Elastic-Remurs: Multi-Linear Regression for the Diagnosis of Autism Spectrum Disorder

Owen Cook
University of Sheffield, UK
oscook1@sheffield.ac.uk

Abstract

Machine learning methods are becoming increasingly applicable to the medical industry. With machines capable of capturing high quality data, the classification accuracy of brain conditions, such as Autism Spectrum Disorder (ASD), is improving. This study applies Regularised multilinear regression and selection (Remurs) to the Autism Brain Imaging Data Exchange (ABIDE) dataset and introduces the Elastic-Remurs regression method, an extension of Remurs, taking inspiration from the elastic net. Elastic-Remurs consistently performed better in the classification task than most other classical machine learning methods tested in this study, including: Remurs, logistic regression (using both L_1-norm and L_2-norm penalties), and the Support Vector Machine (SVM). After comparing different machine learning methods, the effectiveness of Elastic-Remurs and Remurs is compared to state-of-the-art ASD classification methods using a single brain atlas and optionally using the phenotypic data within the ABIDE dataset. Remurs achieves 69.8% accuracy and Elastic-Remurs an accuracy of 70.3% in leave-one-site-out cross-validation, using the Tangent Pearson functional connectivity measure and the CC200 brain atlas.

Introduction

The number of people diagnosed with Autism Spectrum Disorder (ASD) has grown significantly over the last two decades. In many cases, a diagnosis can provide a great benefit and help a person deal with their condition more effectively (Wong et al. 2015). With waiting lists for a diagnosis so long and waiting times on the scale of years, the ASD diagnostic process can be stressful (Crane et al. 2016). PyKale's research surrounding the classification of brain conditions could help supplement diagnoses of neurodivergence, assisting medical professionals to make quicker and more accurate assessments. The quicker doctors can diagnose a condition correctly, the sooner patients can be helped.

This study aims to improve the classification accuracy of ASD, and potentially other brain conditions, by modifying components of the PyKale pipeline (Lu et al. 2022). The key area of focus in this study is in improving the classification stage of the pipeline. The main methods applied in this study are Regularized multilinear regression with embedded feature selection (Remurs) (Song and Lu 2017) and the newly proposed Elastic-Remurs.

With the PyKale team's focus on green machine learning principles (Lu et al. 2022), the pipeline used to classify participants with ASD may be transferable to identifying those with ADHD or other brain conditions. Although this application is beyond the scope of the study, given positive results, further research could be conducted to assess the proficiency of the methods discussed in such classification tasks. The code used to run experiments will be made open-source and available at https://github.com/MiniEggz/asd-diagnosis.

Related Works

Remurs

Regularized multilinear regression with embedded feature selection (Remurs) (Song and Lu 2017) is an extension of Lasso which maintains the structure of the data. With the optimisation problem defining Lasso involving a vector input $\boldsymbol{x}$ and the coefficient vector $\boldsymbol{w}$, the Remurs optimisation problem involves a tensor input $\mathcal{X}$ and a coefficient tensor $\mathcal{W}$:

$$\min_{\mathcal{W}} \frac{1}{2} \sum_{m=1}^{M} (y_m - \langle \mathcal{X}_m, \mathcal{W} \rangle)^2 + \tau ||\mathcal{W}||_* + \gamma ||\mathcal{W}||_1, \quad (1)$$

where

$$||\mathcal{W}||_1 = \sum_{i_1} \sum_{i_2} ... \sum_{i_N} |\mathcal{W}(i_1, i_2, ..., i_N)|, \quad (2)$$

and

$$||\mathcal{W}||_* = \frac{1}{N} \sum_{n=1}^{N} ||\boldsymbol{W}_{(n)}||_*.$$

$||\mathcal{W}||_*$ is the definition of the tensor nuclear norm based on the unfolded matrices, and $||\mathcal{W}||_1$ the entrywise L_1-norm. τ and γ are hyperparameters used to balance the regularisation. Within the regularisation term, there is a trade-off between sparsity and low-rank; sparsity is controlled by γ and low-rank enforced by τ. As with Lasso, the embedded feature selection lies in the regularisation term.

In the experiments carried out by Song and Lu, the Remurs method consistently outperformed other methods with an average accuracy of 78.15%.

Elastic Net

Elastic net regression (Zou and Hastie 2005) combines the strengths of the Lasso (Tibshirani 1996) and Ridge (Hoerl and Kennard 2000) regression methods, balancing the properties of sparsity and smoothness. In the experiments carried out by Song and Lu 2017, elastic net only placed behind Remurs in the classification task on the CMU2008 dataset with an average accuracy of 75.46%.

State of the Art

The Autism Brain Imaging Data Exchange (ABIDE) dataset (Craddock et al. 2013) is a common benchmark for models diagnosing ASD. Previous works have improved classification accuracy through using different, or combined, brain atlases and functional connectivity features, domain adaptation, phenotypic information, and various classification models. Many of these studies employed leave-one-site-out cross-validation (LOSO-CV) to capture the variance in data collection from the different sites within the ABIDE dataset (Kazeminejad and Sotero 2020), (Abbas, Chi, and Chen 2023). Each fold uses one site as the test set and all other sites as the training set.

One study combined the CC200 (Craddock et al. 2012), AAL90 (Tzourio-Mazoyer et al. 2002), and DOS160 (Dosenbach et al. 2010) atlases, feeding them into a multi-input single-output deep neural network (MISO-DNN) (Epalle et al. 2021), obtaining an accuracy of 77.2% using LOSO-CV; this study used 1038 participants from the ABIDE-1 repository. Another study, using 855 participants from the ABIDE-1 repository, developed a deep multimodal neuroimaging framework for diagnosing ASD (Abbas, Chi, and Chen 2023), achieving an accuracy of 78.3%; from surveying the relevant literature, this is the highest classification accuracy achieved with LOSO-CV on the ABIDE dataset. Other studies employing deep-learning approaches use graph convolutional networks (Parisot et al. 2018) and deep neural networks (Heinsfeld et al. 2018). Heinsfeld et al. 2018 obtained 65% using LOSO-CV.

Kunda et al. 2023 introduced a new second-order functional connectivity measure (Tangent Pearson), utilised the phenotypic data within the ABIDE dataset, and applied maximum independence domain adaptation (MIDA). The most performant classifier used in this study was the Ridge classifier. Through applying Tangent Pearson, MIDA, and utilising the phenotypic data, an accuracy score of 71.4% was attained using 1035 subjects, where the data has not been quality checked. On the quality-checked subset of 871 subjects, Kunda et al. 2023 achieved 70.3%; in this case classification using Tangent Pearson without MIDA performed the best.

Analysis

CMU2008 Dataset

To implement the Remurs method in Python, classification was performed on a subset of the CMU2008 dataset (Mitchell et al. 2008) to ensure it matches the MATLAB implementation released alongside the paper introducing Remurs (Song and Lu 2017).

The dataset includes fMRI data of the brain associated with the meanings of nouns. The subset of data contains the rescaled data from only one participant, with each fMRI scan being of size 16 x 19 x 7 (2,128 voxels). As in the MATLAB Remurs code, the data was used for the binary classification task: “animal” vs. “tool”, where “animal” combines the “animal” and “insect” classes, and “tool” combines the “tool” and “furniture” classes.

ABIDE Dataset

The ABIDE dataset (Craddock et al. 2013), available through the nilearn Python package (https://github.com/nilearn/nilearn), provides a large amount of high-quality, preprocessed data, including fMRI brain scans from 505 subjects diagnosed with ASD and 530 controls that can be compared to other Kunda et al. 2023. There is also a quality-checked subset of this data containing 871 of the 1035 total subjects. To match Kunda et al. 2023, this work used the CC200 brain atlas.

The preprocessed fMRI data comes from 20 different sites with variations in size. To correct this, a connectivity measure is applied to ensure the data is of shape 200 x 200, allowing for inter-site comparisons in the classification pipeline. The two different connectivity measures used in this study were Pearson Correlation and Tangent Pearson (Kunda et al. 2023).

Elastic-Remurs

Elastic-Remurs aims to yield the benefits of the elastic net (Zou and Hastie 2005) and Remurs. Elastic net combines the Lasso and Ridge regression penalties, balancing the sparsity and smoothness of the coefficient tensor; Remurs, as discussed above, balances the properties of low-rank and sparsity. Combining the elastic net and Remurs, the task becomes balancing the low-rank, sparsity, and smoothness properties induced by the nuclear norm, L_1-norm, and L_2-norm penalties.

As Elastic-Remurs is an extension of the Remurs problem (1), the new Elastic-Remurs minimisation problem can be written as:

$$\min_{\mathcal{W}} \frac{1}{2} \sum_{m=1}^{M} (y_m - \langle \mathcal{X}_m, \mathcal{W} \rangle)^2 + \tau ||\mathcal{W}||_* + \gamma ||\mathcal{W}||_1 + \lambda ||\mathcal{W}||_2. \tag{3}$$

Due to the time constraints of this study, a formal inclusion of the L_2-norm penalty was not implemented. To provide a proof of concept, a smoothing penalty has been applied (smooth$(\mathcal{V}, \lambda) = \mathcal{V}/(1+\lambda)$, where $\mathcal{V}$ is the input tensor and λ controls the strength of the penalty). As λ increases, the input tensor $\mathcal{V}$ becomes smoother. When $\lambda = 0$, the smoothing penalty is an identity function. Future work may investigate the formalisation of the Elastic-Remurs method. This prototype method will be tested on both the CMU2008 and ABIDE datasets.

Experimental Design

Validation using CMU2008

The subset of the CMU2008 dataset (Mitchell et al. 2008), released with the MATLAB code for the Remurs method (Song and Lu 2017), was used to assess whether the Remurs method had been implemented correctly and that the Elastic-Remurs prototype method was viable. Given the results were at least competitive with the Remurs method, it would be applied to the ABIDE dataset (Craddock et al. 2013).

Testing on ABIDE

After confirming that Elastic-Remurs could be suitably applied to the binary classification problem using the CMU2008 dataset, Elastic-Remurs and Remurs were compared to other machine learning methods (linear regression, logistic regression, the SVM) in classifying data using the ABIDE dataset. The testing pipeline was an adaptation of the ”multisite neuroimg adapt” example within the PyKale repository (https://github.com/pykale/pykale/tree/main/examples/multisite_neuroimg_adapt). LOSO-CV was the chosen method of cross-validation, capturing the variance in data collection from different sites; results from this method of cross-validation are also comparable with other studies, as mentioned in the related works section.

Classification Pipeline The preprocessed ABIDE dataset was loaded using the “nilearn” package (https://github.com/nilearn/nilearn). After loading the dataset, the fMRI data and diagnosis was extracted and the brain networks were standardised using a connectivity measure, ensuring the size of the data was consistent (200 x 200). The cross-validation function was then defined, using LOSO-CV. This cross-validation function takes an estimator object as a parameter. The estimator creates a model of the training data using the preprocessed fMRI brain training data and binary labels from the training set; these one-hot labels represent subjects with ASD and those without ASD. The estimator predicts the output for unlabelled test data. The predicted labels can then be compared to the real values for each participant, producing an accuracy score.

Hyperparameter Selection To select optimal hyperparameters for each regression method, grid search was used. This involved running the cross-validation pipeline on all permutations of the parameters. As the number of hyperparameters, or the number of values being searched for each hyperparameter, increases, the search becomes increasingly computationally expensive. This computational limit led to a strategic selection of values being used within the grid search. In future research, this search process could be scaled horizontally to increase the search space and potentially find more optimal hyperparameter values.

Experimental Results

CMU2008

After verifying that the Python implmentation of Remurs behaved the same as the MATLAB implementation, Remurs and Elastic-Remurs were compared. K-fold cross-validation was used, comparing results on each fold with each model's optimal hyperparameters.

The hyperparameter search space for each method is represented by the two cartesian products below.

$$\text{Remurs} = \{(\alpha, \beta) \mid \alpha, \beta \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 500\}\}$$

$$\text{Elastic-Remurs} = \{(\alpha, \beta, \gamma) \mid \alpha, \beta, \gamma \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 500\}\}$$

Fold	Remurs	Elastic-Remurs
1	91%	94%
2	93%	98%
3	95%	98%
4	99%	100%
5	96%	96%
6	96%	96%

Table 1: Comparison between Remurs and Elastic-Remurs on a subset of the CMU2008 dataset, showing the highest accuracy score achieved by each method on each fold.

From the results presented in Table 1, Elastic-Remurs, with optimal hyperparameter selection, performed as well or better than Remurs in all folds. With Remurs already outperforming other linear regression methods in the classification problem (as shown in the paper introducing Remurs (Song and Lu 2017)), these results were very positive, indicating both methods may successfully be applied to the ABIDE dataset.

ABIDE

After successful applications to the CMU2008 dataset, Elastic-Remurs and Remurs were applied to the ABIDE dataset, comparing their performance against other machine learning methods and other studies achieving state-of-the-art results using only the CC200 brain atlas, optionally using the phenotypic data included in the ABIDE dataset.

The cartesian products below represent the hyperparameter search spaces for each method:

$$\begin{aligned}\text{Elastic-Remurs} = \{(\alpha, \beta, \gamma) \mid \\ \alpha \in \{1, 100, 500\}, \\ \beta \in \{0.1, 1, 5, 1000\}, \\ \gamma \in \{0.1, 1, 100, 1000\} \\ \}\end{aligned}$$

$$\text{Remurs} = \{(\alpha, \beta) \mid \alpha \in \{75, 100, 250, 750, 1000\}, \beta \in \{0.0001, 0.001, 0.01, 0.1, 1\} \}$$

$$\text{Ridge} = \{ \alpha \mid \alpha \in \{1, 250, 500, 750, 1000\} \}$$

$$\text{Logistic (L1)} = \{ \alpha \mid \alpha \in \{0.001, 0.1, 250, 500, 1000, 1500, 2000\} \}$$

$$\text{Logistic (L2)} = \{ \alpha \mid \alpha \in \{0.001, 0.1, 250, 500, 1000, 1500, 2000\} \}$$

$$\text{SVM} = \{ \alpha \mid \alpha \in \{0.1, 1, 10, 50, 100, 500, 1000\} \}$$

Comparing Classical Machine Learning Methods This section compares classical machine learning methods in the ASD classification task, using LOSO-CV and the Pearson Correlation functional connectivity measure. It was hypothesised that Remurs and Elastic-Remurs would perform better than other classical machine learning methods.

Method	Quality Checked	Accuracy
Elastic-Remurs	×	69.2%
Remurs	×	68.9%
Ridge	×	**69.4%**
Logistic (L1)	×	67.6%
Logistic (L2)	×	68.7%
SVM	×	67.4%
Elastic-Remurs	✓	**69.9%**
Remurs	✓	68.8%
Ridge	✓	69.2%
Logistic (L1)	✓	68.3%
Logistic (L2)	✓	68.5%
SVM	✓	67.2%

Table 2: Comparison of machine learning methods using the Pearson Correlation functional connectivity measure and leave-one-site-out cross-validation on the ABIDE dataset.

From the results in Table 2, the Elastic-Remurs and Ridge classifiers were the two most performant, obtaining either the highest or second highest score in both the whole dataset and the quality-checked subset. Linear and logistic regression methods promoting sparsity, were shown to obtain lower accuracy scores than those with a smoothing penalty, including the Elastic-Remurs penalty and L_2 regularisation. It is also worth noting that the Ridge classifier was significantly quicker to train than both Remurs and Elastic-Remurs, due to quicker fitting and a much smaller hyperparameter search space.

Site ID	Scanner	Accuracy
CALTECH	SIEMENS Trio	80.0%
CMU	SIEMENS Verio	72.7%
KKI	Phillips Achieva	72.7%
LEUVEN_1	Phillips Intera	57.1%
LEUVEN_2	Phillips Intera	67.9%
MAX_MUN	SIEMENS Verio	60.9%
NYU	SIEMENS Allegra	77.3%
OHSU	SIEMENS Trio	56.0%
OLIN	SIEMENS Allegra	75.0%
PITT	SIEMENS Allegra	70.0%
SBL	Phillips Intera	61.5%
SDSU	GE MR750	74.1%
STANFORD	GE Signa	72.0%
TRINITY	Phillips Achieva	68.2%
UCLA_1	SIEMENS Trio	68.8%
UCLA_2	SIEMENS Trio	81.0%
UM_1	GE Signa	68.6%
UM_2	GE Signa	73.5%
USM	SIEMENS Trio	61.2%
YALE	SIEMENS Trio	70.7%
Average	-	69.9%

Table 3: Comparison of site classification accuracy, using leave-one-site-out cross-validation, the Elastic-Remurs classifier and Pearson Correlation embedding. The hyperparameters chosen were: $\alpha = 500$, $\beta = 1000$, $\gamma = 1000$.

Table 3 shows the variance in inter-site classification accuracy using LOSO-CV. Despite multiple sites using the SIEMENS Trio scanner, there is large variance in the prediction accuracies between these sites. The classification accuracy at SDSU is notably higher than average, despite being the only site using the GE MR750 scanner. Future research could investigate how much influence the scanners and data collection methods have over the final classification accuracy at each site.

Intercomparison With Other Studies When compared with results from Kunda et al. 2023, shown in Table 4, Remurs and Elastic-Remurs both perform competitively in most cases. On all 1035 subjects, TP MIDA (Kunda et al. 2023) performed best with a score of 71.4%, 0.6% higher than TP Ridge. It is worth noting that TP Ridge and TP raw (Kunda et al. 2023) both used a Ridge classifier and Tangent Pearson embedding; TP raw used phenotypic data and TP Ridge did not. This study and Kunda et al. 2023 used differing hyperparameter search spaces for the Ridge classifier, potentially introducing the difference in accuracy

Method	Quality Checked	Accuracy
Elastic-Remurs	$\times$	69.2%
TP Elastic-Remurs	$\times$	70.3%
Remurs	$\times$	69.2%
TP Remurs	$\times$	69.8%
Ridge	$\times$	69.4%
TP Ridge	$\times$	<u>70.8%</u>
TP MIDA (Kunda)	$\times$	**71.4%**
TP raw (Kunda)	$\times$	70.0%
Elastic-Remurs	✓	<u>69.9%</u>
TP Elastic-Remurs	✓	69.0%
Remurs	✓	68.8%
TP Remurs	✓	68.4%
Ridge	✓	69.2%
TP Ridge	✓	68.8%
TP MIDA (Kunda)	✓	69.3%
TP raw (Kunda)	✓	**70.3%**

Table 4: Comparison of results obtained in this study and Kunda et al. 2023 using both Tangent Pearson (Kunda et al. 2023) and Pearson Correlation embedding in leave-one-site-out cross-validation. TP denotes that Tangent Pearson embedding has been used, otherwise Pearson Correlation has been used.

scores. Future research may aim to understand whether the Ridge, Remurs, and Elastic-Remurs classifiers can be improved by utilising the same hyperparameter search space in this study and the additional phenotypic data available in the ABIDE dataset.

In the quality-checked subset of the ABIDE dataset, TP raw performed the best with 70.3%, and Elastic-Remurs with the Pearson Correlation functional connectivity measure was second best with 69.9% accuracy. An interesting observation is that the Tangent Pearson connectivity measure performs worse than Pearson Correlation for Elastic-Remurs, Remurs, and Ridge using the 871 quality-checked samples.

When compared to other studies using different subsets of the ABIDE dataset, with 1038 (Epalle et al. 2021) or 855 (Abbas, Chi, and Chen 2023) participants, both employing deep learning approaches, the accuracy scores obtained in this study are significantly lower than the 77.2% and 78.3% obtained by Epalle et al. 2021 and Abbas, Chi, and Chen 2023 respectively. While these higher scores may be attributed to the use of deep learning techniques over traditional machine learning methods, it may also be due to the multimodal neuroimaging approaches and the use of multiple brain atlases in these two studies. Future work could investigate the performance of deep learning techniques and machine learning methods, using data from multiple brain atlases.

Discussion

Remurs

Application to ABIDE After showing high levels of performance classifying fMRI data using the CMU2008 dataset (Song and Lu 2017), applying the Remurs method to the ABIDE dataset was a logical next step with the potential to obtain a high classification accuracy. After applying Remurs to the ABIDE dataset, it is clear that it performs better than the majority of classical machine learning methods, including the SVM and logistic regression with L_1 and L_2 regularisation. While Remurs did not quite achieve state-of-the-art classification accuracy, it performed well without the use of the phenotypic data available, domain adaptation, or multiple brain atlases.

Potential Limitations of Sparsity The Remurs method has feature selection embedded through the sparsity inducing feature of the L_1-norm (Song and Lu 2017). On many datasets this can be desirable, as there may be many variables within the input tensor that are not helpful in predicting an output. Sparsity may be a problem, however, when there is high variability within the data. As the fMRI data in the ABIDE dataset comes from different sites and scanners, sparsity could cause issues by enforcing feature selection too strongly, disregarding variables that may carry some weight in classifying a sample correctly. To further support this, machine learning methods using L_2 regularisation performed better in all cases, for both linear and logistic regression methods. The Remurs method, therefore, may achieve better results on data with less variability, as demonstrated with the CMU2008 dataset (Song and Lu 2017).

Elastic-Remurs

Creating Elastic-Remurs Taking inspiration from the elastic net (Zou and Hastie 2005), an L_2-norm penalty was added to the design of Remurs to create Elastic-Remurs. In implementing Elastic-Remurs, however, deriving a mathematically complete solution for a closed-form proximal operator of the L_2-norm was infeasible, given the time constraints of this study, so a simplified shrinkage penalty was used to replace it, maintaining the balance between low-rank, sparsity, and smoothness, as outlined in the design. To be able to fully evaluate the potential of Elastic-Remurs, a complete solution should be implemented in future research. While a more complete solution to the Elastic-Remurs minimisation problem could be developed, it may not necessarily improve upon the results achieved by the implementation in this study. The design of Elastic-Remurs, introducing L_2 regularisation to Remurs, was further supported by the high performance of methods using the L_2-norm penalty.

Another minor issue with the implementation of Elastic-Remurs is that it does not degenerate to the Remurs method when the coefficient of the L_2-norm penalty is set to zero and it does not have the desired effect when the coefficient of the L_1-norm penalty is set to zero, as both the L_1-norm penalty and the shrinkage penalty applied in Elastic-Remurs are identity functions when their coefficients are zero. This indicates that the algorithm for Elastic-Remurs requires modification. Again, it is worth noting that this modification may not improve performance.

Application to ABIDE After competing with and performing better than Remurs on the CMU2008 dataset, the Elastic-Remurs also outperformed Remurs when applied to the ABIDE dataset, nearing state-of-the-art classification accuracy, using one brain atlas. In most cases, it achieved higher than all other linear and logistic classification methods tested within this study, performing slightly worse than the Ridge classifier in others. In some cases, Elastic-Remurs performed better than TP MIDA and TP raw without the additional input of phenotypic data or domain adaptation. As with Remurs, applying Elastic-Remurs in the classification step of the TP MIDA pipeline may lead to improved results.

Data and Preprocessing

Low Dimensionality of Data In this study, the ABIDE dataset was loaded using the nilearn Python package. The data loaded from nilearn was all preprocessed, using connectivity measures rather than the raw fMRI data. Due to this, the dimensionality of the input data was of a lower order (matrix). The ability of both the Remurs and Elastic-Remurs to utilise the underlying structure of the data was therefore significantly lowered. In future research, higher dimensional data could be utilised to assess whether classification accuracy is improved with Remurs and Elastic-Remurs when compared with methods such as the Ridge classifier, which requires input data to be flattened.

Further Research to Investigate Inter-Site Variability Further research should be done to support the effect of the variability between the training data and test data. By running a k-fold cross-validation on the data from each individual site within the dataset and sites using the same scanners, it could be possible to determine whether the data collection methods at certain sites or the type of scanner is the cause for the variability in classification accuracy between sites shown in Table 3.

Real-world Application

Limitations While the classification accuracy in this study and other studies is relatively high, the technology could not be relied upon in clinical practice. Should the classification accuracy of machine learning methods improve significantly, these methods might only ever be used as a supporting tool for doctors. The definition of Autism Spectrum Disorder in the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) (American Psychiatric Association 2013) is based on socially observed behaviours, not simply activities of the brain. While machine learning could be a very useful tool in helping doctors diagnose ASD, it will likely remain as just a helpful tool for some time.

While the development of new methods for diagnosing ASD may be helpful, there are also scenarios where the technology could damage somebody's diagnosis. If a doctor has already seen the output of a machine learning classifier, they may be biased to agree with the computer's conclusion and question their own (Triberti, Durosini, and Pravettoni 2020).

Alternative Use of Linear Regression As machine learning methods will likely only be used as a tool for the foreseeable future, binary classification is not necessarily the goal of the technology available. With the linear regression methods used in this study, the output does not have to be a class, it can be a certainty measure lying somewhere between the two classes. This certainty metric may be of more use to a clinician than a class. To assess the effectiveness of this in medical practice, further research could be conducted, analysing the correctness of samples that are above or below two 'certainty' thresholds and those that lie in the 'uncertain' region between the two 'certainty' thresholds set for classifying a sample as ASD or control. If the classification accuracy is much higher when the output from the model has a more polarised certainty value, these assessments are likely to be more useful to doctors in diagnosing patients.

Novel Contributions

This study has made two key contributions:

- **Applied the Remurs method to the ABIDE dataset:** After achieving high classification accuracy on the CMU2008 dataset, applying Remurs to the ABIDE dataset showed promising results that performed well in comparison to other machine learning methods.
- **Introduced Elastic-Remurs:** As a logical extension to the Remurs method, Elastic-Remurs balances the low-rank and sparse characteristics of Remurs and the smoothing characteristic of Ridge, resembling the balance in the elastic net; this method consistently outperforms most linear machine learning methods in the classification of fMRI data and has the potential to be refined, hopefully leading to improved results.

Conclusion

This study has presented methods capable of achieving strong classification accuracy on the ABIDE dataset. The newly proposed Elastic-Remurs has been shown to perform better than most linear machine learning methods tested in this study, with Remurs also performing well. By using the phenotypic data within the ABIDE dataset, introducing domain adaptation to the classification pipeline, and combining multiple brian atlases, Remurs and Elastic-Remurs could potentially achieve higher accuracies.

With the current state of the art, it is unlikely that technologies for ASD diagnosis could be viably introduced clinically until models are made more accurate and explainable.

Afterword on ChatGPT

When developing Elastic-Remurs, finding a closed-form solution to the proximal operator of the L_2-norm proved particularly challenging. In searching for relevant literature, ChatGPT was used. While a reference to any previous work providing a closed-form solution to the proximal operator of the L_2-norm could not be found, ChatGPT provided a solution: $f(v, \lambda) = v/(1+2\lambda)$. While the correctness of this solution could not be verified as a closed-form solution to the proximal operator of the L_2-norm, it inspired the smoothing penalty $\text{smooth}(\mathcal{V}, \lambda) = \mathcal{V}/(1+\lambda)$, justified in the design of Elastic-Remurs.

References

Abbas, S. Q.; Chi, L.; and Chen, Y.-P. P. 2023. DeepMNF: Deep Multimodal Neuroimaging Framework for Diagnosing Autism Spectrum Disorder. *Artificial Intelligence in Medicine*, 136: 102475.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, DC: American Psychiatric Association, 5th edition.

Craddock, C.; Benhajali, Y.; Chu, C.; Chouinard, F.; Evans, A.; Jakab, A.; Khundrakpam, B. S.; Lewis, J. D.; Li, Q.; Milham, M.; et al. 2013. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Frontiers in Neuroinformatics*, 7: 27.

Craddock, R. C.; James, G. A.; Holtzheimer III, P. E.; Hu, X. P.; and Mayberg, H. S. 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8): 1914–1928.

Crane, L.; Chester, J. W.; Goddard, L.; Henry, L. A.; and Hill, E. 2016. Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. *Autism*, 20(2): 153–162.

Dosenbach, N. U.; Nardos, B.; Cohen, A. L.; Fair, D. A.; Power, J. D.; Church, J. A.; Nelson, S. M.; Wig, G. S.; Vogel, A. C.; Lessov-Schlaggar, C. N.; et al. 2010. Prediction of individual brain maturity using fMRI. *Science*, 329(5997): 1358–1361.

Epalle, T. M.; Song, Y.; Liu, Z.; and Lu, H. 2021. Multi-atlas classification of autism spectrum disorder with hinge loss trained deep architectures: ABIDE I results. *Applied soft computing*, 107: 107375.

Heinsfeld, A. S.; Franco, A. R.; Craddock, R. C.; Buchweitz, A.; and Meneguzzi, F. 2018. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage: Clinical*, 17: 16–23.

Hoerl, A. E.; and Kennard, R. W. 2000. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 80–86.

Kazeminejad, A.; and Sotero, R. C. 2020. The importance of anti-correlations in graph theory based classification of autism spectrum disorder. *Frontiers in neuroscience*, 14: 676.

Kunda, M.; Zhou, S.; Gong, G.; and Lu, H. 2023. Improving Multi-Site Autism Classification via Site-Dependence Minimization and Second-Order Functional Connectivity. *IEEE Transactions on Medical Imaging*, 42(1): 55–65.

Lu, H.; Liu, X.; Zhou, S.; Turner, R.; Bai, P.; Koot, R. E.; Chasmai, M.; Schobs, L.; and Xu, H. 2022. PyKale: Knowledge-Aware Machine Learning from Multiple Sources in Python. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 4274–4278.

Mitchell, T. M.; Shinkareva, S. V.; Carlson, A.; Chang, K.-M.; Malave, V. L.; Mason, R. A.; and Just, M. A. 2008. Predicting human brain activity associated with the meanings of nouns. *science*, 320(5880): 1191–1195.

Parisot, S.; Ktena, S. I.; Ferrante, E.; Lee, M.; Guerrero, R.; Glocker, B.; and Rueckert, D. 2018. Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease. *Medical Image Analysis*, 48: 117–130.

Song, X.; and Lu, H. 2017. Multilinear regression for embedded feature selection with application to fmri analysis. In *Thirty-first AAAI conference on artificial intelligence*.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1): 267–288.

Triberti, S.; Durosini, I.; and Pravettoni, G. 2020. A "third wheel" effect in health decision making involving artificial entities: A psychological perspective. *Frontiers in Public Health*, 8: 117.

Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B.; and Joliot, M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1): 273–289.

Wong, C.; Odom, S. L.; Hume, K. A.; Cox, A. W.; Fettig, A.; Kucharczyk, S.; Brock, M. E.; Plavnick, J. B.; Fleury, V. P.; and Schultz, T. R. 2015. Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of autism and developmental disorders*, 45: 1951–1966.

Zou, H.; and Hastie, T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2): 301–320.