The 2023 IEEE Signal Processing in Medicine and Biology Symposium

Mental Workload Classification from fNIRS Signals by Leveraging Machine Learning

M. Hasan, M. Mahmud, S. Poudel, K. Donthula, and K. Poudel

MIDDLE TENNESSEE STATE UNIVERSITY

Outline

- Introduction and motivation
- Methods & materials
- Results & discussion
- Conclusion

MIDDLE TENNESSEE STATE UNIVERSITY

Introduction and motivation

- What is mental workload (MWL)?
  - Mental workload describe the cognitive demand that places on an individual during a specific
- What research have explored on MWL
  - Some researchers investigated using EEG and fNIRS
  - Other demonstrated the connectivity in LH and RH while classifying low vs. high MWL
  - Researcher examined using overlapping signals with deep learning and found decent result
- What is fNIRS?
  - Functional near-infrared spectroscopy captures the concentration changes of oxygenated hemoglobin and deoxygenated hemoglobin in the brain’s cortical areas

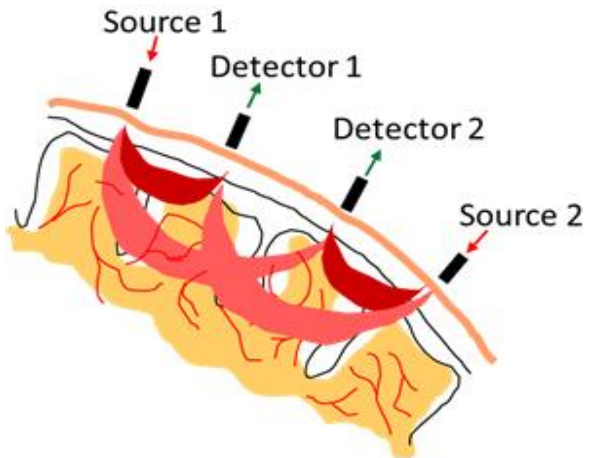

Figure1: fNIRS system [1]

Introduction and motivation

Our main contributions:

1. Identification of low vs. high MWL using whole-brain data
2. Hemisphere (LH vs. RH) analysis to know which hemisphere of the brain dominates in mental workload
3. Find the critical features that are associated with MW classification

Methods & materials

Participants

- 68 participants (32 Asian, 27 White, 3 Black, 2 Hispanic, 1-Pacific Islander, and 3 other race; aged 18 to 44 years). [3]
- All participants were English speakers, and none reported any neurological disease history.
- Participant sat on a standard chair in a quiet room
- The procedures for this study were approved by the IRB at Tuft University.
- Gave written consent about data release for the public.

Methods & materials

Task, procedures, behavioral & fNIRS recoding

- Four n-back (i.e., **0-back, 1-back, 2-back, and 3-back**) stimuli
- Each stimulus was presented for 0.5 seconds, then 1.5 seconds for hidden (**total of 2 seconds**).
- Each stimulus was presented 40 trials (1 to 40)
- Asked to respond to each back via the left or right arrow
- Responses were logged
- fNIRS was recorded using a two-probe headband
- Sampling with 5.2 Hz
- Filtered (0.001-0.2 Hz)

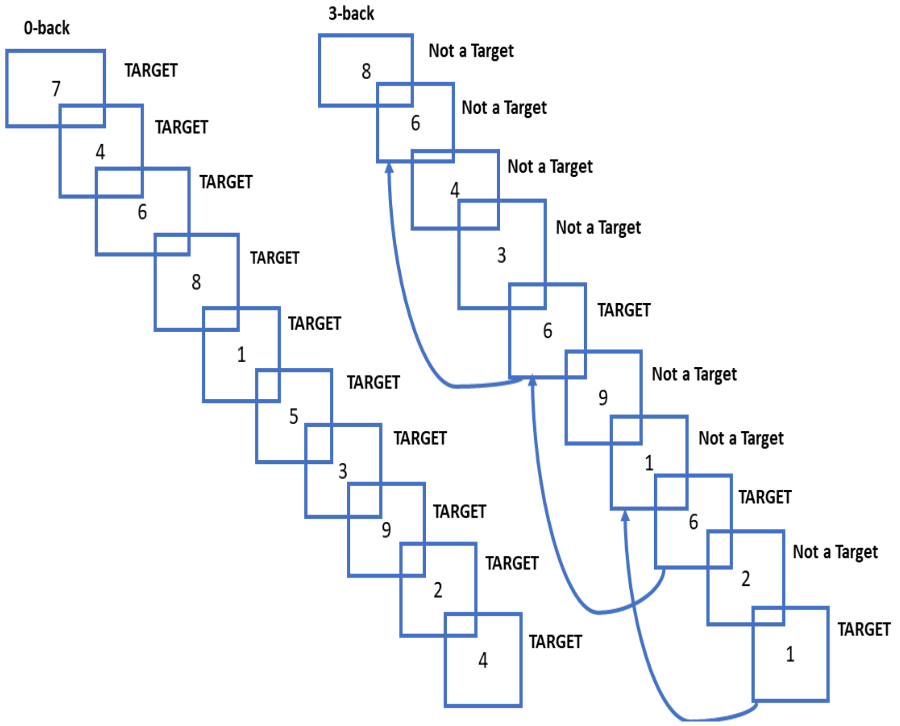

Fig.2: Stimulus presentation (left 0-back; right-3-back)

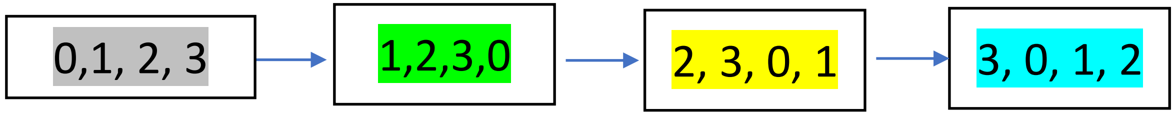

Fig.3: Latino square flatted version of a 4 x4 array

Methods & materials

- Data understanding is the main factor for this project
- We visualize the grand average for 0-back and 3-back data for individuals features and all subjects
- We assume that those features could be useful to classify the MW task (low vs. high)
- (AB or CD) which sensor location on the forehead was used
- which optical measurement type was used (I = intensity, PHI= phase)
- (O = oxygenated hemoglobin; DO = deoxygenated hemoglobin)

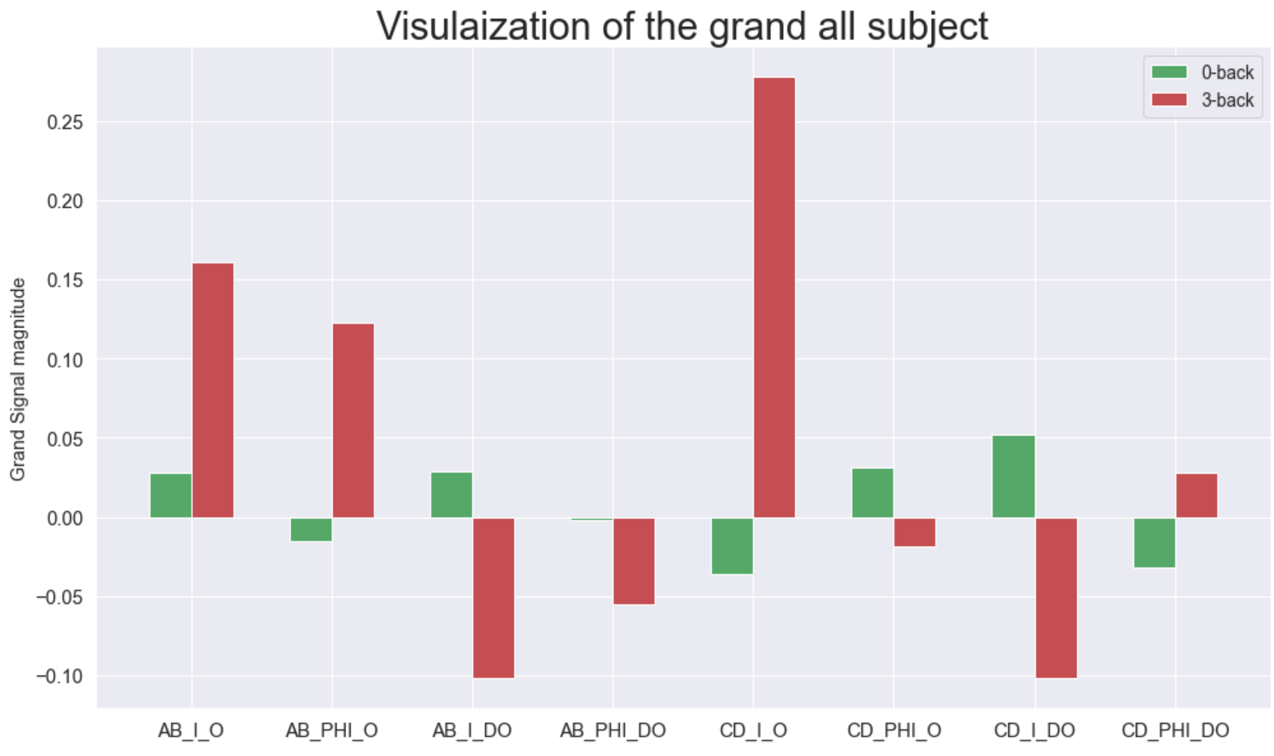

Fig.4: Grand average of each features.

- 0-back task classification accuracy of 96%
- 3-back task classification accuracy of 80%

Methods & materials

Classifiers:

- KNN
- Decision Tree
- XGBoost
- LightGBM
- Random Forest
- SVM

Minkowski distance

$$dist(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{d} |x_i - y_i|^p \right)^{\frac{1}{p}} \qquad (1)$$

Hyper parameter optimization:

- Grid search approach
- Five-fold cross validation
- Select the best model parameters

Methods & materials

Performance metrics formulas:

- Accuracy: (TN+TP)/(TN+TP+FN+FP)
- Precision: TP/(TP+FP)
- Recall: TP/(TP+FN)
- F1 score: Harmonic mean of precision and recall

		Predicted	
		Negative **(N)** **-**	Positive **(P)** **+**
Actual	Negative **-**	True Negative **(TN)**	**False Positive (FP)** **Type I Error**
	Positive **+**	**False Negative (FN)** **Type II Error**	True Positive **(TP)**

Fig.6: Confusion matrix [4].

Feature selection:

- ➢ Permutation features selection
- ➢ We used KNN classifier with permutation feature selection

Results & discussion

TABLE I: KNN, DT, XGBoost, LightGBM, RF, and SVM classifiers' performance metrics (%) for detecting low vs. high MWL.

Classifiers name	Average measure(%)	Whole-brain data	LH's data	RH's data
KNN	Accuracy	98.8	78.5	77.1
	AUC	98.8	78.5	77.1
	Precision	99.0	78.0	77.0
	Recall	99.0	78.0	77.0
	F1-score	99.0	78.0	77.0
DT	Accuracy	87.9	71.4	69.4
	AUC	87.9	71.4	69.4
	Precision	88.0	71.0	69.0
	Recall	88.0	71.0	69.0
	F1-score	88.0	71.0	69.0
XGBoost	Accuracy	91.6	65.3	61.7
	AUC	91.6	65.3	61.7
	Precision	92.0	65.0	62.0
	Recall	92.0	65.0	62.0
	F1-score	92.0	65.0	62.0
LighGBM	Accuracy	77.3	68.7	65.2
	AUC	77.3	68.7	65.2
	Precision	77.0	69.0	65.0
	Recall	77.00	69.0	65.0
	F1-score	77.0	69.0	65.0
RF	Accuracy	96.7	78.8	77.5
	AUC	96.7	78.8	77.5
	Precision	97.0	79.0	78.0
	Recall	97.0	79.0	78.0
	F1-score	97.0	79.0	78.0
SVM	Accuracy	69.5	62.7	59.9
	AUC	69.5	62.7	59.9
	Precision	70.0	63.0	60.0
	Recall	69.0	63.0	60.0
	F1-score	69.0	63.0	60.0

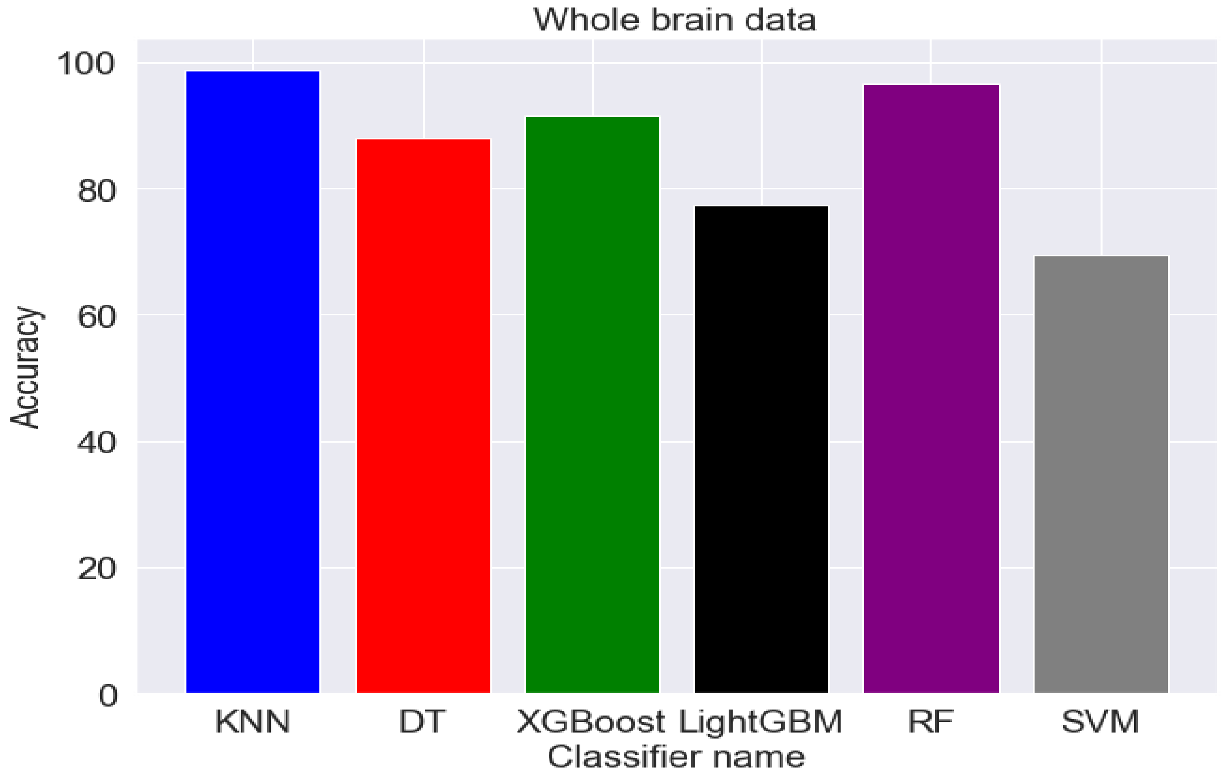

Fig 7. shows the classification of 0-back vs. 3-back using full brain data

Results & discussion

Hemisphere (LH vs. RH) analysis:

- RF classifier provided the best classification accuracy.
- LH: 78.8% accuracy; RH: 77.5% accuracy
- KNN showed 78% on LH and 77% on RH
- SVM exhibited the lowest accuracy (LH: accuracy 62.7%; RH 59.9%)

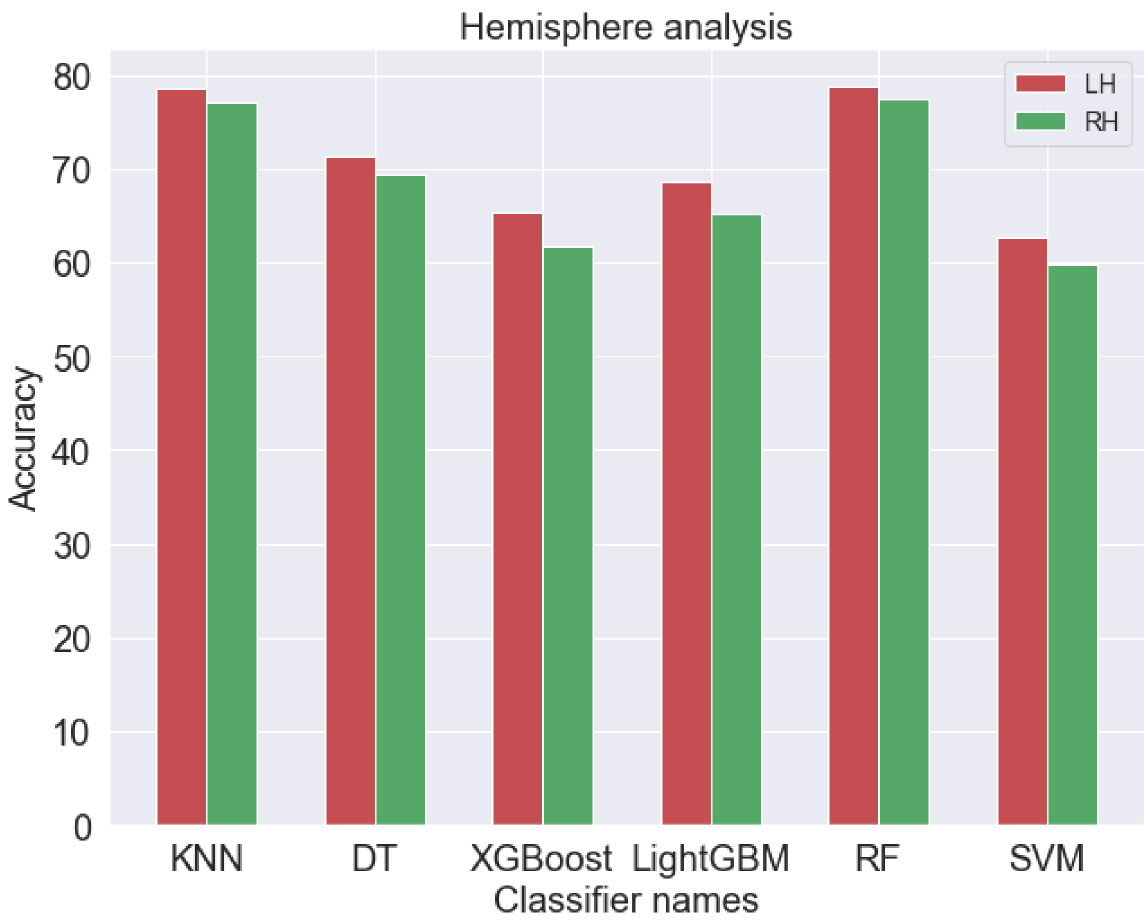

Fig 8. Shows the classification of 0-back vs. 3-back using hemisphere data.

Results & discussion

Features Selection:

- We used the Permutation feature selection with KNN Classifier
- The most important feature of “CD_I_O” is that score of 0.33
- CD_PHI_DO yielded the lowest score.
- We used the top six features out of eight features.

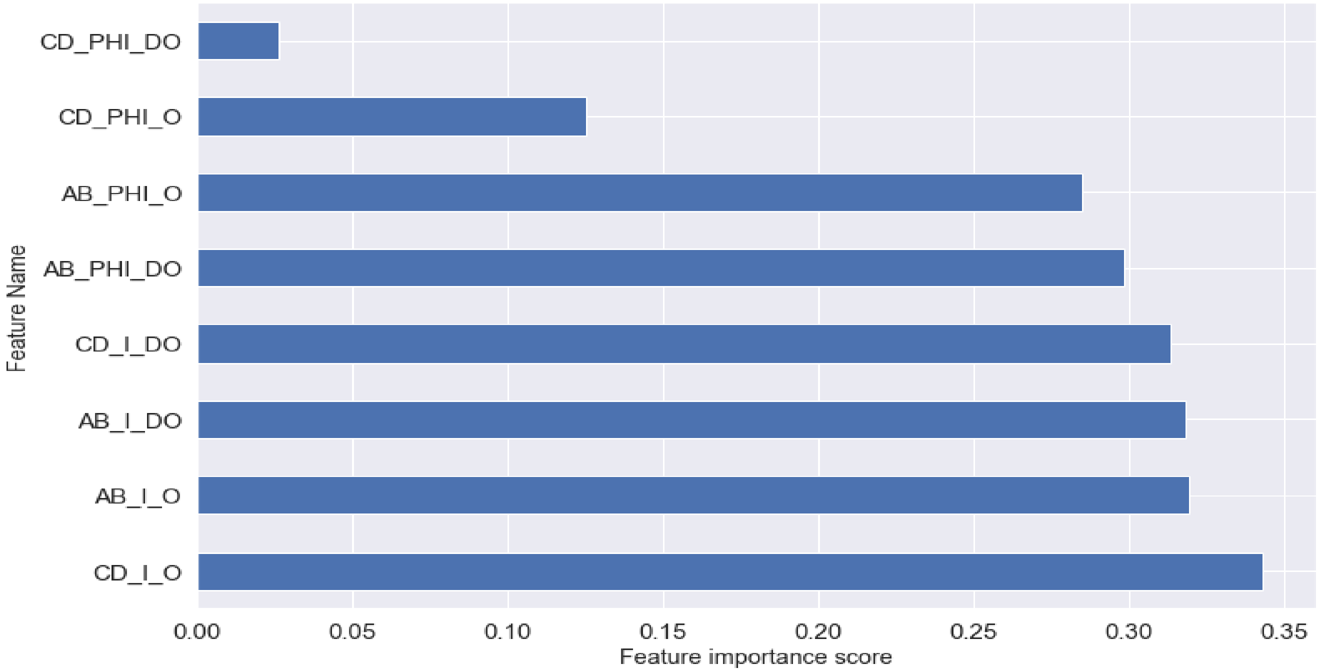

Fig. 9: Visualization of features with their importance ranked. The Y-axis represents the features’ names; the X-axis represents the feature score.

Results & discussion

Classification using the top six important features:

- **KNN** Classifier's accuracy of 97.4%, AUC 97.4 precision, recall, and F1-score 97.0%
- **RF:** accuracy 93.3%, AUC 93.3%, precision, recall, and F1-score 93.0%
- **XGBoost**: accuracy 85.6%, AUC 85.6%, precision, recall, and F1-score 86.0%
- **SVM:** accuracy 67.1%, AUC 67.1%, precision, recall, and F1-score 67.0%

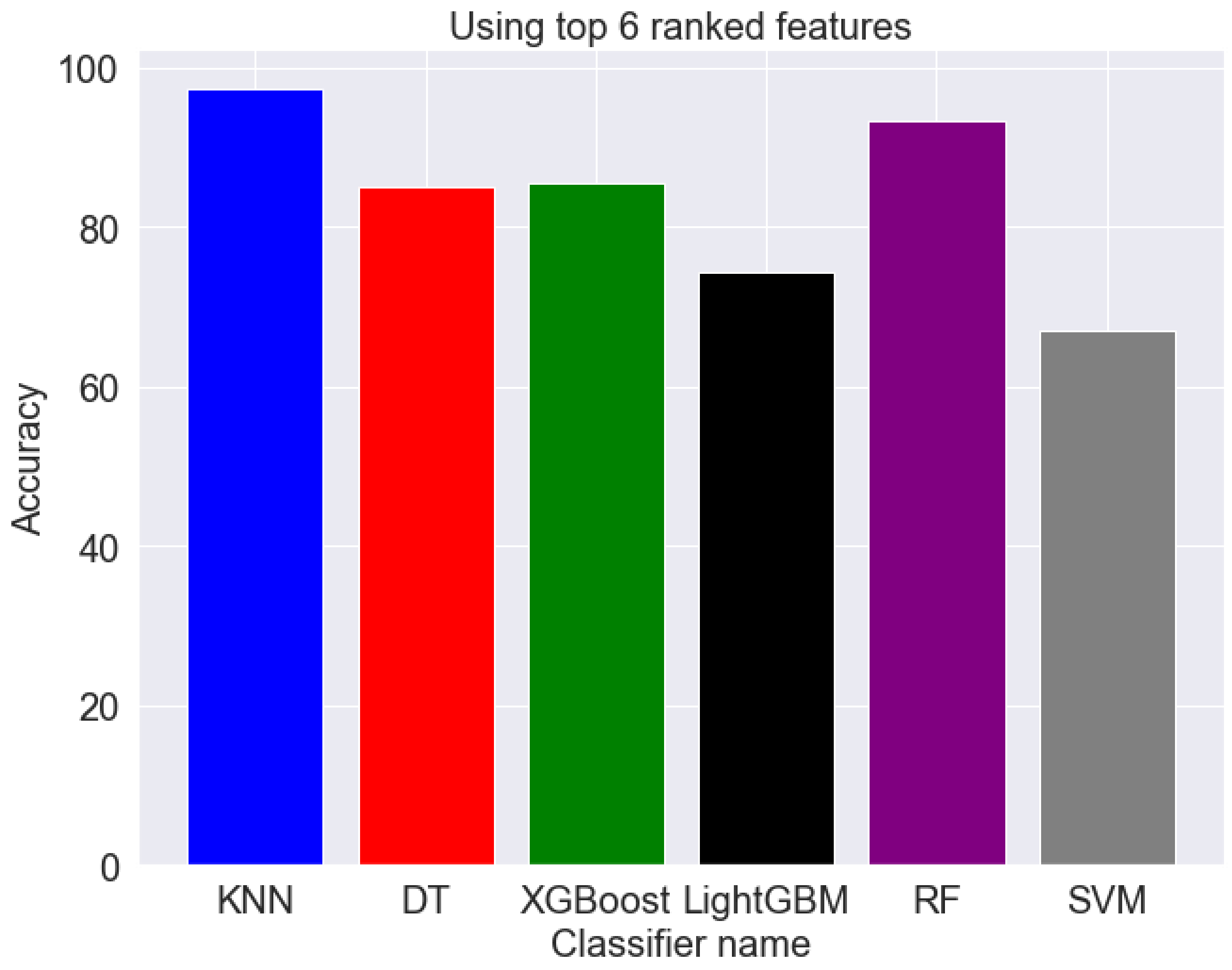

Fig. 10: Classification based on the top six ranked features.

Discussion

- MWL can be classified with an accuracy of 98.8% using whole-brain data
- Six critical features can classify with an accuracy of 97.4%
- LH showed better classification accuracy as compared the RH
- Our result collaborates with previous findings of LH dominance in MWL classification

Conclusion

- This can lead to the design of user-friendly interfaces, automation, and stress-reduced application to enhance safety and performance.
- we used only two probes headband fNIRS system
- In future work, we will explore high, medium, and low MWL classification
- We will use multiprobes headband fNIRS system data.

Acknowledgments

- This work was supported by the Department of Computational & Data Science, and Computer Science at the Middle Tennessee State University, Murfreesboro, TN, USA.

MIDDLE TENNESSEE STATE UNIVERSITY

References

1. https://nirx.net/our-story
2. https://tufts-hci lab.github.io/code_and_datasets/fNIRS2MW.html
3. Huang, Z., Wang, L., Blaney, G., Slaughter, C., McKeon, D., Zhou, Z., ... & Hughes, M. C. (2021). The Tufts fNIRS mental workload dataset & benchmark for brain-computer interfaces that generalize.
4. https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5

Thank you!

MIDDLE TENNESSEE STATE UNIVERSITY