International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST2310161

Machine Learning based Feature Selection Approaches for Early Prediction of Autism Spectrum Disorder - Review

Asif Mohamed H B [1*],Dr.Md. Sameeruddin Khan [2], Dr. Mohan K G[3] , Dr. Parashuram Baraki[4]

[1]Research Scholar, School of CSE, Presidency University, Bengaluru, Karnataka, India,

[2]Professor & Dean, School of CSE, Presidency University, Bengaluru, Karnataka, India,

[3]Professor, Department of CSE, GITAM University, Bengaluru, Karnataka, India.

[4]Professor, Department of CSE, Smt. Kamala and Sri Venkappa M Agadi College of Engineering and Technology, Lakshmeshwar Dist : Gadag, India

Article Info

Publication Issue

Volume 10, Issue 1

January-February-2023

Page Number

451-455

Article History

Accepted: 03 Feb 2023

Published: 16 Feb 2023

ABSTRACT

Feature selection is the task of selecting a small subset of the original features that can achieve maximum classification accuracy. This subset of features has some very important benefits: This makes feature selection an essential task for classification tasks.

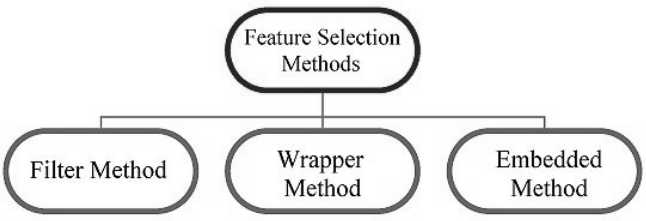
Keywords: Machine Learning, Feature Selection, Autism Spectrum Disorder

I. INTRODUCTION

Supervised methods can be classified into her three main categories depending on their reliance on classifiers: filter methods, wrapper methods and embedding methods. Filtering methods improve classification performance by examining only the inherent properties of the data to assess feature relevance without receiving feedback from the classifier. Estimate. The wrapper method relies on the classifier. These methods assess the 'goodness' of a selected feature subset directly from classifier feedback in terms of classification accuracy. In the embedding method, searching for the best subset of features is built into the construction of the classifier and can be viewed as searching in the combined space of feature subsets and hypotheses. Embedding methods are also specific to a particular classifier, but are cheaper than wrapper methods. According to the results reported in the literature, the performance of the filtering method alone is lower than that of the wrapper method, which, due to its numerical and computationally intensive nature, is not suitable for large datasets. may not be possible. Functional.

II. Feature Selection Methods

Machine Learning based Supervised methods can be classified into her three main categories depending on their reliance on classifiers: filter methods, wrapper methods and embedding methods.



Filter method

Filtering methods improve classification performance by examining only the inherent properties of the data to assess feature relevance without receiving feedback from the classifier. Estimate.. It's a faster and usually better approach if the number of features is huge. Avoids overfitting, but may not select the best features.

• Information gain – defined as the amount of information provided by a function that measures the decrease in entropy value given the target value. The information gain for each attribute is calculated considering the target value for feature selection.

• Chi-Square Test – The Chi-Square (X2) method is commonly used to test relationships between categorical variables. Compare the observed values of various attributes of the dataset to the expected values.

Chi-square Formula

$$X^2 = \sum \frac{(Observed\ value - Expected\ value)^2}{Expected\ value}$$

1. Fisher's Score –
2. Correlation Coefficient –
3. Variance Threshold –.
4. Mean Absolute Difference (MAD) –
5. Dispersion Ratio –
6. Mutual Dependence –
7. Relief

Fisher's **score** – Fisher's score selects each feature **separately from the** score based on Fisher's criteria, which produces many non-optimal features. The higher the Fisher score, the better the selected trait.

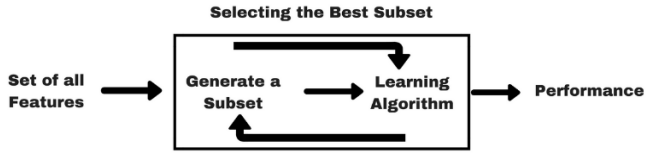
Correlation Coefficient – Pearson's correlation coefficient is a measure that quantifies the association and direction of the relationship between two continuous variables with values ranging from -1 to 1.

Variance Threshold – By default, this method removes features with zero variance. This method assumes that features with higher variance are more likely to contain more information.

Mean Absolute Difference (MAD) – Mean Absolute Difference (MAD) – This method is similar to the Threshold Variance method except that the MAP is not squared. This method calculates the mean absolute difference from the mean.
Volatility – Volatility is defined as the ratio of the arithmetic mean (AM) to the geometric mean (GM) for a given characteristic. Values range from +1 to ∞ because AM ≥ GM for a particular property. A higher variance ratio means more relevant features

Wrapper method

The wrapper method relies on the classifier. These methods assess the 'goodness' of a selected feature subset directly from classifier feedback in terms of classification accuracy.



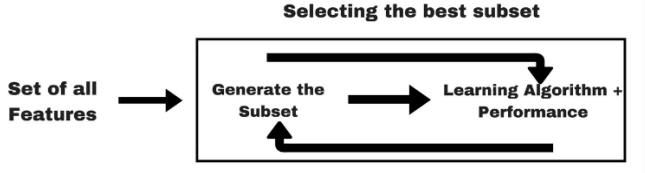
- Forward Selection — This method is an iterative approach that starts with an empty feature set and adds features to optimize the model at each iteration. The stopping criterion is until adding

new variables no longer improves the model's performance.

- Backward Elimination – This method is also an iterative approach, starting with all features first and removing less important features after each iteration. The stopping criterion is until no improvement in model performance is observed after removing features.
- Bidirectional Elimination – This method uses forward selection and backward elimination techniques simultaneously to arrive at a single solution.
- Exhaustive Selection - This technique is considered a brute-force approach to evaluating a subset of features. Create all possible subsets, create a learning algorithm for each subset, and select the subset that gives the best model performance.
- Recursive elimination – This greedy optimization technique selects features by recursively considering a smaller set of features. The estimator is trained on the first set of features and its importance is determined using the feature's feature importance attribute. Then less important features are removed from the current feature set until the required number of features remains.

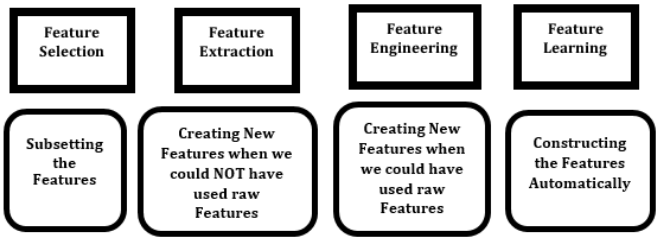
Embedded method

In the embedding method, searching for the best subset of features is built into the construction of the classifier and can be viewed as searching in the combined space of feature subsets and hypotheses. Embedding methods are also specific to a particular classifier, but are cheaper than wrapper methods.



Regularization - This method adds penalties to various parameters of the machine learning model to avoid overfitting the model. This feature selection approach uses lasso (L1 regularization) and elastic mesh (L1 and L2 regularization). A penalty is applied to the coefficients, making some coefficients zero. Features with zero coefficients can be removed from the data set.

Tree-based methods – methods such as random forests, gradient boosting. To select features, specify the feature importance. Feature importance indicates which feature is more important in influencing the target feature. Figure depicts the feature selection methods.



Feature Selection Methods

III. Conclusion

The more features, the faster and more convenient the filter method
In addition to the methods described above, there are many other ways to select features. Using a hybrid technique for feature selection can reduce the shortcomings of the algorithm by choosing advantages over other techniques.

IV. REFERENCES

[1] Alfaras, M., Soriano, M. C., and Ortín, S. (2019). A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. Front. Phys. 7:103. doi: 10.3389/fphy.2019.00109

[2] Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., and Cheng, X. (2020). Artificial intelligence and machine learning to

fight COVID-19. Physiol. Gen. 52, 200–202. doi: 10.1152/physiolgenomics.00029.2020

[3] Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., et al. (2020). An introduction to machine learning. Clin. Pharmacol. Therapeut. 107, 871–885. doi: 10.1002/cpt.1796

[4] Chan, H. L., Chou, W. S., Chen, S. W., Fang, S. C., Liou, C. S., and Hwang, Y. S. (2005). Continuous and online analysis of heart rate variability. J. Med. Eng. Technol. 29, 227–234. doi: 10.1080/03091900512331332587

[5] Chen, M., Hao, Y., Hwang, K., Wang, L., and Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. IEEE Access 5, 8869–8879. doi: 10.1109/ACCESS.2017.2694446

[6] R. Chinnaiyan and D. Stalin Alex, "Early Analysis and Prediction of Fetal Abnormalities Using Machine Learning Classifiers," 2021 2nd International Conference on Smart Electronics andCommunication(ICOSEC),2021,pp.1764-1767,doi: 10.1109/ICOSEC51865.2021.9591828. https://ieeexplore.ieee.org/abstract/document/9591828)

[7] R. Chinnaiyan and S. Alex, "Machine Learning Approaches for Early Diagnosis and Prediction of Fetal Abnormalities," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-3, doi: 10.1109/ICCCI50826.2021.9402317. https://ieeexplore.ieee.org/abstract/document/9402317)

[8] R.Chinnaiyan, Dr.Stalin Alex (2021). Optimized Machine Learning Classifier for Early Prediction of Fetal Abnormalities . International Journal of Computational Intelligence in Control 13(2)

[9] Manoj Challa , Dr.R.Chinnaiyan (2019) "Optimized Machine Learning Approach for the Prediction Of Diabetes-Mellitus (Recommended for Scopus Indexed Publication in Springer - Advances in Intelligent Systems and Computing Series)

[10] Preetika B, ; M. Latha; M. Senthilmurugan; R. Chinnaiyan, , "MRI Image based Brain Tumour Segmentation using Machine Learning Classifiers," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-9, doi: 10.1109/ICCCI50826.2021.9402508

[11] Dr.R.Chinnaiyan, Sabarmathi.G (2017)," Investigations on Big Data Features , Research Challenges and Applications", IEEE International Conference on Intelligent Computing and Control Systems, ICICCS 2017, 782 – 786

[12] Dr.S.Nirmala ,R.Chinnaiyan et.al. (2021). Blockchain based Secured Framework for Road Traffic Management using Fog Computing . International Journal of Computational Intelligence in Control 13(2)

[13] G. Sabarmathi and Dr.R. Chinnaiyan, "Reliable Machine Learning Approach to Predict Patient Satisfaction for Optimal Decision Making and Quality Health Care," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1489-1493.

[14] G. Sabarmathi and Dr.R. Chinnaiyan, "Big Data Analytics Framework for Opinion Mining of Patient Health Care Experience" International Conference on Computing Methodologies and Communication (ICCMC 2020), IEEE Xplore Digital Library

[15] G. Sabarmathi and R. Chinnaiyan, "Reliable feature selection model for evaluating patient home health care services opinion mining systems," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-4, doi: 10.1109/ICAECA52838.2021.9675485.

[16] G.Sabarmathi , Dr.R.Chinnaiyan (2018), "Envisagation and Analysis of Mosquito Borne Fevers – A Health Monitoring System by Envisagative Computing using Big Data Analytics" in ICCBI 2018 – Springer on 19.12.2018 to 20.12.2018 (Recommended for Scopus Indexed Publication IEEE Xplore digital library)

[17] G.Sabarmathi , Dr.R.Chinnaiyan, Reliable Data Mining Tasks and Techniques for Industrial Applications, IAETSD JOURNAL FOR ADVANCED RESEARCH IN APPLIED SCIENCES, VOLUME 4, ISSUE 7, DEC/2017,PP- 138-142, ISSN NO: 2394-8442

[18] G.Sabarmathi, Dr.R.Chinnaiyan "Mining Patient Health Care Service Opinions for Hospital Recommendations" International Journal of Engineering Trends and Technology 69.9(2021):161-167

[19] Hari Pranav A; M. Latha; Ashwin. M. S; R. Chinnaiyan, "BlockchainAs a Service (BaaS) Framework for Government Funded Projects e-Tendering Process Administration and Quality Assurance using Smart Contracts," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-4, doi: 10.1109/ICCCI50826.2021.9402348.

[20] Hari Pranav A;M. Senthilmurugan;Pradyumna Rahul K;R. Chinnaiyan , "IoT and Machine Learning based Peer to Peer Platform for Crop Growth and Disease Monitoring System using Blockchain," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402435.

[21] M. Latha, M. Senthilmurugan and R. Chinnaiyan, "Brain Tumor Detection and Classification using Convolution Neural Network Models," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-5, doi: 10.1109/ICAECA52838.2021.9675585.

[22] M. Senthilmurugan, M. Latha and R. Chinnaiyan, "Analysis and Prediction of Tuberculosis using Machine Learning Classifiers," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-4, doi: 10.1109/ICAECA52838.2021.9675482.

Cite this article as :

Mohamed H B, Dr. Md. Sameeruddin Khan, Dr. Mohan K G, Dr. Parashuram Baraki, "Machine Learning based Feature Selection Approaches for Early Prediction of Autism Spectrum Disorder - Review", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 1, pp. 451-455, January-February 2023. Available at doi : https://doi.org/10.32628/IJSRST2310161
Journal URL : https://ijsrst.com/IJSRST2310161