Alzheimer's disease

Complexity-Loss-as-a-Biomarker-for-Alzheimer-s-Disease

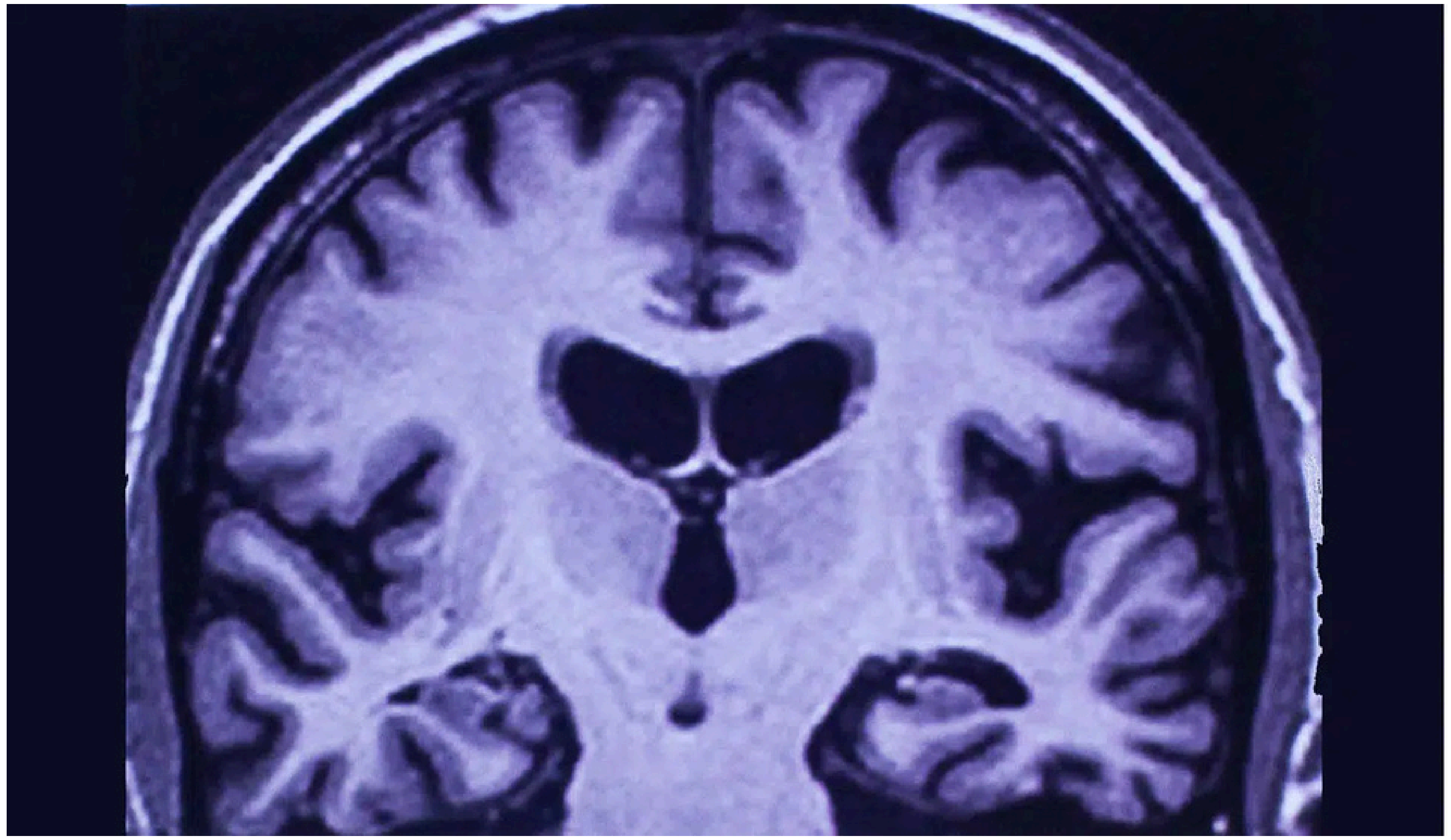
CSC_5AI25_TP

ZAIT Nawel
CLERVILSSON Christelle

Overview

Introduction

1. **Theoritical fundation**
2. **Dataset and methodology**
3. **Classification results**
4. **Regional analysis**
5. **Key contribution and future work**



Conclusion

Introduction

Context:

- Alzheimer's affects 50M people worldwide, causing progressive brain atrophy
- Current diagnosis relies on cognitive tests and expensive PET scans

Challenge:

- Can we detect neurodegeneration through information-theoretic analysis of MRI scans?

Methodology: Use compression algorithms to quantify structural complexity.

1 - Theoretical Foundation

Our approach is grounded in Algorithmic Information Theory

- $K(x) \approx |\text{compressed}(x)| / |x|$
- Practical: Use gzip/bz2/lzma as complexity proxies
- Healthy brain = high complexity (rich, irregular structure)
- Alzheimer's brain = low complexity (atrophy, ventricle enlargement)

2 - Dataset & Methodology

Dataset: 11,519 brain MRI scans (128×128 pixels)

- 4-class problem initially -> Merged to 2-class (Healthy vs Impaired)

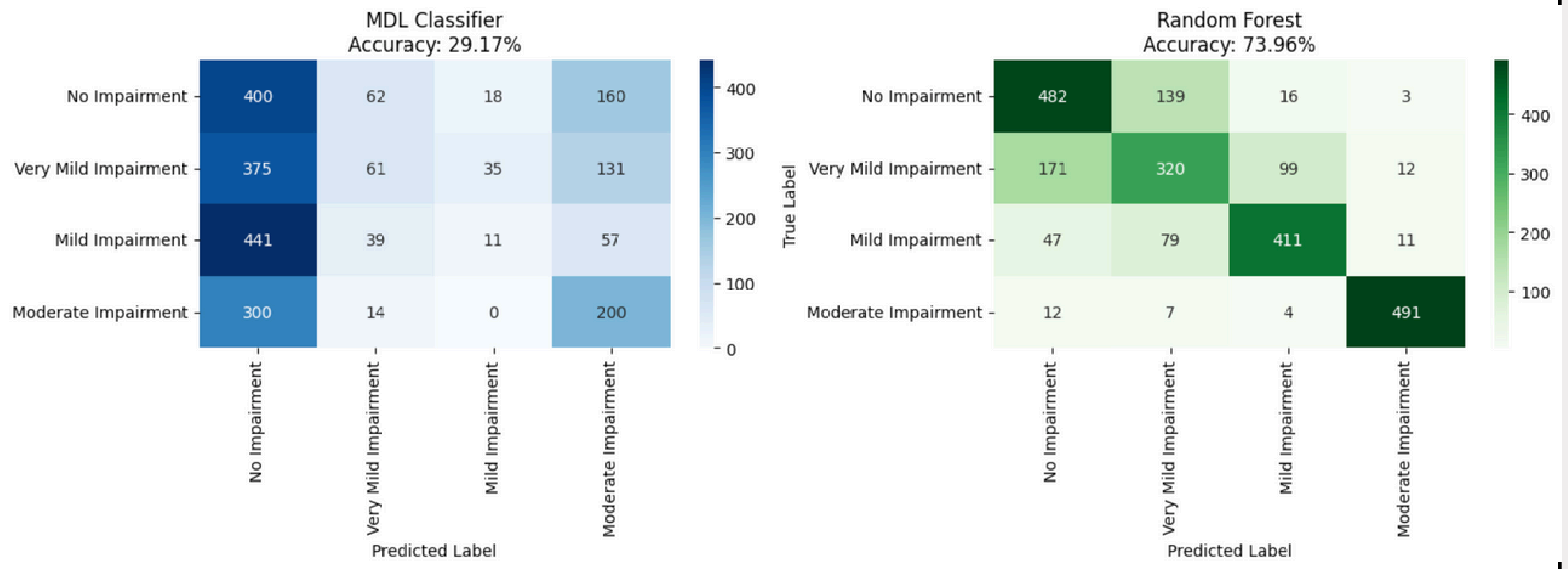
285 features extracted:

- Multi-scale compression (3 scales × 9 features)
- Anatomical regions (hippocampus, cortex, ventricles)
- Grid patches (4×4 + 8×8) with compression + entropy
- Advanced: Wavelet energy, fractal dimension, Laplacian

3 - Classification Results

Class Classification:

- Random Forest: 87.41% (best)
- SVM/LR: ~64-65%
- MDL (AIT): 29.17%
- **MDL (AIT with features): 33.85 %**



Binary Classification (Healthy vs Impaired):

- Random Forest: 93.79% (+6.4%)
- Logistic Regression: 90.41% (+25.8%!)
- SVM: 89.37% (+24.7%)
- **MDL (AIT with features): 59.16% (+30% improvement!)**
- **MDL (with 100 best features) : 66.23% (+37% improvement!)**

3 - Classification Results

Performance Hierarchy:

1. ML on AIT Features (90-94%)
   - ✅ Best accuracy
   - ✅ Leverages non-linear interactions
   - ❌ Black box (limited interpretability)

1. Distance-based AIT (59-69%)
   - ✅ Theoretically grounded
   - ✅ Moderate interpretability
   - ⚠️ Computationally expensive (BNN: $O(n^2)$)

1. Simple AIT (24-29%)
   - ✅ Fully interpretable
   - ✅ No hyperparameters
   - ❌ Too simplistic for complex patterns

4 - Regional Analysis

Regional Complexity Maps:

- Computed 16×16 grid complexity per image
- Averaged across 30 samples per class
- Visualized difference maps (Healthy - Impaired)

Key Findings:

- Complexity loss concentrated in central regions (hippocampus)
- Ventricle areas show paradoxical patterns (uniform expansion)
- Matches known neuroanatomical changes in Alzheimer's

5 - Key Contributions & Future Work

Main Contributions:

a. Validated complexity loss hypothesis ($p < 10^{-6}$, d up to 0.95)
b. Created 285-feature AIT-based representation
c. Achieved 93.79% accuracy (binary classification)
d. Demonstrated synergy between AIT theory and ML practice

Future Directions:

- Test on longitudinal data (track individual progression)
- Integrate clinical metadata (age, APOE genotype, MMSE scores)

Limitations:

- Compression ≠ true Kolmogorov Complexity (halting problem)
- Algorithm-dependent (gzip biased toward certain patterns)
- Dataset imbalance (12 moderate vs 640 healthy in test)

Conclusion

Key takeaways:

- AIT provides a rigorous theoretical framework"
- Compression-based biomarkers are statistically validated and clinically relevant
- The future lies in hybrid approaches combining theoretical foundations with ML power

Thank you !