

# Linguistic features and Automatic Classifiers for identifying Mild Cognitive Impairment and Dementia

Laura Calzà, Gloria Gagliardi, Rema Rossini Favretti, Fabio Tamburini

## BACKGROUND

Early recognition of cognitive impairment is a growing medical need due to the increasing incidence of cognitive decline related to neurological and non-neurological conditions.

Classical **pen-and-paper neuropsychological tests** are costly and time consuming

Sensitive and reliable **new instruments are needed** for monitoring patients with cognitive frailty and for **screening** the general population.

## CORPUS

**96 participants** (48 M, 48 F; age range 50–75, Italian mother-tongue):

- 48 healthy controls
- 48 cognitively impaired participants:
  - 16 Single-Domain Amnesic Mild Cognitive Impairment
  - 16 Multiple-Domain MCI
  - 16 early Dementia

These cohorts were **tested twice**, at a one-year interval

## RESULTS

Statistically significant indexes:

- **Acoustic level:** fluency (e.g. silence segments duration, speech segments duration, transformed phonation rate, standardized pause rate) and spectral properties of the voice (Higuchi Fractal dimension and Spectral Centroid)
- **Lexical level:** Content density
- **Syntactic level:** utterance length, global dependency distance, syntactic embeddedness

## OPLON Care & Cure project

(Opportunities for active and healthy LONgevity)

an interdisciplinary study funded by the Italian Ministry of University and Research (Smart Cities and Communities and Social Innovation program)

**AIM:** to apply **Natural Language Processing methods** for the detection of **early linguistic signs of cognitive decline** due to dementia in the elderly

## METHODS

- conventional **neuropsychological evaluation**
- **automatic extraction of 87 linguistic features** (i.e. acoustic, rhythmic, lexical, and morpho-syntactic indexes)
- **statistical analysis**
- application of **Machine Learning classifiers** (i.e. Support Vector Machines, Random Forest Classifiers) to distinguish healthy controls from MCI subjects

## RESULTS

- a multidimensional spoken language analysis using NLP techniques can identify **subtle language modifications in preclinical stages of dementia**
- **Machine Learning classifiers** can distinguish between controls and MCI subjects displaying **high F1 scores** (around 0.70)
- this **methodology is robust** also from a **longitudinal point of view**