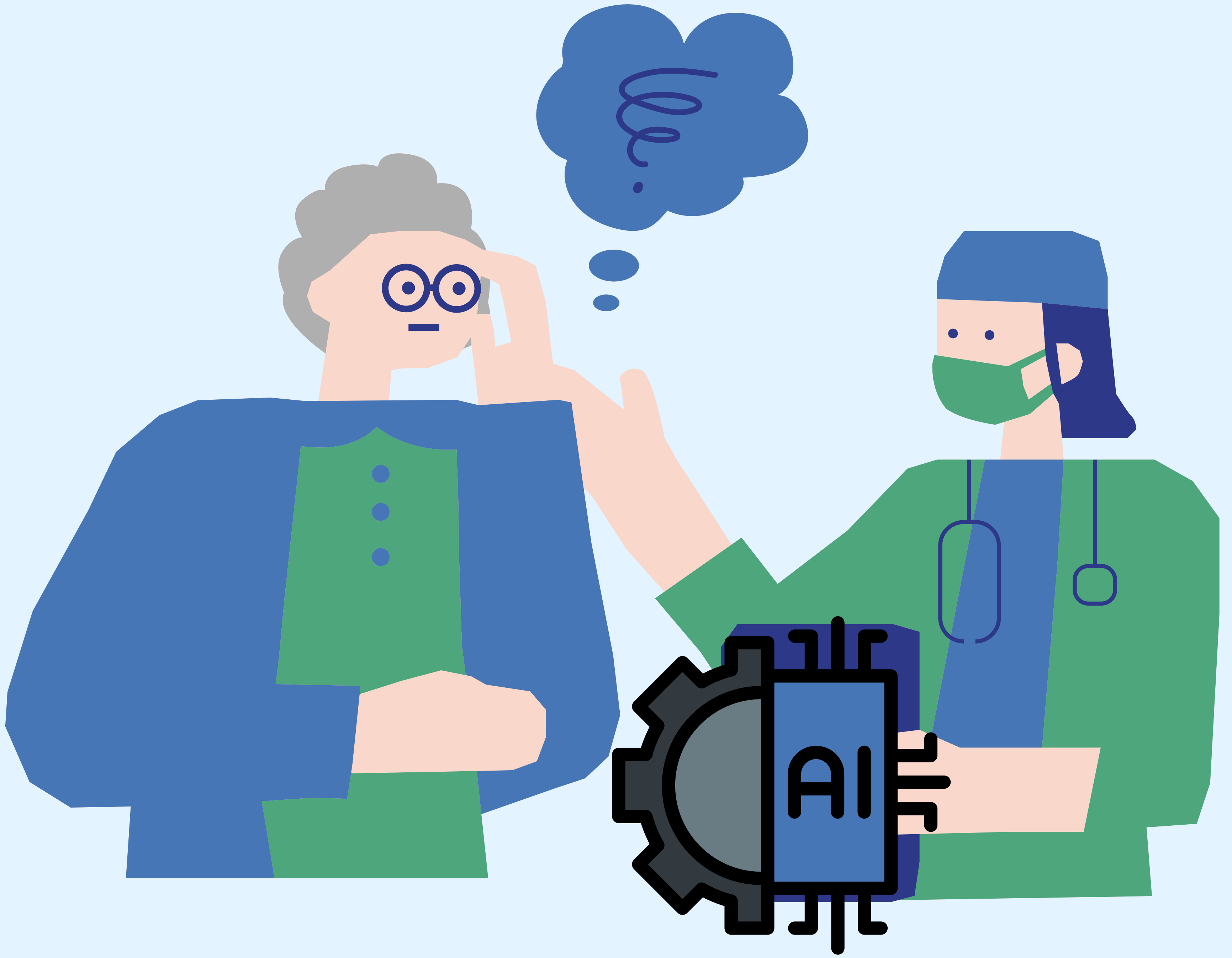
Prediction Model for Alzheimer's Disease

Team 60: Xinyu Chen, Kyla Gabriel, Sofia Rojas, and Tilly Rigby

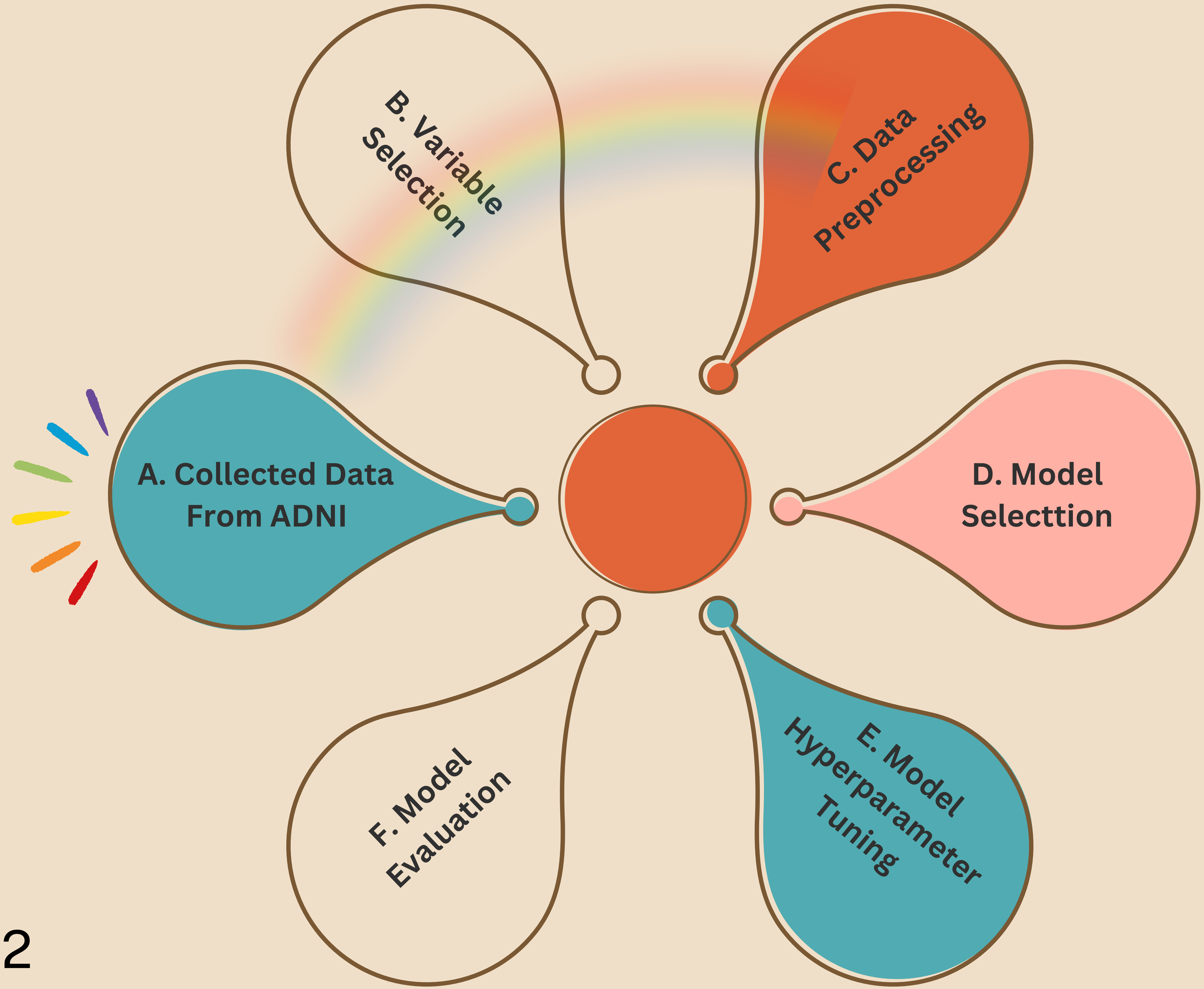
CompSci 109A Final Project, Fall 2023



Problem Statement

Can we accurately predict a **dementia diagnosis** in patients using demographic and clinical factors from the ADNI dataset?

01 Problem Statement
02 EDA
03 Modeling
04 Training details
05 Results
06 Conclusion
07 Future Work



01 Problem Statement

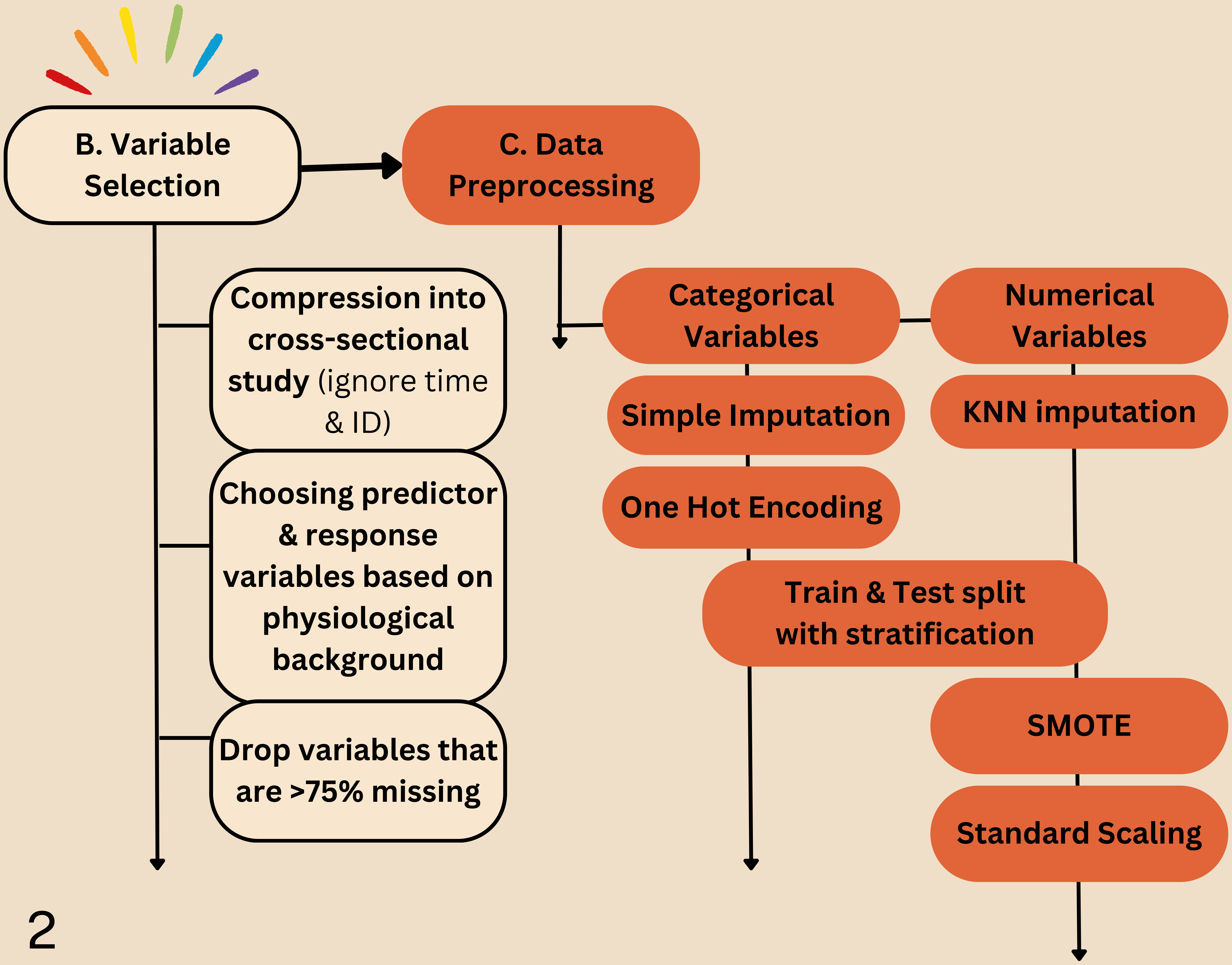
02 EDA

03 Modeling, Training

04 Results

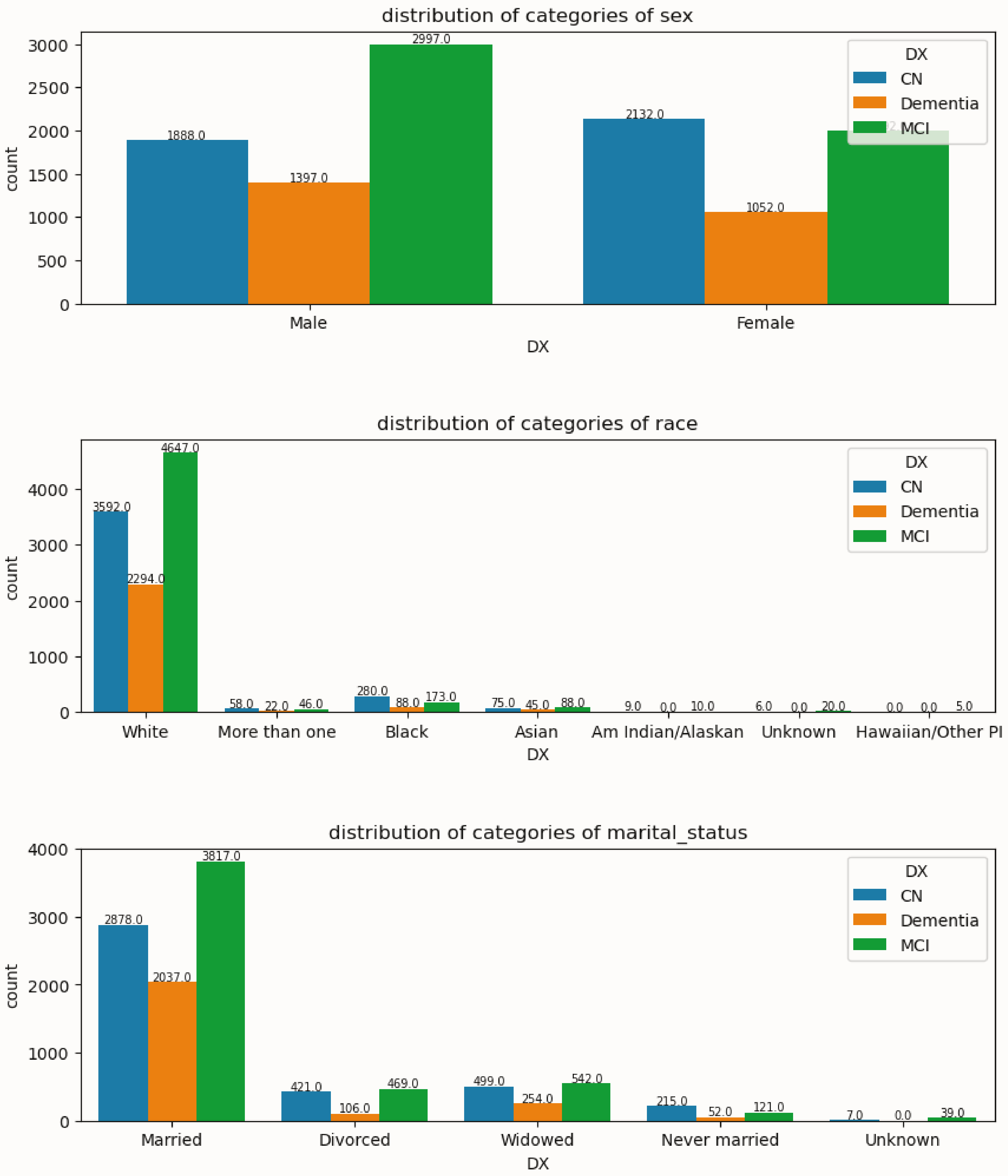
05 Conclusion

06 Future Work



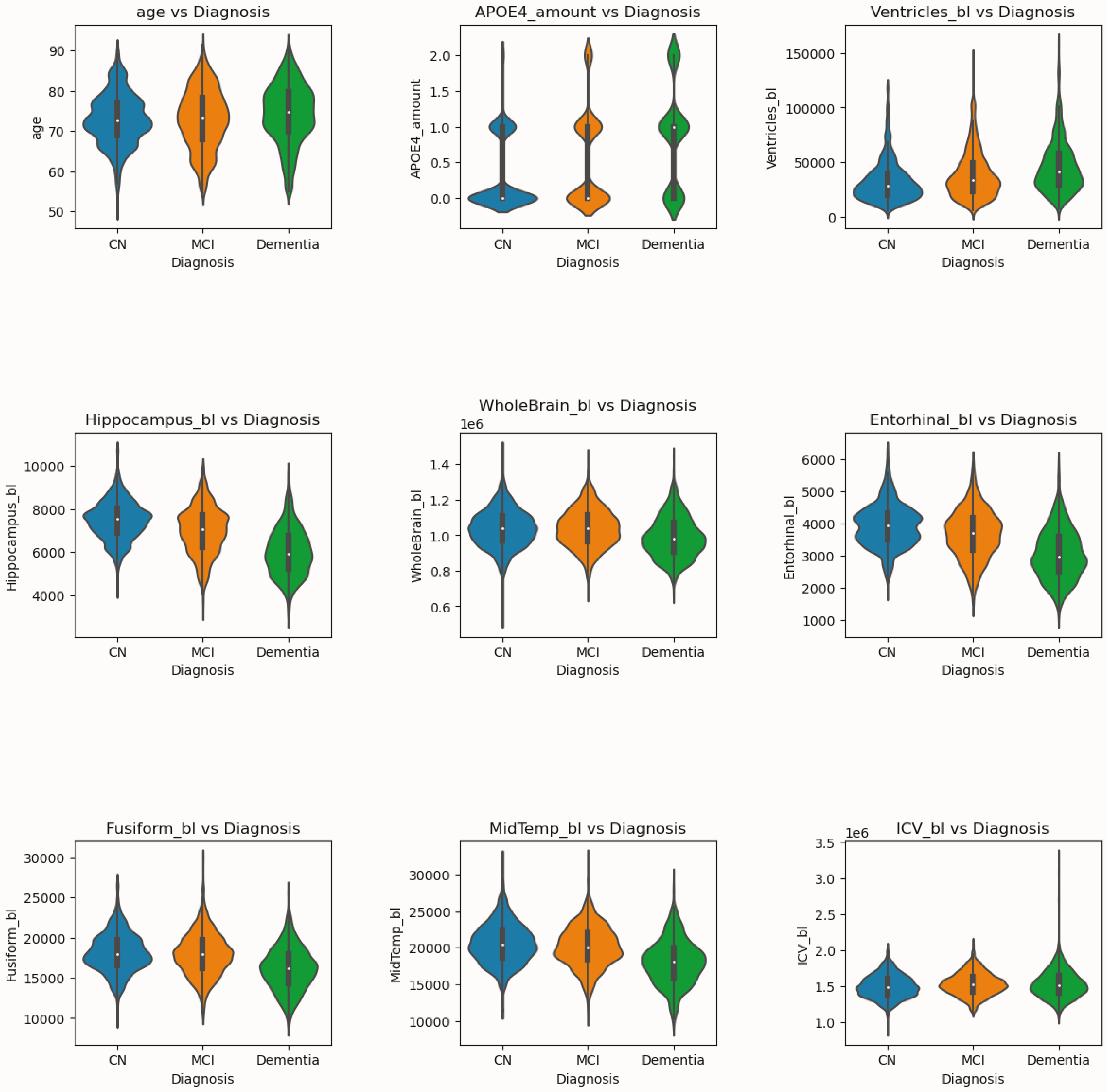
01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

Variable Distribution

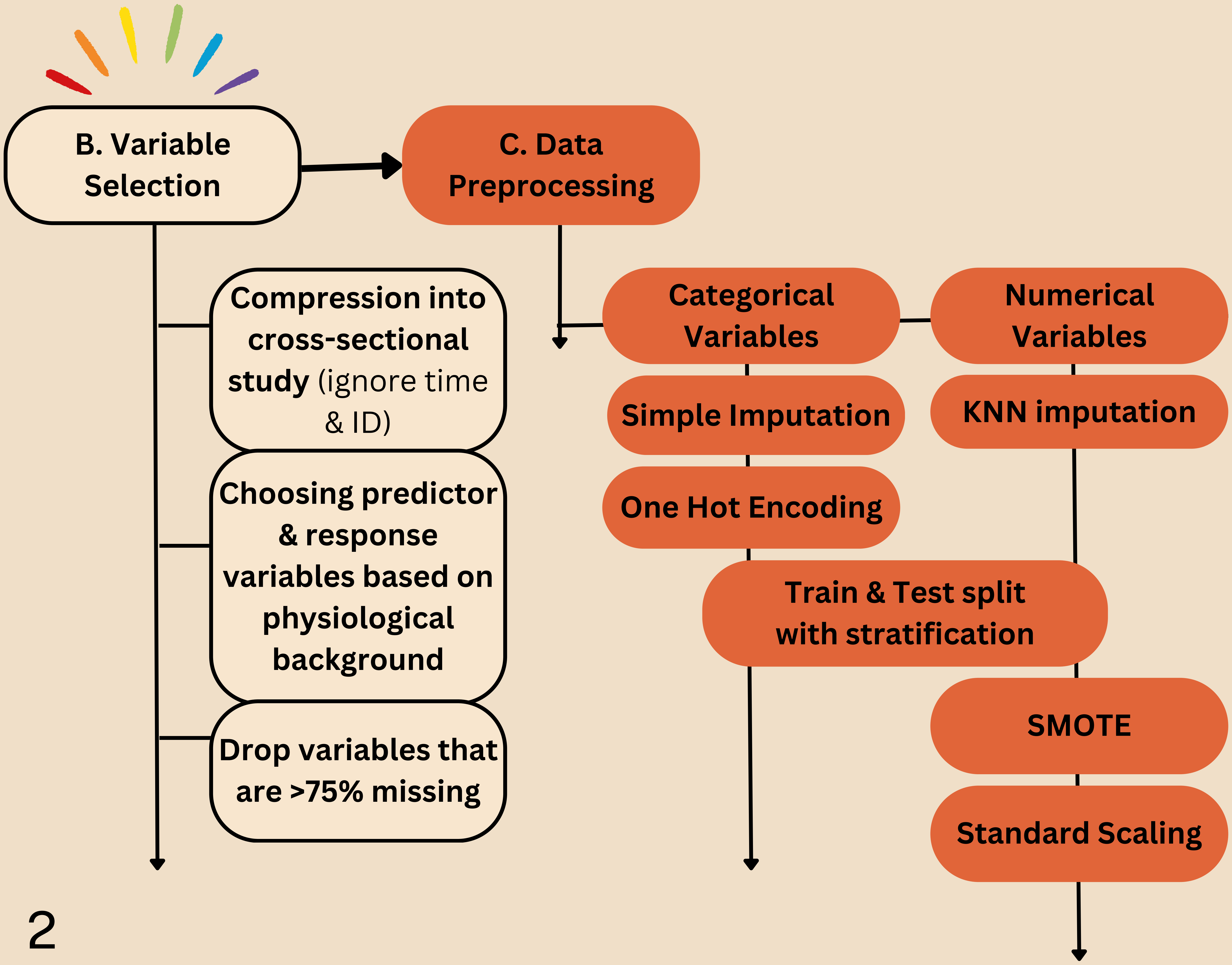


01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

Variable Distribution

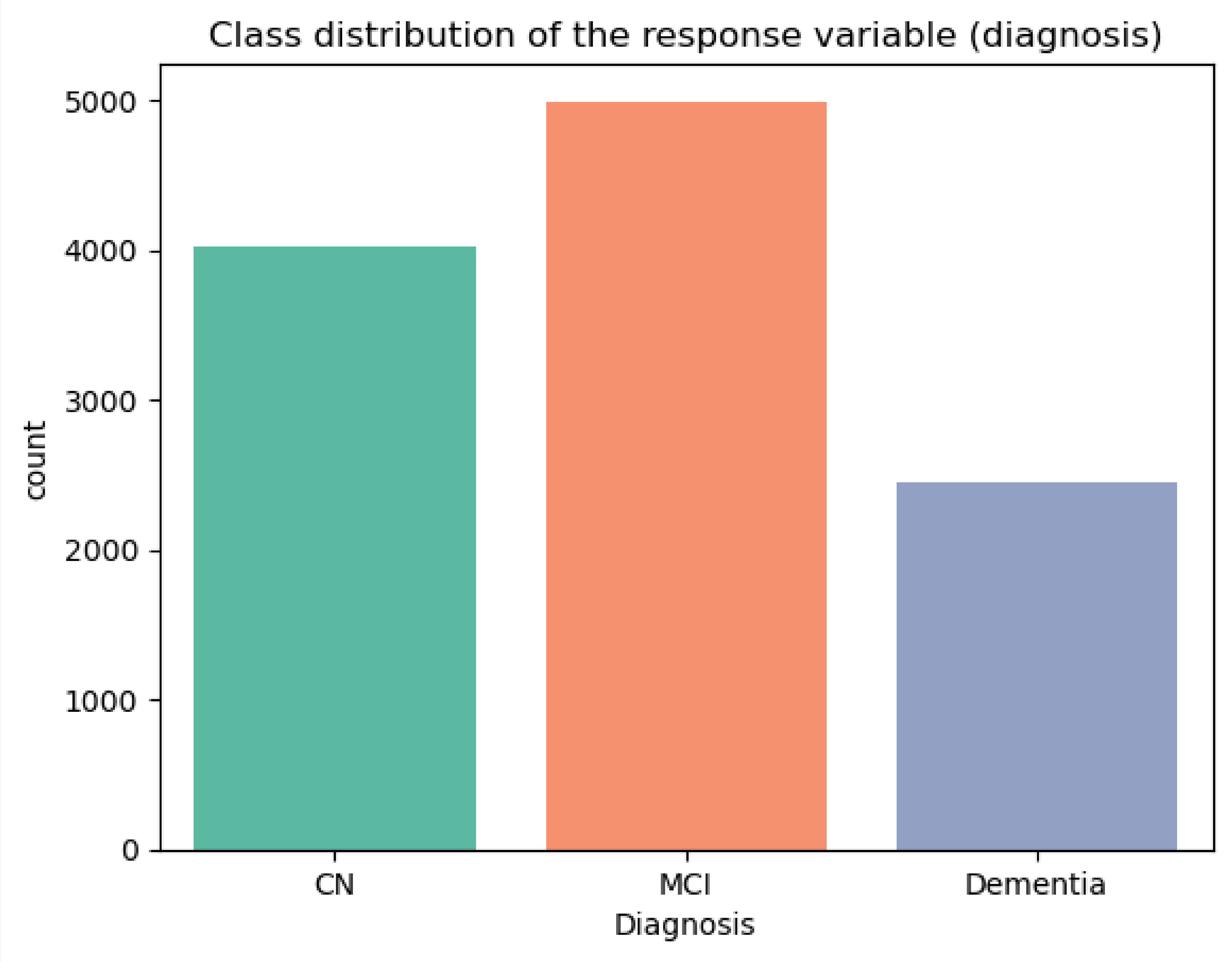


01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

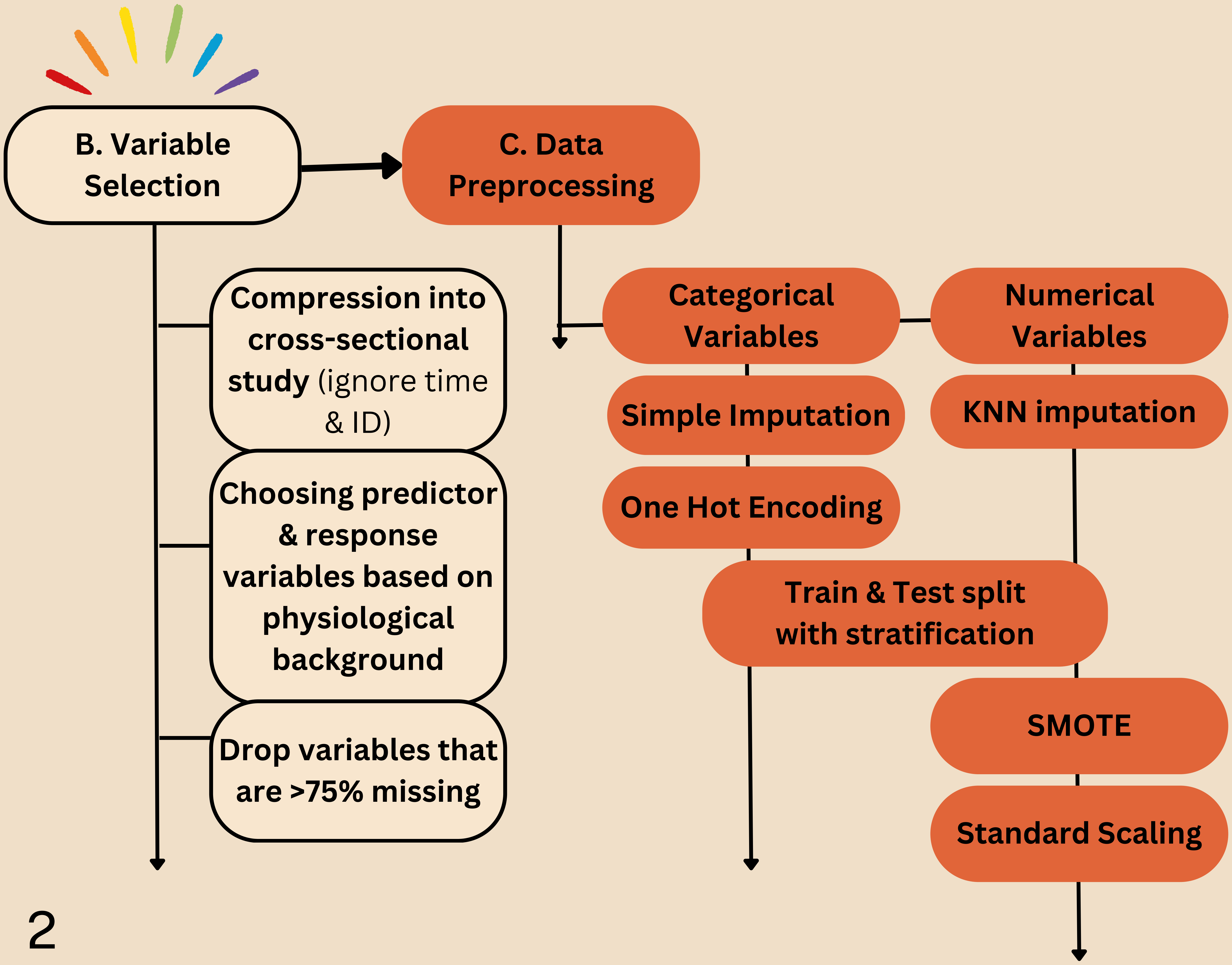


01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

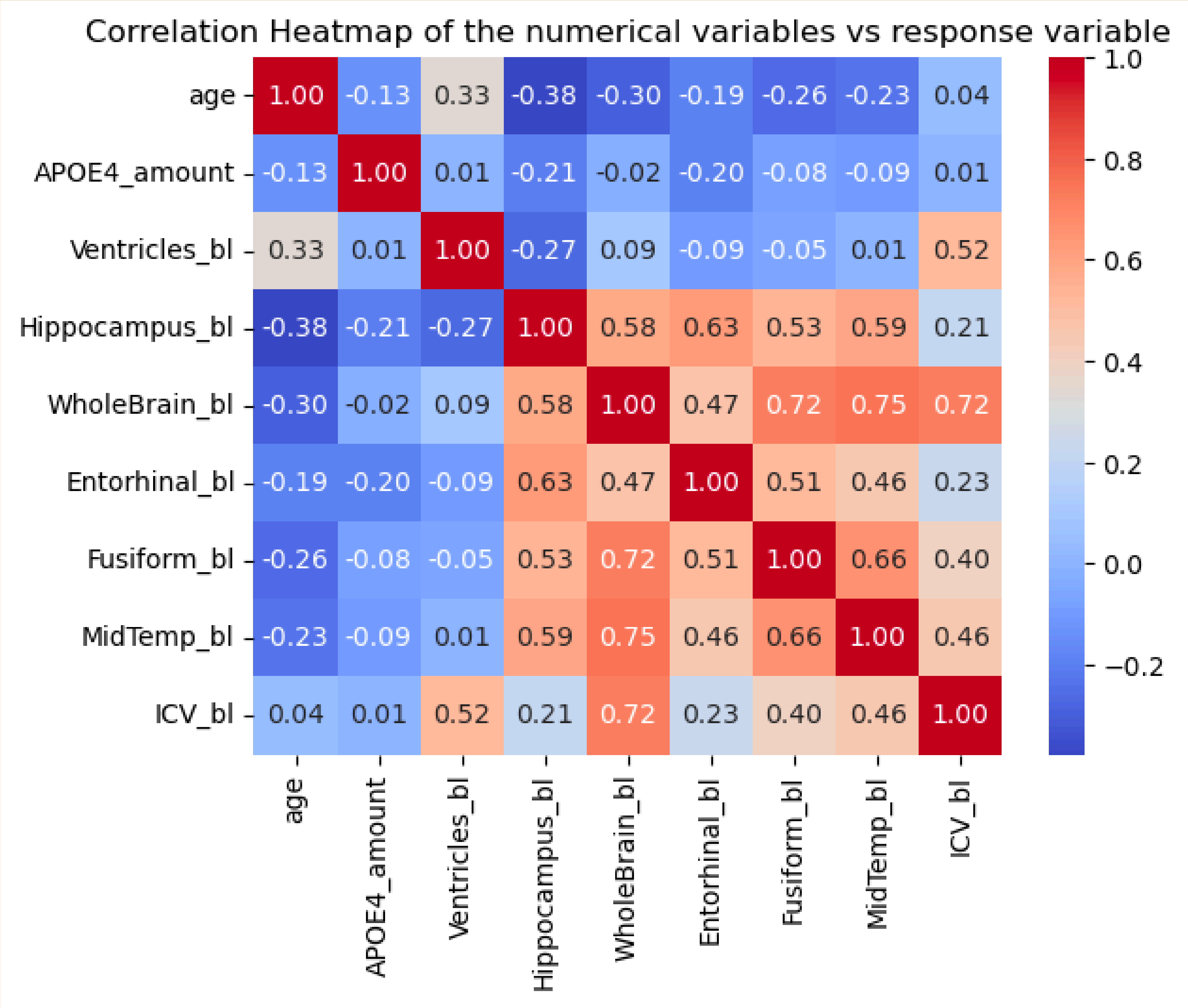
Variable Distribution



01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

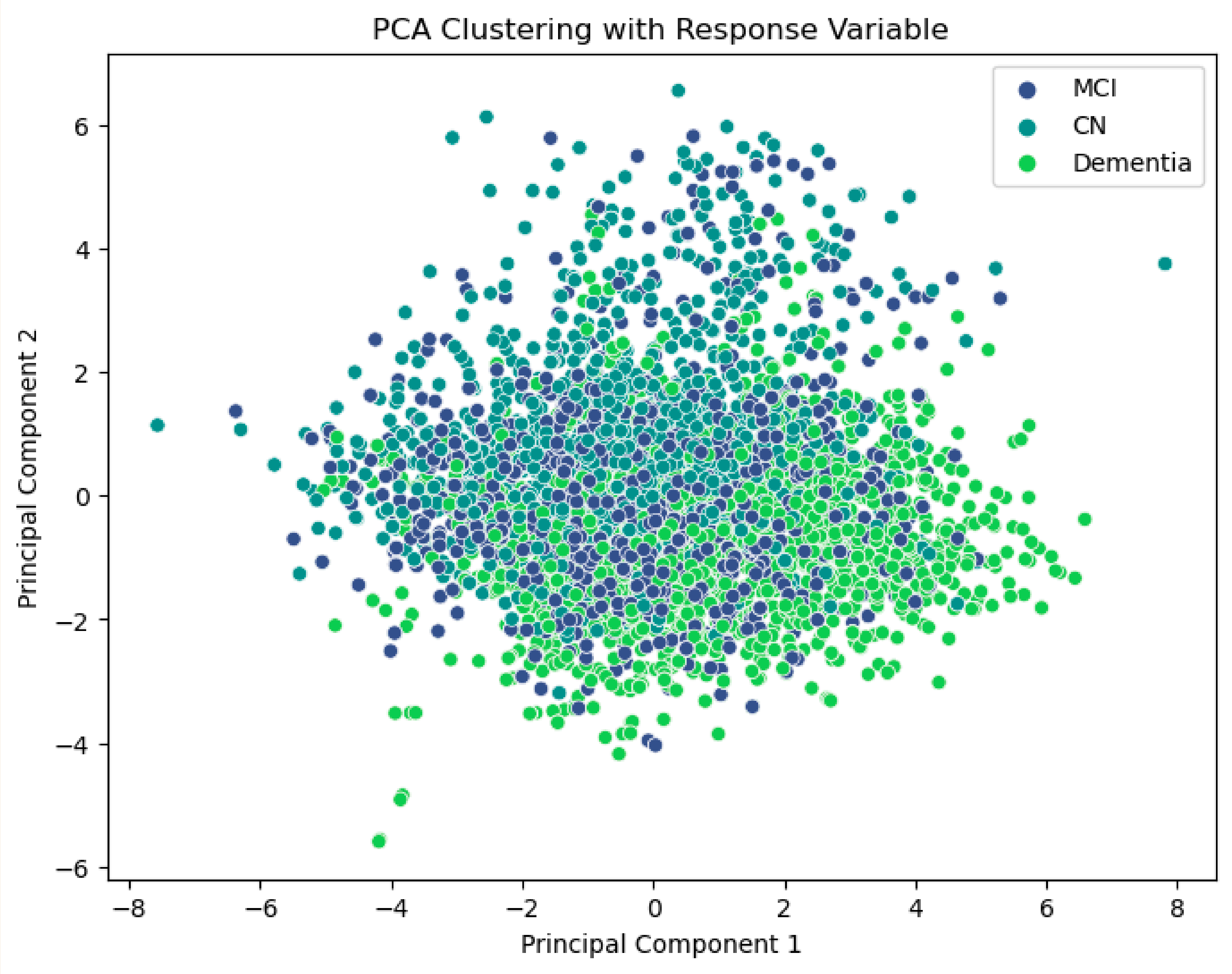


01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work



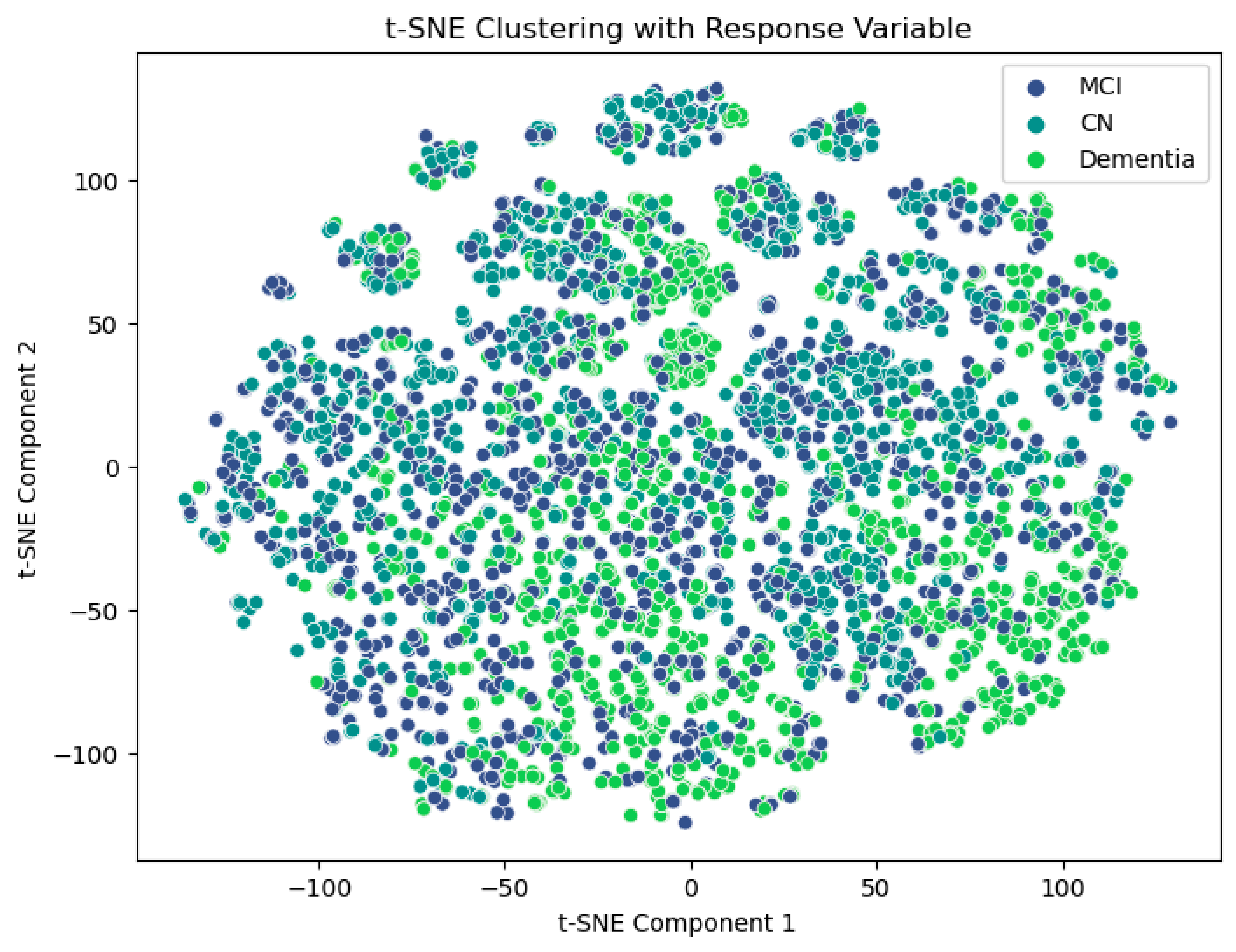
Confounding Variables

01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work



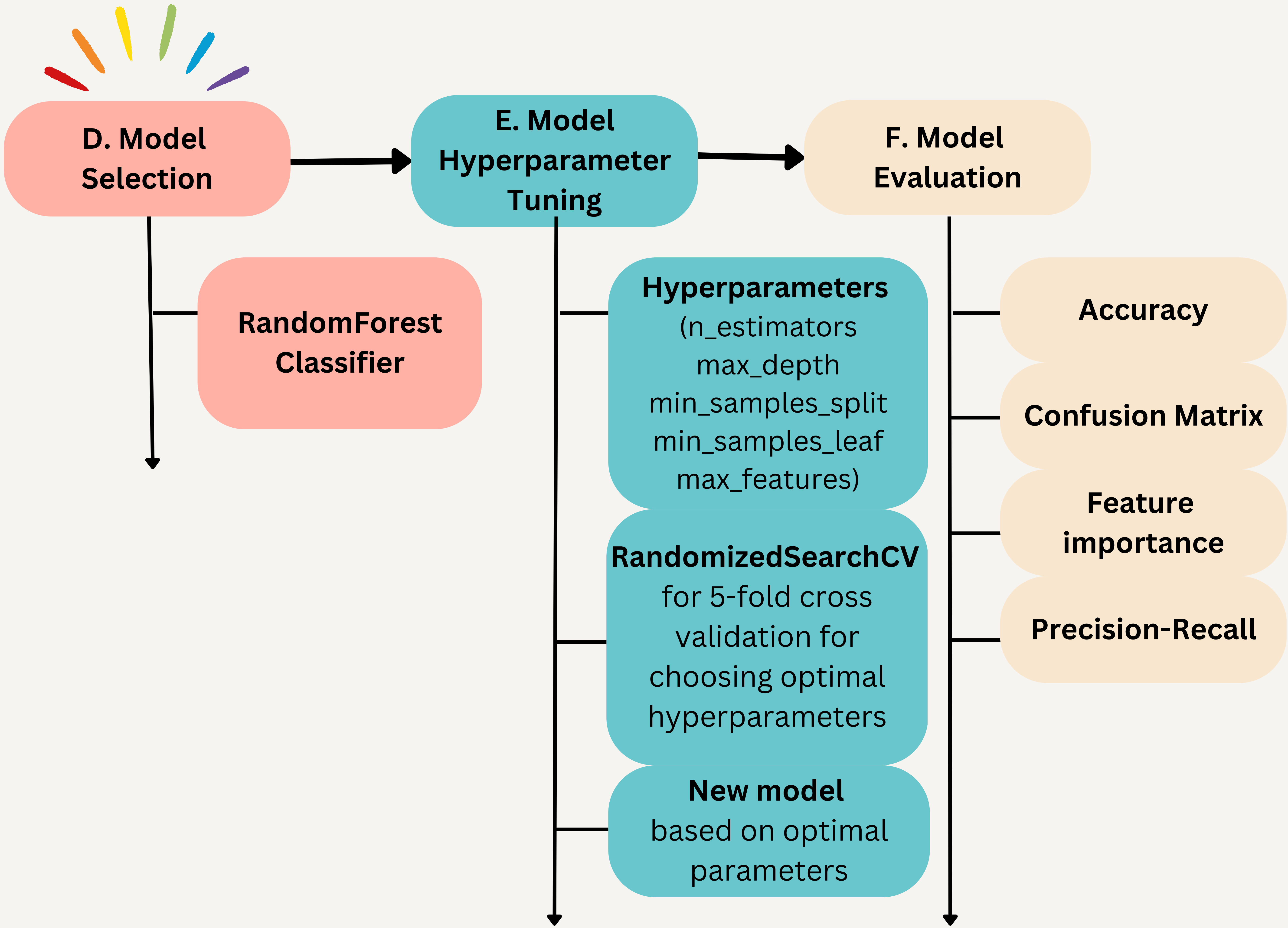
PCA/t-SNE

01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work



PCA / t-SNE

01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work



01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

94 seconds

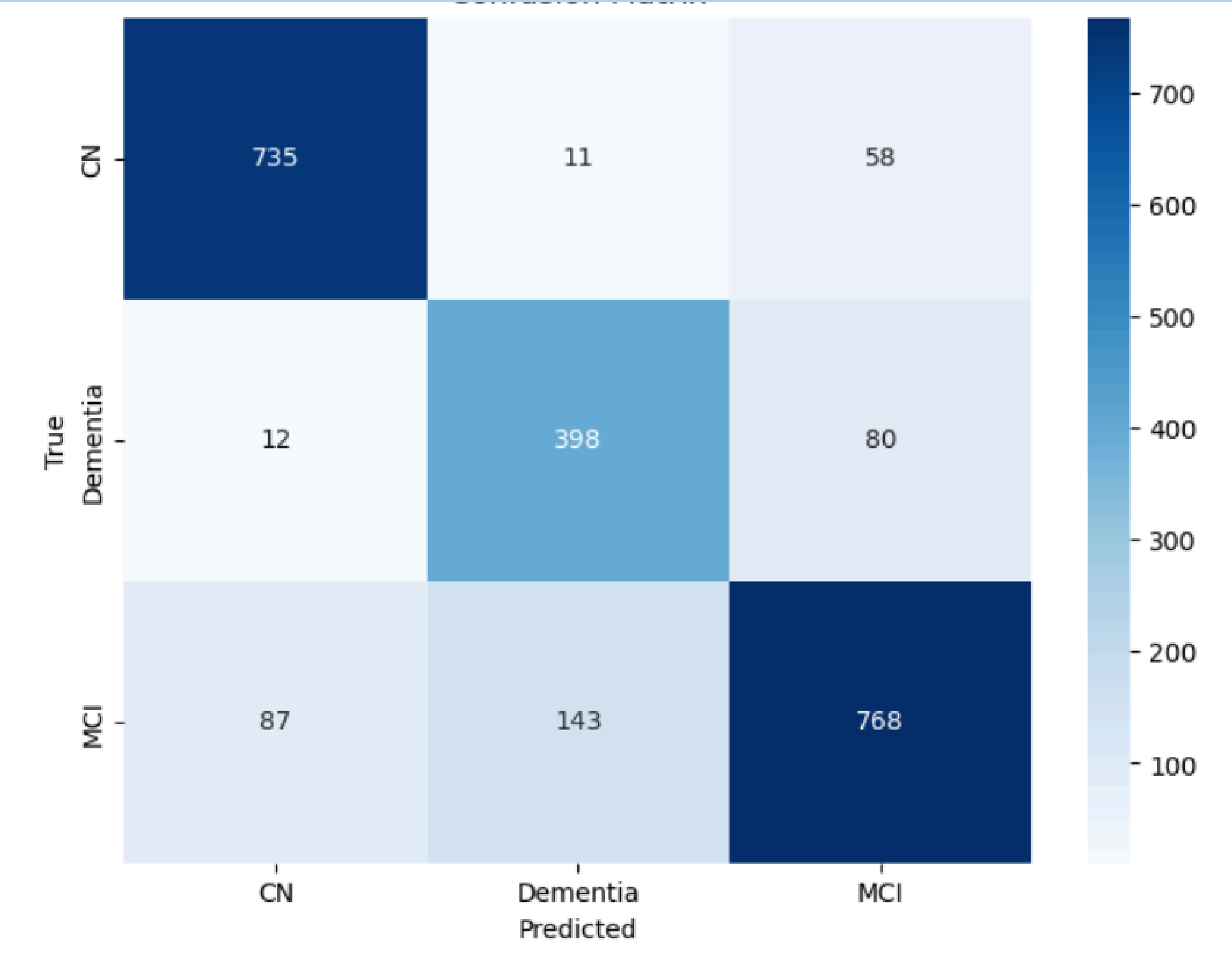
Hyperparameters

n_estimators	50
max_depth	None
min_samples_split	55
min_samples_leaf	1
max_features	sqrt

	F1 Score
Train	0.91
Test	0.83

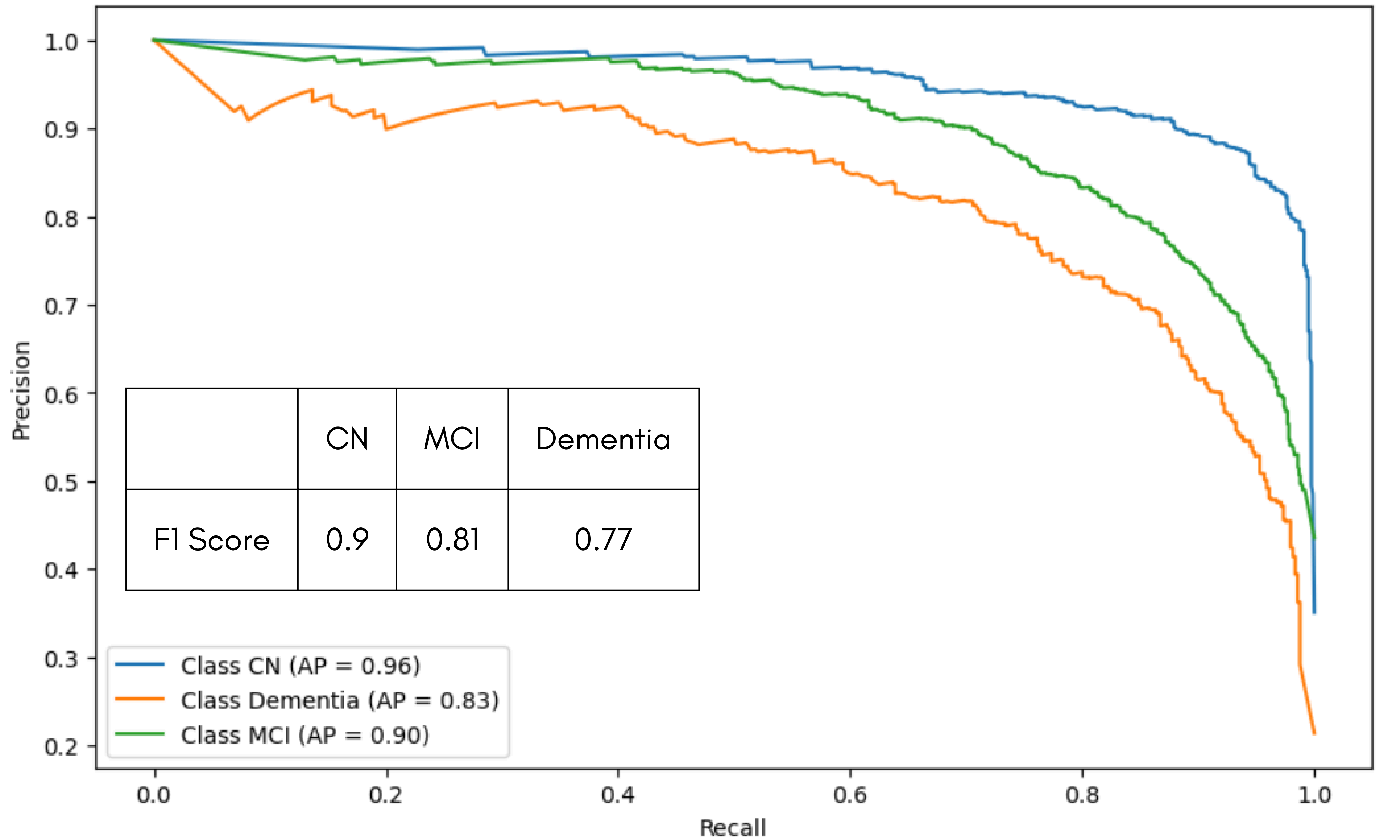
01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

Confusion Matrix



01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

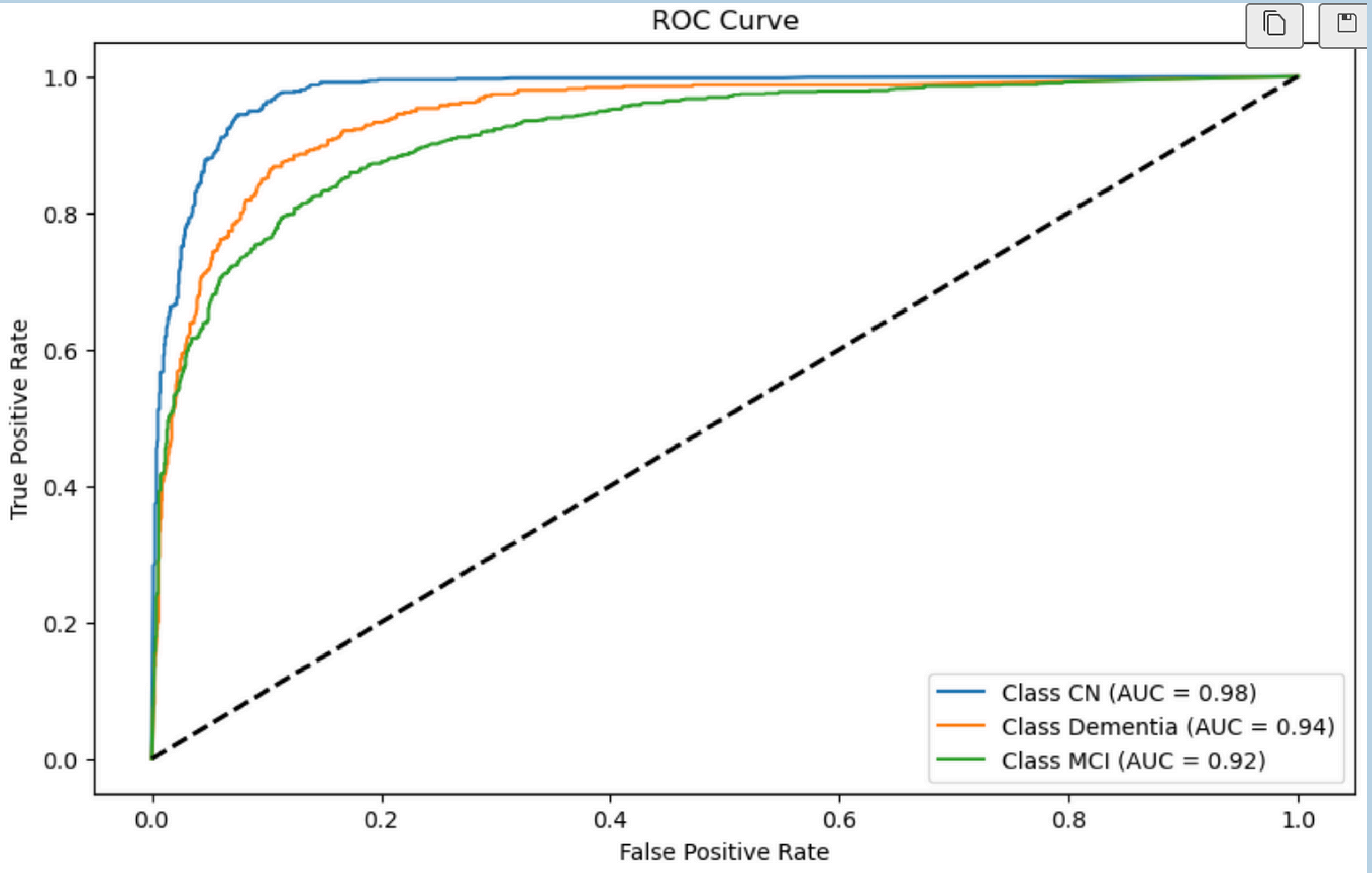
Precision Recall Curve



	CN	MCI	Dementia
F1 Score	0.9	0.81	0.77

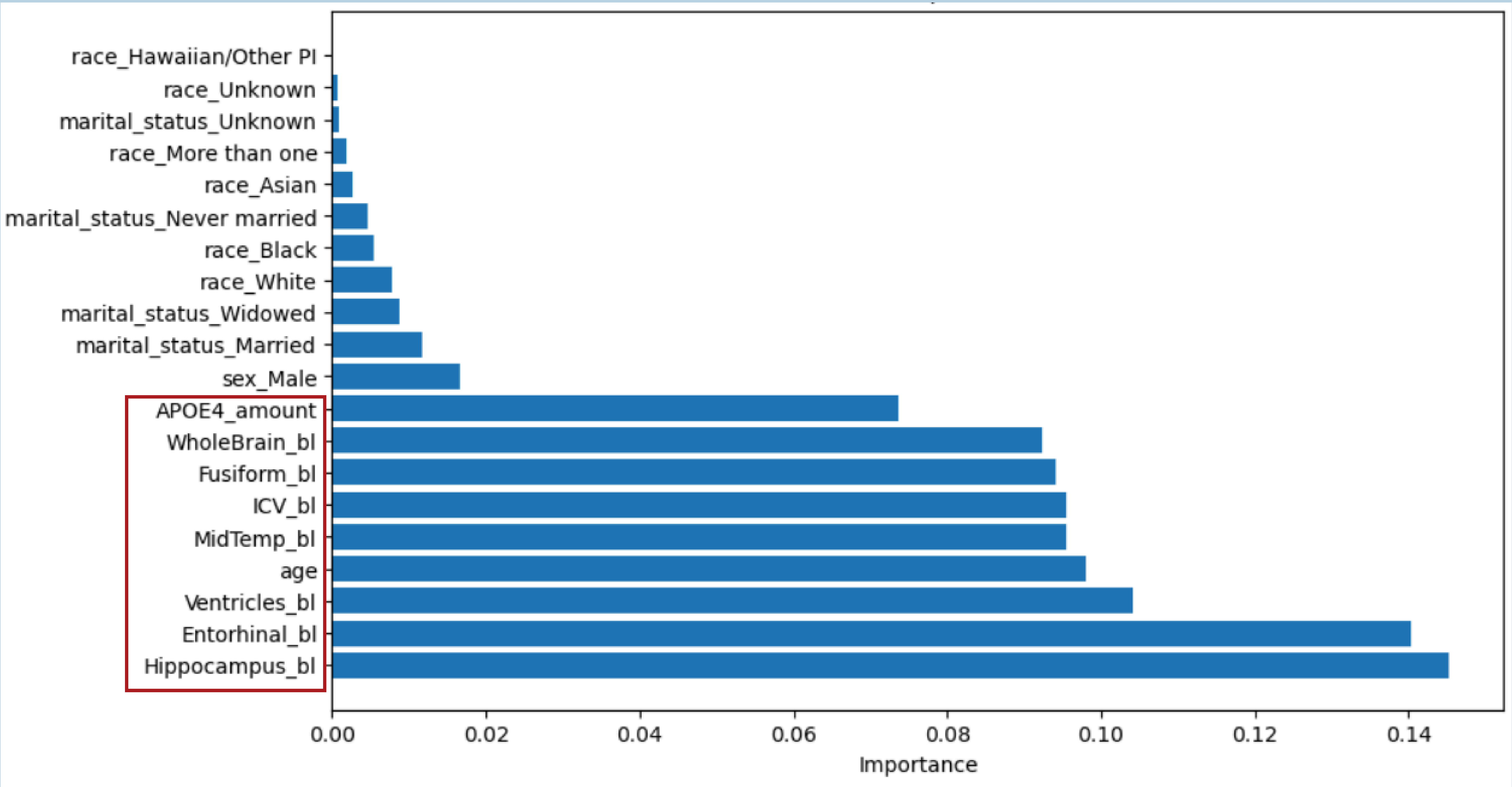
01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

ROC Curve



01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

Feature Importance



01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

Main Takeaways

01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

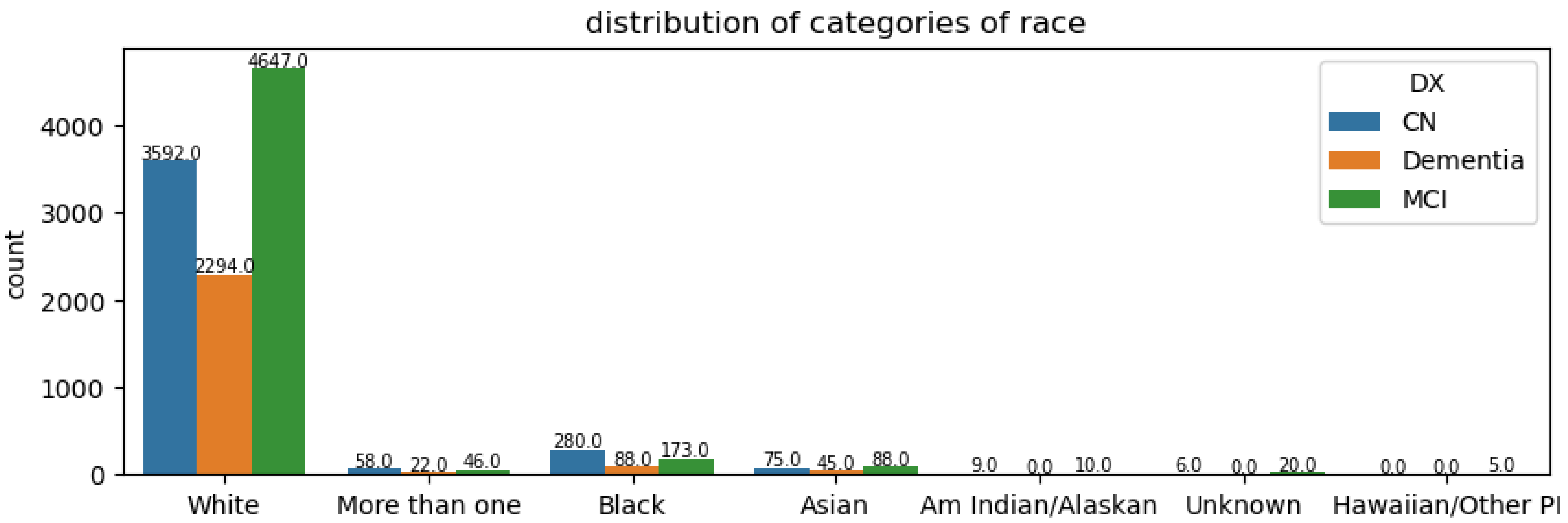
→ Most important features are brain region size, age, and APOE4 expression

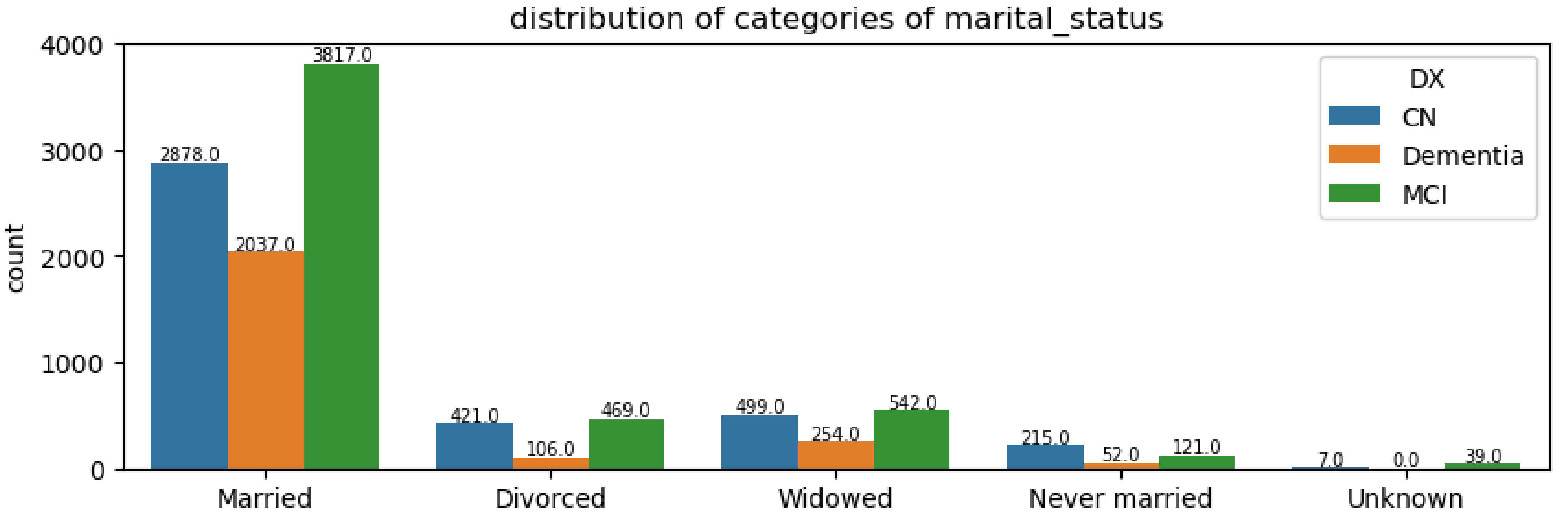
→ Model should include APOE4 gene expression when predicting diagnosis

→ Model had less efficient with balancing precision and recall for dementia (compared to MCI and CN)

Model was a ***very good*** classifier based on dementia ROC AUC score of 0.94

LIMITATIONS





“white” and “married” are both **overrepresented** and have the highest feature importance of their groups.

01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

SAMPLING TECHNIQUES

Cluster-Based Oversampling to **augment data**

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008).
Jo, T., & Japkowicz, N. (2004).

BIAS MITIGATION ALGORITHMS

Pre-processing and in-processing techniques to **reduce bias**, such as reweighting and adversarial debiasing

Kamiran, F., & Calders, T. (2012).
Zhang, B. H., Lemoine, B., & Mitchell, M. (2018).

MODEL INTERPRETATION

Methods like SHAP for model **output explanations**

Lundberg, S. M., & Lee, S.-I. (2017).

01 Problem Statement
02 EDA
03 Modeling, Training
04 Results
05 Conclusion
06 Future Work

References

- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). IEEE. IEEE Xplore.
- Jo, T., & Japkowicz, N. (2004). Class imbalances versus small disjuncts. ACM SIGKDD Explorations Newsletter, 6(1), 40-49. ACM Digital Library.
- Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems, 33(1), 1-33. Springer Link.
- Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. ACM. ACM Digital Library.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874. arXiv.

THANK YOU