs as ordered by chromosome number and position, referred to as set I SNVs, to make a 64 × 64 pixel AIO for each of the individuals in the MGS, SCCSS, and CATIE datasets. To evaluate the effect of allele frequency on the AIO classification, we further selected SNVs with minor allele frequency $\geq$0.05 from the 51,855 SNVs with $p \leq 5 \times 10^{-5}$, this produced 44,320 SNVs. From these frequency-filtered SNVs, we selected the first 44,100 SNVs as ordered by chromosome number and position, referred to as set II SNVs, to make a 210 × 210 pixel AIO for the subjects in the 3 datasets. For all the AIOs, the genotypes of the SNVs, i.e., AA, Aa, and aa, were converted to the values of 0, 154, and 254, respectively, the AIOs made from set I and set II SNVs could be considered grayscale images (Figure 1B). The values used for the 3 genotypes were intended to approximate pixel intensity range (1 byte), other values, such as 0, 0.5, and 1, could be used as long as the three genotypes could be separated clearly.

PRS calculation

We calculated PRS at $p = 5 \times 10^{-5}$ for all the subjects in the MGS, SCCSS, and CATIE samples using the PRSice2 method[26,27] and the same GWAS summary statistics[2] we used to select SNVs for AIOs. The PRSice2 allowed users to specify parameters to clump LD, genetic model and scoring algorithm. We used these default parameters to calculate the PRS: –clump-kb 250, –clump-p 0.500, –clump-r2 0.100, –interval 0.001, –lower 5×10^{-5}, –upper 0.5, –model add, –score avg. With these parameters, the best p value threshold identified was 0.3171, explaining 0.5145 (R^2) of the phenotype variations and containing 235,931 SNVs. At a p value threshold of 5×10^{-5}, 1,397 SNVs were included in the PRS calculation, explaining 0.1468 (R^2) of the phenotype variations. The PRS medians of the case and control in the training (MGS and SCCSS combined) samples were 0.0011 and 0.0017, and that for the validation (CATIE) samples were 0.0014 and 0.0020. The distributions of the PRSs for the training and validation samples were shown in Figure S1.

AIO classification with CNN algorithms

In this study, we used the TensorFlow (www.tensorflow.org/),[11,28] keras (https://keras.io/api/), and the CNN architecture[29,30] to classify and predict AIOs generated from selected SNV data. In these analyses, we used two phenotypes or labels. One was the diagnosis provided by the investigators from the MGS, SCCSS, and CATIE studies. The normal controls (ctrl), were coded as 0, and the diagnosed patients (case) were coded as 1. The second phenotype was based on PRS stratification. We calculated the mean (M) and standard deviation (SD) for the diagnosed subjects in the MGS and SCCSS datasets. The new phenotype, or PRS-stratified classes was defined as the following:

Class I: $PRS_i \leq (M - 0.25 \times SD)$, $T_i = 0$;
Class II: $(M + 0.25 \times SD) < PRS_i > (M - 0.25 \times SD)$, $T_i = 1$;
Class III: $PRS_i \geq (M + 0.25 \times SD)$, $T_i = 2$.

where PRS_i is the individual's PRS and T_i is the assigned phenotype. Please note that the diagnosed individuals from the CATIE dataset were not included in the calculation of the mean and SD for this new phenotype. For this new phenotype, referred to as PRS-stratified classes, there were 5,000 (41.4%) class I subjects, 2,854 (23.7%) class II subjects, and 4,211 (34.9%) class III subjects in the combined MGS and SCCSS datasets (train data). For the CATIE dataset (test data), there were 483 (32.4%) class I subjects, 325 (21.8%) class II subjects, and 684 (45.8%) class III subjects, respectively. There were some differences in class distribution between the training (MGS and SCCSS) and the testing (CATIE) samples. Since the differences were relatively small, they would not have too much impact on model performance.

Once the AIOs were made, and phenotypes or labels were assigned to the subjects in the three datasets, we used the TensorFlow and keras packages to conduct image classification analyses. For these analyses, we combined the MGS and SCCSS datasets and used them to train the models. Since the possible combinations of parameters for a CNN model were extremely large, we started out with the architecture and hyperparameters from the examples reported in recent literature.[31,32] We first tested the CNN architectures with varying numbers of convolutional layers, fully connected layers, and the number of filters (neurons) in each layer. We then tested initial learning rate (0.1–0.0001), epsilon value ($1.05–10^{-10}$), kernel regularizer value (0.1–0.00001), and kernel size value (5–15). Based on these screening of architectures and hyperparameters, we settled on the architectures that consisted of three convolutional layers and three fully connected layers. Each convolutional layer was followed with a batch normalization layer and average pooling layer. The specific parameters were detailed in the scripts (https://github.com/mdsamchen/AIO_scripts). After the training, we used the CATIE samples to evaluate the performance of the models. For the binary phenotype, the diagnosis, we reported the binary accuracy ([true positive + true negative]/[true positive + false positive + true negative + false negative]), precision (true positive/[true positive + false positive]), recall or sensitivity (true positive/[true positive + false negative]), f1-score ([2 × precision × recall]/[precision + recall]) and the AUC of the ROC for the training processes as defined in the scikit-learn library.[33] For the multi-label PRS classes, we reported categorical accuracy and class-specific AUC.

Classification of diagnosis with RF and SVM

To compare the performance of CNN classification of AIOs, we used the PRSs calculated at the same threshold, i.e., $p \leq 5 \times 10^{-5}$ to classify schizophrenia diagnosis with RF and SVM. The classifications were conducted with the R packages randomForest and e1071. For these analyses, we calculated 20 principal components for the 3 datasets using the PLINK program,[24,25] and used the principal components and sex along with the PRS in the classification models. For RF analyses, we set the number of trees to 2,000, and conducted grid search for optimal values of the mtry (number of variables at the split, 6–12), node size (number of samples at terminal node, 20–40), and sample size (the fraction of samples used in training, 0.55–0.80) with the R package Ranger, which was a fast version of RF. Based on the results of grid search,

we set on an RF model with mtry of 6, node size of 28, sample size of 0.80. For SVM, we did a grid search for the penalty factor C (10^{-3} to 10^{3}) and gamma (2^{-10} to 2^{8}) with the RBF kernel. Based the results of the grid search, we set on the C of 3 and gamma of 0.0175, and used 5-fold cross-validation to build the models with the training data (MGS and SCCSS combined). Once the models were established, we run the model 5 times to evaluate its performance on the independent CATIE samples.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2021.100303.

ACKNOWLEDGMENTS

This work was supported in part by NIH grant P20GM121325.

The genetic and clinical data for the MGS, SCCSS, and CATIE studies were obtained from the Genomics Repository of National Institute of Mental Health (https://www.nimhgenetics.org/). We thank the patients, control subjects, and the investigators involved in these studies. The investigators and co-investigators for the MGS were: ENH/Northwestern University, Evanston, IL, MH059571, Pablo V. Gejman, M.D. (Collaboration Coordinator; PI), Alan R. Sanders, M.D.; Emory University School of Medicine, Atlanta, GA, MH59587, Farooq Amin, M.D. (PI); Louisiana State University Health Sciences Center; New Orleans, LA, MH067257, Nancy Buccola APRN, B.C., M.S.N. (PI); University of California-Irvine, Irvine, CA, MH60870, William Byerley, M.D. (PI); Washington University, St. Louis, MO, U01, MH060879, C. Robert Cloninger, M.D. (PI); University of Iowa, Iowa, IA, MH59566, Raymond Crowe, M.D. (PI), Donald Black, M.D.; University of Colorado, Denver, CO, MH059565, Robert Freedman, M.D. (PI); University of Pennsylvania, Philadelphia, PA, MH061675, Douglas Levinson, M.D. (PI); University of Queensland, QLD, Australia, MH059588, Bryan Mowry, M.D. (PI); Mt. Sinai School of Medicine, New York, NY, MH59586, Jeremy Silverman, Ph.D. (PI).

The SCCSS was supported by funding provided by the NIMH (R01 MH077139 to Patrick F. Sullivan and R01 MH095034 to Pamela Sklar), the Stanley Center for Psychiatric Research, the Sylvan Herman Foundation, the Friedman Brain Institute, Icahn School of Medicine at Mount Sinai at the Mount Sinai School of Medicine, the Karolinska Institutet, Karolinska University Hospital, the Swedish Research Council, the Swedish County Council, the Söderström Königska Foundation, and the Netherlands Scientific Organization (NWO 645-000-003). Co-principal investigators involved in this study were Pamela Sklar (Mount Sinai School of Medicine), Christina M. Hultman (Karolinska Institutet, Stockholm, Sweden), and Patrick F. Sullivan (University of North Carolina and Karolinska Institutet, Stockholm, Sweden). We are deeply grateful for the participation of all subjects contributing to this research and to the collection team that worked to recruit them.

The principal investigators of the CATIE (Clinical Antipsychotic Trials of Intervention Effectiveness) trial were Jeffrey A. Lieberman, M.D., T. Scott Stroup, M.D., M.P.H., and Joseph P. McEvoy, M.D. The CATIE trial was funded by a grant from the National Institute of Mental Health (N01 MH900001) along with MH074027 (PI PF Sullivan). Genotyping was funded by Eli Lilly and Company.

AUTHOR CONTRIBUTIONS

X.C. conceived the concept, designed the study, analyzed the data, and wrote the manuscript. D.G.C. was involved in the model training and analyses. J.C. was involved in study design and data selection and analyses and also reviewed and commented on the manuscript. C.J., Z.Z., and J.Z. were involved in discussion of study design and model training, and reviewed and commented on the manuscript.

DECLARATION OF INTERESTS

X.C. has filed a USPTO and PCT patent application for the AIO technology. The patent is currently under examination by the agencies.

Received: October 20, 2020
Revised: March 17, 2021
Accepted: June 7, 2021
Published: June 30, 2021

REFERENCES

1. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, D1005–D1012.
2. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.
3. Timpson, N.J., Greenwood, C.M.T., Soranzo, N., Lawson, D.J., and Richards, J.B. (2018). Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat. Rev. Genet. *19*, 110–124.
4. O'Connell, K.S., McGregor, N.W., Lochner, C., Emsley, R., and Warnich, L. (2018). The genetic architecture of schizophrenia, bipolar disorder, obsessive-compulsive disorder and autism spectrum disorder. Mol. Cell. Neurosci. *88*, 300–307.
5. Torkamani, A., Wineinger, N.E., and Topol, E.J. (2018). The personal and clinical utility of polygenic risk scores. Nat. Rev. Genet. *19*, 581–590.
6. Janssens, A.C.J.W. (2019). Validity of polygenic risk scores: are we measuring what we think we are? Hum. Mol. Genet. *28*, R143–R150.
7. Domingue, B.W., Belsky, D.W., Harris, K.M., Smolen, A., McQueen, M.B., and Boardman, J.D. (2014). Polygenic risk predicts obesity in both white and black young adults. PLoS ONE *9*, e101596.
8. Escott-Price, V., Sims, R., Bannister, C., Harold, D., Vronskaya, M., Majounie, E., et al. (2015). Common polygenic variation enhances risk prediction for Alzheimer's disease. Brain *138*, 3673–3684.
9. Rawat, W., and Wang, Z. (2017). Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. *29*, 2352–2449.
10. Sultana, F., Sufian, A., and Dutta, P. (2018). Advancements in Image Classification using Convolutional Neural Network. 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN) 122–129.
11. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). TensorFlow: a system for large-scale machine learning. ArXiv, 1605.08695.
12. Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., et al. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature *460*, 753–757.
13. Bergen, S.E., O'Dushlaine, C.T., Ripke, S., Lee, P.H., Ruderfer, D.M., Akterin, S., et al. (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol. Psychiatry *17*, 880–886.
14. Stroup, T.S., McEvoy, J.P., Swartz, M.S., Byerly, M.J., Glick, I.D., Canive, J.M., et al. (2003). The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol development. Schizophr Bull. *29*, 15–31.
15. Sullivan, P.F., Lin, D., Tzeng, J.-Y., van den Oord, E., Perkins, D., Stroup, T.S., et al. (2008). Genomewide association for schizophrenia in the CATIE study: results of stage 1. Mol. Psychiatry *13*, 570–584.
16. Simonyan, K., Vedaldi, A., and Zisserman, A. (2013). Deep inside convolutional networks: visualising image classification models and saliency maps. ArXiv, arXiv:1312.6034.
17. Bracher-Smith, M., Crawford, K., and Escott-Price, V. (2021). Machine learning for genetic prediction of psychiatric disorders: a systematic review. Mol. Psychiatry *26*, 70–79.
18. Chen, J., Wu, J.-S., Mize, T., Shui, D., and Chen, X. (2018). Prediction of schizophrenia diagnosis by integration of genetically correlated conditions and traits. J. Neuroimmune Pharmacol. *13*, 532–540.

19. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat. Genet. *50*, 1219–1224.

20. Elliott, J., Bodinier, B., Bond, T.A., Chadeau-Hyam, M., Evangelou, E., Moons, K.G.M., et al. (2020). Predictive accuracy of a polygenic risk score-enhanced prediction model vs a clinical risk score for coronary artery disease. JAMA *323*, 636–645.

21. Perkins, D.O., Olde Loohuis, L., Barbee, J., Ford, J., Jeffries, C.D., Addington, J., et al. (2019). Polygenic risk score contribution to psychosis prediction in a target population of persons at clinical high risk. AJP *177*, 155–163.

22. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.

23. Ware, J.J., Chen, X., Vink, J., Loukola, A., Minica, C., Pool, R., et al. (2016). Genome-wide meta-analysis of cotinine levels in cigarette smokers identifies locus at 4q13.2. Sci. Rep. *6*, 20092.

24. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.

25. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience *4*, 7.

26. Euesden, J., Lewis, C.M., and O'Reilly, P.F. (2015). PRSice: polygenic risk score software. Bioinformatics *31*, 1466–1468.

27. Choi, S.W., and O'Reilly, P.F. (2019). PRSice-2: polygenic Risk Score software for biobank-scale data. Gigascience *8*. https://doi.org/10.1093/gigascience/giz082.

28. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). TensorFlow: large-scale machine learning on heterogeneous distributed systems. ArXiv, 1603.04467.

29. Ciresan, D.C., Meier, U., Gambardella, L.M., and Schmidhuber, J. (2011). Convolutional Neural Network Committees for Handwritten Character Classification. In 2011 International Conference on Document Analysis and Recognition, pp. 1135–1139.

30. Chen, X., Xiang, S., Liu, C., and Pan, C. (2013). Vehicle Detection in Satellite Images by Parallel Deep Convolutional Neural Networks. In 2013 2nd IAPR Asian Conference on Pattern Recognition, pp. 181–185.

31. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., and Mougiakakou, S. (2016). Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Trans. Med. Imaging *35*, 1207–1216.

32. Anwar, S.M., Majid, M., Qayyum, A., Awais, M., Alnowami, M., and Khan, M.K. (2018). Medical image analysis using convolutional neural networks: a review. J. Med. Syst. *42*, 226.

33. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.