**CONCLUSION**

Thus, we were able to give doctors a more interpretable machine learning model with 5 interaction pairs. These 5 pairs were confirmed by expert neonatologists as risk factors that were most significant that lead to severe RoP. Confirming our results with domain experts, we have shown that a generalized additive model with pairwise interactions was increasing the interpretability of the model. Our approach was consistent with the research suggestions in the cited papers on interpretability of machine learning in healthcare. The adding of these 5 pairwise interactions increased the accuracy of our model from 33.0 to 33.5 We had chosen the cut-off for 0-1 classification to be 0.05 instead of the 0.5 value, because we wanted to minimize the type II error. We have also not used the ROC metric with various classification thresholds, because classification threshold invariance was not desirable. Minimizing the type II error is critical in healthcare, where a missed diagnosis could lead to full blindness of the patient for lifetime. There was no practical benefits to explore cases where type II error was not zero, but when we relaxed our type II error being zero constraint, we could see a small increase of accuracy, from 75% up to 76.6%, by the adding of most significant pairwise interactions. Thus, we concluded that GA2M approach was helpful mostly with interpretability of the machine learning model by confirming hypothesis of risk factor two-way interactions. However, comparing our RoP data set with other diseases in healthcare machine learning, it should be noted that the size of our data set is relatively small and also many variables are categorical, rather than numerical. Thus, the accuracy effect of GA2M approach over GAM was rather minimal. Further work should focus on collecting numeric data and turning some of the categorical variables into numeric ones, such as blood transfusion as number of times rather than a binary value..