How You Split Matters: Data Leakage and Subject Characteristic Studies in Longitudinal Brain MRI Analysis

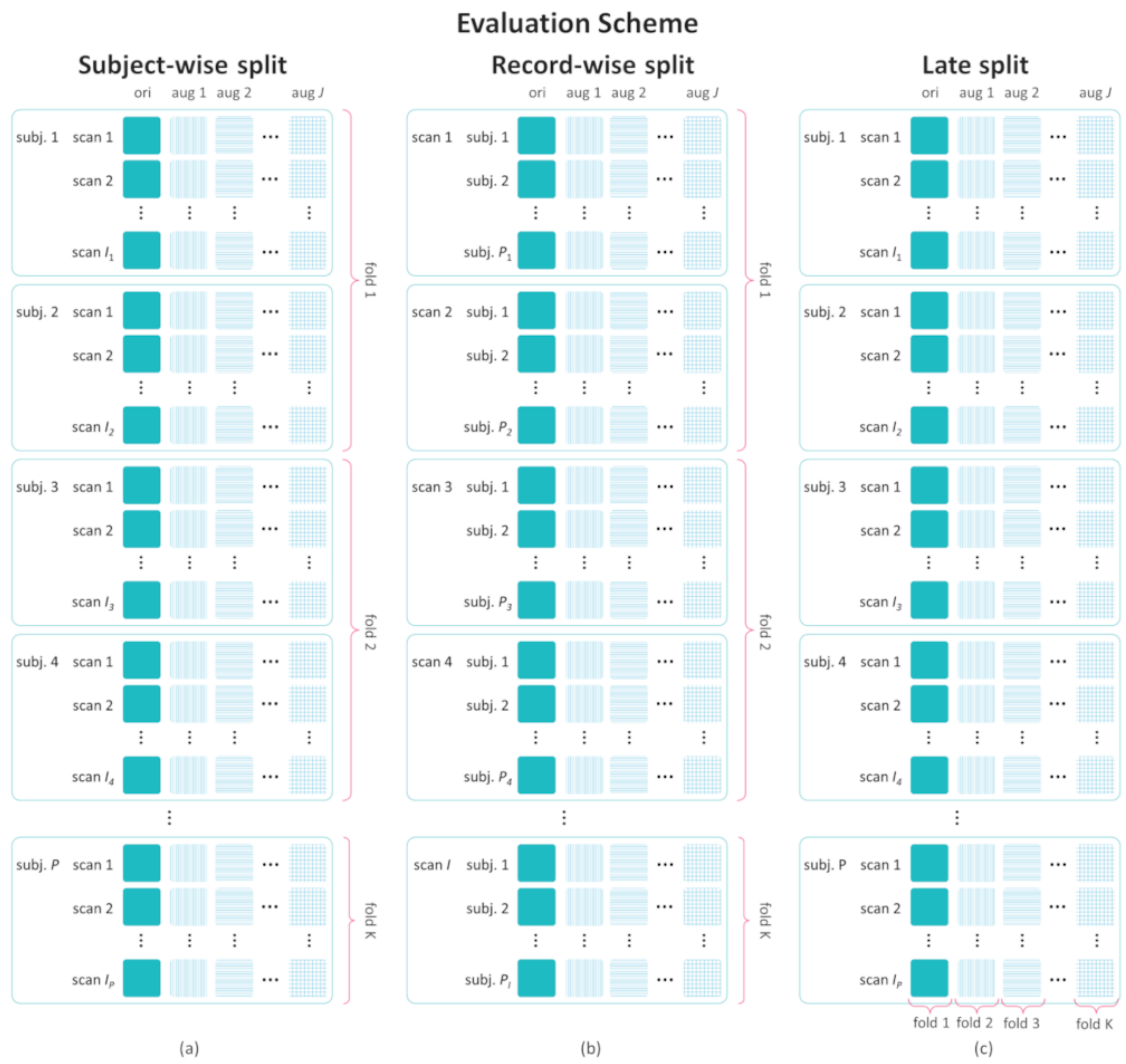
Dewinda J. Rumala

Introduction

- Longitudinal data offers critical insights into disease progression and treatment efficacy
- Improper handling of longitudinal data poses issues even in 3D-based medical image analysis
- The reliability of deep learning models can be jeopardized by biases, such as data leakage

Methods

- 3D CNN for Alzheimer's Disease (AD) diagnosis with longitudinal brain MRI data from ADNI



Discussion

- **How You Split Matters**
  The choice of data splitting strategy during CV significantly influences the performance of AI models
- **Data Leakage and Identity Confounding**
  Improper data splitting can lead to data leakage, affecting model generalization and causing identity confounding within the models
- **Shortcut Learning Revealed by GradCAM**
  GradCAM visualization highlights potential shortcut learning in models from record-wise and late splitting strategies possibly due to identity confounding
- **Validating Robustness with Subject-Wise Split**
  This study validates previous findings suggesting subject-wise split as a less data leakage-prone approach
- **Future Directions**
  Promoting Subject-Wise Split: future research should consider subject-wise split for more reliable model evaluation and development
  Investigating Data Variance and Sensitive Attributes: Further research should delve into the correlation between data splitting strategies and data variance

Data	Scheme	Acc	Prec	Rec	F1-score
Cross-Validation	Subject-wise	67.11±6.11	69.38±6.02	67.11±6.12	68.28±5.63
	Record-wise	97.33±1.86	97.54±1.66	97.33±1.86	97.34±1.85
	Late split	81.33±12.37	89.45±8.31	79.31±13.29	89.44±77.64
Hold-out	Subject-wise	42.15±5.45	38.71±7.54	42.12±5.50	38.57±4.99
	Record-wise	38.71±7.75	37.48±9.20	38.63±7.72	35.68±7.37
	Late split	40.43±8.95	37.62±13.31	40.43±8.95	39.92±4.80

Results

Data Splitting Strategy Impact

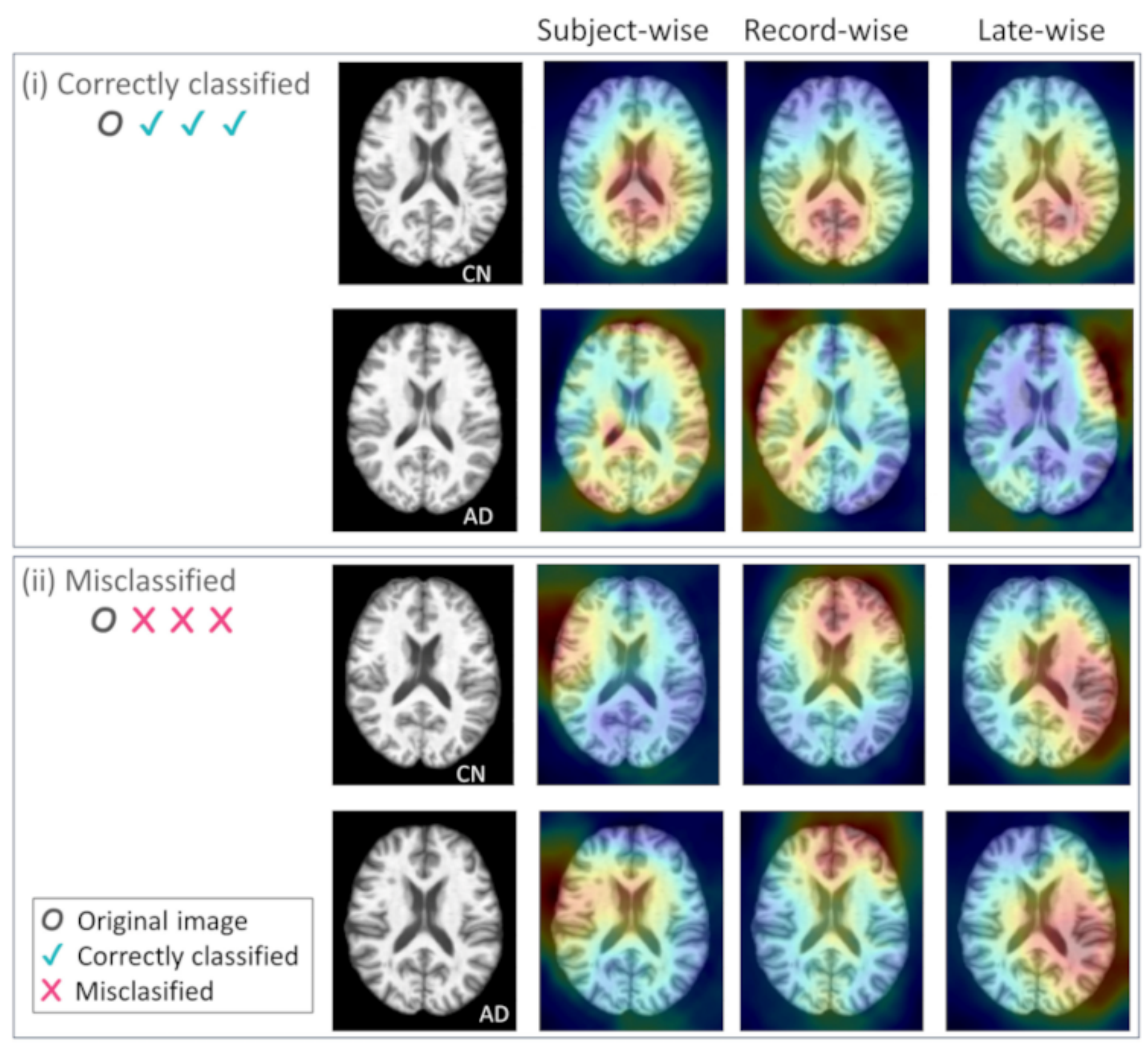
- Record-wise split excels during CV, closely followed by late split, but performs the worst on hold-out data
- Subject-wise split performs poorest during CV but generalizes best to hold-out data
- Data splitting strategy influences model performance (P=0.0389)

MRI Sequence Influence

- The choice of T1 or T2 MRI sequences has no significant impact on classification performance (P=0.7921)

Insights from GradCAM Visualization

- Shortcut learning was observed in record-wise and late splits



References

1. Neto, E., et al., "Detecting the impact of subject characteristics on machine learning-based diagnostic applications," npj Digital Medicine 2(1), 2019.
2. Yagis, E., et al., "Effect of data leakage in brain MRI classification using 2D convolutional neural networks," Scientific Reports 11(1), 2021.

Special thanks to:
DGHERT Ministry of Higher Education, Indonesia
Prof. I Ketut Eddy Purnama (Institut Teknologi Sepuluh Nopember, Indonesia)
Prof. Tae-Seong Kim (Kyung Hee University, South Korea)

Contact information:
dewinda.207022@mhs.its.ac.id
djrumala.github.io