International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Review Article)

A literary review of machine learning sex classification methods

Nirav Adavikolanu *

The Harker School, San Jose, CA, USA.

International Journal of Science and Research Archive, 2025, 17(01), 092-096

Publication history: Received on 22 August 2025; revised on 28 September 2025; accepted on 01 October 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.17.1.2723

Abstract

Being able to determine a person's biological sex from brain scans offers valuable insight into how male and female brains differ in structure and function. These differences are linked to variations in how certain neurological and psychiatric disorders, such as Alzheimer's disease, autism spectrum disorders, and schizophrenia, develop and progress. Traditional methods for sex classification have often relied on comparing overall brain volumes, which can miss subtle and complex patterns in brain connectivity and shape. Various technical advances in neuroimaging now make it possible to examine the brain from multiple perspectives, capturing both fine-grained structural details and dynamic functional activity. The scale and complexity of these data require more powerful analytical tools, and recent work has turned to machine learning (ML) and artificial intelligence (AI) models. For such models to be useful, they must not only achieve robust and generalizable performance but also provide meaningful insights into sex-related brain differences and their behavioral implications. In this review, we critically analyze recent AI-based studies according to criteria including model performance and generalizability, uni- versus multi- modal approaches, identification of biomarkers, and brain–behavior associations. We further discuss the relative strengths and limitations of different methods within these frameworks.

Keywords: Machine learning (ML); Artificial Intelligence (AI); Neuroimaging; Biomarker; Sex Classification

1. Introduction

The ability to identify the sex of a brain based on imaging scans holds significant importance in both clinical and research contexts. Sex differences in brain structure and function are associated with variations in the prevalence, progression, and manifestation of numerous neurological and psychiatric disorders, including Alzheimer's disease, autism spectrum disorders, and schizophrenia [1, 2]. If researchers can determine what causes certain neurological disorders to appear more in one sex than another, it will likely lead to progress in understanding these disorders and their cures [3].

Noninvasive neuroimaging techniques, such as electroencephalography (EEG), structural magnetic resonance imaging (sMRI), functional magnetic resonance imaging (fMRI), and diffusion tensor imaging (DTI), can help achieve this goal. Each modality provides a different perspective on brain organization: EEG and fMRI provides information about brain function [4, 5], sMRI focuses on structural anatomy [6], and DTI examines white matter integrity [7]. These complementary views allow researchers to identify both global and localized neurological features that contribute to sex-based differences.

Traditional statistical models used for this task often analyze only macro-level brain volume differences between males and females. These models typically ignore the rich spatiotemporal and functional information embedded in brain networks, and lack the ability to capture subtle, nonlinear patterns [8]. By contrast, AI and ML models are capable of identifying more nuanced relationships in high-dimensional data, making them well-suited for this classification task [9].

* Corresponding author: Nirav Adavikolanu

Considering recent advances in AI technologies, several studies have attempted to develop machine learning models that can accurately predict sex based on brain scans. For such models to be useful, they must not only achieve robust and generalizable performance but also provide meaningful insights into sex-related brain differences and their behavioral implications. In this review, we critically analyze recent AI-based studies according to criteria including model performance and generalizability, uni- versus multi- modal approaches, identification of biomarkers, and brain–behavior associations. We further discuss the relative strengths and limitations of different methods within these frameworks.

2. Methods

We review the recent AI or machine learning approaches for classifying the biological sex based on the neuroimaging datasets. A generic work flow for these approaches is shown in **Figure 1** which mainly consists of the data modalities used, preprocessing, extraction of relevant features, ML/AI methods used, performance analysis, identification of biomarkers, and brain-behavior analysis. Below, we present a brief overview of the recent methods used for sex classification and discuss their relative merits.

van Putten et al. [10] developed a six-layer convolutional neural network to identify the sex by brain imaging. Specifically, their model was built to take in electroencephalograms (EEG), which are recordings of brain activity. The EEG data was translated into a 2D image and also represented spatial and temporal dimensions. To do this, they created an input matrix of 24 EEG channels, with a total of 26 channels being recorded. The model managed to achieve the highest accuracy, 81% (p<10^-5), of any EEG machine learning model when it was created. The testing can be further corroborated by the training set of 1000 individuals, and the separate, independent testing data set of 308. One major success of this paper was that, using a visualization technique similar to "deep dreaming", the researchers were able to determine that fast beta activity (20-25 Hz) was one of the main differentiators between sex in brains.

Kim et al.'s [11] model to classify brain images by sex was a combination of multiple ML algorithms: GLM, GBM, and xgboost. In order to find the best model, a ML pipeline was implemented in H2O's Driverless AI package. The pipeline included random hyperparameter tuning, along with feature selection and generation. A stacked ensemble model was finally chosen, and it combined the three previously mentioned models to provide the optimal classification. As for inputs to the model, it used brain morphometric and white matter structural connectomic estimates. The data came from the Adolescent Brain Cognitive Development (ABCD) study release 2.0, with about 9,658 participants used. In terms of accuracy, the model was able to predict sex with a success rate of 93.32%, while also performing better than TabNet, a deep neural network with high feature interpretability and efficient learning. As a note, the testing was not done on an independent dataset; they instead used 20% of the ABCD data. Using a machine learning interpretation framework called K-Local Interpretable Model-agnostic Explanation (K-LIME), these researchers were able to find that, out of the 100 most important brain features, 41 related to diffusion white matter connectivity, while a large number of the remaining ones were grey matter morphometric features.

The approach taken for *"Functional connectomics from resting-state fMRI"* by Smith et al. [12] was to map the macroscopic functional connectome using resting-state functional MRI (R-fMRI), which involved studying spontaneous fluctuations in brain activity. The authors first identified node networks and estimated functional connections between those nodes. The data used came from the Human Connectome Project (HCP), which consists of one hour of whole-brain R-fMRI information per subject. This also means they did not use multimodal data. A significant amount of effort was placed on achieving accurate alignment of functional data across subjects, using methods like Multimodal Surface Matching (MSM). They also used multivariate models and statistics to relate connectomes. The data came in the form of R-fMRI and MRI from the HCP, which consisted of 1000 data points. In terms of accuracy, they were able to predict the sex of a person with an 87% accuracy.

Casanova et al. [13] combine machine learning methods with graph theoretical analysis to determine sex differences in resting-state functional brain connectivity. Using correlation coefficients derived from R-fMRI time series, a Random Forest (RF) and a Ensemble Method based on Least Angle Shrinkage and Selection Operator (Lasso) Regressors (ELRC) were both utilized. The models' direct input were the correlation matrices, so there was no multimodal data used. In terms of accuracy, the ELRC reached 62.3% while the RF achieved 65.4%. These models have not been generalized outside of their given dataset, the 1000 Functional Connectome Project (FCP), but are quite robust internally. In addition, the model does not appear to have used multimodal data. This study did indeed find what can be considered biomarkers, as gender-discriminative edges and brain regions were pinpointed as sources of difference. As a note, these biomarkers were not linked to individual cognitive measures within the models, but based on known behavioral and cognitive sex differences, these were claimed to be the neural source.

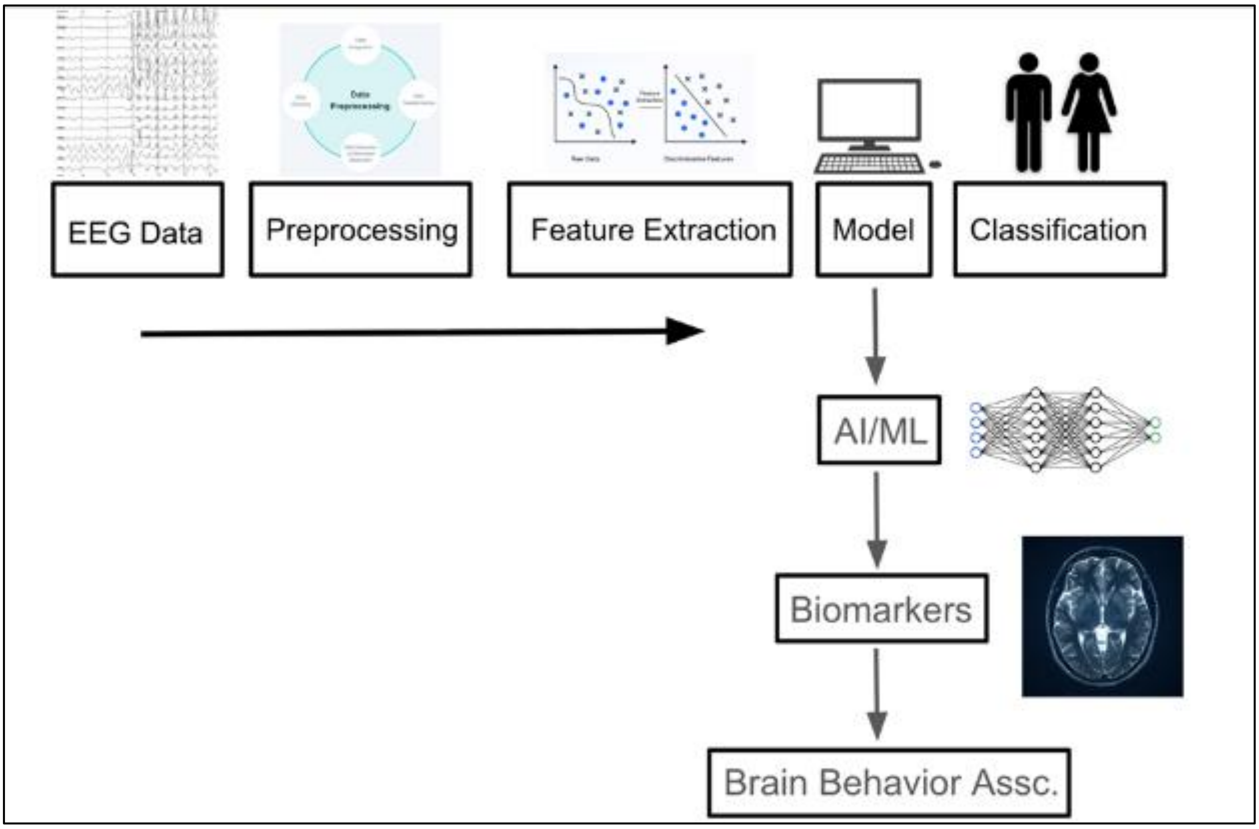

Figure 1 Schema for Sex Classification Model

3. Discussion

We now discuss the relative merits and demerits of the six ML or AI methods for sex classification presented in the previous section.

Van Putten et al. developed a six-layer convolutional neural network to classify sex using EEG data. A visualization technique called as deep dreaming is used to explain the differences in the brain function of these two sexes. They trained the model using the data from 1000 individuals and reported an 81% classification accuracy on an independent dataset that is not used in training the model. They identified the fast beta band (20-25 Hz) activity as a key differentiator in discriminating the two sexes. In this approach, the 24 channel multivariate temporal EEG data was represented as an image and 2D convolutions were used to model it. However, these EEG channels are permutable and therefore can be arranged in any order. Therefore, the results with one ordering of the channels may differ from the other. Moreover, the EEG data is characterized by rich temporal characteristics of the data which may not be adequately modeled by representing the data as an image. Such datasets can be effectively modeled by methods such as recurrent or 1D convolutional neural networks. They have also not analyzed how the biomarkers that they extracted inform the behavioral or cognitive measures of the participants.

Kim et al. utilized a stacked ensemble model combining GLM, GBM, and XGBoost classifiers, trained on morphometric and white matter connectomic features from nearly 10,000 adolescent participants in the ABCD study. They reported accuracy of 93.32% and using feature identification methods they found 41 brain features that helped in classification. They performed hyperparameter tuning to get optimal performance. However, the hyperparameter tuning requires a careful implementation of hyperparameter tuning using a nested validation. Otherwise, the reported high accuracies could be overoptimistic. They have also not performed any brain behavioral analysis in this work. The model was trained primarily on data from adolescents, which may limit its applicability to adult populations due to developmental neuroplasticity. Further testing is required to validate its effectiveness across the lifespan.

Smith et al. took a different approach, leveraging resting-state fMRI (R-fMRI) data from the Human Connectome Project (HCP) and applying Multimodal Surface Matching (MSM) to enhance spatial alignment between subjects. This led to more accurate cross-subject comparisons and better parcellation of functional brain regions. Their multivariate model achieved 87% classification accuracy, supported by the high quality and large volume of the HCP dataset. A particular strength of this study is its emphasis on functional connectivity patterns, which are increasingly recognized as key indicators of sex differences. However, the model relied on static correlation matrices and did not account for dynamic functional connectivity, potentially overlooking meaningful time-varying neural interactions that could further improve classification performance.

Casanova et al. combined machine learning and graph-theoretical approaches by feeding resting-state fMRI-derived correlation matrices directly into a Random Forest (RF) and an Ensemble Lasso Regression Classifier (ELRC). The RF achieved 65.4% accuracy, and the ELRC 62.3%, lower than other methods but notable for its ability to identify sex-discriminative brain edges without extensive feature engineering or reduction. This graph-based method provided interpretable insights into network-level sex differences and proposed candidate biomarkers. However, the models were validated only within a single dataset (the Functional Connectome Project), limiting its generalizability. Additionally, while the ELRC approach is promising for sparse, interpretable models, its lower performance and lack of linkage to behavioral or cognitive outcomes highlight the need for more comprehensive, multimodal analysis.

4. Conclusion

In this review, we examined six recent studies that applied machine learning methods to sex classification using brain imaging data. We evaluated their strengths and limitations with respect to key criteria, including model generalizability on independent datasets, the use of uni- versus multi- modal approaches, the identification of meaningful brain biomarkers, and the extent to which these findings inform our understanding of human behavior. Our analysis highlights important shortcomings in current approaches and underscores the need for more robust, interpretable, and generalizable models. We hope this review will guide future research toward developing AI methods that not only improve classification performance but also advance our understanding of sex-related brain differences and their implications for health and disease.

Compliance with ethical standards

Acknowledgements

We would like to thank Dr. Srikanth Ryali, senior scientist at Stanford school of medicine, Stanford, CA, USA for his guidance in reviewing the AI methods for sex differentiation.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Cosgrove, K. P., Mazure, C. M., & Staley, J. K. (2007). Evolving knowledge of sex differences in brain structure, function, and chemistry. Biological Psychiatry, 62(8), 847–855. https://doi.org/10.1016/j.biopsych.2007.03.001

[2] Cahill, L. (2006). Why sex matters for neuroscience. Nature Reviews Neuroscience, 7(6), 477–484. https://doi.org/10.1038/nrn1909

[3] Joel, D., & McCarthy, M. M. (2017). Incorporating sex as a biological variable in neuropsychiatric research: Where are we now and where should we be? Neuropsychopharmacology, 42(2), 379–385. https://doi.org/10.1038/npp.2016.79

[4] Niedermeyer, E., & da Silva, F. L. (2005). Electroencephalography: Basic principles, clinical applications, and related fields (5th ed.). Lippincott Williams & Wilkins.

[5] Smith, S. M. (2012). The future of FMRI connectivity. NeuroImage, 62(2), 1257–1266. https://doi.org/10.1016/j.neuroimage.2012.01.022

[6] Fischl, B. (2012). FreeSurfer. NeuroImage, 62(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

[7] Basser, P. J., & Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. Journal of Magnetic Resonance, Series B, 111(3), 209–219. https://doi.org/10.1006/jmrb.1996.0086

[8] Ingalhalikar, M., Smith, A., Parker, D., et al. (2014). Sex differences in the structural connectome of the human brain. PNAS, 111(2), 823–828. https://doi.org/10.1073/pnas.1316909110

[9] Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik, 29(2), 102–127. https://doi.org/10.1016/j.zemedi.2018.11.002

[10] van Putten, M. J., Olbrich, S., & Arns, M. (2018). Predicting sex from brain rhythms with deep learning. Scientific Reports, 8, 3069. https://doi.org/10.1038/s41598-018-21467-5

[11] Kim, J., Yoon, H., & Park, H. (2021). The sexual brain, genes, and cognition: A machine-predicted brain sex score explains individual differences in cognitive intelligence and genetic influence in young children. Scientific Reports, 11(1), 18349. https://doi.org/10.1038/s41598-021-97696-w

[12] Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., ... & Nichols, T. E. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nature Neuroscience, 18(11), 1565–1567. https://doi.org/10.1038/nn.4125

[13] Casanova, R., Srikanth, R., Baer, A., Laurienti, P. J., Burdette, J. H., & Hayasaka, S. (2007). Biological parametric mapping: A statistical toolbox for multimodality brain image analysis. *NeuroImage, 34*(1), 137–143. https://doi.org/10.1016/j.neuroimage.2006.09.011