

Predicting Cognitive States Using Machine Learning Fusion Paradigms to Reduce Model Uncertainty

Victoria R. Rodrigues, Mudit Paliwal, Nicholas J. Napoli,

Industrial and Systems Engineering, University of Florida, Gainesville, FL, 32611, USA

Human Informatics and Predictive Performance Optimization (HIPPO) Lab, Gainesville, FL, 32611, USA

Angela Harrivel, Kellie D. Kennedy, Chad L. Stephens

NASA Langley Research Center, Hampton, VA 23681

I. Introduction

The development of a synergetic system between humans and technology is a challenge that the scientific community has been facing for many years. Our aeronautic research aims to enhance this synergy between humans and machines through predictive human performance modeling for systems to mitigate high-risk situations. By being able to predict and anticipate human states, the crew monitoring system should be able to adjust and support the pilot for aviation safety. Our work focusing on attention-related human performance-limiting states (AHPLS) that impact a pilot's performance and introduce high-risk catastrophic situations [1]. For example, AHPLS has been cited as a causal factor in more than 50% of all loss control in flight and thus contributes significantly toward commercial aviation fatalities [1, 2]. Cognitive state and its physiological fingerprint can be valuable information for this detecting AHPLS, but human cognitive state detection is still a major limitation for these crew monitoring systems.

Prior Work. Despite the considerable amount of literature that discusses human predictive modeling, some papers do not actually use “predictive modeling” but rather use “explanatory modeling”, where only statistical significance is examined. These two terms are frequently conflated making the literature a bit ambiguous at first glance [3]. The majority of the other cognitive state prediction literature focuses on examining a single physiological modality (e.g., EEG, ECG) to infer a cognitive state [4]. Thus the research literature for predicting task-related errors caused by cognitive impairment is very sparse, especially when complex multi-class human state prediction problems arise. Even from the standpoint of AHPLS, which assesses engagement, fatigue, emotion, and workload, there is scant research for real-time classification performance for such multi-class problems [5] and there is no accepted generalized approach to predict psychological states [6]. One major reason there is no generalized predictive approach for cognitive psychophysiological research is many of the published work for multi-class problems have not demonstrated exceptional performance on any particular dataset as the gold standard.

From the perspective of predictive modeling, the purpose of a classifier in machine learning (ML) is to characterize patterns within data in order to recognize and detect pre-determined classes or labels. As ML problems have increased in their difficulty for classification, a single classical ML architecture has been shown to be a sub-optimal ML approach for classification. However, utilizing a diverse set of ML classifiers (e.g., a committee or ensemble), we can increase the detection performance when compared to a single classifier [7]. One of the first contributions of ensemble approaches was Dasarthy et al., who proposed the partitioning of the feature space using two or more classifiers [8]. The main idea behind ensemble methods is to combine multiple individual classifiers in order to reduce the mean squared prediction error (the uncertainty) of the classifier [7]. Over the years the research has grown exponentially, leading to many different ensemble approaches (e.g., dynamic classifier selection, committee of classifiers, voting pool of classifiers) all with the intention to reduce the model's uncertainty or means squared error of the prediction [9–13]. This work was extended to concepts like “ADA boosting” to focus on developing ensemble classifiers that adaptively weight erroneous training data to be re-utilize in the next iteration of the ensemble. This type of ensemble approach sub-samples the training data to design stronger classifiers from weak classifiers by focusing on reducing model uncertainty through the data [14]. While ensemble classifiers have proved to be a better approach than classical classifiers in many different applications, there are still some potential research gaps to improve [15].

Within human performance monitoring and cognitive state detection, predictive models generate uncertainty within the model from various avenues such as large class imbalances, conflicting ground truth labels, subject-to-subject variability, and generate wide data sets due to the diverse amount of sensors [16–18]. Thus, this raises two major research questions regarding how we can better reduce uncertainty: 1) How can we reduce model uncertainty at the foundational level from the perspective of data or input level of the model to optimally define our decision boundaries?

2) How can we improve a model’s decision boundaries to better handle uncertainty at the model level and between ensemble models?

Insights. To address these issues, we propose two methods to handle uncertainty at the data level and the model level. First, we focus on data level uncertainty. A major problem with class imbalanced datasets is classes with lower samples usually have a concentrated representation in the feature space. This adds uncertainty to the data, as it is much harder to generalize the decision boundaries learned during the training phase to the testing phase. Since the model is most likely to make incorrect predictions in the region surrounding its decision boundary, adding new synthetic data points from this area may reduce uncertainty, thus reducing prediction error. For this task, we will utilize an oversampling technique known as SMOTE: Synthetic Minority Oversampling Technique [19]. The main idea behind SMOTE is it incorporates new minority instances between existing (real) minority instances, with no change to the majority classes. We hypothesize this will aid in addressing uncertainty at the data level.

Next, we will propose a novel methodology called “Naive Adaptive Probabilistic Sensor” (NAPS) Fusion framework to improve the uncertainty at the model level and between models. The NAPS Fusion is a framework that utilizes bagged models, each being a different and random subspace of the feature space, which augments the response variables randomly cover the full feature space of the dataset. By taking advantage of Dempster-Shafer (DS) Theory and its ability to handle and quantify uncertainty for distributions that are not fully defined due to incomplete information, NAPS Fusion aims to properly handle between-subject variability and reduce model uncertainty that is caused by the data structure. Its approach consists of first sampling the feature space into organized sub-sets of the feature space and augmenting the response variables, which aids with model sparsity. Then bagging approach is applied to the sensors and uncertainty assignment is utilized as a means for determining which models to choose. Finally, the selected models are combined together to account for a large feature space, thus acting like a larger complex model. The cognitive data we will analyze are from two similar NASA Langley Research Center experiments that induced AHPLS, Scenarios for Human Attention Restoration Using Psychophysiology (SHARP-1), and SHARP-2.

Contributions. Based on our insights, this study aims to investigate the following hypothesis:

1. We will investigate how additional data sets combined with SMOTE can improve model uncertainty and predictive performance for cognitive state detection.
2. We will investigate how does the NAPS framework dynamic model selection based on model uncertainty reduces prediction error when compared to a static ensemble and classical ML methods.

II. Methods

Experimental Design. We will work with data from two similar experiments designed and executed by NASA Langley Research Center, Scenarios for Human Attention Restoration Using Psychophysiology (SHARP-1), and SHARP-2. Both experiments were designed to induce AHPLS. For SHARP-1, twenty-four commercial aviation pilots, two per crew, were asked to perform tasks (resting tasks, benchmark tasks designed to induce AHPLS and low/high workload conditions, and experimental flight scenarios) in a motion-based flight simulator while wearing psychophysiological sensors. Task load, attentional demand, and startle/surprise variations occurred throughout the experimental flight scenarios to induce AHPLS. Each task was used to induce AHPLS under controlled conditions for 6 minutes, except for the Startle/Surprise. After each task, participants were given three self-report questionnaires: one to evaluate workload (NASA Task Load Index), one to assess situation awareness (NASA Situation Awareness Rating Technique), and one to evaluate to which extent each AHPLS was experienced. Benchmark tasks used to induce each AHPL state include Tetris for Channelized Attention, movie scene observation for Startle/Surprise, and Multi-Attribute Task Battery (MATB) for High Workload and Low Workload [20]. For SHARP-2 the experimental design is very similar. The same tasks designed for SHARP-1 were performed by 10 commercial aviation pilots. Most of the task performed in these both studies have been applied to previous task-oriented studies [20, 21]

Data Collection and Study Design. For SHARP-1, EEG data was measured using the Advanced Brain Monitoring electrode head net. ECG, GSR, and respiration were measured using the Mind Media, B.V. Nexus Mark II system. All modalities mentioned above were recorded using a sampling rate of 256Hz. Additionally, Smart Eye eye-tracking system data were recorded. Data from the self-report questionnaires were collected. For SHARP-2, EEG data was measured and recorded using the Advanced Brain Monitoring B-Alert X24 with a sampling frequency of 256 Hz. The MindMedia NeXus-10 system was used to collect electrocardiogram and respiration data at 256 Hz. Functional Near Infrared Spectroscopy (fNIRS) signals were measured for the captain’s seat only using the Imagent by ISS, Inc. Blood volume pulse, galvanic skin response and skin temperature, heart rate, and acceleration of the sensor was collected using Empatica at 24Hz, 4Hz, 1.3Hz, and 32Hz respectively. For Muse, electroencephalogram

data was collected at 500Hz and the acceleration of the sensor at 50 Hz. The electroencephalogram and respiration data were collected using Nexus-10 Stream at 256Hz. The respiration force and acceleration of the sensor were collected using Spire at 473Hz. Lastly, gaze location and left and right pupil diameter was collected using Tobii 2 eye tracker at 50Hz. The software used to record all the time series was MAPPS (EyesDx, Inc., Coralville, IA). This software is designed to collect aircraft and simulator state, event markers, video, and pilot physiological and behavioral data and the software time synchronizes all data channels for real-time review [22].

A. Algorithmic Design and Data Analysis

Research Question (RQ)-1: How additional data sets combined with SMOTE can improve model uncertainty and predictive performance for cognitive state detection?

To answer this question we will compare the prediction accuracy when using the NAPS framework and classical ML approaches on six different scenarios. The first two scenarios would be to predict on SHARP-1 and SHARP-2 separately. In addition to that, the next step would be to predict on both datasets separately but using SMOTE as an oversampling technique to adjust the class imbalance. Our next step would be to join both datasets and use this joint data as input to our prediction model. Finally, we would apply SMOTE on our joint dataset and use it as data to the NAPS framework.

We hypothesize that SMOTE will produce inaccuracies within the synthetic data around the minority class, but the SMOTE approach will improve as the additional SHARP data is amended into the SMOTE model. When the original dataset is too small we are unable to fully define the probability mass function and appropriately capture the variance necessary to discriminate the class labels. This leads to uncertainties into the model and improper predictions will be formed. It is hypothesized that the greater modeling improvement will come when SMOTE is combined with larger SHARP-1 and SHARP-2. SMOTE will never replace real data and if the original data is too small this will cause the model to fail even worse. Thus, the lower the number of samples, the more we are uncertain on how to fully define the data distribution. It is important to also take into account that these two datasets are very similar, they share the same modalities, they aim to induce the same cognitive states, and they apply mostly the same benchmark tests. This is of major import assumption in supporting our hypothesis.

RQ-2: How does the NAPS framework dynamic model selection based on model uncertainty reduces prediction error when compared to a static ensemble and classical ML methods?

To answer this, we will join our two datasets together and apply SMOTE an up-sampling method to reduce class imbalances and help improve the decision. The resultant dataset will be used as input into three different ML learning approaches. The first approach would be to use Support Vector Machines (SVM), which was used before with this SHARP-1 and SHARP-2, and obtained favorable results [5, 22]. The second approach would be a static ensemble, where a fraction of the best performing classifiers on the testing data is selected to compose the ensemble. Finally, the third approach would be the NAPS Fusion framework, which dynamically decides on the classifiers based on the certainty produced by the sample for that particular model. We hypothesize that the NAPS framework will achieve the best results because of its "top-down" approach on how to manage model uncertainty.

III. Conclusion and Future Work

Cognitive state prediction has the potential to detect AHPLS and is paramount for aviation crew monitoring systems to avoiding loss of control and accidents in flight. However, cognitive modeling is still a complex problem with many limitations due to its experimental data collection constraints, and modeling methods for human performance. This work aims to address these challenges by utilizing the NAPS Fusion framework, which implements concepts of Dempster-Shafer Theory, stacking of ML models, and bagging to address uncertainty for multi-classification ML problems. This framework is applied to data of two experiments designed to induce AHPLS, SHARP-1, and SHARP-2. This study answers two important questions: how additional data sets combined with SMOTE can improve model uncertainty and predictive performance for cognitive state detection and how does the NAPS framework dynamic model selection based on model uncertainty reduces prediction error when compared to a static ensemble and classical ML methods. This work is an important step to better understand cognitive modeling and prediction. Future work can focus on a better understanding of how the dimensionality reduction aspect of NAPS Fusion aids prediction accuracy in both datasets and how NAPS Fusion performs on different datasets.

References

- [1] Harrivel, A. R., Liles, C., Stephens, C. L., Ellis, K. K., Prinzel, L. J., and Pope, A. T., "Psychophysiological Sensing and State Classification for Attention Management in Commercial Aviation," *AIAA Infotech @ Aerospace*, American Institute of Aeronautics and Astronautics, 2016.
- [2] Reveley, M. S. B., "Causal Factors and Adverse Conditions of Aviation Accidents and Incidents Related to Integrated Resilient Aircraft Control," , 2010.
- [3] Shmueli, G., "To Explain or to Predict?" *Statistical Science*, Vol. 25, No. 3, 2010, pp. 289–310.
- [4] Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., and Olejarczyk, J., "Predicting Cognitive State from Eye Movements," *PLoS ONE*, Vol. 8, No. 5, 2013.
- [5] Harrivel, A. R., Stephens, C. L., Milletich, R. J., Heinich, C. M., Last, M. C., Napoli, N. J., Abraham, N., Prinzel, L. J., Motter, M. A., and Pope, A. T., "Prediction of Cognitive States during Flight Simulation Using Multimodal Psychophysiological Sensing," *AIAA Information Systems-AIAA Infotech @ Aerospace*, American Institute of Aeronautics and Astronautics, 2017.
- [6] Novak, D., Mihelj, M., and Munih, M., "A Survey of Methods for Data Fusion and System Adaptation Using Autonomic Nervous System Responses in Physiological Computing," *Interacting with Computers*, Vol. 24, No. 3, 2010, pp. 154–172.
- [7] Tresp, V., "Committee Machines," *Handbook for neural network signal processing*, 2001, pp. 1–18.
- [8] Dasarathy, B., and Sheela, B., "A Composite Classifier System Design: Concepts and Methodology," *Proceedings of the IEEE*, Vol. 67, No. 5, 1979, pp. 708–713.
- [9] Woods, K., Kegelmeyer, W., and Bowyer, K., "Combination of Multiple Classifiers Using Local Accuracy Estimates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 4, 1997, pp. 405–410.
- [10] Ho, T. K., Hull, J., and Srihari, S., "Decision Combination in Multiple Classifier Systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, 1994, pp. 66–75.
- [11] Rogova, G., "Combining the Results of Several Neural Network Classifiers," *Neural Networks*, Vol. 7, No. 5, 1994, pp. 777–781.
- [12] Drucker, H., Cortes, C., Jackel, L. D., LeCun, Y., and Vapnik, V., "Boosting and Other Ensemble Methods," *Neural Computation*, Vol. 6, No. 6, 1994, pp. 1289–1301.
- [13] Battiti, R., and Colla, A. M., "Democracy in Neural Nets: Voting Schemes for Classification," *Neural Networks*, Vol. 7, No. 4, 1994, pp. 691–707.
- [14] Freund, Y., and Schapire, R. E., "Experiments with a New Boosting Algorithm," *Proceedings of the Thirteenth International Conference on Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1996, p. 148–156.
- [15] Dietterich, T. G., "Ensemble Methods in Machine Learning," *Multiple Classifier Systems*, Springer, 2000, pp. 1–15.
- [16] Mandryk, R. L., and Atkins, M. S., "A Fuzzy Physiological Approach for Continuously Modeling Emotion during Interaction with Play Technologies," *International Journal of Human-Computer Studies*, Vol. 65, No. 4, 2007, pp. 329–347.
- [17] Seghier, M. L., and Price, C. J., "Dissociating Functional Brain Networks by Decoding the Between-Subject Variability," *NeuroImage*, Vol. 45, No. 2, 2009, pp. 349–359.
- [18] Shilaskar, S., Ghatol, A., and Chatur, P., "Medical Decision Support System for Extremely Imbalanced Datasets," *Information Sciences*, Vol. 384, 2017, pp. 205–219.
- [19] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, Vol. 16, 2002, pp. 321–357.
- [20] Santiago-Espada, Y., Myer, R. R., Latorella, K. A., and Comstock, J. R., "The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide," *NASA Langley Research Center*, 2011, p. 269.
- [21] Gross, J. J., and Levenson, R. W., "Emotion Elicitation Using Films," *Cognition & Emotion*, Vol. 9, No. 1, 1995, pp. 87–108.
- [22] Napoli, N., Adams, S., Harrivel, A. R., Stephens, C., Kennedy, K., Paliwal, M., and Scherer, W., "Exploring Cognitive States: Temporal Methods for Detecting and Characterizing Physiological Fingerprints," *AIAA Scitech 2020 Forum*, 2020.