Reviewer: 1

The authors have provided reasonable explanations and the revisions and proposed approach are commendable. One minor issue is that accuracy while popular in the deep learning community is not an objective and correct metric to assess a learning or decision making function but rather also consider the trust given that there be nonidealities in the training datasets (errors, misclasifications, misinterpretations, subjectivity, etc). There have been seminal work efforts in clarifying this issue of accuracy versus trust (see There Is Hope After All: Quantifying Opinion and Trustworthiness in Neural Networks, Front. Artif. Intell., 31 July 2020) and this could be discussed as a potential future research direction.

RESPONSE:
We have noted this caveat and potential hurdle to adoption in the conclusion

Reviewer: 2
Having reviewed this manuscript again, I continue to believe it reports key information that will be impactful for applied and basic research alike, including with relatively short-term potential relevance for clinical practice. In the current revision, the authors have addressed my comments from the last round of review. They have also introduced additional text, most of which I find helpful. (I have a few minor comments on the new text, included at the bottom of this review). However, I have identified one more critical issue that I believe the authors need to acknowledge in the manuscript and/or supplement.
This one critical comment, which I do believe the authors can and should readily address by noting the issue in their main text and/or supplement at relevant parts of the manuscript, is that the negative associations observed between certain diagnoses and a subsequent diagnosis of autism may in fact derive from collider biases (the collider here being "is being seen for a medical issue.") If I am correct that this is an important caveat to the interpretation of these findings, and in particular to some of these salient observations that might otherwise be read as 'protective' for autism, then this issue should be noted and cited clearly in the manuscript. The existence of this issue does not change my insistence on the importance of the observations either for basic or applied research. To the contrary, the impact of the manuscript will be increased to the extent that the authors can convey their awareness of Berkson's paradox/collider bias as it may manifest here, because it is likely to enhance the design and conduct of the basic research that is now needed to better understand the novel associations reported here.
My more minor comment pertains to new text comparing the current method with neural network approaches. As a neural network researcher, I take exception to some of the phraseology used to compare the current method with neural networks. I think these exceptions are fundamentally immaterial to the main contribution of the manuscript, and so I find these issues distracting. (I also fear they may act as a lightning rod for irrelevant criticism and competition between subfields of machine learning). Consider this section: "Importantly, unlike neural network architectures with chosen and often fixed topologies, the number of states and connectivities in the PFSA models are inferred from data,

implying a more adaptive framework; with the algorithms generating higher resolution models when needed, and opting for a lower resolution otherwise." While it is true that neural networks often (not always) have chosen and fixed topologies, the connectivity is of course definitionally learned, and therefore the effective topology at the end of training can often represent a different topology than was designed at the start (e.g., by setting numerous weights to zero). In addition, as neural networks are provably universal function approximators, the whole discussion seems unnecessary. I think it more than sufficient to walk back many of the more speculative claims about these comparisons and simply say that derivation of a neural network equivalent of these methods is not obvious, and that the methods used in the current manuscript involve fewer nominal free parameters. I do not think this takes anything away from the importance of the approach or findings, whereas the current discussion may (simply by distracting the reader with what to me is at best an irrelevant and at worst trivial yet contentious point).

RESPONSE:

1. We have discussed collider bias as part of limitations, and noted that the associations we find needs further study in future
2. We have softened the discussion on neural networks following reviewer 2, noting that we have fewer free parameters, which aids training, and might explain our performance boost.