



HAL
open science

AI-based multimodal ICF classification: application to Alzheimer disease

Feten Skhiri Gabbouj, Laura Wallard, Mathias Blandeau, Virginie Vandebulcke, Bernard Gosselin, Matei Mancias

► **To cite this version:**

Feten Skhiri Gabbouj, Laura Wallard, Mathias Blandeau, Virginie Vandebulcke, Bernard Gosselin, et al.. AI-based multimodal ICF classification: application to Alzheimer disease. Conférence en TélésANTé et Dispositifs Médicaux, Université de Technologie de Troyes [UTT], Jun 2025, Troyes, France. hal-05384680

HAL Id: hal-05384680

<https://hal.science/hal-05384680v1>

Submitted on 27 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AI-based multimodal ICF classification: application to Alzheimer disease

Feten Skhiri Gabbouj^{1,2}, Laura Wallard², Mathias Blandeau², Bernard Gosselin¹,
Virginie Vandenbulcke¹, Matei Mancas¹

¹ ISIA lab - Université de Mons (UMONS), Mons, Belgium

² Université Polytechnique Hauts-de-France, CNRS, UMR 8201 LAMIH, F-59313, Valenciennes, France

Correspondance: Feten.SKHIRIGABBOUJ@umons.ac.be

Abstract

The International Classification of Functioning, Disability, and Health (ICF) is an important tool as a holistic approach. within this preliminary study we show that ICF can be (semi)automatically filled by AI models from multimodal data (text and images here) avoiding time-consuming manual filling. We use LLMs to process textual data and CNNs and VIT models for image data and fuse evrything into a single ICF.

Keywords— e-Health, AI, ICF , LLM , CNN

1 Introduction

The International Classification of Functioning, Disability, and Health (ICF), developed by the World Health Organization (WHO), offers a comprehensive framework for describing health and disability, emphasizing not only medical conditions but also functional abilities and contextual factors.

Despite its clinical relevance, the ICF remains underused in practice due to the time required for manual annotation. This pilot study explores the potential of artificial intelligence (AI) to assist in completing ICF-based assessments and proposes its use as a common structure for integrating multimodal data, particularly medical images and clinical text.

In the context of Alzheimer’s disease, most AI research focuses on detection—determining whether a patient is cognitively impaired. In contrast, this

study adopts a functional perspective by using the ICF to describe how Alzheimer’s impacts an individual’s everyday life, providing a richer and more patient-centered understanding.

2 Literature Review

2.1 International Classification of functioning and disease(ICF)

The International Classification of Functioning, Disability, and Health (ICF), developed by the World Health Organization (WHO), complements the International Classification of Diseases (ICD). While the ICD focuses on disease diagnosis, the ICF offers a broader view of health by classifying functioning, disability, and contextual factors. It includes four key components: body functions,



body structures, environmental and personal factors, and activity and participation. A major strength of the ICF is its holistic approach, recognizing that individuals with the same diagnosis can face different challenges depending on their environment and personal context [1, 2]. However, its manual application in clinical settings can be time-consuming, limiting its use. Automated or semi-automated methods could improve usability and reduce the burden on healthcare providers. Additionally, the ICF's structured format is well-suited for AI-driven multimodal research, integrating data from medical records, imaging, and physiological signals. Leveraging AI to support ICF use could enhance decision-making and better connect structured medical knowledge with real-world patient experiences [3].

2.2 Overview of Alzheimer's disease

Alzheimer's disease (AD) is a chronic, progressive neurodegenerative disorder and the most common cause of dementia worldwide, affecting millions of individuals, particularly older adults. It is characterized by a gradual but irreversible decline in cognitive functions, including memory, executive function, language, attention, and visuospatial skills, ultimately interfering with daily activities and diminishing quality of life [4].

Neuropathologically, AD is defined by the accumulation of beta-amyloid plaques and tau neurofibrillary tangles, leading to widespread neuronal loss and synaptic dysfunction, particularly in the hippocampus and cerebral cortex regions crucial for memory and higher cognitive functions. The disease follows a continuum, beginning with preclinical changes that are asymptomatic but detectable through biomarkers, progressing from early to mild cognitive impairment, and eventually culminating in severe dementia [5]. However, symptom presentation and progression are highly heterogeneous,

varying significantly between individuals and affecting multiple domains beyond cognition, including motor function, behavior, and social interactions.

Given this complexity, diagnosing and monitoring AD requires a multimodal approach that incorporates neuropsychological assessments such as the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA), neuroimaging techniques like magnetic resonance imaging (MRI) and positron emission tomography (PET), as well as behavioral observations and lifestyle changes [6]. Despite extensive research, no cure currently exists for AD, and current treatment strategies focus on symptom management and slowing disease progression through pharmacological and non-pharmacological interventions. To improve care and disease monitoring, it is essential to develop a comprehensive framework capable of systematically tracking disease progression across diverse modalities including imaging, cognitive assessments, behavioral changes, and lifestyle adaptations allowing for a more holistic and dynamic understanding of the disease over time.

2.3 Alzheimer AI Classification Methods

Recent advances in artificial intelligence have demonstrated substantial potential for Alzheimer's disease detection and diagnosis, offering promising tools for clinical applications [7]. The literature reveals two predominant approaches: traditional machine learning algorithms, such as Random Forest [8], alongside deep learning techniques, particularly Convolutional Neural Network (CNN) architectures that excel at processing the neuroimaging data central to Alzheimer's assessment. CNNs have consistently outperformed conventional machine learning methods in this domain, leveraging their hierarchical feature extraction capabilities to identify subtle patterns in MRI scans that may

elude human detection. Notably, transfer learning approaches utilizing pre-trained CNN models such as VGG16 and ResNet have yielded remarkable results, achieving classification accuracies exceeding 95% in some studies [9, 10, 11, 12], particularly when fine-tuned on Alzheimer’s specific datasets. These pre-trained models benefit from knowledge acquired on larger datasets, which proves particularly valuable given the limited availability of labeled Alzheimer’s neuroimaging data. Additional promising developments include multimodal approaches that integrate imaging with clinical, genetic, and cognitive assessment data, potentially providing more robust diagnostic frameworks that mirror the multifaceted nature of Alzheimer’s disease progression. Challenges remain in the standardization of preprocessing techniques, addressing dataset biases, and ensuring robust external validation across demographically diverse populations [13, 14, 15].

3 Methodology

In this research, we aim to obtain an ICF classification for each patient using both textual and imaging data. After preprocessing, an LLM-generated ICF classification was derived from the patient text, while a separate ICF classification was obtained from medical images. These two classifications were then combined into a unified ICF representation. Given that the input images clearly depict brain structures, we prioritized qualification over classification when it comes to images, focusing on the severity of cognitive function impairment. The severity was categorized into four levels: 0 (normal), 1 (early impairment), 2 (cognitive impairment) and 3 (Alzheimer’s disease).

3.1 Data Description

In this study, we utilized the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset, a

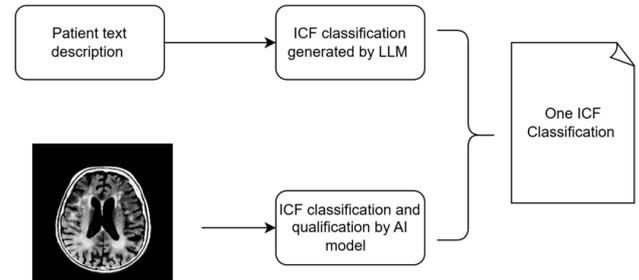


Figure 1: From patient description, a LLM-based ICF classification is first done. Additional classification and qualification is aggregated to the same ICF based on image data.

widely recognized and extensively used resource in Alzheimer’s disease research [16, 17]. The ADNI dataset provides a comprehensive collection of multimodal data, including structural and functional MRI scans, as well as other imaging modalities. In addition to imaging data, ADNI includes detailed clinical assessments, cognitive evaluations, and demographic information, allowing for a more holistic understanding of disease progression. The dataset encompasses multiple diagnostic categories, including Alzheimer’s disease (AD), mild cognitive impairment (MCI), early mild cognitive impairment (EMCI) and cognitively normal (CN) individuals. This diversity enables the development and evaluation of models capable of distinguishing between different stages of cognitive decline, making ADNI a crucial resource for advancing AI-driven analysis in neurodegenerative research. For this study, we focused on a subset of the ADNI dataset tailored to our specific research objectives. Regarding imaging data, we utilized axial FLAIR and T2-weighted axial MRI scans, as these modalities are well-suited for detecting structural abnormalities relevant to our task. From the available patient population,

we selected 225 individuals spanning four diagnostic categories: Alzheimer’s disease (AD), mild cognitive impairment (MCI), early mild cognitive impairment (EMCI), and cognitively normal (CN) subjects. In addition to imaging data, we incorporated clinical and neuropsychological assessments to enhance our analysis. Whenever available, we selected the corresponding assessments for our 225 patients, including diagnostic summaries and neuropsychological evaluations. Specifically, we extracted data from key tables such as the Diagnostic Summary, Neuropsychiatric Inventory Questionnaire, Montreal Cognitive Assessment, and Cognitive Change Index, among others.

3.2 Preprocessing

3.2.1 Image Preprocessing

To prepare the MRI scans for analysis, we applied a structured preprocessing pipeline (Figure 2) to normalize image intensity, enhance contrast, and focus on diagnostically relevant brain regions. The steps included Z-score normalization, skull stripping, histogram equalization, resizing, and format conversion. Finally, slice selection was guided by a ventricle-based similarity approach to retain the most informative slices for classification.

3.2.2 Text Preprocessing

To ensure structured representation of patient data, we created individual text files combining relevant ADNI dataset information for each patient. These files included diagnostic classifications (CN, MCI/EMCI, or AD), reported symptoms (vomiting, dizziness, headaches, etc.), psychological assessments (particularly anxiety levels recorded with qualitative descriptors like "Not at all" or "Somewhat"), and neuropsychological evaluation results from cognitive tasks such as drawing a cube or clock and other memory or visuospatial tests that are crucial for detecting early cognitive

MRI Preprocessing Pipeline

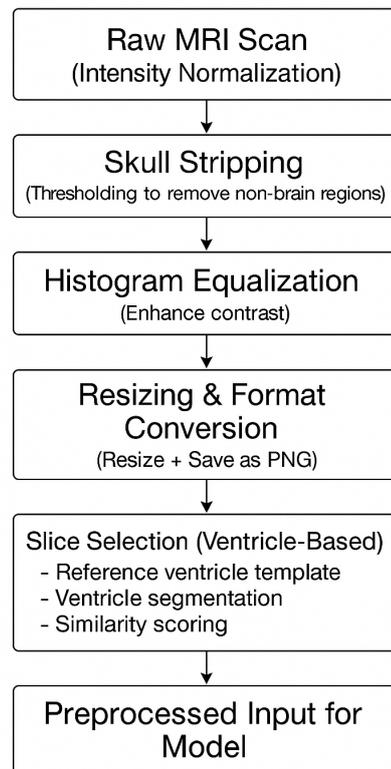


Figure 2: Image Preprocessing steps



decline. Our preprocessing strategy preserved full textual descriptions of variables and values rather than using encoded labels or keyword-based representations. This approach enhanced readability and provided contextually rich representations, making the data more interpretable for large language models while maintaining clinical relevance - similar to how physicians document and analyze patient medical records. This decision ensures the textual data remains optimally structured for subsequent natural language processing tasks.

3.3 Experiment: automatic ICFs

3.3.1 Image data

In this study, we employed transfer learning, a technique that utilizes pre-trained deep learning models to adapt to a new task [18] where less data is available. Instead of training a model from scratch, transfer learning allows us to leverage knowledge from models trained on large-scale datasets, improving efficiency and generalization. We used several well-established convolutional neural networks (CNNs), including MobileNet, VGG16, and ResNet-50, as well as a Vision Transformer (ViT) model [19].

To maintain the general feature extraction capabilities learned from their original training, we froze the initial layers of each model, which capture fundamental patterns such as edges and textures. Training was then performed on the later layers, which were fine-tuned to classify medical images specific to our dataset. Each model was trained for 40 epochs. We also applied data augmentation techniques, including rotation, zoom, flipping, and contrast adjustment, ensuring greater variability in the training data.

For patient-level classification, rather than making predictions based on a single image slice, we adopted a majority-vote approach. Each patient's classification was determined by aggregating predictions from all their slices, assigning the final la-

bel based on the most frequently predicted class. This approach improves robustness by reducing the impact of potential misclassifications at the slice level.

3.3.2 text data

After preprocessing, each patient's text file was paired with ground-truth ICF classifications provided by a specialist. These classifications followed the standard ICF framework at the category level (e.g., **b** for body functions, **s** for body structures, **e** for environmental factors). To generate ICF classifications using large language models (LLMs), we designed a simple prompt introducing the context and giving clear classification instructions. Due to time constraints and the complexity of the ICF framework, only six ground-truth classifications were available, as producing them requires significant time and medical expertise.

After testing several LLMs, including domain-specific models, Claude [20] was selected for its superior understanding of ICF classes. This choice was based on preliminary evaluations of coherence, completeness, and alignment with ground truth. The model's outputs were compared to expert annotations using the following metrics:

- **True Positive (TP)**: ICF class present in both the ground truth and LLM output (text and code).
- **False Negative (FN)**: ICF class in the ground truth but missing in the LLM output.
- **False Positive (FP)**: ICF class in the LLM output but absent from the ground truth.

True Negatives (TN) were excluded due to the large number of ICF classes unlikely to appear in the texts, which could bias the results. From TP, FN, and FP values, we calculated key performance metrics: Accuracy, Precision, Recall, and F-Score. Final results were averaged across all tests to provide an overall performance assessment.



4 Results and discussion

The classification results for four-class (CN, EMCI, MCI, and AD) are presented in Table 1, first line. Overall, the models exhibit relatively low performance, which can be attributed to the limited dataset size and the high similarity between certain classes, particularly between EMCI and MCI. Given this challenge, we restructured the classification task into a binary problem, grouping all pathological cases (EMCI, MCI, and AD) under a single "unhealthy" category and comparing them against the "healthy" (CN) class. This adjustment led to a moderate improvement in performance in all models (Table 1, second line).

In further refinement of the classification, we focused solely on distinguishing AD from CN. In this scenario, the models achieved their best performance (Table 1, third line). This can be explained by the clearer distinction between AD and CN, whereas the finer differences between EMCI and MCI remain difficult for the models to generalize due to limited training examples. To achieve more detailed and accurate multi-class classification, a larger dataset with more balanced samples per class is essential.

Classes	Vgg16	Mobilenet	Resnet50	VIT
4 classes	40.90	54.54	54.54	50.00
CN/NOT	68.10	66.00	59.90	63.64
CN/AD	80.00	68.18	66.67	73.33

Table 1: Average Accuracy for image classification: 4-class (line 1), binary CN vs. non-CN (line 2), and CN vs. AD (line 3).

The evaluation of the ICF classifications generated by the selected LLM was carried out by comparing them to the ground-truth classifications. The average results of the comparison between the ICF classifications generated by Claude and the ground truth annotations are presented in Table 2.

The results indicate that Claude achieved an

LLM	Accuracy	Precision	Recall	F-Score
Claude	85	98	86	91

Table 2: Average results of Claude in ICF classification.

overall accuracy of 85%, demonstrating its ability to correctly classify ICF categories in patient descriptions. The model exhibited particularly high precision (98%), reflecting its tendency to generate classifications with minimal false positives. Beyond classification performance, it is important to highlight the nature of the ICF outputs produced. Our objective was not to simply reproduce or confirm diagnostic labels, but rather to generate a more functional and human-centered view of the patient's condition. In line with the ICF's philosophy, we focused on representing the *impact of cognitive decline on daily activities and social participation*. This includes classes from the **Activities and Participation** component of the ICF, such as: **d160** – Focusing attention **d175** – Solving problems **d230** – Carrying out daily routine **d240** – Handling stress and other psychological demands **d310** – Communicating with – receiving – spoken messages **d350** – Conversation **d450** – Walking **d570** – Looking after one's health

We also included relevant classes from the **Body Functions** component where applicable, especially when image-based qualification was involved: **b114** – Orientation functions **b117** – Intellectual functions **b140** – Attention functions **b144** – Memory functions **b164** – Higher-level cognitive functions

This broader coverage enables the system to go beyond binary or categorical labels and reflect how Alzheimer's manifests itself in real-life limitations such as difficulties in communication, personal care, or handling routine tasks, which are directly observable in clinical notes or imaging data. As such, this approach provides more actionable and individualized insight for clinicians, caregivers,



and potentially automated decision support tools.

In our future work, we will focus on more complex fusion techniques to enhance the integration of text and image data. Currently, due to the limited number of annotated samples, we manually combine the ICF classifications produced by the LLM with the output of the image-based classifier. For each patient, we identify relevant ICF classes, particularly those related to cognitive functions, such as *b114* (memory functions) and *b117* (intellectual functions)—and assign a qualifier value based on the diagnosis predicted from medical images. These qualifiers reflect the degree of impairment and follow a scale where: 0 = CN , 1 = MCI, 2 = EMCI , 3 = AD

This manual fusion step allows us to enrich the ICF representation with both class and severity information, ensuring consistency across modalities, reducing classification errors, and improving the precision of the final output. With larger and more diverse datasets, this process can be further refined and automated.

5 Conclusion

Our results suggest that AI can effectively help in the classification of ICFs by integrating textual and imaging data into a unified framework. This study demonstrates the potential of combining LLM-based text classification with CNN-based image analysis to automate ICF coding, offering a more comprehensive and efficient representation of a patient's condition. In the case of Alzheimer's disease, our approach enables a multimodal evaluation that would be time-consuming and labor-intensive for humans. Although limited by the size of the dataset, this proof of concept shows that with larger datasets, generalizability could improve, ultimately supporting the creation of consistent ICF profiles and reducing the workload of clinicians. This work opens promising directions for enhancing AI-driven clinical decision support.

References

- [1] J. N. Vargus-Adams and A. Majnemer, "International classification of functioning, disability and health (icf) as framework for change: revolutionizing rehabilitation," *Journal of Child Neurology*, vol. 29, no. 8, pp. 1030–1035, 2014.
- [2] O. Svestkova, "International classification of functioning, disability and health of world health organization (icf)," *Prague Med Rep*, vol. 109, no. 4, pp. 268–274, 2008.
- [3] A. Cieza, R. Hilfiker, S. Chatterji, N. Kostanjsek, B. T. Üstün, and G. Stucki, "The international classification of functioning, disability, and health could be used to measure functioning," *Journal of Clinical Epidemiology*, vol. 62, no. 9, pp. 899–911, 2009.
- [4] A. Kumar, J. Sidhu, F. Lui, *et al.*, "Alzheimer disease," Feb 2024. Updated 2024 Feb 12.
- [5] R. J. Castellani, R. K. Rolston, and M. A. Smith, "Alzheimer disease," vol. 56, pp. 484–546, 9 2010.
- [6] F. Hasan Saif, M. N. Al-Andoli, and W. M. Y. W. Bejuri, "Explainable ai for alzheimer detection: A review of current methods and applications," *Applied Sciences*, vol. 14, no. 22, 2024.
- [7] C. Fabrizio, A. Termine, C. Caltagirone, and G. Sancesario, "Artificial intelligence for alzheimer's disease: Promise or challenge?," *Diagnostics*, vol. 11, no. 8, 2021.
- [8] M. Alvarado, D. Gómez, A. Nuñez, A. Robles, H. Marecos, and W. Ticona, "Implementation of an early detection system for neurodegenerative diseases through the use of artificial intelligence," in *2023 IEEE XXX International Conference on Electronics, Elec-*

- trical Engineering and Computing (INTERCON), pp. 1–6, 2023.
- [9] D. Ebrahim, A. M. T. Ali-Eldin, H. E. Moustafa, and H. Arafat, “Alzheimer disease early detection using convolutional neural networks,” in *2020 15th International Conference on Computer Engineering and Systems (ICCES)*, pp. 1–6, 2020.
- [10] A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhyay, “A cnn model: Earlier diagnosis and classification of alzheimer disease using mri,” in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 156–161, 2020.
- [11] S. Kumar, N. P. Singh, and B. Brahma, “Ai-based model for detection and classification of alzheimer disease,” in *2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*, pp. 1–6, 2023.
- [12] P. Krishnamoorthy, D. Swetha, P. S. Geetha, K. Karunambiga, R. K. Ayyasamy, and A. Kiran, “Revolutionizing medical diagnostics: Exploring creativity in ai for biomedical image analysis,” in *2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IconSCEPT)*, pp. 1–7, 2024.
- [13] M. Leela and K. Helenprabha, “Incremental learning based two-level multimodal data fusion model for alzheimer disease prediction on different data modalities,” *Connection Science*, vol. 37, no. 1, p. 2458501, 2025.
- [14] G. Singh Chhabra, A. Guru, B. J. Rajput, L. Dewangan, and S. K. Swarnkar, “Multimodal neuroimaging for early alzheimer’s detection: A deep learning approach,” in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–5, 2023.
- [15] A. S. Teaima, M. A. Wahed, and S. A. Sami, “Multimodal-based classification paradigm for alzheimer disease,” in *2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 430–433, 2023.
- [16] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner, “Alzheimer’s disease neuroimaging initiative (adni): Clinical characterization,” *Neurology*, vol. 74, no. 3, pp. 201–209, 2010.
- [17] Alzheimer’s Disease Neuroimaging Initiative (ADNI), “Alzheimer’s disease neuroimaging initiative,” 2025. Accessed: 2025-03-03.
- [18] J. Gupta, S. Pathak, and G. Kumar, “Deep learning (cnn) and transfer learning: A review,” *Journal of Physics: Conference Series*, vol. 2273, p. 012029, may 2022.
- [19] A. Steiner, A. Kolesnikov, X. Zhai, R. Wightman, J. Uszkoreit, and L. Beyer, “How to train your vit? data, augmentation, and regularization in vision transformers,” 2022.
- [20] Anthropic, “Claude ai,” 2024. Accessed: March 5, 2025.