Are screening methods useful in feature selection? An empirical study

Mingyuan Wang¹, Adrian Barbu¹

¹ Statistics Department, Florida State University, Tallahassee, Florida, U.S.A

* abarbu@stat.fsu.edu * mw15m@my.fsu.edu

Abstract

Filter or screening methods are often used as a preprocessing step for reducing the number of variables used by a learning algorithm in obtaining a classification or regression model. While there are many such filter methods, there is a need for an objective evaluation of these methods. Such an evaluation is needed to compare them with each other and also to answer whether they are at all useful, or a learning algorithm could do a better job without them. For this purpose, many popular screening methods are partnered in this paper with three regression learners and five classification learners and evaluated on ten real datasets to obtain accuracy criteria such as R-square and area under the ROC curve (AUC). The obtained results are compared through curve plots and comparison tables in order to find out whether screening methods help improve the performance of learning algorithms and how they fare with each other. Our findings revealed that the screening methods were only useful in one regression and three classification datasets out of the ten datasets evaluated.

Introduction

For the past few decades, with the rapid development of online social platforms and information collection technology, the concept of big data grew from a novel terminology in the past to one of the most powerful resources in present day. Especially in recent years, the sample sizes and feature space dimensions of datasets rose to levels beyond precedent. This development poses great challenges for machine learning in extracting the relevant variables and in building accurate predictive models on such large datasets.

One of the most popular machine learning tasks is feature selection, which consists of extracting meaningful features (variables) from the data, with the goal of obtaining better prediction on unseen data, or obtaining better insight on the underlying mechanisms driving the response.

Feature selection methods have grown into a large family nowadays, with T-score [1], Mutual Information [2], Relief [3], Lasso [4], and MRMR [5] as some of the more popular examples.

There are three categories of feature selection methods: screening methods (a.k.a. filter methods), wrapper methods and embedded methods. The screening methods are independent of the model learned, whereas the other two methods build the model while selecting the features. In this study, we focus our attention on screening methods and would like an unbiased answer to the following questions:

- Do screening methods help to build good predictive models, or comparable models can be obtained without them?
• How do the existing screening methods compare with each other in terms of predictive capabilities, which one is the best and which one is the worst?

To answer these questions, we evaluated different screening methods (three for regression and seven for classification) on ten real datasets, five for regression and five for classification. The screening methods and the datasets will be described in the Methods section, but here we present our main findings.

The screening methods themselves cannot provide predictive models. For that purpose, different supervised learning algorithms such as SVM, Feature Selection with Annealing (FSA), Boosted Trees, and Naive Bayes were employed to construct the predictive models on the features selected by the screening methods.

For regression, all but one of the datasets evaluated point to the conclusion that the screening methods are not very useful, in the sense that similar or worse predictive models are obtained using them than without them. The findings are summarized in Table 1 which shows the number of datasets where a filter method helps an algorithm perform significantly better, and the number of datasets where the screening method helps the best performing learning algorithm for that dataset.

Table 1. Ranking of feature screening methods for regression by the number of datasets where the screening method performed significantly better than no screening for different learning algorithms and for the best algorithm (larger numbers are better).

| Screening Method | FSA | Ridge | Boost Tree | Best |
|------------------|-----|-------|------------|------|
| RReliefF [6]     | 0   | 2     | 3          | 1    |
| Mutual Information [2] | 1   | 2     | 3          | 1    |
| Correlation      | 0   | 2     | 2          | 0    |

For classification, the experiments indicate that the screening methods are useful, in the sense that better predictive models can be obtained using them than without them. The findings are summarized in Table 2 which shows the number of datasets on which a filter method helps an algorithm perform significantly better, and the number of datasets on which the filter method helps the best performing learning algorithm perform even better. We see that the chi-square score, Gini index and Relief are the three methods that performed best on our classification datasets.

Table 2. Ranking of feature screening methods for classification by the number of datasets where the screening method performed significantly better than no screening for different learning algorithm and the best algorithm (larger numbers are better).

| Screening Method | Boost Tree | FSA | SVM | NB | Logistic | Best |
|------------------|------------|-----|-----|----|----------|------|
| Chi-square Score [7] | 1          | 1   | 2   | 5  | 2        | 2    |
| Gini Index [8]     | 1          | 1   | 2   | 5  | 5        | 2    |
| Relief [3]         | 2          | 2   | 1   | 5  | 4        | 2    |
| Mutual Information | 1          | 1   | 2   | 5  | 5        | 2    |
| T-score            | 1          | 1   | 2   | 5  | 5        | 2    |
| MRMR [5]           | 1          | 0   | 1   | 5  | 1        | 2    |
| Fisher Score [9]   | 1          | 1   | 2   | 3  | 4        | 2    |

Through our comparison study, we intend to provide researchers with a clear understanding of some of the well known screening (filter) methods and their performance of handling high-dimensional real data.

Related Work

The focus of this study is to examine the effect of screening (filter) methods on obtaining good predictive models on high-dimensional datasets. There are several works
that compare feature selection methods in the past. A recent feature selection survey \cite{10} from Arizona State University (ASU) shows a comprehensive feature selection contents, studying feature selection methods from different data type perspectives. The survey is very broad, examining both supervised and unsupervised learning using binary and multi-class data, whereas our study focuses on supervised learning on regression and binary classification problems. The ASU study evaluates many classification datasets, but it does not have our goal of comparing feature screening methods and testing whether they are useful in practice or not. In this respect, we found some issues with the ASU study and we corrected them in this paper. First, the ASU study uses the misclassification error as a measure of the predictive capability of a classifier. The misclassification error is sensitive to the choice of threshold, and is a more noisy measure than the AUC (area under the ROC curve). In our work we used the AUC instead obtained performance curves that have less noise, as it will be seen in experiments. Second, the ASU study obtains the results with 10-fold cross-validation, and it is not averaged over multiple independent runs. In our work we used 40 random splits to further increase the power of our statistical tests. Third, we draw our comparisons and conclusions using statistical methods based on paired \(t\)-tests to obtain groups of similarly performing methods. An earlier version of the ASU report is \cite{11}, which is an overview of different types of feature selection methods for classification.

In \cite{12} are evaluated feature selection methods for flat features including filter methods, wrapper methods and embedded methods. However, tests are only conducted on low-dimensional datasets. In contrast we evaluate the filter methods on high dimensional datasets with 500-20,000 features and in many instances with more features than observations. Moreover, our goal is to compare filter methods themselves not the filter-learning algorithm combination, since different datasets could have different algorithms that are appropriate (e.g. linear vs nonlinear). We achieve this goal by employing many learning algorithms and choosing the best one for each filter method and each dataset.

Two other studies of feature selection methods are \cite{13} and \cite{14}. In contrast to our study, they solely focus on unsupervised learning. More studies exist in this field (Guyon and Elisseeff, 2003 \cite{15}, Sanchez-Marono et al.,2007 \cite{16}, Saeys et al.,2007 \cite{17}).

**Methods**

Experiments were conducted separately for regression and classification. For regression, the screening methods were Correlation, Mutual Information \cite{2}, and RReliefF \cite{6}. These screening methods were combined with learners including Feature Selection with Annealing (FSA) \cite{18}, Ridge Regression, and Boosted Regression Trees.

For classification, the screening methods were T-score \cite{1}, Mutual Information \cite{2}, Relief \cite{3}, Minimum Redundancy Maximum Relevance (MRMR) \cite{5}, Chi-square score \cite{7}, Fisher score \cite{9}, and Gini index \cite{8}. They were combined with learners including FSA \cite{18}, Logistic Regression, Naive Bayes, SVM, and Boosted Decision Trees.

Among these screening methods, Mutual information, Correlation, Gini index, Fisher-score, Chi-square score and T-score select features individually. In contrast, MRMR requires to calculate the redundancy between the already selected features and the current feature, and Relief requires to calculate the distance between two observations using Euclidean norm so that one can determine the nearest neighbor with the same label and with different labels. The calculation of Euclidean norm involves all the feature value. Consequently, these two methods select features in combination and are slower than the other methods.
Evaluation of Screening Methods

For each dataset, experimental results were obtained as the average of 40 independent runs. For each run, a random subset of 10% of instances was selected as test set and the rest as training set. One of the screening methods mentioned above was used to reduce the dimension of the feature space, then a learning algorithm using different tuning parameters was applied on the selected features to obtain the model.

To insure the consistency of the comparison, the number of features that were selected by each screening method was kept the same for each dataset. For each dataset (except Wikiface), 30 different values of the number of selected features were assigned. The $R^2$ was used as an indicator of performance for the regression problems and the AUC (Area under the ROC curve) was used as an indicator of performance for the classification problems. Plots were used to compare the average performance over of 40 runs of different combinations of screening method and learner. Also for each combination, the optimal number of selected features was selected based on the maximum of the average performance. Pairwise t-tests at the significance level $\alpha = 0.05$ were used to compare between different combinations to see if they are statistically different.

Construction Tables of Groups

Groups of screening method-learner combinations that are not significantly different from each other were constructed as follows (we use paired t-tests to obtain p-values when comparing different methods combinations and set 0.05 as the significance level in our experiment). The screening method-learner combinations are first sorted in descending order of their peak performance. Then starting from the first combination F downward, the last combination in the sequence that is not statistically significantly different from combination F is marked as combination L. All combinations between F and L are put into the same group. The same procedure was used for other combinations along the sequence. All these tables of groups are provided in the Supporting information section.

Construction Comparison Tables

Comparison tables were established based on how many times each screening method-learner combination appeared of in the group tables. Three kinds of counting methods were applied.

1) The number of datasets where the screening method performed significantly better than no screening for different learning algorithms. For each learning algorithm, it is the number of datasets on which the screening method appeared in higher group tiers than the same learning algorithm without screening. This counting method is used to construct Table 1 and Table 2 except the “Best” column.

2) The number of datasets where the screening method was significantly better than the best performing algorithm with no screening (usefulness per dataset). For each dataset, we checked for each screening method whether it appeared with a learning algorithm in a higher group tier than the best learning algorithm without screening. This counting method is used to construct Table 4 and Table 7 and the “Best” column of Tables 1 and 2. Column with name “Total Count” is generated from the summation of counts across all datasets for each screening methods.

3) The number of datasets where each filter-learning algorithm combination was in the top performing group (top performing). This counting method is used in Table 5 and Table 8.
Table 3. The datasets used for evaluating the screening methods. The parameter $\tau$ controls the number of selected features as $(4^t \tau^t)$, $t = 1, 30$.

| Dataset      | Learning type | Feature type | Number of features | Number of observations | $\tau$ |
|--------------|---------------|--------------|--------------------|------------------------|--------|
| Mouse BMI    | Regression    | Continuous   | 21575              | 294                    | 1.825  |
| Tumor        | Regression    | Continuous   | 16790              | 1750                   | 1.825  |
| Indoorloc    | Regression    | Continuous   | 520                | 19937                  | 1.25   |
| Wikiface     | Regression    | Continuous   | 4096               | 53040                  | 1.65   |
| CoEPrA2006   | Regression    | Continuous   | 5787               | 133                    | 1.68   |
| Gisette      | Binary Classification | Continuous | 5000               | 7000                   | 1.73   |
| Dexter       | Binary Classification | Continuous | 20000              | 600                    | 1.78   |
| Madelon      | Binary Classification | Continuous | 500                | 2600                   | 1.25   |
| SMK_CAN_187  | Binary Classification | Continuous | 19993              | 187                    | 1.78   |
| GLI_85       | Binary Classification | Continuous | 22283              | 85                     | 1.78   |

Results

Data sets

Five datasets were used for regression and five datasets for classification, with the specific dataset details given in Table 3.

The regression dataset Indoorloc is available on the UCI Machine Learning Repository [27]. The original dataset has longitude and latitude as response. In our study, we only use latitude as response. The dataset Tumor was extracted from TCGA (The Cancer Genome Atlas). The response of this dataset is the survival time (in days) of the patient, and the predictors represent gene expression levels. The classification datasets Gisette, Dexter, Madelon are part of the NIPS 2003 Feature selection challenge [24] and are also available on the UCI Machine Learning Repository.

The dataset Wikiface is a regression problem of predicting the age of a person based on the person’s face image, and was obtained from the Wikiface images [22]. A CNN (Convolutional Neural Network) vgg-face [28] pre-trained for face recognition was applied to each face and the output of the 34-th layer was used to generate a 4096 feature vector for each face. This 4096 dimensional vector was used as the feature vector for age regression, with the age value from the original Wikiface data as the response.

Regression Results

The following results are based on the output generated using a Matlab 2016b [29] implementation of all three screening methods and three learning algorithms.

Performance Plots

For the regression datasets, the plots from Fig. 1 and 2 show the $R^2$ value vs. the number $M_i$ of selected features, where $M_i = \left(4^t \tau^t\right)$, $t = 1, ..., 30$. The value of $\tau$ for each dataset is given in Table 3.

In Fig 1 left, are shown the $R^2$ of the best learning algorithm vs. the number of features selected by a screening method for the BMI and tumor datasets. Observe that these datasets are both gene expression datasets with many features and few observations. In Fig 1 right, are shown the $R^2$ of FSA (the best overall learning algorithm) vs. the number of features selected by a screening method. We see that the screening methods do not improve the performance of the optimal regression learners for
the BMI and tumor dataset. The plots on the right show that the screening methods even need to select more features to obtain similar performance to FSA without screening.

In Fig 2, left, are shown the $R^2$ of the best learning algorithm vs. the number of features selected by a screening method for the other three regression datasets. In Fig 2, right, are shown the $R^2$ of the best overall learning algorithm in each case (ridge for CoEPrA and Wikiface, boosted trees for Indoorloc) vs. the number of features selected by a screening method. From the plots we observe that the correlation method gives slightly better results than the learning algorithms without screening on the CoEPrA dataset and shows no clear advantage on the other two datasets. The statistical significance of the improvement can be seen in the table of groups from the supplementary information or in the comparison tables below.

Comparison Tables

The counts in the comparison table are based on the table of groups from the supplementary information.

Table 4. Ranking of feature screening methods for regression by number of datasets where screening method was significantly better than the best performing no screening method. (larger numbers are better)

| Screening Method | BMI | Tumor | CoEPrA | Indoorloc | Wikiface | Total Count |
|-----------------|-----|-------|--------|-----------|----------|-------------|
| RReliefF        |     |       | **    |           |          | 1           |
| Mutual Information       | |       |   *    |           |          | 1           |
| Correlation            |   |       |        |           |          | 0           |

In Table 4 is shown a "*" for each dataset and each screening method whether it has a learning algorithm that obtains significantly better performance than the best learning algorithm without screening. We can see that RReliefF and Mutual Information methods worked on Indoorloc dataset. However, screening methods didn’t provide improvement of performance on the other four regression datasets.

In Table 5 are shown the counts of screening method-learner combinations that are in the top group. Combination of FSA with screening methods worked on the most
Fig 2. Performance plots of methods with and without feature screening. Left: for each screening method are shown the maximum $R^2$ value across all learners. Right: $R^2$ of the screening methods with the best learner for this data (ridge).

regression datasets 3 out of 5. However it was not a significant improvement compared to FSA without screening method which also worked on 3 out of 5 datasets.

In Table 6 is shown the number of times each screening method was in the top performing group. In the first column, these methods were counted together with the learning algorithms they were applied. So there can be at most 15 counts (For each screening method there are three learning algorithms and five datasets total) in each cells. The second column shows the counts regardless learning algorithms. So there can be at most 5 counts in each cells. RReliefF and Mutual Information have the highest counts. But they only have 0 to 1 more count than Correlation and non screening method.

Table 5. Number of datasets each combination was in the top performing group.

| Filter            | Learners | FSA | Ridge | Boost Tree |
|-------------------|----------|-----|-------|------------|
| RReliefF          | 3        | 1   | 1     |            |
| Mutual Information| 3        | 1   | 1     |            |
| Correlation       | 3        | 1   | 0     |            |
| —                 | 3        | 1   | 0     |            |
Table 6. Ranking of feature screening methods for regression by the number of times each was in the top performing group. (larger numbers are better)

| Screening Method | Top performing Method-Algorithm | Method |
|------------------|--------------------------------|--------|
| RReliefF         | 5                              | 4      |
| Mutual Information | 5                        | 4      |
| Correlation      | 4                              | 3      |
| No Screening     | 4                              | 4      |

Classification Results

The following results are based on the output generated by Matlab 2016b. Methods including Relief, T-score, chi-square score, logistic regression, naive Bayes, SVM, boosted decision trees are applied on datasets using their Matlab 2016b implementation. Methods including MRMR, Fisher score, Gini index are applied on datasets using the ASU repository implementation[^1]. Some of the implementations only accept discrete predictors, so quantile discretization method is used based on the implementation[^30] from Matlab File Exchange.

Performance Plots

In Fig 3, left are shown the AUC of the best learning algorithm vs. the number of features selected by a screening method for four of the classification datasets. In Fig 3, right, are shown the AUC of the best overall learning algorithm in each case (Boosted Trees for SMK_CAN_187, Madelon and Dexter, FSA for Gisette) vs. the number of features selected by a screening method.

The plots show that all screening methods give higher best results on the SMK_CAN_187 and Madelon datasets. Observe that on the SMK_CAN_187 dataset all screening methods select a higher number of selected features than FSA when they reach their optimal values. For the Gisette dataset, all methods except RRelief show slightly better results, and only the Chi-square method shows a slight improvement on the Dexter dataset.

In Fig 4, left are shown the AUC of the best learning algorithm vs. the number of features selected by a screening method on the GLI_85 dataset. In Fig 4, right, are shown the AUC of FSA (for GLI_85 ) vs. the number of features selected by a screening method. We can observe that there is no screening method that can improve the performance of learning algorithms on this dataset.

Comparison Tables

In Table 7 is shown a “*” for each dataset and each screening method if it has a learning algorithm that obtains significantly better performance than the best learning algorithm without screening. We observe that Relief worked on two datasets (Madelon and SMK_CAN_187 ). MRMR method only worked on SMK_CAN_187 dataset. The other screening method worked on both Gisette and SMK_CAN_187 datasets. Overall, except MRMR, the screening methods have similar performance on five classification datasets. In Table 8 is shown for each screening method-learning algorithm combination the number of datasets for which it was in the top performing group. We can observe that the screening methods with boosted trees have the overall best performance on more than half of five classification datasets. And they have slight advantage compared to boosted trees without screening. FSA and SVM worked moderately well with several

[^1]: http://featureselection.asu.edu/old/software.php
methods (including Chi-square Score, Gini Index, Relief, Mutual Information and T-score) on 2-3 out of the 5 datasets. Naive Bayes and Logistic regression didn’t perform well on these five datasets.

In Table 9 are shown the number of times each screening method was in the top performing group. In the first column, these methods were counted with respect to the learning algorithms they were applied. So there can be at most 25 counts (for each screening method there are five learning algorithms and five datasets) in each cell. The second column shows the counts with the best learning algorithm, so there can be at most 5 counts in each cell. The Chi-square score has the highest count. It’s significantly higher than no screening. Gini Index, Relief, Mutual Information and T-score also have relative high counts when count including different learning algorithms. Each of these five screening methods has good performance on two more dataset than using no screening method when considering only the best learning algorithm for each method and each dataset.
Fig 4. Performance plots of methods with and without feature screening. Left: for each screening method are shown the maximum $R^2$ value across all learners. Right: $R^2$ of the screening methods with the best learner for this data.

Table 7. Ranking of feature screening methods for classification by number of datasets where screening method was significantly better than the best performed no screening method. (larger numbers are better. * indicates appearance.)

| Screening Method   | Dexter | Gisette | Madelon | SMK_CAN_187 | GLI_85 | Total count |
|--------------------|--------|---------|---------|-------------|--------|-------------|
| Chi-square Score   | *      | *       | *       |             |        | 2           |
| Gini Index         | *      | *       | *       |             |        | 2           |
| Relief             | *      | *       | *       |             |        | 2           |
| Mutual Information | *      | *       | *       |             |        | 2           |
| T-score            | *      | *       | *       |             |        | 2           |
| Fisher Score       | *      | *       | *       |             |        | 2           |
| MRMR               |        |         |         |             | *      | 1           |

Conclusion

The screening methods that were evaluated in this paper show an improvement for most classification experiments. At the same time, the screening methods perform poorly for most regression tasks. In the classification tasks, the screening methods with boosted tree give the best overall results. All seven classification screening methods we evaluated more or less improve the performance of learner. The Chi-square Score, Gini Index, Relief and Mutual Information work slightly better than other methods. It also can be seen from tables that the screening methods work well especially on learning algorithms that give poor results on their own. Compared to classification, there are fewer screening methods for regression problems. From our experiments, there is no outstanding screening method that can significantly improve the performance of regression.

Table 8. Number of datasets where each combination was in the top performing group.

| Filter                | Learners | Boost Tree | FSA | SVM | NB | Logistic |
|-----------------------|----------|------------|-----|-----|----|----------|
| Chi-square Score      | 3        | 3          | 3   | 1   | 0  |          |
| Gini Index            | 3        | 2          | 3   | 1   | 0  |          |
| Relief                | 4        | 2          | 2   | 1   | 0  |          |
| Mutual Information    | 3        | 2          | 2   | 0   | 0  |          |
| T-score               | 3        | 2          | 1   | 0   | 0  |          |
| MRMR                  | 3        | 1          | 1   | 0   | 0  |          |
| Fisher Score          | 3        | 1          | 1   | 0   | 0  |          |
| —                     | 2        | 1          | 1   | 0   | 0  |          |
Table 9. Ranking of feature screening methods for classification by the number of times each was in the top performing group. (larger numbers are better)

| Screening Method     | Top performing Method-Algorithm | Method |
|----------------------|---------------------------------|--------|
| Chi-square Score     | 10                              | 4      |
| Gini Index           | 9                               | 4      |
| Relief               | 9                               | 4      |
| Mutual Information   | 7                               | 4      |
| T-score              | 6                               | 4      |
| MRMR                 | 5                               | 3      |
| Fisher Score         | 5                               | 3      |
| No Screening         | 4                               | 2      |

Acknowledgments

The authors declare that there are no conflicts of interest.

References

1. Davis JC, Sampson RJ. Statistics and data analysis in geology. vol. 646. Wiley New York et al.; 1986.

2. Lewis DD. Feature selection and feature extraction for text categorization. In: Proceedings of the workshop on Speech and Natural Language. Association for Computational Linguistics; 1992. p. 212–217.

3. Kira K, Rendell LA. The feature selection problem: Traditional methods and a new algorithm. In: AAAI. vol. 2; 1992. p. 129–134.

4. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996; p. 267–288.

5. Han C, Chris D, Fu H. Minimum redundancy maximum relevance feature selection [J]. IEEE Intelligent Systems. 2005;20(6):70–71.

6. Robnik-ˇSikonja M, Kononenko I. An adaptation of Relief for attribute estimation in regression. In: Machine Learning: Proceedings of the Fourteenth International Conference (ICML’97); 1997. p. 296–304.

7. Liu H, Setiono R. Chi2: Feature selection and discretization of numeric attributes. In: Tools with artificial intelligence, 1995. proceedings., seventh international conference on. IEEE; 1995. p. 388–391.

8. Gini C. Variability and mutability, contribution to the study of statistical distribution and relatons. Studi Economico-Giuricici della R. 1912;.

9. Duda RO, Hart PE, Stork DG. Pattern classification. John Wiley & Sons; 2012.

10. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: A data perspective. ACM Computing Surveys (CSUR). 2017;50(6):94.

11. Tang J, Alelyani S, Liu H. Feature selection for classification: A review. Data Classification: Algorithms and Applications. 2014; p. 37.
12. Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering. 2014;40(1):16–28.

13. Alelyani S, Tang J, Liu H. Feature Selection for Clustering: A Review. Data Clustering: Algorithms and Applications. 2013;29:110–121.

14. Talavera L. An evaluation of filter and wrapper methods for feature selection in categorical clustering. Advances in Intelligent Data Analysis VI. 2005; p. 742–742.

15. Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003;3(Mar):1157–1182.

16. Sánchez-Marono N, Alonso-Betanzos A, Tombilla-Sanromán M. Filter methods for feature selection—a comparative study. Intelligent Data Engineering and Automated Learning-IDEAL 2007. 2007; p. 178–187.

17. Saey Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. bioinformatics. 2007;23(19):2507–2517.

18. Barbu A, She Y, Ding L, Gramajo G. Feature Selection with Annealing for Computer Vision and Big Data Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(2):272–286.

19. Wang S, Yehya N, Schadt EE, Wang H, Drake TA, Lusis AJ. Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. PLoS genetics. 2006;2(2):e15.

20. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. New England Journal of Medicine. 2016;375(12):1109–1112.

21. Torres-Sospedra J, Montoliu R, Martínez-Usó A, Avariento JP, Arnau TJ, Benedito-Bordonau M, et al. UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. In: Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on. IEEE; 2014. p. 261–270.

22. Rothe R, Timofte R, Gool LV. Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision (IJCV). 2016;

23. Ivanciuc O. CoEPrA 2006 Round 3 Comparative Evaluation of Prediction Algorithms; 2006. Available from: http://www.coepra.org/

24. Guyon I, Gunn S, Ben-Hur A, Dror G. Result analysis of the NIPS 2003 feature selection challenge. In: Advances in neural information processing systems; 2005. p. 545–552.

25. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nature medicine. 2007;13(3):361.

26. Freije WA, Castro-Vargas FE, Fung Z, Horvath S, Cloughes T, Lian LM, et al. Gene expression profiling of gliomas strongly predicts survival. Cancer research. 2004;64(18):6503–6510.

27. Lichman M. UCI Machine Learning Repository; 2013. Available from: http://archive.ics.uci.edu/ml
Supporting information

In this section we present the summary of the performance of each screening method-learning algorithm combination and their division into groups such that the difference between the best method and the worst method in each group is not significant at the 0.05 level.

Table of groups

Table 10. Table of groups, BMI dataset. SE is the standard error of mean estimation, \( \omega \) is the number of features selected by the screening method, \( \kappa \) is the number of features selected by FSA.

| Group | Screening Methods | Learner | Mean  | SE   | \( \omega \) | \( \kappa \) |
|-------|-------------------|---------|-------|------|-------------|-------------|
| A     | RReliefF          | FSA     | 0.7409| 0.0168| 998         | 692         |
| A     | Mutual Information| FSA     | 0.7405| 0.0176| 5493        | 1170        |
| A     | —                 | FSA     | 0.7395| 0.0174| —           | 1354        |
| A     | Correlation       | FSA     | 0.7392| 0.0170| 5493        | 1170        |
| B     | Correlation       | Ridge   | 0.7017| 0.0176| 3537        | —           |
| B     | RReliefF          | Ridge   | 0.6988| 0.0185| 5140        | —           |
| B     | Mutual Information| Ridge   | 0.6867| 0.0188| 6230        | —           |
| B     | —                 | Ridge   | 0.6851| 0.0188| —           | —           |
| C     | Correlation       | Boosted Reg. Trees | 0.4754 | 0.0249 | 13   | —         |
| C     | Mutual Information| Boosted Reg. Trees | 0.4673 | 0.0297 | 13   | —         |
| C     | RReliefF          | Boosted Reg. Trees | 0.4473 | 0.0288 | 93   | —         |
| D     | —                 | Boosted Reg. Trees | 0.2877 | 0.0285 | —   | —         |

In Table 10 are shown the groups, the mean \( R^2 \) of test data and standard error of mean estimation obtained over all the runs for the BMI dataset. Also shown are the number of features \( \omega \) selected by the screening method and the number of features \( \kappa \) selected by the learning algorithm where the average \( R^2 \) is maximum. From Table 10 we see that the best learner is FSA and that the FSA results with and without screening methods belong to the same group indicating that the screening methods don’t improve the performance of FSA significantly. The same holds for ridge regression. For boosted regression trees, the screening methods do provide a significant improvement. Also we can see the number of features selected by FSA is smaller than the number of features selected by the screening methods. So for FSA, the features selected by screening methods can still be reduced in order to get the best result.
Table 11. Table of groups, tumor dataset. SE is the standard error of mean estimation, $\omega$ is the number of features selected by the screening method, $\kappa$ is the number of features selected by FSA.

| Group | Screening Methods | Learner | Mean   | SE     | $\omega$ | $\kappa$ |
|-------|-------------------|---------|--------|--------|----------|----------|
| A     | —                 | FSA     | 0.3431 | 0.0161 | —        | 692      |
| A     | RReliefF          | FSA     | 0.3413 | 0.0165 | 6230     | 1170     |
| A     | Correlation       | FSA     | 0.3403 | 0.0164 | 6230     | 1758     |
| B     | Mutual Information| FSA     | 0.3345 | 0.0170 | 6230     | 1550     |
| B     | Correlation       | Ridge   | 0.2961 | 0.0159 | 13       | —        |
| B     | RReliefF          | Ridge   | 0.2931 | 0.0166 | 13       | —        |
| C     | Mutual Information| FSA     | 0.2924 | 0.0171 | 13       | —        |
| C     | —                 | —       | 0.2824 | 0.0171 | —        | —        |

The same types of results are shown in Table 11 for the tumor dataset. Again, the best results are obtained with FSA and the FSA results with and without screening methods belong to the same group. So screening methods do not improve the performance of FSA in this case either. For ridge regression and boosted regression trees, the results with screening methods belong to higher tier groups than results without screening method, which means the screening methods work on those two learners. Also for FSA, the features selected by screening methods are further reduced in order to get the maximum result.

Table 12. Table of groups, CoEPrA2006.3 dataset. SE is the standard error of mean estimation, $\omega$ is the number of features selected by the screening method, $\kappa$ is the number of features selected by FSA.

| Group | Screening Methods | Learner | Mean   | SE     | $\omega$ | $\kappa$ |
|-------|-------------------|---------|--------|--------|----------|----------|
| A     | Correlation       | Ridge   | 0.2567 | 0.0370 | 971      | —        |
| A     | Correlation       | FSA     | 0.2538 | 0.0329 | 1575     | 270      |
| A     | —                 | FSA     | 0.2475 | 0.0329 | —        | 208      |
| A     | RReliefF          | FSA     | 0.2431 | 0.0302 | 2447     | 153      |
| A     | Mutual Information| Ridge   | 0.2427 | 0.0372 | 865      | —        |
| A     | Mutual Information| FSA     | 0.2416 | 0.0345 | 1198     | 153      |
| A     | RReliefF          | Ridge   | 0.2354 | 0.0324 | 971      | —        |
| B     | Correlation       | Boosted Reg. Trees | 0.1484 | 0.0354 | 153      | —        |
| B     | RReliefF          | Boosted Reg. Trees | 0.1302 | 0.0273 | 338      | —        |
| B     | Mutual Information| Boosted Reg. Trees | 0.1275 | 0.0312 | 1319     | —        |
| B     | —                 | Boosted Reg. Trees | 0.1082 | 0.0271 | —        | —        |
| C     | —                 | Ridge   | 0.0618 | 0.0229 | —        | —        |

In Table 12 are shown the results for the CoEPrA2006.3 dataset. Again the FSA with or without screening are in the same (top) group. Only the results with screening methods for Ridge regression belong to higher tier group than without screening. So in this case, the screening methods only improve the performance of Ridge regression.
Table 13. Table of groups, Indoorloc dataset. SE is the standard error of mean estimation, $\omega$ is the number of features selected by the screening method, $\kappa$ is the number of features selected by FSA.

| Group | Screening Methods | Learner       | Mean   | SE    | $\omega$ | $\kappa$ |
|-------|-------------------|---------------|--------|-------|----------|----------|
| A     | Mutual Information| Boosted Reg. Trees | 0.9719 | 0.0004 | 195      | —        |
| A     | RReliefF          | Boosted Reg. Trees | 0.9715 | 0.0006 | 348      | —        |
| B     | Correlation       | Boosted Reg. Trees | 0.9699 | 0.0005 | 397      | —        |
| B     | —                 | Boosted Reg. Trees | 0.9697 | 0.0004 | —        | —        |
| C     | Mutual Information| FSA           | 0.9212 | 0.0005 | 397      | 397      |
| C     | —                 | Ridge         | 0.9212 | 0.0005 | —        | —        |
| D     | Mutual Information| Ridge         | 0.9210 | 0.0005 | —        | 397      |
| E     | —                 | FSA           | 0.9207 | 0.0005 | —        | 397      |
| F     | Correlation       | FSA           | 0.9204 | 0.0005 | 397      | 397      |
| G     | Correlation       | Ridge         | 0.9202 | 0.0005 | —        | —        |
| H     | RReliefF          | FSA           | 0.9169 | 0.0006 | 397      | 397      |
| I     | RReliefF          | Ridge         | 0.9167 | 0.0006 | 397      | —        |

In Table 13 are shown the results for the Indoorloc dataset. Here we see that two results with screening methods for Boosted Trees belong to higher tier group than without screening. For FSA, there is only one such screening method. Therefore for this dataset, some screening methods help when it comes to Boosted Trees and FSA.

Table 14. Table of groups, Wikiface dataset. SE is the standard error of mean estimation, $\omega$ is the number of features selected by the screening method, $\kappa$ is the number of features selected by FSA.

| Group | Screening Methods | Learner       | Mean   | SE    | $\omega$ | $\kappa$ |
|-------|-------------------|---------------|--------|-------|----------|----------|
| A     | —                 | Ridge         | 0.5462 | 0.0018 | —        | —        |
| B     | Correlation       | Ridge         | 0.5437 | 0.0018 | 1739     | —        |
| B     | C                 | RReliefF      | 0.5434 | 0.0019 | 1739     | —        |
| C     | Mutual Information| Ridge         | 0.5432 | 0.0019 | 1739     | —        |
| D     | —                 | FSA           | 0.5338 | 0.0019 | —        | 304      |
| E     | RReliefF          | FSA           | 0.5318 | 0.0020 | 1739     | 244      |
| E     | F                 | Mutual Information| 0.5312 | 0.0020 | 1739     | 304      |
| F     | Correlation       | FSA           | 0.5307 | 0.0019 | 1739     | 244      |
| G     | RReliefF          | Boosted Reg. Trees | 0.5108 | 0.0020 | 1616     | —        |
| G     | H                 | Boosted Reg. Trees | 0.5100 | 0.0021 | —        | —        |
| H     | Mutual Information| Boosted Reg. Trees | 0.5095 | 0.0021 | 1269     | —        |
| H     | Correlation       | Boosted Reg. Trees | 0.5092 | 0.0021 | 1381     | —        |

In Table 14 are shown the results for the Wikiface dataset. For each learner, the results without screening belong to higher group tiers than results with screening methods, which means screening methods do not improve the performance for this dataset.
In Table 15 are shown the results for Dexter, a classification dataset. We see that some results with screening methods for Naive Bayes and Logistic Regression belong to higher tier groups than those without screening. The results with and without screening methods belong to the same groups respectively for the other learners. For some of the FSA combinations, the number of selected features by screening methods and number of selected features by FSA are the same, meaning the screening methods already selected features that can give the best result.

In Table 16 are shown the results for Gisette. The results with and without screening belong to the same group for Boosted Decision Trees. So the screening methods can’t improve the performance of Boosted Decision Trees in this case. A large portion of the results with screening for Naive Bayes, FSA, SVM and Logistic Regression belong to higher tier groups than those without screening, meaning that those screening methods can improve the performance of corresponding learning algorithms. For some of the FSA combinations, the number of selected features by screening methods and number of selected features by FSA are the same, meaning the screening methods already selected the features that can give the best result.

In Table 17 are shown the results for the SMK_CAN_187 dataset. The results with screening for Naive Bayes and Logistic Regression belong to higher tier groups than those without screening. Results with and without screening belong to the same groups for the other learning algorithms respectively, except for the Relief/FSA combination and Relief/Boosted Decision Trees combination. Relief improves the performance of these two learning algorithms.

In Table 18 are shown the results for Madelon. The results with screening for Naive Bayes, SVM, Boosted Decision Trees and Logistic Regression belong to higher tier groups than those without screening. For FSA, only the result of Relief/FSA belongs to higher tier group than FSA without screening.

In Table 19 are shown the results for the GLI_85 dataset. The results with and without screening belong to the same groups for FSA, SVM and Boosted Decision Trees respectively. Screening methods can’t improve these learning algorithms for this dataset. Results with screening for Naive Bayes and Logistic Regression belong to higher tier groups than those without screening, except for the Fisher-score method. The Fisher-score method doesn’t improve the performance of Naive Bayes.
Table 15. Table of groups, Dexter dataset. SE is the standard error of mean estimation, \( \omega \) is the number of features selected by the screening method, \( \kappa \) is the number of features selected by FSA.

| Group | Screening Methods | Learner       | Mean   | SE     | \( \omega \) | \( \kappa \) |
|-------|-------------------|---------------|--------|--------|-------------|-------------|
| A     | Chi-square Score  | Boosted Decision Trees | 0.9879 | 0.0017 | 139         | —           |
| A B   | Mutual Information| Boosted Decision Trees | 0.9869 | 0.0016 | 2023        | —           |
| A B C | MRMR              | Boosted Decision Trees | 0.9865 | 0.0018 | 207         | —           |
| A B C | Gini Index        | Boosted Decision Trees | 0.9865 | 0.0018 | 207         | —           |
| A B C | T-score           | Boosted Decision Trees | 0.9865 | 0.0017 | 207         | —           |
| A B C | —                 | Boosted Decision Trees | 0.9864 | 0.0018 | —           | —           |
| A B C D| Relief           | Boosted Decision Trees | 0.9858 | 0.0019 | 1463        | —           |
| A B C D| Fisher Score     | Boosted Decision Trees | 0.9857 | 0.0017 | 4442        | —           |
| A B C D| Chi-square Score | SVM           | 0.9853 | 0.0019 | 3376        | —           |
| A B C D| Gini Index       | SVM           | 0.9851 | 0.0019 | 3376        | —           |
| A B C D| Mutual Information| SVM           | 0.9846 | 0.0019 | 2023        | —           |
| A B C D| Mutual Information| FSA          | 0.9842 | 0.0018 | 2023        | 1294        |
| A B C D| Chi-square Score | FSA          | 0.9840 | 0.0019 | 711         | 711         |
| A B C D| Gini Index       | FSA          | 0.9839 | 0.0019 | 711         | 711         |
| A B C D| —                | SVM          | 0.9835 | 0.0019 | —           | —           |
| A B C D| Relief           | FSA          | 0.9834 | 0.0020 | 4442        | 3631        |
| A B C D| MRMR             | SVM          | 0.9831 | 0.0018 | 5023        | —           |
| A B C D| Fisher Score     | FSA          | 0.9826 | 0.0020 | 4729        | 983         |
| A B C D| T-score          | FSA          | 0.9826 | 0.0020 | 1463        | 1294        |
| A B C D| —                | FSA          | 0.9824 | 0.0021 | —           | 842         |
| A B C D| T-score          | SVM          | 0.9809 | 0.0023 | 5023        | —           |
| A B C D| Relief           | SVM          | 0.9780 | 0.0021 | 41          | —           |
| A B C D| Fisher Score     | SVM          | 0.9728 | 0.0028 | 5023        | —           |
| A B C D| Chi-square Score | Logistic Reg. | 0.9420 | 0.0041 | 589         | —           |
| A B C D| Gini Index       | Logistic Reg. | 0.9413 | 0.0045 | 478         | —           |
| A B C D| MRMR             | Logistic Reg. | 0.9396 | 0.0046 | 286         | —           |
| A B C D| T-score          | Logistic Reg. | 0.9395 | 0.0049 | 286         | —           |
| A B C D| MRMR             | Naive Bayes  | 0.9391 | 0.0039 | 41          | —           |
| A B C D| Mutual Information| Logistic Reg. | 0.9388 | 0.0049 | 286         | —           |
| A B C D| T-score          | Naive Bayes  | 0.9377 | 0.0039 | 41          | —           |
| A B C D| Relief           | Logistic Reg. | 0.9304 | 0.0039 | 842         | —           |
| A B C D| Gini Index       | Naive Bayes  | 0.9284 | 0.0038 | 41          | —           |
| A B C D| —                | Logistic Reg. | 0.9214 | 0.0053 | —           | —           |
| A B C D| Chi-square Score | Naive Bayes  | 0.9206 | 0.0044 | 41          | —           |
| A B C D| Relief           | Naive Bayes  | 0.9176 | 0.0078 | 83          | —           |
| A B C D| Mutual Information| Naive Bayes  | 0.9175 | 0.0054 | 41          | —           |
| A B C D| Fisher Score     | Logistic Reg. | 0.7817 | 0.0073 | 5023        | —           |
| A B C D| Fisher Score     | Naive Bayes  | 0.6554 | 0.0078 | 12          | —           |
| A B C D| —                | Naive Bayes  | 0.6420 | 0.0093 | —           | —           |
Table 16. Table of groups, Gisette dataset. SE is the standard error of mean estimation, $\omega$ is the number of features selected by the screening method, $\kappa$ is the number of features selected by FSA.

| Group | Screening Methods | Learner               | Mean   | SE     | $\omega$ | $\kappa$ |
|-------|-------------------|-----------------------|--------|--------|----------|----------|
| A     | Mutual Information| FSA                   | 0.9977 | 0.0002 | 1480     | 1480     |
| A     | Gini Index        | FSA                   | 0.9977 | 0.0002 | 1960     | 1960     |
| A     | Chi-square Score  | FSA                   | 0.9976 | 0.0002 | 1634     | 1634     |
| B     | T-score           | FSA                   | 0.9975 | 0.0003 | 1960     | 1960     |
| B     | Fisher Score      | SVM                   | 0.9973 | 0.0003 | 1960     | —        |
| B     | C                 | SVM                   | 0.9973 | 0.0003 | 1960     | —        |
| A     |                   | SVM                   | 0.9973 | 0.0003 | —        | 1058     |
| A     | MRMR              | FSA                   | 0.9973 | 0.0003 | 1960     | 1058     |
| B     | C                 | SVM                   | 0.9973 | 0.0003 | 1634     | —        |
| B     | C                 | SVM                   | 0.9967 | 0.0003 | 1634     | —        |
| C     |                   | SVM                   | 0.9967 | 0.0003 | 1634     | —        |
| D     |                   | SVM                   | 0.9965 | 0.0002 | —        | —        |
| D     | MRMR              | SVM                   | 0.9964 | 0.0003 | 1960     | —        |
| E     | Relief            | Boosted Decision Trees| 0.9952 | 0.0004 | 1058     | —        |
| E     | Chi-square Score  | Boosted Decision Trees| 0.9951 | 0.0003 | 2312     | —        |
| E     | Fisher Score      | Boosted Decision Trees| 0.9951 | 0.0004 | 1634     | —        |
| E     | Mutual Information| Boosted Decision Trees| 0.9951 | 0.0003 | 1333     | —        |
| E     | T-score           | Boosted Decision Trees| 0.9951 | 0.0003 | 1333     | —        |
| E     | Gini Index        | Boosted Decision Trees| 0.9951 | 0.0003 | 2884     | —        |
| E     |                   | Boosted Decision Trees| 0.9950 | 0.0003 | —        | —        |
| F     | Gini Index        | Logistic Reg.          | 0.9815 | 0.0009 | 1960     | —        |
| F     | T-score           | Logistic Reg.          | 0.9814 | 0.0009 | 1794     | —        |
| F     | Chi-square Score  | Logistic Reg.          | 0.9813 | 0.0009 | 1794     | —        |
| F     | Mutual Information| Logistic Reg.          | 0.9813 | 0.0008 | 2133     | —        |
| F     | Fisher Score      | Logistic Reg.          | 0.9808 | 0.0010 | 1960     | —        |
| H     | Relief            | Logistic Reg.          | 0.9795 | 0.0009 | 2497     | —        |
| H     | MRMR              | Logistic Reg.          | 0.9795 | 0.0008 | 1794     | —        |
| I     |                   | Logistic Reg.          | 0.9764 | 0.0009 | —        | —        |
| J     | MRMR              | Naive Bayes            | 0.9675 | 0.0010 | 178      | —        |
| K     | Mutual Information| Naive Bayes            | 0.9589 | 0.0012 | 2312     | —        |
| K     | Chi-square Score  | Naive Bayes            | 0.9559 | 0.0012 | 1794     | —        |
| K     | T-score           | Naive Bayes            | 0.9588 | 0.0012 | 1333     | —        |
| L     | Gini Index        | Naive Bayes            | 0.9587 | 0.0012 | 1794     | —        |
| M     | Relief            | Naive Bayes            | 0.9482 | 0.0014 | 2887     | —        |
| M     | Fisher Score      | Naive Bayes            | 0.9476 | 0.0014 | 2887     | —        |
| N     |                   | Naive Bayes            | 0.9369 | 0.0019 | —        | —        |
Table 17. Table of groups, SMK_CAN_187 dataset. SE is the standard error of mean estimation, $\omega$ is the number of features selected by the screening method, $\kappa$ is the number of features selected by FSA.

| Group | Screening Methods | Learner       | Mean   | SE    | $\omega$ | $\kappa$ |
|-------|------------------|---------------|--------|-------|----------|----------|
| A     | Relief           | FSA           | 0.8298 | 0.0142| 3631     | 83       |
| A B   | Relief           | Boosted Decision Trees | 0.8206 | 0.0118| 2228     | —        |
| A B C | Gini Index       | Boosted Decision Trees | 0.8192 | 0.0144| 589      | —        |
| A B C | MRMR             | SVM           | 0.8104 | 0.0134| 2892     | —        |
| A B C | Fisher Score     | Boosted Decision Trees | 0.8085 | 0.0122| 4442     | —        |
| A B C | Gini Index       | SVM           | 0.8074 | 0.0158| 5023     | —        |
| A B C | Chi-square Score | FSA           | 0.8061 | 0.0152| 2023     | 1641     |
| A B C | Chi-square Score | SVM           | 0.8057 | 0.0154| 2662     | —        |
| A B C | Chi-square Score | Boosted Decision Trees | 0.8043 | 0.0134| 711      | —        |
| A B C | Mutual Information | Boosted Decision Trees | 0.8029 | 0.0133| 2023     | —        |
| A B C | T-score          | Boosted Decision Trees | 0.8020 | 0.0135| 4729     | —        |
| B C D | MRMR             | SVM           | 0.8006 | 0.0133| 3376     | —        |
| B C D | Mutual Information | FSA           | 0.7995 | 0.0155| 4729     | 139      |
| B C D | T-score          | SVM           | 0.7992 | 0.0131| 4729     | —        |
| B C D | Fisher Score     | SVM           | 0.7970 | 0.0145| 5023     | —        |
| B C D | Mutual Information | SVM           | 0.7955 | 0.0168| 4729     | —        |
| B C D | Gini Index       | FSA           | 0.7955 | 0.0146| 5023     | 2441     |
| C D   | —                | FSA           | 0.7953 | 0.0155| —        | 207      |
| C D   | —                | Boosted Decision Trees | 0.7952 | 0.0133| —        | —        |
| C D   | —                | SVM           | 0.7888 | 0.0148| —        | —        |
| C D   | —                | SVM           | 0.7888 | 0.0148| —        | —        |
| C D   | Fisher Score     | FSA           | 0.7858 | 0.0158| 2662     | 207      |
| D     | MRMR             | FSA           | 0.7840 | 0.0139| 1641     | 983      |
| D     | T-score          | FSA           | 0.7824 | 0.0140| 1641     | 842      |
| E     | Mutual Information | Naive Bayes | 0.7458 | 0.0183| 41       | —        |
| E F   | MRMR             | Naive Bayes   | 0.7421 | 0.0163| 12       | —        |
| E F   | Gini Index       | Naive Bayes   | 0.7417 | 0.0176| 41       | —        |
| E F   | Chi-square Score | Naive Bayes   | 0.7393 | 0.0182| 41       | —        |
| E F   | T-score          | Naive Bayes   | 0.7370 | 0.0169| 12       | —        |
| E F G | Relief           | Logistic Reg. | 0.7325 | 0.0162| 1828     | —        |
| E F G | Relie            | Naive Bayes   | 0.7317 | 0.0184| 83       | —        |
| E F G | Chi-square Score | Logistic Reg. | 0.7177 | 0.0155| 1641     | —        |
| E F G | Fisher Score     | Logistic Reg. | 0.7172 | 0.0137| 1463     | —        |
| E F G | T-score          | Logistic Reg. | 0.7163 | 0.0142| 3893     | —        |
| E F G | MRMR             | Logistic Reg. | 0.7157 | 0.0163| 3376     | —        |
| E F G | Gini Index       | Logistic Reg. | 0.7128 | 0.0173| 589      | —        |
| F G   | Mutual Information | Logistic Reg. | 0.7077 | 0.0183| 2441     | —        |
| G     | Fisher Score     | Naive Bayes   | 0.7043 | 0.0175| 83       | —        |
| H     | —                | Naive Bayes   | 0.6576 | 0.0159| —        | —        |
| H     | —                | Logistic Reg. | 0.6479 | 0.0203| —        | —        |
Table 18. Table of groups, Madelon dataset. SE is the standard error of mean estimation, $\omega$ is the number of features selected by the screening method, $\kappa$ is the number of features selected by FSA.

| Group | Screening Methods | Learner         | Mean  | SE   | $\omega$ | $\kappa$ |
|-------|-------------------|-----------------|-------|------|----------|---------|
| A     | Relief            | Boosted Decision Trees | 0.9520| 0.0019| 22       | —       |
| B     | T-score           | Boosted Decision Trees | 0.9468| 0.0019| 13       | —       |
| B     | Fisher Score      | Boosted Decision Trees | 0.9468| 0.0019| 13       | —       |
| B     | Chi-square Score  | Boosted Decision Trees | 0.9464| 0.0019| 13       | —       |
| B     | MRMR              | Boosted Decision Trees | 0.9464| 0.0020| 13       | —       |
| B     | Gini Index        | Boosted Decision Trees | 0.9462| 0.0019| 13       | —       |
| B     | Mutual Information| Boosted Decision Trees | 0.9450| 0.0024| 13       | —       |
| C     | —                 | Boosted Decision Trees | 0.8683| 0.0034|          |         |
| D     | Relief            | Naive Bayes      | 0.6888| 0.0051| 22       | —       |
| E     | Gini Index        | Naive Bayes      | 0.6850| 0.0050| 22       | —       |
| E     | F                 | Mutual Information| Naive Bayes | 0.6834| 0.0053| 32       | —       |
| F     | Chi-square Score  | Naive Bayes      | 0.6828| 0.0052| 13       | —       |
| F     | T-score           | Naive Bayes      | 0.6824| 0.0052| 13       | —       |
| F     | Fisher Score      | Naive Bayes      | 0.6824| 0.0052| 13       | —       |
| F     | MRMR              | Naive Bayes      | 0.6821| 0.0052| 13       | —       |
| G     | Relief            | SVM              | 0.6390| 0.0057| 13       | —       |
| G     | T-score           | SVM              | 0.6367| 0.0056| 13       | —       |
| G     | Fisher Score      | SVM              | 0.6367| 0.0056| 13       | —       |
| G     | MRMR              | SVM              | 0.6367| 0.0057| 13       | —       |
| G     | Chi-square Score  | SVM              | 0.6366| 0.0057| 13       | —       |
| G     | Gini Index        | SVM              | 0.6363| 0.0057| 13       | —       |
| G     | Mutual Information| SVM              | 0.6362| 0.0056| 6        | —       |
| G     | Relief            | FSA              | 0.6334| 0.0056| 6        | 6       |
| G     | —                 | Naive Bayes      | 0.6320| 0.0050|          |         |
| H     | Chi-square Score  | FSA              | 0.6294| 0.0057| 32       | 32      |
| H     | Gini Index        | FSA              | 0.6293| 0.0045| 6        | —       |
| H     | Mutual Information| FSA              | 0.6282| 0.0057| 64       | 53      |
| H     | MRMR              | FSA              | 0.6278| 0.0056| 53       | 22      |
| H     | T-score           | FSA              | 0.6278| 0.0056| 53       | 22      |
| H     | Fisher Score      | FSA              | 0.6278| 0.0056| 53       | 22      |
| H     | —                 | FSA              | 0.6270| 0.0056|          | 22      |
| I     | Relief            | Logistic Reg.    | 0.6162| 0.0052| 13       | —       |
| I     | J                 | Gini Index       | 0.6155| 0.0045| 6        | —       |
| I     | J                 | Chi-square Score | Logistic Reg. | 0.6115| 0.0047| 13       | —       |
| I     | J                 | Mutual Information| Logistic Reg. | 0.6115| 0.0048| 13       | —       |
| I     | J                 | T-score          | Logistic Reg. | 0.6112| 0.0047| 13       | —       |
| I     | J                 | Fisher Score     | Logistic Reg. | 0.6112| 0.0047| 13       | —       |
| I     | J                 | MRMR             | Logistic Reg. | 0.6111| 0.0049| 13       | —       |
| K     | —                 | Logistic Reg.    | 0.5617| 0.0053|          | —       |
| L     | —                 | SVM              | 0.5445| 0.0057|          | —       |
**Table 19.** Table of groups, GLL85 dataset. SE is the standard error of mean estimation, \( \omega \) is the number of features selected by the screening method, \( \kappa \) is the number of features selected by FSA.

| Group | Screening Methods | Learner           | Mean     | SE      | \( \omega \) | \( \kappa \) |
|-------|------------------|-------------------|----------|---------|-------------|-------------|
| A     | —                | FSA               | 0.9646   | 0.0138  | —           | 377         |
| A     | Relief           | Boosted Decision Trees | 0.9627   | 0.0138  | 83          | —           |
| A B   | Relief           | SVM               | 0.9596   | 0.0151  | 377         | —           |
| A B   | T-score          | FSA               | 0.9579   | 0.0149  | 5023        | 478         |
| A B   | Mutual Information | FSA            | 0.9578   | 0.0154  | 4164        | 286         |
| A B   | Mutual Information | SVM            | 0.9576   | 0.0161  | 4729        | —           |
| A B   | Relief           | FSA               | 0.9574   | 0.0166  | 4729        | 377         |
| A B   | MRMR             | FSA               | 0.9561   | 0.0158  | 4164        | 377         |
| A B   | Gini Index       | FSA               | 0.9561   | 0.0158  | 4164        | 377         |
| A B   | —                | SVM               | 0.9543   | 0.0151  | —           | —           |
| A B   | MRMR             | SVM               | 0.9543   | 0.0161  | 4729        | —           |
| A B   | Chi-square Score | SVM               | 0.9543   | 0.0161  | 4164        | —           |
| A B   | Chi-square Score | FSA               | 0.9541   | 0.0174  | 4164        | 377         |
| A B   | Gini Index       | SVM               | 0.9539   | 0.0163  | 3631        | —           |
| A B   | —                | SVM               | 0.9509   | 0.0163  | 4442        | —           |
| A B   | Fisher Score     | FSA               | 0.9493   | 0.0151  | 4729        | 377         |
| A B   | Gini Index       | Boosted Decision Trees | 0.9493   | 0.0162  | 139         | —           |
| A B   | Mutual Information | Boosted Decision Trees | 0.9478   | 0.0149  | 207         | —           |
| A B   | MRMR             | Boosted Decision Trees | 0.9399   | 0.0177  | 41          | —           |
| A B   | T-score          | Boosted Decision Trees | 0.9395   | 0.0174  | 41          | —           |
| A B   | Chi-square Score | Boosted Decision Trees | 0.9380   | 0.0176  | 2441        | —           |
| A B   | Fisher Score     | Boosted Decision Trees | 0.9340   | 0.0187  | 3130        | —           |
| A B   | Fisher Score     | SVM               | 0.9298   | 0.0198  | 4442        | —           |
| A B C | Relief           | Naive Bayes       | 0.9262   | 0.0189  | 41          | —           |
| A B C | —                | Boosted Decision Trees | 0.9252   | 0.0213  | —           | —           |
| A B C D| Chi-square Score | Naive Bayes       | 0.9165   | 0.0228  | 12          | —           |
| A B C D E| Gini Index    | Naive Bayes       | 0.9160   | 0.0279  | 12          | —           |
| A B C D E| T-score         | Naive Bayes       | 0.9146   | 0.0211  | 12          | —           |
| A B C D E| Mutual Information | Naive Bayes   | 0.9142   | 0.0202  | 139         | —           |
| A B C D E| MRMR            | Naive Bayes       | 0.9051   | 0.0211  | 12          | —           |
| F C D E| MRMR            | Logistic Reg.     | 0.8944   | 0.0221  | 12          | —           |
| F C D E| T-score         | Logistic Reg.     | 0.8921   | 0.0230  | 12          | —           |
| F D E | Relief           | Logistic Reg.     | 0.8872   | 0.0176  | 2662        | —           |
| F D E | Mutual Information | Logistic Reg. | 0.8783   | 0.0184  | 3376        | —           |
| F E   | Gini Index       | Logistic Reg.     | 0.8756   | 0.0177  | 2662        | —           |
| F E   | Chi-square Score | Logistic Reg.     | 0.8745   | 0.0196  | 3893        | —           |
| F E   | Fisher Score     | Logistic Reg.     | 0.8601   | 0.0217  | 4729        | —           |
| G     | —                | Logistic Reg.     | 0.8055   | 0.0153  | —           | —           |
| G H   | Fisher Score     | Naive Bayes       | 0.7308   | 0.0418  | 1134        | —           |
| H     | —                | Naive Bayes       | 0.7173   | 0.0281  | —           | —           |