Genetic Risk Score for Predicting Schizophrenia Using Human Chromosomal-Scale Length Variation.

Christopher Toh (tohc@uci.edu)
University of California Irvine https://orcid.org/0000-0003-3298-9055
James P. Brody
University of California Irvine

Primary research

Keywords: Predicting Schizophrenia, Chromosomal-Scale, Biobank dataset

Posted Date: March 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-268559/v1

Abstract

Introduction

Schizophrenia is a neurological disorder that often manifests itself as a combination of psychotic symptoms such as delusions, hallucinations, and disorganized cognitive functions. Several lines of evidence indicate that schizophrenia has a genetic component, however it cannot be isolated to a single gene. We set out to determine how well one could predict that a person will develop schizophrenia based on their germ line DNA.

Methods

We compared 1129 people from the UK Biobank dataset who had a diagnosis of schizophrenia to an equal number of age matched people drawn from the general UK Biobank population. For each person, we constructed a profile consisting of a sequence of numbers. Each number characterized the length of a segment of one of their chromosomes. We tested several machine learning algorithms using the h2o.ai framework to determine which was most effective in predicting schizophrenia. We also tested whether there was any improvement in prediction by breaking the chromosomes into smaller chunks. We used SHAP values to better understand features important to the predictive model.

Results

We found that the stacked ensemble, a combination of four different machine learning algorithms, performed best with an area under the receiver operating characteristic curve (AUC) of 0.583 (95% CI 0.581-0.586). We noted an increase in the AUC by breaking the chromosomes into smaller chunks for analysis. Using SHAP values, we identified the X chromosome as the most important contributor to the predictive model.

Conclusion

We conclude that germ line chromosomal scale length variation data can provide an effective genetic risk score for schizophrenia. Length variations of several regions of the X Chromosome are the greatest contributing factor.

Introduction

Schizophrenia is a highly heritable, complex psychiatric disorder[1,2]. Genome wide association studies have identified over one hundred genetic loci that contribute to its heritability[2–6]. However, these loci still account for less than half of the genetic risk for schizophrenia[3]. Environmental exposure to chemicals appears to play almost no role in the development of schizophrenia, but different forms of trauma experienced during development does appear to be a risk factor[7] . Twin studies have consistently shown a significant non-zero genetic contribution to schizophrenia, and many, but not all, twin studies find that the environmental contribution to schizophrenia is consistent with zero[8].

Genetic risk scores [9–11] have been developed for many different forms of disease, including breast cancer[12], coronary artery disease[13], and stroke[14]. Polygenic risk scores based on SNPs clearly can predict schizophrenia. One study measured an odds ratio of about 8 (95% CI 4-14) for the highest decile compared to the lowest decile[15]. A second study found that polygenic risk scores for schizophrenia (and bipolar disorder) are also associated with creativity[16]. A review of polygenic risk scores for schizophrenia highlighted the difficulty these studies had finding a consistent diagnosis of schizophrenia[17].

We have previously shown that chromosome-scale length variation is a powerful tool to predict phenotypes from a person's genome [18]. This method is particularly appealing for genetic risk scores because it includes epistatic effects that might be missed with conventional genome wide association studies, which use logistic regression—a linear combination of SNP scores.

The purpose of this paper is to evaluate how well a genetic risk score based on chromosome-scale length variation and machine learning classification algorithms can predict schizophrenia in individuals. We evaluated this approach on a dataset of 1129 patients who had schizophrenia in the UK Biobank dataset. These patients were previously genotyped as part of the UK Biobank project.

Methods

Data was obtained from the UK Biobank under Application Number 47850. The UK Biobank project collected extensive data from about 500,000 people who were between the ages of 40 and 69 during the 2006-2010 recruitment years. This data included genotyping data and medical records. In addition, most of the participants' medical records are linked, through the National Health Service, to the UK Biobank records. This linkage provides for ongoing follow-up of health conditions [19,20]..

First, we downloaded the "l2r" files from the UK Biobank. Each chromosome has a separate "l2r" file. Each "l2r" file contained 488,377 columns and a variable number of rows. Each column represented a unique patient in the dataset, who can be identified with an encoded ID number. Each row represented a different location in the genome. The values in the file represent the log base 2 ratio of intensity relative to the expected two copies measured at the SNP location.

After downloading the "l2r" data from the UK Biobank, we computed the mean l2r value for different portions of each chromosome for each patient in the dataset. We created three different datasets, which we refer to as "splits". We split each chromosome into either 1, 4, or 8 nominally equal parts. Then, we compute the actual length for each person's chromosome split using the l2r files by taking the average of all l2r values measured within that portion of the chromosome split. A value of 0 represents the nominal average length of that portion of the particular chromosome. We call this dataset the chromosome-scale length variation (CSLV) dataset. For each person, we have 1 split, 4 split, and 8 split datasets. The 1 split data consists of 23 numbers, one for each of the autosomes and one for the X chromosome (Y Chromosome is excluded in UK Biobank for anonymization). The 4 split data consists of 92 numbers and the 8 split data has 184 numbers.

This CSLV dataset was matched with the UK Biobank Health records dataset. UK Biobank matched the person in the Public Health England data with UK Biobanks internal records to produce the person's encoded participant ID. The dataset we have, provided by UK Biobank, contains the participant ID and date the patient was diagnosed by a doctor as having schizophrenia.

Using the CSLV-Schizophrenia dataset, we selected all people who had a diagnosis of schizophrenia and labelled them in the dataset. We constructed an age-matched control group of the same size that had an identical age profile as those in the schizophrenia group. The age-matched control group was selected from all those in the UK Biobank dataset having no indication of schizophrenia. Since only a small fraction of the people in the UK Biobank had a schizophrenia diagnosis, we could rerun the analysis with a different age-matched control group many times to build up statistics.

We used the H2O machine learning package in R[21,22]. We created 100 machine learning models that were trained to classify a person in the dataset, consisting of those who had schizophrenia and age-matched controls, based solely on their chromosome scale length variation data. Each model was trained with 5-fold cross-validation. Each model had a distinct set of controls. These models were trained to perform a binary classification, distinguishing between those who had been diagnosed with schizophrenia and those who did not have schizophrenia. The models were evaluated by measuring the area under the curve of the receiver operating characteristic curve, known as the AUC.

The H2O package implements several common machine learning algorithms. Distributed Random Forest (drf) is based on an algorithm originally called "Extremely randomized trees" [23]. The Gradient Boosting Machine algorithm (gbm) builds regression trees in parallel[24,25]. The generalized linear model (glm) is implemented using an augmented linear model[26–28]. XGBoost is a refinement to the general Gradient Boosting Machine algorithm [29]. Ensembles are a combination of these other machine learning algorithms. This combination often provides superior results to any particular algorithm[30,31]. The H2O package implements stacked ensembles as super learner algorithms[32]. The H2O package also uses SHAP values to interpret the models[33].

Our computer analysis system is a Linux server running Ubuntu 18.04. The system is a 64-bit system running two Intel Xeon E5-2690 2.90 GHz CPUs. It also has a GeForce GT 710 NVIDIA GPU. 32 GBs of RAM were also available with a 10 TB HDD.

Results

Figure 1 presents results showing the performance of different machine learning algorithms. We found that the stacked ensemble models consistently performed best. As Figure 1 shows, we found a slight difference between algorithms and their performance. But all algorithms could predict schizophrenia significantly better than chance (AUC=0.50). This finding indicates that germ line genetics of the patient, as represented by the set of chromosome-scale length variation numbers, demonstrates predictability of schizophrenia.

The AUC (area under the curve of the receiver operating characteristic curve) for the machine learning classification models was 0.583 (standard deviation 0.014, 95% confidence interval of 0.581-0.586). A classification model with an AUC of 0.50 is equivalent to random guessing. The measured AUC differs from 0.50 with $p<0.00001$.

We also tested how well each model could predict schizophrenia on a holdout set of validation data. The holdout set was 30% of the original test data and was not included in the training of the models. The AUC of the holdout set was 0.5734 with a 95% confidence interval of 0.569-0.578.

We then tested whether increasing the number of splits improves model performance. We constructed three overlapping datasets with 1 split, 4 splits, and 8 splits. The phrase “1 split” represents the average l2r value measured across an entire chromosome for all 23 chromosomes giving a total of 23 numbers, “4 splits” represents the average of each quarter of the 23 chromosomes l2r values for a total of 92 numbers, and “8 splits” represent the average of each eighth of the 23 chromosomes’ l2r values for a total of 184 numbers.

Figure 2 shows how models compare on the 3 different split datasets. Overall, a stacked ensemble had the best performance, however a general linear model (glm) was most often the best candidate model.

In all models, increasing splits improves model performance for the same runtime. Figure 3 demonstrates the difference of all models for 1 split, 4 splits, and 8 splits datasets. We tested whether finer splits of the dataset provided significantly improved AUCs. As shown in Table 1, the p-value of the 4 splits model compared to the 1 split model is $p = 1 \times 10^{-24}$. *Comparing the mean AUC for the 8 splits model to the 1 split model gave a p-value of* $p = 3 \times 10^{-30}$. *indicating that finer splits significantly improved the predictive ability of the models. The 4 splits and 8 splits models performed better than the 1 split models by a significant amount.*

Table 1. The mean and standard deviation of the cross validated AUCs of 1 split, 4 splits, and 8 splits datasets of 150 models each.

Dataset	Mean AUC	Standard Deviation	P-value vs 1 split
1 split	0.5614	0.0148	
4 splits	0.5807	0.0146	1×10^{-24}
8 splits	0.5838	0.0141	3×10^{-30}

We then calculated the odds ratio (OR) of our predictions drawn from the cross-validated model. Table 2 shows that a patient in the upper quintile is approximately twice as likely to have schizophrenia when compared to the lower quintile.

Table 2. This table represents the odds ratio between the quintiles of predicted results from our cross-validated results. The result indicates that the top quintile is twice as likely to have an accurate prediction

for Schizophrenia as the bottom quintile.

Quintile	Normal	Schizophrenia	Odds Ratio	Count	95% CI
1	185	123	0.67	308	0.51-0.85
2	156	152	0.97	308	0.76-1.24
3	153	155	1.0	308	0.79-1.3
4	142	165	1.2	307	0.91-1.5
5	133	174	1.3	307	1.0-1.7

In order to understand, how our models came to their conclusions we created several plots to explain them from H2O's "explainability" framework. The first is a variable importance heatmap across the generated models which is shown in Figure 4. Our analysis here indicated that chromosome X was one of the highest contributing variables in predicting Schizophrenia, expecially in tree models such as GBM and XGBoost. We then confirmed this with a Shapley Additive exPlanation or SHAP plot in Figure 5. This plot also indicates that chromosome X was the leading factor in our leading model for predicting schizophrenia.

Utilizing our findings above, we then proceeded to train new models from scratch using only CSLV values from chromosome X but with 64 CSLV splits. This model did not contain any information from the 22 autosomes but instead relied solely on CNVs in the X chromosome and our aim was to see if the model would be comparable to our previous 4-split and 8-split models. We found that on average these models had a comparable performance of about 0.58 with the highest being around 0.627 as shown in Figure 6.

We then again performed a variable importance heatmap analysis to get greater granularity of our understanding of the contributing CSLVs in chromosome X. We found that this was again consistent with the previous findings from the 4-split model. Figure 7 indicates that the top features of variable importance are again being found in the first and last regions of chromosome X. As such it appears that the majority of the predictive power of any model trained with CSLV and when predicting schizophrenia in an individual is a result of CNVs on chromosome X. We also report corresponding estimates of hg38 coordinates in Table 3.

Table 3. This table shows the estimated hg38 coordinates for the corresponding CSLV splits with high variable importance as shown in Figure 7.

CSLV Split	Estimated hg38 Coordinates
1	chrX:60425-634774
6	chrX:5651118-7792613
9	chrX:11426091-13234434
13	chrX:20912585-22990332
42	chrX:107331058-110669244
50	chrX:128031497-130523635
58	chrX:145709120-147908169

Discussion

These results indicate that germline genetic variation contributes to the onset of schizophrenia in individuals. Our results indicate that genetic structural variation across the global chromosomal scope is sufficient to predict, better than guessing, whether or not an individual will have schizophrenia. Further analysis revealed that length variation in a handful of regions of the X chromosome was sufficient to reproduce the predictive model. Recently, there has been revived discussion of copy number variations as a large contributing factor to several neurological ailments including schizophrenia [34]. Additionally, hypotheses about sex chromosome links to schizophrenia inheritance have been discussed for several decades and our findings lend support to this idea [35].

On average, a stacked ensemble is the best approach to creating a predictive model for the prediction of schizophrenia. However, all models that were tested still created models with predictive power better than chance. Since H2O's AutoML performs a grid-search of all the possible datasets and each trial we ran included the same disease group but with a different control groups, we can see in Figure 1 that a general linear model (GLM) oftentimes was still the best option. Gradient Boosted Machines (GBM) and XGBoost also typically performed the same as GLM.

Utilizing a more granularized dataset by splitting the autosomes into quarters and eighths performs significantly better than using a CSLV averaged across an entire chromosome. This observation suggests we can increase performance by increasing splits. In the future, we plan on exploring the trade off in run time and computational resources required by increasing splits.

The CSLV values are averages of copy number variation (CNV) measured at each SNP location. Simply using every single CNV value introduces a dimensionality problem as our dataset only has roughly 488,000 individuals while the total number of CNV values is 764,257 across the 22 autosomes and an additional 18,857 CNV values for the X Chromosome. This means there is likely diminishing returns for using more splits unless it can be offset with increased data.

This approach has several limitations. First, CSLV is an averaged measure of copy-number variations across a large section of the entire chromosome. We used SHAP values to highlight the regions that seem to be more important, but this does not provide a mechanistic explanation. Second, the dataset lacks diversity. The UK Biobank population is primarily Caucasian individuals in the United Kingdom (although not exclusively). Finally, the diagnosis of schizophrenia in an individual is difficult to quantify and the disease might consist of a heterogeneous group of underlying biological processes.

Conclusion

We were able to create machine learning models for prediction of schizophrenia in patients. These models perform better than chance with an average AUC of 0.583. Prediction was performed with only chromosomal scale length variation measurements as the input variables. Further analysis of the SHAP values suggests that the length variation of several regions of the X chromosome are sufficient to reproduce this predictive value.

Declarations

Ethics approval and consent to participate:

Ethics approval and participant consent was collected by UK Biobank at the time participants enrolled. This paper is an analysis of anonymized data provided by UK Biobank. According to UC Irvine's IRB, analysis of anonymized data does not constitute Human Subjects Research.

Consent for publication:

Not applicable.

Availability of data and materials:

The datasets analyzed during the current study are available from UK Biobank at https://www.ukbiobank.ac.uk/

Competing interests:

The authors declare that they have no competing interests.

Funding:

No external funding supported this research.

Author Contributions:

CT conceived of the study. CT and JB analyzed the UK Biobank data. CT and JB contributed to the manuscript. All authors read and approved the final manuscript.

Acknowledgements:

The data used in this study was obtained from the UK Biobank under Application Number 47850.

Author Information:

Department of Biomedical Engineering, University of California, Irvine, USA

Christopher Toh & James P. Brody

Corresponding author:

Correspondence to Christopher Toh.

References

1. Flint J, Munafò M. Genesis of a complex disease. Nature. 2014;511: 412–413. doi:10.1038/nature13645
2. Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511: 421–427. doi:10.1038/nature13595
3. Lee SH, Decandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME, et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet. 2012. doi:10.1038/ng.1108
4. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet. 2013. doi:10.1038/ng.2742
5. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. Nat Genet. 2011. doi:10.1038/ng.940
6. Farrell MS, Werge T, Sklar P, Owen MJ, Ophoff RA, O'donovan MC, et al. Evaluating historical candidate genes for schizophrenia. Mol Psychiatry. 2015;20: 555–562. doi:10.1038/mp.2015.16
7. Van Os J, Kenis G, Rutten BPF. The environment and schizophrenia. Nature. 2010. doi:10.1038/nature09563
8. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a Complex Trait: Evidence from a Meta-analysis of Twin Studies. Arch Gen Psychiatry. 2003. doi:10.1001/archpsyc.60.12.1187
9. Sugrue LP, Desikan RS. What Are Polygenic Scores and Why Are They Important? JAMA. 2019;321: 1820. doi:10.1001/jama.2019.3893
10. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nature Reviews Genetics. 2018. doi:10.1038/s41576-018-0018-x
11. Lello L, Raben TG, Yong SY, Tellier LCAM, Hsu SDH. Genomic Prediction of 16 Complex Disease Risks Including Heart Attack, Diabetes, Breast and Prostate Cancer. Sci Rep. 2019. doi:10.1038/s41598-019-51258-x

12. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. Am J Hum Genet. 2019. doi:10.1016/j.ajhg.2018.11.002
13. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. Nat Commun. 2020. doi:10.1038/s41467-020-17374-3
14. Abraham G, Malik R, Yonova-Doing E, Salim A, Wang T, Danesh J, et al. Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. Nat Commun. 2019. doi:10.1038/s41467-019-13848-1
15. Agerbo E, Sullivan PF, Vilhjálmsson BJ, Pedersen CB, Mors O, Børglum AD, et al. Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: A Danish population-based study and meta-analysis. JAMA Psychiatry. 2015. doi:10.1001/jamapsychiatry.2015.0346
16. Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM, et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. Nat Neurosci. 2015. doi:10.1038/nn.4040
17. Mistry S, Harrison JR, Smith DJ, Escott-Price V, Zammit S. The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: Systematic review. Schizophrenia Research. 2018. doi:10.1016/j.schres.2017.10.037
18. Toh C, Brody JP. Analysis of copy number variation from germline DNA can predict individual cancer risk. bioRxiv. 2018; 303339. doi:10.1101/303339
19. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015. doi:10.1371/journal.pmed.1001779
20. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562: 203–209. doi:10.1038/s41586-018-0579-z
21. Click C, Malohlava M, Candel A, Roark H, Parmar V. Gradient Boosting Machine with H2O. Https://WwwH2OAi/Resources/. 2017; 30.
22. Aiello S, Eckstrand E, Fu A, Landry M, Abouyon P. Machine Learning with R and H2O. 2015. Available: http://h2o-release.s3.amazonaws.com/h2o/master/3283/docs-website/h2o-docs/booklets/R_Vignette.pdf
23. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006. doi:10.1007/s10994-006-6226-1
24. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat. 2001;29: 1189–1232. doi:10.1214/aos/1013203451
25. Friedman JH. Stochastic gradient boosting. Comput Stat Data Anal. 2002;38: 367–378. doi:10.1016/S0167-9473(01)00065-2

26. Lee Y, Nelder JA. Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. Biometrika. 2001. doi:10.1093/biomet/88.4.987
27. Lee Y, Nelder JA. Hierarchical Generalized Linear Models. J R Stat Soc Ser B. 1996. doi:10.1111/j.2517-6161.1996.tb02105.x
28. Lee Y, Nelder JA, Pawitan Y. Generalized linear models with random effects: Unified analysis via H-likelihood. Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood. 2006. doi:10.1111/j.1467-985x.2007.00485_4.x
29. Chen T, Guestrin C. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16. New York, New York, USA: ACM Press; 2016. pp. 785–794. doi:10.1145/2939672.2939785
30. Wolpert DH. Stacked generalization. Neural Networks. 1992. doi:10.1016/S0893-6080(05)80023-1
31. Breiman L. Stacked Regressions. Mach Learn. 1996. doi:10.1007/bf00117832
32. Van Der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol. 2007. doi:10.2202/1544-6115.1309
33. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems. 2017.
34. Zarrei M, Burton CL, Engchuan W, Young EJ, Higginbotham EJ, MacDonald JR, et al. A large data resource of genomic copy number variation across neurodevelopmental disorders. npj Genomic Med. 2019;4. doi:10.1038/s41525-019-0098-3
35. Bache WK, DeLisi LE. The Sex Chromosome Hypothesis of Schizophrenia: Alive, Dead, or Forgotten? A Commentary and Review. Mol Neuropsychiatry. 2018;4: 83–89. doi:10.1159/000491489

Figures

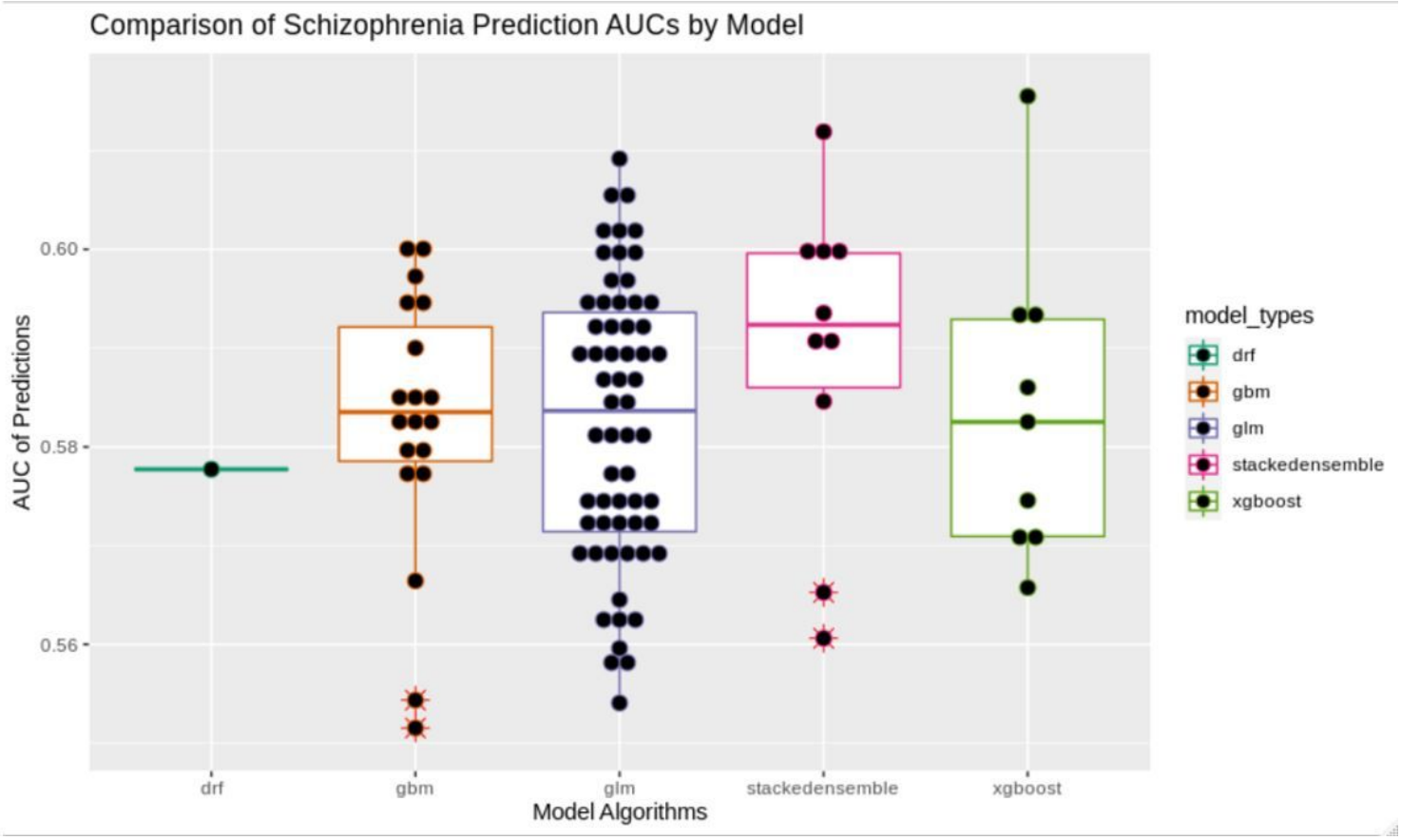

Figure 1

This boxplot figure presents the results of the machine learning predictions. We created 100 different datasets. For each dataset, we used the same set of schizophrenia patients with a distinct set of age matched people from the general UK Biobank population as controls. Then H2O was used to perform a grid-search of possible best algorithms. The best performing algorithm was then reported with an AUC. The differences between algorithms is reported here. The machine learning algorithms tested were distributed random forests (drf), gradient boosting machine (gbm), general linear model (GLM), stacked ensemble (a combination of the other four algorithms) and XGBoost (XGBoost).

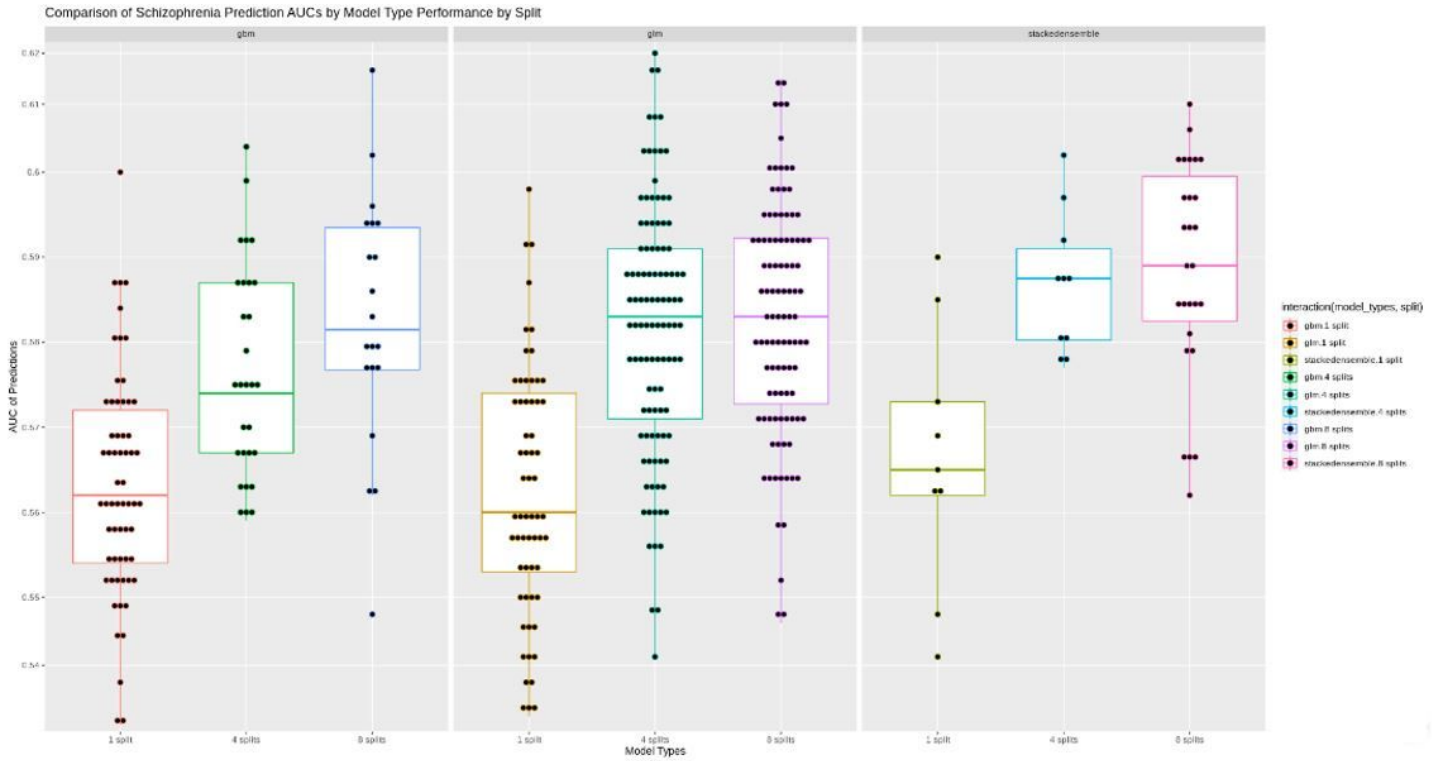

Figure 2

We tested whether finer splits of each chromosome lead to better predictability. We split each chromosome into either one, four, or eight subsections. We computed the chromosome scale length variation for each of these subsections for each person. This set of numbers was used to predict whether patients had schizophrenia. The quality of this prediction was characterized by the AUC. This plot demonstrates how the quality of these predictions increase with finer information on chromosome length variation. The Stacked Ensemble algorithm performs the best across all split variations.

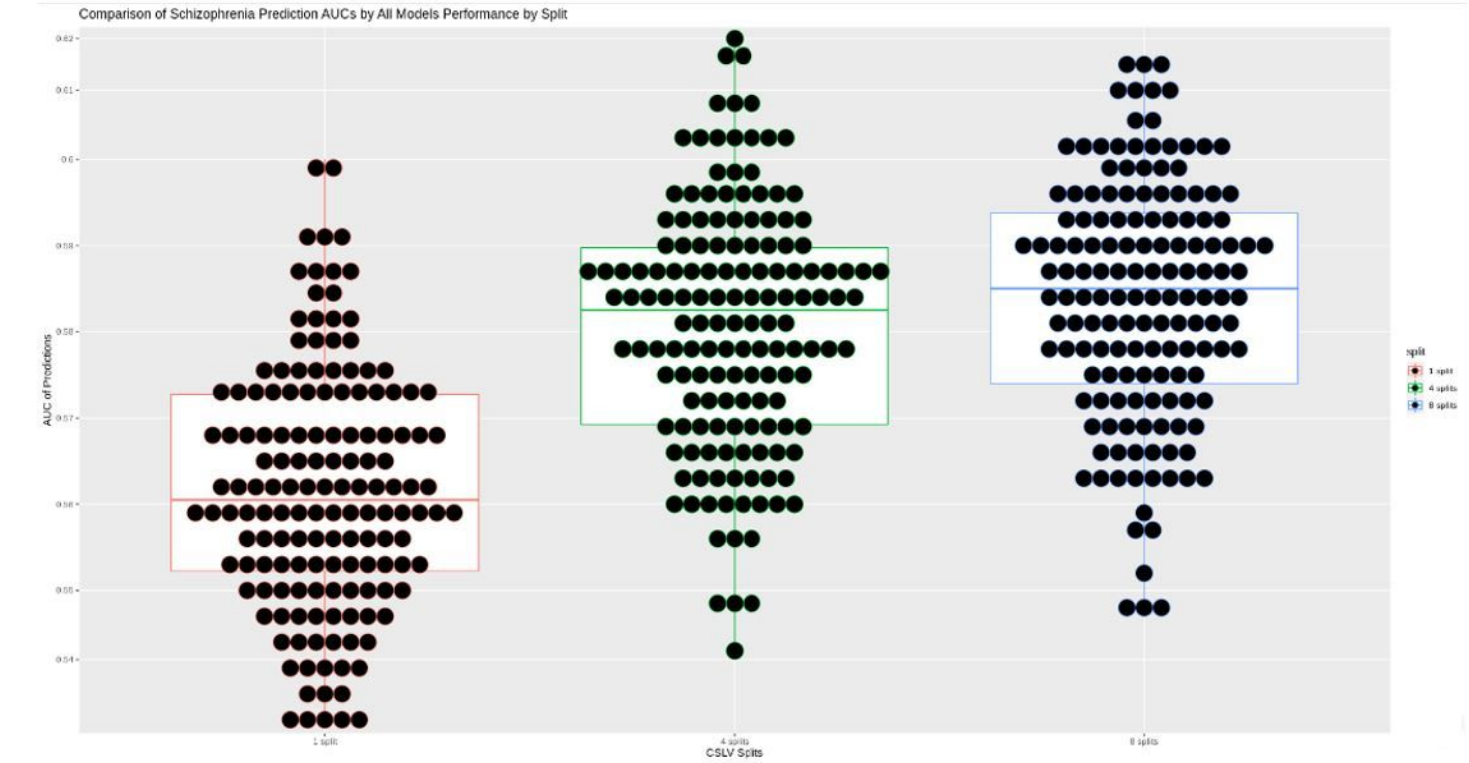

Figure 3

This plot represents the average performance of 150 models for each split type for a total of 450 models.

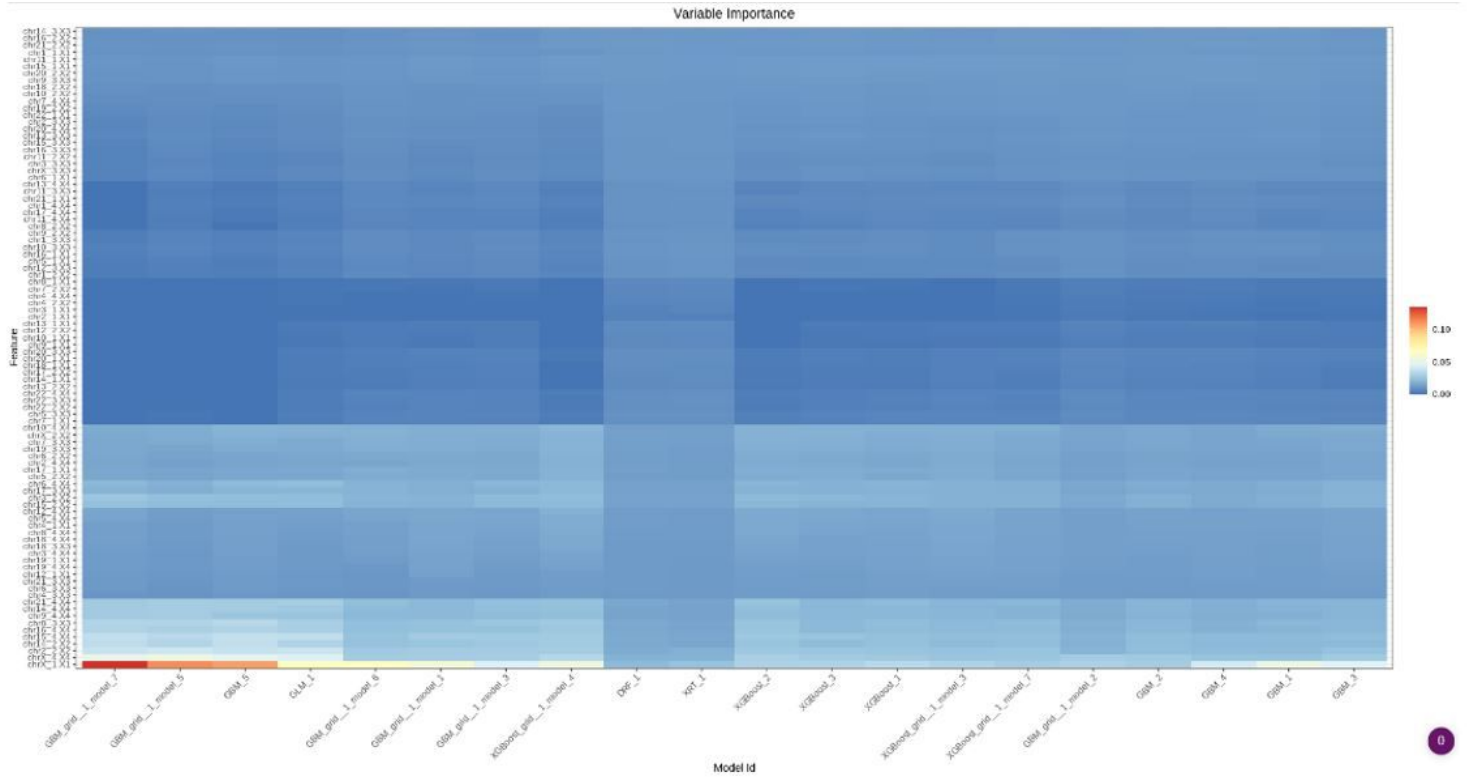

Figure 4

This variable importance heatmap shows the variables which most affected the performance and outcome of decisions made by the specified model. A value closer to 1.0 indicates higher importance of that variable. In most tree-based models the CSLV values for chromosome X have the highest importance.

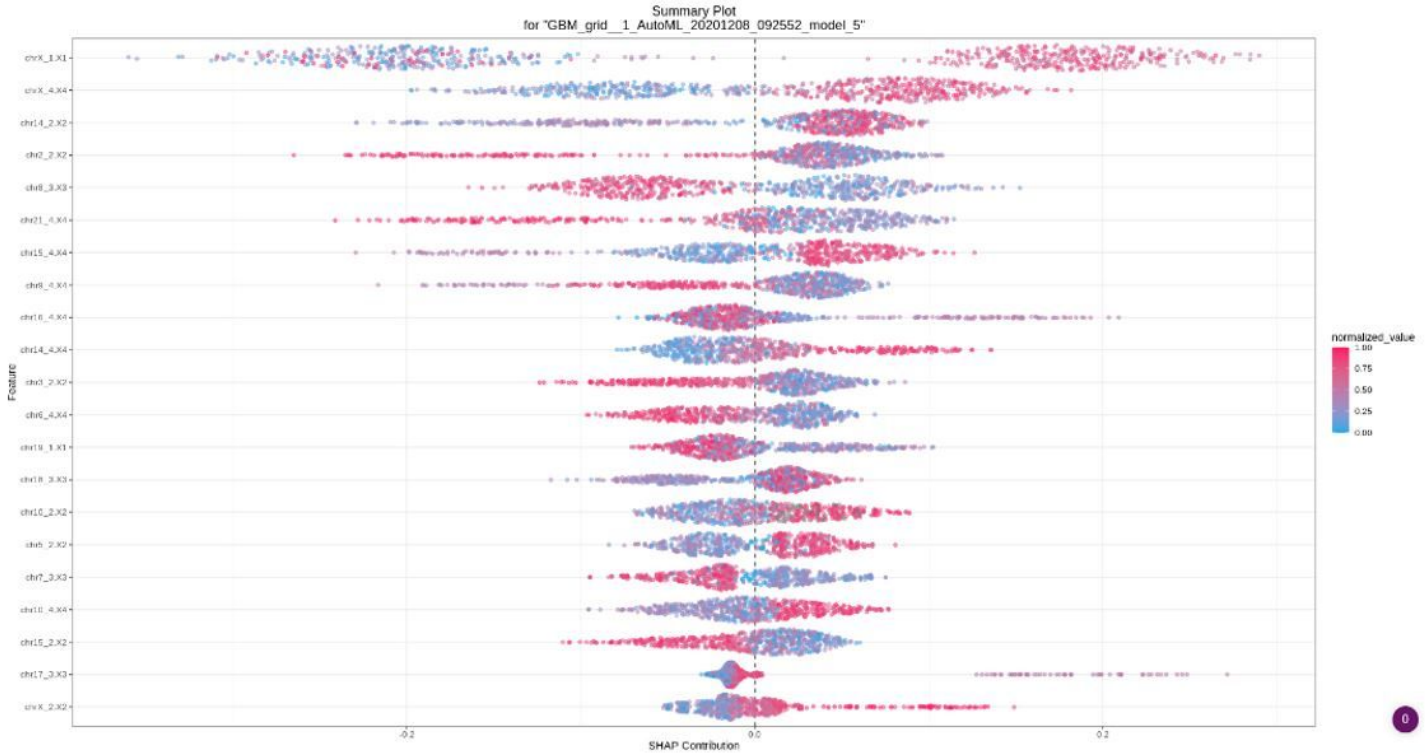

Figure 5

This SHAP plot indicates that the leading model for our 4-splits model relied heavily on the first quarter and last quarter value of chromosome X with some contribution from other regions and the second quarter of chromosome X.

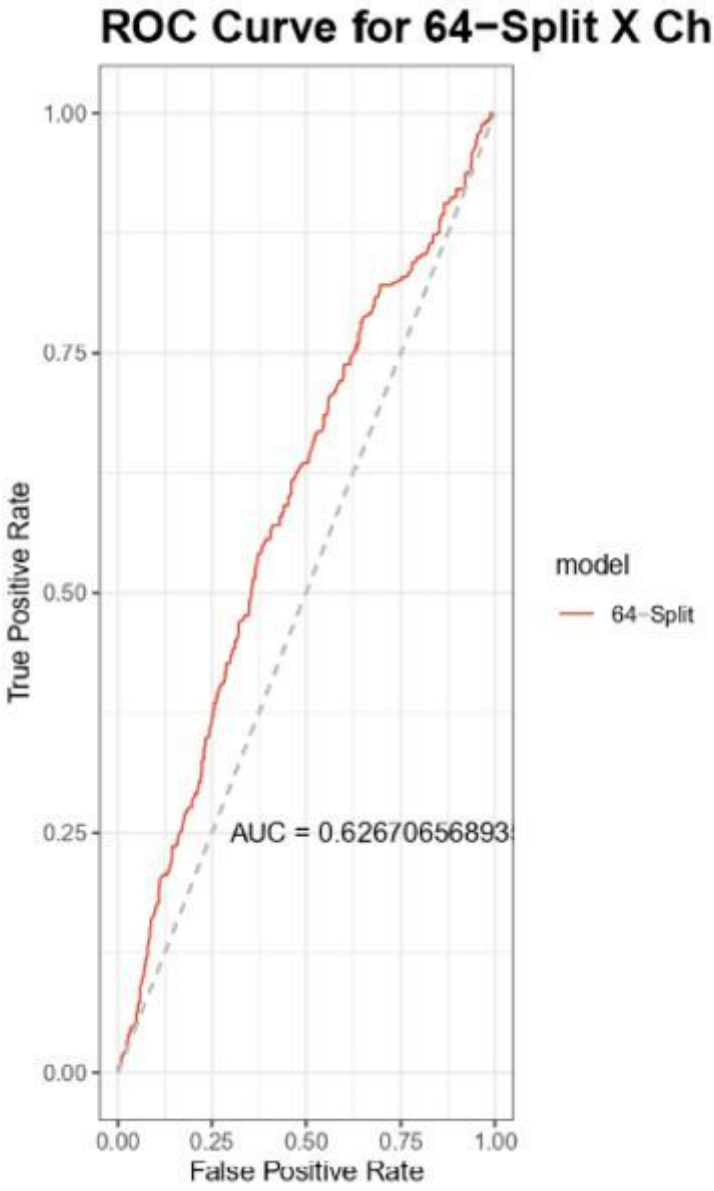

Figure 6

This ROC Curve for a schizophrenia prediction model utilizing 64-splits or 64 CSLVs of chromosome X only. The reported AUC is 0.627

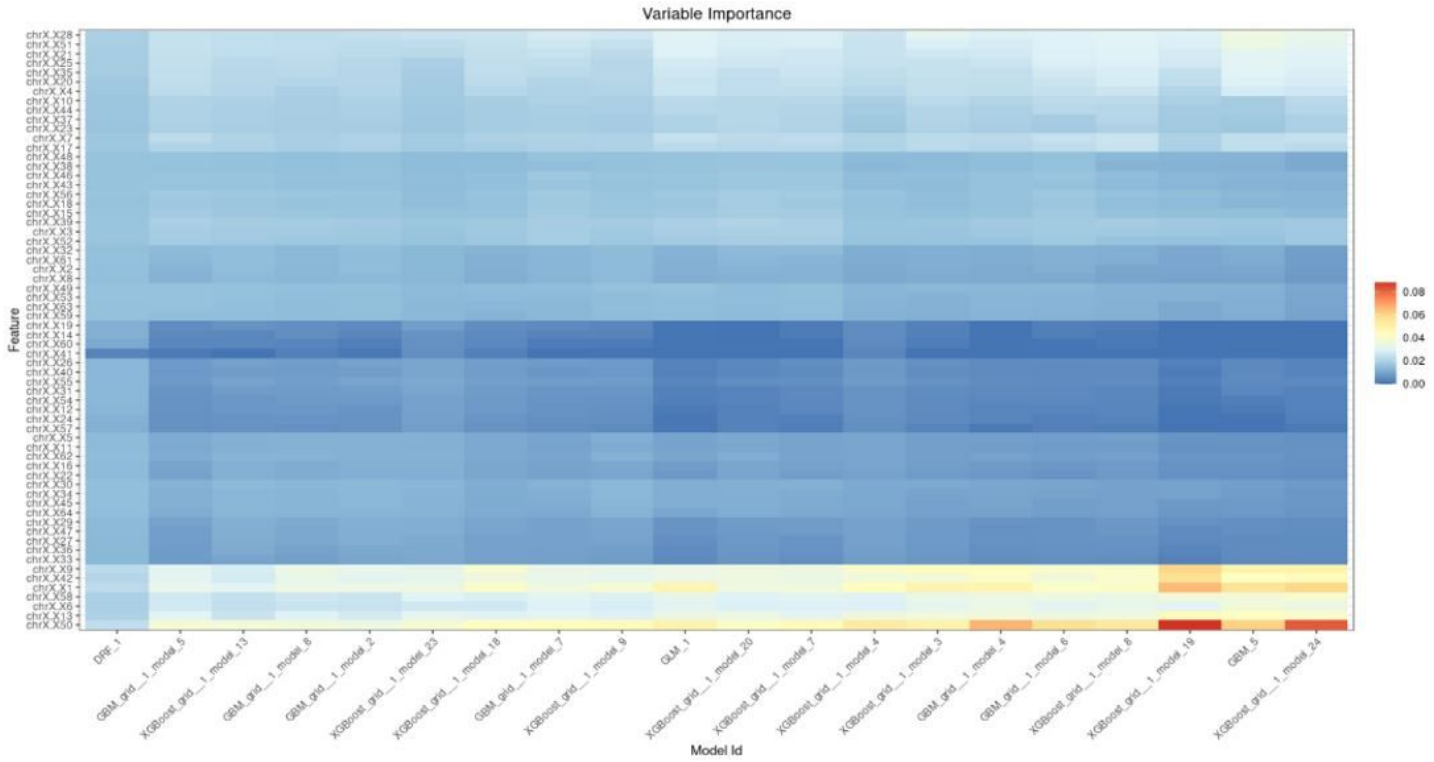

Figure 7

This variable importance heatmap shows the variables which most affected the performance and outcome of decisions made by the specified model. A value closer to 1.0 indicates higher importance of that variable. In most of the models we find that the CSLV values were mostly centered around split 50, 1, 9, 42, 13, 58, and 6. This is consistent with Figure 4.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementalfigures.docx