November, 2025

CSC 5AI25 TP

Algorithmic Information & Artificial Intelligence

Micro-study

teaching.dessalles.fr/FCI

Name: Zait Nawel, Clervilsson Christelle

Alzheimer's disease

Abstract

Alzheimer's disease accounts for 60–70% of dementia cases worldwide. While AI models can detect Alzheimer's from MRI scans, they lack interpretability. We apply Algorithmic Information Theory (AIT) to test a novel hypothesis: that Alzheimer's-affected brains exhibit lower complexity than healthy ones. Using multi-stage MRI data, we provide an interpretable framework for disease detection based on information-theoretic measures.

Problem

Alzheimer's disease (AD) is the leading cause of dementia worldwide, affecting millions of people. Early detection is one of the most critical challenges in managing the disease. Structural brain changes, such as hippocampal atrophy and ventricular enlargement, begin years or even decades before clinical symptoms become severe. MRI is an essential diagnostic tool because it captures these morphological changes. However, interpreting subtle structural differences remains extremely difficult, especially in the early stages, which is precisely when intervention is most effective.

Deep Learning has become the dominant computational approach for diagnosing Alzheimer's disease from MRI data. Multiple models have achieved impressive accuracy, sometimes exceeding 90%. Despite this, these models suffer from important limitations. First, deep learning systems operate largely as black boxes, which is problematic in clinical settings where decisions must be transparent and justifiable. Second, these models are highly dependent on large, well-annotated training datasets, which introduces issues of bias,

reproducibility, and generalization. These factors make deep learning models less suitable for direct medical deployment despite their high performance.

Algorithmic Information Theory (AIT) provides a fundamentally different perspective. AIT defines the Kolmogorov complexity of an object as the length of the shortest program that can generate it, offering a universal and objective measure of information content. We introduce the hypothesis that MRI images of patients with Alzheimer's disease are algorithmically simpler than those of healthy individuals. Neurodegeneration causes neuronal loss, tissue homogenization, and structural simplification—changes that could make diseased images more compressible. This hypothesis is biologically grounded, as Alzheimer's progression reduces the amount of structural information in the brain. While Kolmogorov complexity is theoretically uncomputable, practical approximations can be obtained using lossless compression algorithms such as gzip, bz2, and lzma.

This project aims to address the following key questions:

- Can AIT-based complexity measures reliably distinguish between AD patients and healthy controls?
- Which brain regions show the strongest complexity changes during disease progression?
- How do complexity-based models compare to traditional machine learning approaches?

Method

Dataset

We chose a dataset from Kaggle, which has previously enabled Deep Learning models to reach up to 99% accuracy. The dataset consists of 11,519 MRI scans distributed across four classes: No Impairment (3,200 images, 27.8%), Very Mild Impairment (3,008 images, 26.1%), Mild Impairment (2,739 images, 23.8%), and Moderate Impairment (2,572 images, 22.3%). Each class contains both real and synthetically augmented scans. While the distribution is relatively balanced overall, the test set exhibits some class imbalance. This reflects real-world clinical distributions where advanced AD cases are less frequently scanned, but may introduce bias in evaluation metrics, particularly affecting sensitivity for moderate impairment detection.

Complexity measure

We used multiple methods to quantify the complexity of MRI images:

Kolmogorov Complexity (KC)

Formally, KC is the length of the shortest program capable of reproducing an object. Because it is uncomputable, we approximate it using lossless compression algorithms such as gzip, bz2, and lzma. The compressed size provides a practical proxy for the information content of an image. Lower compressed size indicates lower algorithmic complexity.

Conditional Complexity (Normalized Compression Distance, NCD)

NCD measures the informational similarity between two objects. To compute it, images are converted into bytes and compressed using gzip. The NCD formula combines the compressed lengths of the individual images and their concatenation to produce a normalized distance in the range [0, 1]. Low NCD between two images suggests they share similar algorithmic structure, making them informationally similar regardless of pixel-level differences.

Four Class Classification

Pure AIT Classification

We first implemented two classification methods based exclusively on algorithmic information theory, without traditional machine learning optimization:

1.1 - Minimum Description Length (MDL) Classifier

The MDL principle selects the model that minimizes the combined cost of encoding both the model itself and the data given that model. Our implementation proceeds in three steps: first, we compute a prototype for each class by averaging the training images. Then, for each test image, we calculate its NCD to each class prototype. Finally, we assign the test image to the class with the minimum NCD. This approach provides a fully interpretable classifier rooted in algorithmic information theory, as classification decisions are based purely on informational similarity without learned parameters.

1.2 - Binary Neural Network-inspired classifier (BNN)

This classifier is similar to MDL but does not compute class prototypes. Instead, each test image is compared to all training images using NCD, and the label of the closest image (minimum NCD) is assigned. While conceptually straightforward, the computational cost is extremely high ($O(n^2)$), making it impractical for large datasets. In practice, the BNN was not fully executed due to excessive runtime.

AIT-based Features

To evaluate how Algorithmic Information Theory (AIT) could improve classical machine learning models, we first needed to extract meaningful features from MRI images. The goal was to convert each MRI into a structured feature vector that encodes algorithmic complexity, local variability, and anatomical information.

Feature Extraction Pipeline

We first developed a rich feature extraction function composed of several elements:

Multi-scale Compression: Images were resized to 0.25×, 0.5×, and 1× of their original size. For each scale, we calculated the compression-based complexity (KC) using lossless compressors. This captures how image complexity changes at different spatial resolutions, as coarser scales may reveal global patterns while finer scales capture local details.

Approximation of Anatomical Regions: We focused on biologically relevant regions known to be affected by Alzheimer's, such as the hippocampus and cortex. These regions were approximated by cropping fixed subregions of the MRI. Compression ratios were calculated for each region, capturing region-specific complexity loss that may correlate with known patterns of neurodegeneration.

Local Variability: The image was divided into small patches (32×32 pixels). For each patch, compression was computed. We then calculated the mean and standard deviation of patch complexities, reflecting heterogeneity across the brain. High variability suggests diverse tissue types, while low variability may indicate homogenization due to atrophy.

Gradient Complexity: We applied a Sobel filter to capture edges and structural variations. The compression ratio of the edge map quantifies structural irregularity, as sharp boundaries between tissue types produce complex edge patterns.

Shannon Entropy: Calculated on the full image to capture the overall randomness of pixel intensity distribution. While Shannon entropy measures statistical randomness, it complements Kolmogorov complexity by providing a probability-theoretic perspective.

Multi-compressor Ensemble: We used gzip, bz2, and lzma to calculate compression ratios of the full image. Different compressors capture different statistical regularities, enhancing robustness to algorithm-specific biases.

We then used ML classifiers like SVMs, Random Forests, Gradient Boosting, and Logistic Regression to test the results obtained with these features. Initial results were not satisfactory, prompting further refinement.

Improved Feature Extraction

To improve classification performance, we developed a richer feature extraction pipeline that generates approximately 285 features per image. Compared to the previous version, which only considered global compression, a few anatomical regions, and basic local variability, the new extraction captures multi-scale, multi-region, and multi-metric information. Specifically, for each image scale (0.25×, 0.5×, 1×), we compute compression, Shannon entropy, and edge complexity. We extract more precise anatomical regions (hippocampus, frontal cortex, temporal cortex, parietal cortex, and ventricles) rather than just hippocampus and cortex. Instead of simple local patch statistics, we divide the image into multiple grid sizes (4×4 and 8×8) and compute compression, entropy, and edge complexity for each patch. Overall, this richer representation captures both global and localized structural variations, providing more discriminative features for machine learning models and increasing the potential to detect subtle changes associated with Alzheimer's disease.

Features Classification

Since the results with the pure AIT approach were not fully satisfactory, we implemented a feature-based classification pipeline. The underlying idea remains the same as pure AIT: we aim to compare new images to reference patterns. However, instead of computing conditional complexity directly, we first extract rich, informative features from each image and then perform classification using standard distance measures in feature space.

Binary Classification

We reformulated the original four-class problem into a binary classification task (Healthy vs. Impaired) for several converging reasons. Preliminary analysis revealed that AIT-based complexity measures produced clear separation between healthy and impaired brains, but consecutive impairment stages (Very Mild, Mild, Moderate) showed substantially smaller complexity differences with significant distributional overlap. This reflects the gradual nature

of neurodegeneration, where structural simplification occurs most dramatically in the transition from healthy to diseased states, while subsequent progression involves more subtle quantitative changes. Additionally, the test set's class imbalance particularly affected severe stages, introducing evaluation bias. By aggregating all impaired classes (Mild + Moderate) into a single category against healthy controls (No Impairment + Very Mild), we maximize signal-to-noise ratio, improve statistical power, and address the most clinically critical question: detecting the presence of Alzheimer's disease, which is the actionable threshold for early intervention, regardless of precise staging. The final binary split resulted in 6,208 healthy images (53.9%) and 5,311 impaired images (46.1%), providing a reasonably balanced dataset.

For the binary classification experiments, we focused primarily on the MDL classifier as it yielded the best results among pure AIT methods, while also benchmarking against classical machine learning approaches to establish performance baselines.

Results

Pure AIT Classification (Four Classes)

The pure AIT methods performed poorly on the four-class problem. The MDL classifier achieved only 34.2% accuracy, with precision of 36.5%, recall of 34.2%, and F1-score of 33.8%. This poor performance demonstrates that raw complexity measures, while theoretically grounded, lack the discriminative power needed for fine-grained classification when disease stages show overlapping complexity distributions.

Feature Classification with Initial Feature Extraction (Four Classes)

Classifier	Accuracy	Precision	Recall	F1-score
Logistic Regression	0.65	0.64	0.66	0.65
Random Forest	0.74	0.74	0.74	0.74
Gradient Boosting	0.68	0.68	0.68	0.68
SVM	0.65	0.65	0.65	0.65

The initial feature extraction pipeline showed substantial improvement over pure AIT methods, with Random Forest achieving the best performance at 74% accuracy. However, these results remained below clinical requirements.

MDL Classifier with Updated Features (Four Classes)

After implementing the enriched 285-feature extraction pipeline, the MDL classifier improved significantly to 55.9% accuracy (precision: 57.6%, recall: 55.9%, F1-score: 55.5%). While

still modest, this represented a 21.7 percentage point improvement over the basic MDL approach, demonstrating that richer feature representations enhance AIT-based methods.

Binary Classification Results

Model	Features	Accuracy	Precision	Recall	F1-Score
MDL	All	57.60%	59.31%	57.60%	57.16%
MDL	Top 100	65.54%	67.02%	65.54%	65.39%
Logistic Regression	All	90.58%	90.94%	90.58%	90.52%
Logistic Regression	Top 100	89.15%	89.41%	89.15%	89.09%
Random Forest	All	94.05%	94.63%	94.05%	94.01%
Random Forest	Top 100	93.88%	94.43%	93.88%	93.83%
SVM	All	89.89%	90.73%	89.89%	89.77%
SVM	Top 100	96.18%	96.23%	96.18%	96.17%

Cross-Validation Results (5-Fold):

- Random Forest: 92.77% ± 0.76%
- Logistic Regression: 89.91% ± 0.47%
- SVM: 88.74% ± 0.76%

The binary classification task yielded dramatically improved results. SVM with feature selection achieved the highest test accuracy at 96.18%, while Random Forest demonstrated the most stable cross-validation performance. Feature selection using SelectKBest (top 100 features) had mixed effects: MDL benefited substantially (+7.94%), and SVM improved significantly (+6.29%), while Random Forest and Logistic Regression showed slight decreases, suggesting they were already utilizing the full feature space effectively.

Complexity Analysis

Compression ratio analysis across disease stages revealed unexpected patterns:

Four-Class Analysis:

- No Impairment: 0.5746 ± 0.0208
- Very Mild Impairment: 0.5823 ± 0.0251
- Mild Impairment: 0.5947 ± 0.0210
- Moderate Impairment: 0.5879 ± 0.0182

Binary Analysis:

- Healthy (No + Very Mild): 0.5783 ± 0.0233
- Impaired (Mild + Moderate): 0.5914 ± 0.0200

Contrary to our initial hypothesis, impaired brains showed higher compression ratios (lower complexity) than healthy brains, though the difference was modest (0.0131). Importantly, the progression across four stages was non-monotonic, with Moderate Impairment showing lower compression than Mild Impairment. This suggests that the relationship between neurodegeneration and algorithmic complexity is more nuanced than initially anticipated.

Discussion

At the beginning of the project, our objective was to evaluate whether Algorithmic Information Theory could extract meaningful structural differences between MRI images of healthy and impaired individuals. Our working hypothesis was that Alzheimer-affected brains should appear algorithmically simpler than healthy brains, leading to measurable reductions in compressibility-based complexity. The biological rationale was straightforward: neurodegeneration causes tissue loss, ventricular enlargement, and structural homogenization, all of which should reduce the information content encoded in brain images.

Our results provided partial validation of this hypothesis. The binary complexity analysis showed that impaired brains indeed exhibited higher compression ratios (0.5914) compared to healthy brains (0.5783), indicating lower algorithmic complexity. However, this difference was relatively small (0.0131), and the four-class analysis revealed non-monotonic progression, with complexity not decreasing uniformly across disease stages. This suggests that the relationship between neurological deterioration and algorithmic simplicity is more complex than a simple linear decline. Possible explanations include compensatory mechanisms in early stages, heterogeneity in disease progression patterns, or artifacts introduced by synthetic data augmentation in the dataset.

When used as the basis for classification, pure AIT models performed far below expectations. The MDL classifier achieved only 34.2% accuracy in four-class classification and 57.6% in binary classification, demonstrating that raw complexity measures lack sufficient discriminative power for clinical deployment. This poor performance can be attributed to the fact that complexity differences between classes, while statistically significant, are too subtle to serve as reliable decision boundaries when used in isolation.

In stark contrast, after designing a rich 285-feature AIT-inspired representation, classical machine learning approaches performed dramatically better. Random Forest achieved 94.05% accuracy in binary classification, while SVM with feature selection reached 96.18%. Cross-validation results confirmed the robustness of these models, with Random Forest showing remarkable stability at 92.77% ± 0.76%. This demonstrates that while raw complexity metrics are insufficient as standalone classifiers, they become highly effective when transformed into engineered features capturing multi-scale, multi-region, and multi-metric information. The synergy between AIT-derived features and traditional machine learning methods proved far more powerful than either approach alone.

Despite these encouraging results, the approach presents several important limitations that must be acknowledged. First, compression algorithms only approximate Kolmogorov complexity and remain bound by algorithm-dependent biases. For instance, gzip favors repeated byte patterns and may not capture structural regularities that other compressors would detect. This means our complexity measures are implementation-dependent rather than universal. Second, compression metrics proved insufficiently discriminative across disease stages. The small differences in compression ratios between classes, combined with substantial variance within each class, limited the separability achievable through methods like NCD and MDL. Third, applying compression-based distances to high-dimensional flattened MRI images led to significant computational challenges, including heavy memory usage, long computation times, and numerical instability. The BNN-like classifier became completely infeasible for our dataset size. Fourth, dataset imbalance in certain test subsets introduced evaluation bias, potentially inflating accuracy for overrepresented classes while underestimating performance on rare categories. Finally, and perhaps most fundamentally, compression focuses on pixel-level redundancy and byte-pattern regularities, whereas medically meaningful biomarkers such as hippocampal atrophy and cortical thinning are spatially localized anatomical changes. These may not correlate directly with overall image compressibility, as critical pathological changes could be masked by the dominance of non-affected regions in global compression measures.

Perspectives and Future Directions

Several promising research directions emerge naturally from this work. First, testing the approach on longitudinal MRI data would enable evaluation of algorithmic complexity as a biomarker of progressive cognitive decline. Tracking complexity changes within individual patients over time might reveal patterns invisible in cross-sectional comparisons and could provide early warning signals for accelerated deterioration. Second, integrating clinical metadata such as age, APOE genotype, and MMSE cognitive scores could enable the construction of truly multimodal predictive models that combine structural complexity with established risk factors and functional assessments.

From a methodological perspective, improving the MDL prototype construction represents an accessible enhancement. Rather than using simple arithmetic means of training images, employing medoid-based prototypes or multiple centroid representations through clustering could better capture intra-class variability and improve classification robustness. More ambitiously, applying AIT principles to learned representations rather than raw pixels may unlock new potential. Computing NCD or compression ratios on feature embeddings extracted by pretrained MRI models such as ResNet or DenseNet could reveal whether algorithmic complexity aligns better with high-level semantic features than with low-level pixel patterns. This would bridge the gap between end-to-end deep learning and interpretable AIT-based analysis.

Developing hybrid AIT-ML architectures represents another fertile direction. Rather than treating AIT features as static inputs to classical models, we could design systems where complexity growth, multiscale compression dynamics, entropy evolution, and patch variability serve as complementary signals in ensemble frameworks. Such systems might leverage the interpretability of AIT while exploiting the pattern recognition strengths of modern machine learning. Finally, shifting from whole-image analysis to region-based complexity measurement could dramatically improve biological relevance. By computing complexity measures specifically within anatomically defined regions of interest, we could

better capture the spatially heterogeneous nature of neurodegeneration and potentially identify novel imaging biomarkers tied directly to the known neuropathology of Alzheimer's disease.

Bibliography

Dataset: Best Alzheimer's MRI Dataset (Kaggle) - https://www.kaggle.com/datasets/lukechugh/best-alzheimer-mri-dataset-99-accuracy

Github Repository: https://github.com/Nawel08/Complexity-Loss-as-a-Biomarker-for-Alzheimer-s-Disease