Predicting Schizophrenia

Henning Heyen

University College London

Abstract. In this study, we examine a linear, tree-based and non-linear machine learning model to predict schizophrenia based on grey matter in high an low-dimensional structural MRI images. We find that logistic regression with l1 regularization outperforms the other models even at low computational costs.

Keywords: schizophrenia prediction · MRI · machine learning

1 Introduction

Motivation: Schizophrenia is a chronic and severe mental disorder that affects a person's thinking, emotions, and behaviour. It is characterized by a range of symptoms, including delusions, hallucinations and disorganized speech or thinking [7]. The lifetime prevalence of schizophrenia is about 1,25% [2]. The disease decreases life expectancy by approximately 20 years [6] and onsets typically in adolescence or early adulthood [12]. Failing to address prolonged untreated psychosis can result in worse outcomes for those affected [5]. Thus, early detection of schizophrenia could significantly enhance the quality of life. There is evidence from Magnetic Resonance Imaging (MRI) data that brain volumes differ for schizophrenic people [4, 15], which motivates predicting schizophrenia from structural MRI data using machine learning.

Background to this project: The objective of this study is to utilize grey matter measurements obtained from structural MRI brain scans to create a predictor that can differentiate between schizophrenic patients and healthy controls. Previous work in machine learning for schizophrenia prediction from MRI data has shown promising results [10, 16, 14]. In this report, we compare the predictive and computational performance of a linear (Logistic Regression), a tree-based (Random Forest) and a non-linear model (Multi-Layer Perceptron).

Dataset: The dataset was obtained from a RAMP challenge[1]. The MRI scans are already preprocessed with the cat12 software. There are 410 samples in the training set and 103 samples in the test set. The data comes in two different formats: (1) Regions Of Interest (ROI) of grey matter and (2) Voxel-based morphometry (VBM) 3D images. The ROI data has 284 features, while the flattened VBM data has 331695 features (voxels). 54% of the participants are diagnosed with schizophrenia and 46% are healthy controls. The mean age of the participants is 33 (std: 12.5). There is an imbalance in gender (63%: 0, 37%: 1). The encoding for gender was not documented.

2 Methods

Models: (1) *Logistic Regression* is a linear model that is commonly used for binary classification problems. It computes the log odds of the event as a linear function of the input features. Each class is assigned to a probability by applying the sigmoid function. L1 and L2 regularization can help to improve the generalizability of logistic regression. L1 penalty can be of particular interest when working with high dimensional feature spaces as it can help to select a subset of features by setting less important feature weights to zero. Since the feature-sample ratio is low in this task (especially for the VBM

[1] https://ramp.studio/events/brain_anatomy_schizophrenia_UCL_2023

data, 513/331695 = 0.001), logistic regression is a promising candidate. Despite its simplicity, logistic regression has previously been shown to outperform more complex models for mental illness prediction based on neuroimages [17]. (2) *Random Forest* is an ensemble learning method that builds a collection of decorrelated decision trees, and aggregates their predictions (majority vote for classification) to obtain a final prediction. The decision trees are based on the random selection of features and samples, which can help to reduce variance and prevent overfitting. In contrast to linear models, random forests can capture non-linear relationships, which can be useful as the relationship between brain grey matter volume and schizophrenia is believed to be nonlinear [8]. Additionally, random forests can handle high dimensional data better than other tree-based models (e.g. gradient boosting was not feasible for the VBM data due to its memory intensity [9]). Random forests have previously been used for childhood-onset schizophrenia prediction using structural MRI [3]. (3) *Multi-Layer-Perceptron* (MLP) is a type of artificial neural network that consists of layers of interconnected nodes. The first layer receives input data, and the last layer produces output predictions. In between, there can be one or more hidden layers, each composed of a set of neurons that transform the input using activation functions and weights. Neural networks have been shown to be successful in detecting tumours and multiple sclerosis from MRI data [13, 11]. Neural networks tend to perform better on high dimensional data, which motivates the usage of MLP, especially for the VBM data.

Pipeline: For the low dimensional data (ROI), all relevant hyperparameters of the three models were tuned using GridSearch with cross-validation. As a cross-validation strategy, Stratified K-fold (SKF) with five folds was chosen to ensure the same class distribution for each fold, which can lead to a more reliable and robust model evaluation. The data was randomly shuffled before splitting to reduce biases. Given that there is evidence of gender differences for schizophrenia in symptoms and the age of onset [1], Stratified Group K-fold (SGKF) was additionally performed considering sex as a group on the ROI data. The dataset is also imbalanced in gender, and we found that both groups are differently affected by schizophrenia, which further motivates SGKF. To decorrelate both groups, SGKF ensures that each group is represented in a different fold. Therefore SGKF uses two splits. Unfortunately, performing cross-validation and hyperparameter tuning on the high dimensional VBM data was computationally too expensive to run locally (2020 M1 8GB RAM Macbook Air). Therefore default parameters were used to evaluate the three models on the high dimensional data.

Metrics: The metrics being used in this report correspond the metrics in the RAMP challenge, i.e. balanced accuracy (BACC) and area under the curve (AUC). Both are common metrics in the domain of machine learning in neuroimaging. For hyperparameter tuning AUC was used as the primary metric. Additionally, training times are reported to compare the models with respect to computational costs. AUC measures the ability of a classification model to distinguish between positive and negative cases across different threshold values. It is calculated as the area under the Receiver Operating Characteristic (ROC) curve, which is a plot of the true positive rate (TPR) against the false positive rate (FPR) for different threshold values. The AUC can take values between 0 and 1, where an AUC of 1 represents a perfect classifier, and an AUC of 0.5 represents a random classifier. Since there is a slight imbalance in the target variable (236 schizophrenia, 277 control), BACC can give more reliable results than accuracy. It is defined as the arithmetic mean of the true positive rate (TPR) and true negative rate (TNR):

$$\text{BACC} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP}+\text{FN}} + \frac{\text{TN}}{\text{TN}+\text{FP}}\right)$$

where TP, TN, FP, and FN represent the number of True Positives, True Negatives, False Positives, and False Negatives, respectively. Finally, to compare the models from a computational costs perspective, the mean fit time from the cross-validation is reported for the ROI data. For the VBM data, the single fit time on the training data is provided.

3 Results

Low Dimensional Data (ROI) : The results for the ROI data are summarized in table 1 for both stratified KFold and stratified group KFold. The table contains train and test evaluation scores, mean validation scores from cross-validation and mean training times. Standard deviations are reported in parentheses. In general, we find that the results are very similar across both cross-validation strategies. Logistic regression is outperforming the other models (test AUC≈0.85). Grid search resulted in strong l1 regularization (C=0.1). MLP performs slightly worse (test AUC≈0.82) followed by random forest (test AUC≈0.80). Despite hyperparameter tuning, both MLP and random forest are strongly overfitting, indicated by the train performance. The best MLP uses a single 100-neuron hidden layer, and the best random forest uses 300 random trees with a max depth of 5. While all models take a similar amount of time to train, random forest showed the lowest mean fit time for both cross-validation strategies (7.9s and 7.7s, respectively). Table 1a illustrates the ROC curves. Again, we can see that the performance is very similar for both cross-validation strategies and that logistic regression performs best.

CV	model	test auc	test bacc	train auc	train bacc	mean val auc	mean val bacc	mean fit time
SKF	LR	**0.844**	**0.781**	0.882	0.801	0.823 (0.047)	0.757 (0.06)	9.5s (0.1s)
	RF	0.798	0.757	0.996	0.951	0.786 (0.058)	0.716 (0.065)	**7.9s (0.5s)**
	MLP	0.823	0.731	0.906	0.821	0.823 (0.033)	0.718 (0.042)	26.1s (0.9s)
SGKF	LR	**0.855**	**0.781**	0.929	0.827	0.793 (0.034)	0.699 (0.009)	8.9s (0.0s)
	RF	0.804	0.748	0.995	0.951	0.78 (0.089)	0.697 (0.077)	**7.7s (0.2s)**
	MLP	0.827	**0.781**	1.0	1.0	0.808 (0.048)	0.706 (0.035)	8.8s (0.2s)

Table 1: Results on Low Dimensional Data (ROI)

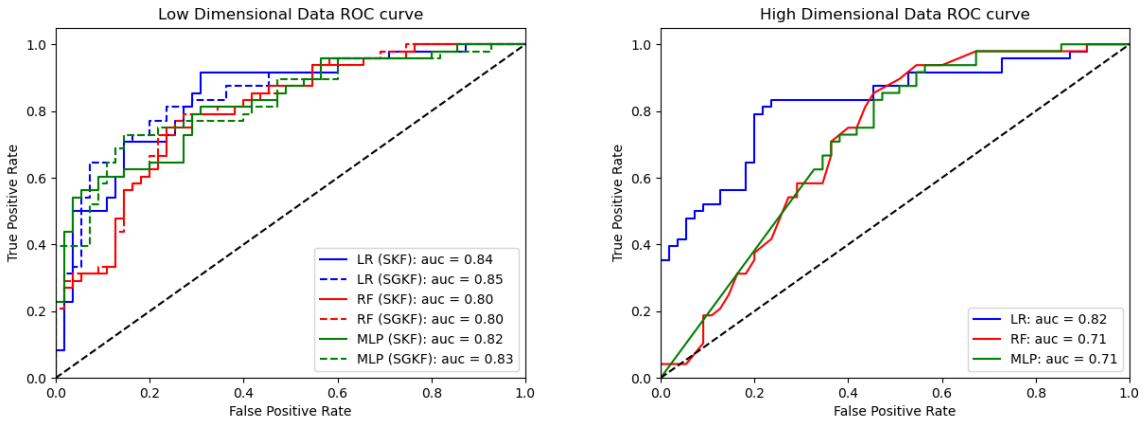
High Dimensional Data (VBM) For the high dimensional data, cross-validation and hyperparameter tuning were not feasible due to computational constraints. The results are depicted in table 2. Running all three models with default configuration, we find that the logistic regression is again outperforming the other two models (test AUC≈0.80). Random forest and MLP perform much worse on the VBM data compared to the ROI data (test AUC≈0.71 for both). All three models are clearly overfitting (train AUC=1.0). While both logistic regression and random forest were trained in less than a minute, MLP fit time is about 4 minutes. However, changing the hyperparameter varied the fit time dramatically (e.g. using the liblinear solver instead of the default lbfgs in logistic regression resulted in train times of around 6 minutes). Interestingly, all three models perform similarly for a high TPR threshold (>0.8), while logistic regression clearly dominates for lower TPR thresholds (see table 1b).

model	test auc	test bacc	train auc	train bacc	fit time
LR	**0.809**	**0.719**	1.0	1.0	13.7s
RF	0.712	0.612	1.0	1.0	**10.4s**
MLP	0.713	0.669	1.0	1.0	4min 36s

Table 2: Results on High Dimensional Data (VBM)

4 Discussion

Overall we find that reducing the number of features with regions of interest helped to improve the model's performance and drastically reduces the computational costs. Furthermore, the high dimensional data



(a) Low dimensional data (ROI)

(b) High dimensional data (VBM)

Fig. 1: ROC curves

was more prone to overfitting than the low dimensional data. This emphasises the necessity of adjusted hyperparameter tuning for the VBM data. However, more computational resources are needed for cross-validation and hyperparameter tuning to prevent overfitting. Interestingly, despite its simplicity, logistic regression performed best on both the high and low-dimensional data in the prediction of schizophrenia. L1 regularization seems to be effective in dealing with the high number of features. The good performance of logistic regression also indicates that the relationship between brain grey matter and schizophrenia can be partly reduced to a simple linear relation. The l1 regularized logistic regression onn the RIO data was finally submitted to the RAMP challenge (AUC$\approx$ 0.84). The MLP was the most computationally intensive model, which is a common drawback for neural networks. In comparison with logistic regression and random forests, neural networks are also considered to be black box models meaning that their predictions are hard to interpret. Logistic regression and random forest, on the other hand, can be explained by their feature weights and feature importances, respectively. As far as stratified group cross-validation is concerned, we find that there was no improvement in performance using gender as a group. The models actually tend to overfit more. The fact that gender was considered as a group also resulted in only two folds which reduces the robustness of the cross-validation results. The gender of the participants, however, was highly imbalanced, which can motivate the usage of group cross-validation to improve the generalizability of unseen data.

5 Conclusion

This report examined three different machine learning models for schizophrenia prediction based on structural MRI data. We found that a simple logistic regression with l1 regularization generalizes very well for both low-dimensional (ROI) and high-dimensional (VBM) images while being computationally less expensive than neural networks. We utilized the ability of l1 regularization to naturally select features by setting some feature weights to zero. Random forest was computationally efficient but failed to generalize well. Future research should also tune models on the VBM images using more powerful hardware resources, which will probably leverage the performance of the non-linear models. Furthermore, the number of training samples is very limited. Especially the performance of neural networks will improve as more training data is available. This study used a simple single hidden layer neural network. More sophisticated architectures (e.g. CNN) with regularization (e.g. dropout layers) can potentially lead to much better results. Considering gender as a group for cross-validation did not improve the performance but may help to generalize better as neuroimages might be correlated depending on gender, especially as the distribution of gender is highly imbalanced in this dataset.

References

1. Abel, K.M., Drake, R., Goldstein, J.M.: Sex differences in schizophrenia. International review of psychiatry **22**(5), 417–428 (2010)
2. Chang, W.C., Wong, C.S.M., Chen, E.Y.H., Lam, L.C.W., Chan, W.C., Ng, R.M.K., Hung, S.F., Cheung, E.F.C., Sham, P.C., Chiu, H.F.K., et al.: Lifetime prevalence and correlates of schizophrenia-spectrum, affective, and other non-affective psychotic disorders in the chinese adult population. Schizophrenia Bulletin **43**(6), 1280–1290 (2017)
3. Greenstein, D., Malley, J.D., Weisinger, B., Clasen, L., Gogtay, N.: Using multivariate machine learning methods and structural mri to classify childhood onset schizophrenia and healthy controls. Frontiers in psychiatry **3**, 53 (2012)
4. Haijma, S.V., Van Haren, N., Cahn, W., Koolschijn, P.C.M., Hulshoff Pol, H.E., Kahn, R.S.: Brain volumes in schizophrenia: a meta-analysis in over 18 000 subjects. Schizophrenia bulletin **39**(5), 1129–1138 (2013)
5. Howes, O.D., Whitehurst, T., Shatalina, E., Townsend, L., Onwordi, E.C., Mak, T.L.A., Arumuham, A., O'Brien, O., Lobo, M., Vano, L., et al.: The clinical significance of duration of untreated psychosis: an umbrella review and random-effects meta-analysis. World Psychiatry **20**(1), 75–95 (2021)
6. Laursen, T.M., Nordentoft, M., Mortensen, P.B.: Excess early mortality in schizophrenia. Annual review of clinical psychology **10**, 425–448 (2014)
7. McCutcheon, R.A., Marques, T.R., Howes, O.D.: Schizophrenia—an overview. JAMA psychiatry **77**(2), 201–210 (2020)
8. Mechelli, A., Friston, K.J., Frackowiak, R.S., Price, C.J.: Structural covariance in the human cortex. Journal of Neuroscience **25**(36), 8303–8310 (2005)
9. Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Frontiers in neurorobotics **7**, 21 (2013)
10. Oh, J., Oh, B.L., Lee, K.U., Chae, J.H., Yun, K.: Identifying schizophrenia using structural mri with a deep learning algorithm. Frontiers in psychiatry **11**, 16 (2020)
11. Shoeibi, A., Khodatars, M., Jafari, M., Moridian, P., Rezaei, M., Alizadehsani, R., Khozeimeh, F., Gorriz, J.M., Heras, J., Panahiazar, M., et al.: Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review. Computers in Biology and Medicine **136**, 104697 (2021)
12. Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., Il Shin, J., Kirkbride, J.B., Jones, P., Kim, J.H., et al.: Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Molecular psychiatry **27**(1), 281–295 (2022)
13. Ullah, F., Ansari, S.U., Hanif, M., Ayari, M.A., Chowdhury, M.E.H., Khandakar, A.A., Khan, M.S.: Brain mr image enhancement for tumor segmentation using 3d u-net. Sensors **21**(22), 7528 (2021)
14. Vieira, S., Gong, Q.y., Pinaya, W.H., Scarpazza, C., Tognin, S., Crespo-Facorro, B., Tordesillas-Gutierrez, D., Ortiz-García, V., Setien-Suero, E., Scheepers, F.E., et al.: Using machine learning and structural neuroimaging to detect first episode psychosis: reconsidering the evidence. Schizophrenia bulletin **46**(1), 17–26 (2020)
15. Vita, A., De Peri, L., Deste, G., Sacchetti, E.: Progressive loss of cortical gray matter in schizophrenia: a meta-analysis and meta-regression of longitudinal mri studies. Translational psychiatry **2**(11), e190–e190 (2012)
16. Xiao, Y., Yan, Z., Zhao, Y., Tao, B., Sun, H., Li, F., Yao, L., Zhang, W., Chandan, S., Liu, J., et al.: Support vector machine-based classification of first episode drug-naïve schizophrenia patients and healthy controls using structural mri. Schizophrenia research **214**, 11–17 (2019)
17. Yassin, W., Nakatani, H., Zhu, Y., Kojima, M., Owada, K., Kuwabara, H., Gonoi, W., Aoki, Y., Takao, H., Natsubori, T., et al.: Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis. Translational psychiatry **10**(1), 278 (2020)