Francesco Sigona, Barbara Gili Fivela
UNIVERSITÀ DEL SALENTO

Pietro Vigorelli
GRUPPO ANCHISE

Andrea Bolioli
maize.

A COMPUTATIONAL ANALYSIS OF SPEECH PATTERNS IN DEMENTIA: THE "ANCHISE 2022" CORPUS

UNIVERSITÀ DEGLI STUDI DI ENNA "KORE" · UNIVERSITÀ DEL SALENTO · UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

IL PARLATO IN AMBITO MEDICO:
ANALISI LINGUISTICA, APPLICAZIONI TECNOLOGICHE E STRUMENTI CLINICI
15-17 febbraio 2023
Università del Salento, Lecce

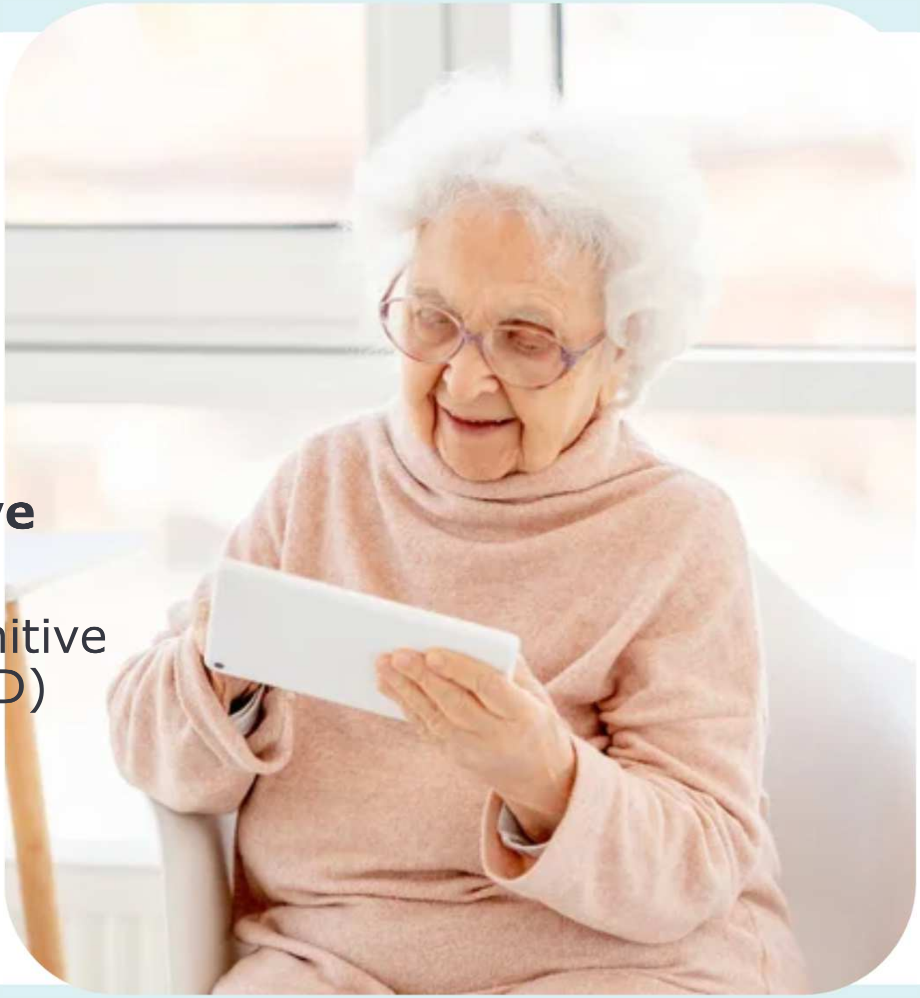
AISV

Outline

This study aims at **exploring connections** (in a wide sense) **between** Mini-mental Score Examination scores (**MMSE**, Folstein et al., 1975) and **some linguistic features** derived from a Computational Analysis of speech from people with dementia, within the "Anchise 2022" Corpus

- The «Anchise 2022» Corpus
- Methods
- Results
- Discussion
- Future directions

Speech as a promising way to digitally assess cognitive abilities



Speech is easy to record

- Smart devices with high-quality microphones
- No need for additional hardware or sensors.

Significance of speech in detecting cognitive impairment and dementia

- Speech patterns become changed in Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD)

Language changes in AD

Function	Early Stages	Moderate to Severe Stages
Spontaneous Speech	Fluent, grammatical	Non-fluent, echolalic
Paraphrastic errors	Semantics	Semantic and phonetic
Repetition	Intact	Very affected
Naming objects	Slightly affected	Very affected
Understanding the words	Intact	Very affected
Syntactical understanding	Intact	Very affected
Reading	Intact	Very affected
Writing	± Intact	Very affected
Semantic knowledge of words and objects	Difficulties with less used words and objects.	Very affected

Language changes in AD (adapted from Ferris and Farlow 2013; Vigo et al., 2022)

Reduced verbal fluency
Reduced speech rate
Increased word retrieval difficulty
Increased use of filler sounds

Reduced coherence
Implausible or irrelevant details
Prosodic modifications

The «Anchise 2022 Corpus»

- Latest extension of the «Anchise 320 Corpus»
  - (Bolioli et al. 2020; Benvenuti et al., 2020)

- **First Italian** corpus of **free conversations** between healthcare professionals (HPs) and people with **dementia**, since 2007 (**ecological** conditions)
  - by the Anchise Group, an association of experts for the research, training and care of the elderly with dementia

ApproccioCapacitante®

- HPs are trained with the «**Enabling approach**» http://www.formalzheimer.it
- Not a speech task, but a way of relating to the patient that aims to facilitate communication by establishing a positive and constructive relationship between the elderly and the healthcare professional

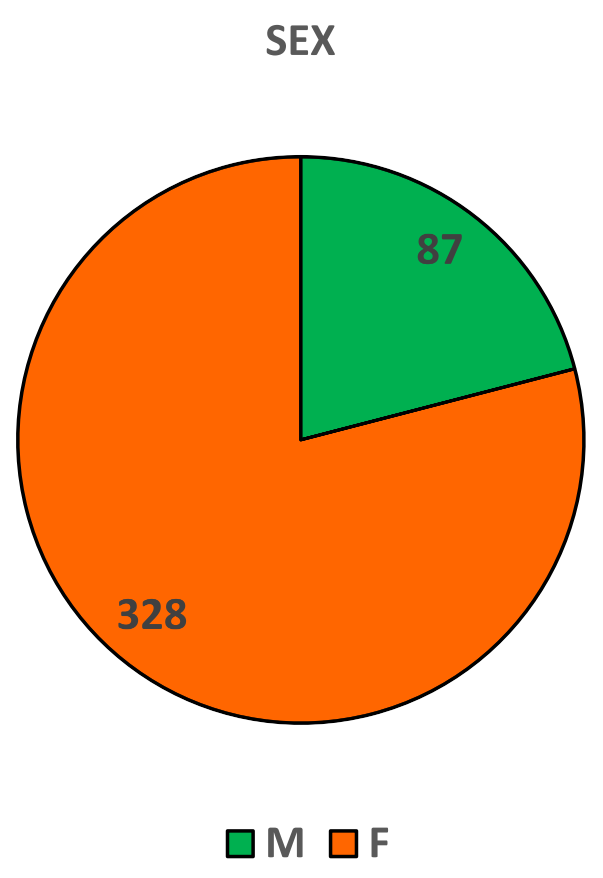
Conversations characteristics and pre-processing

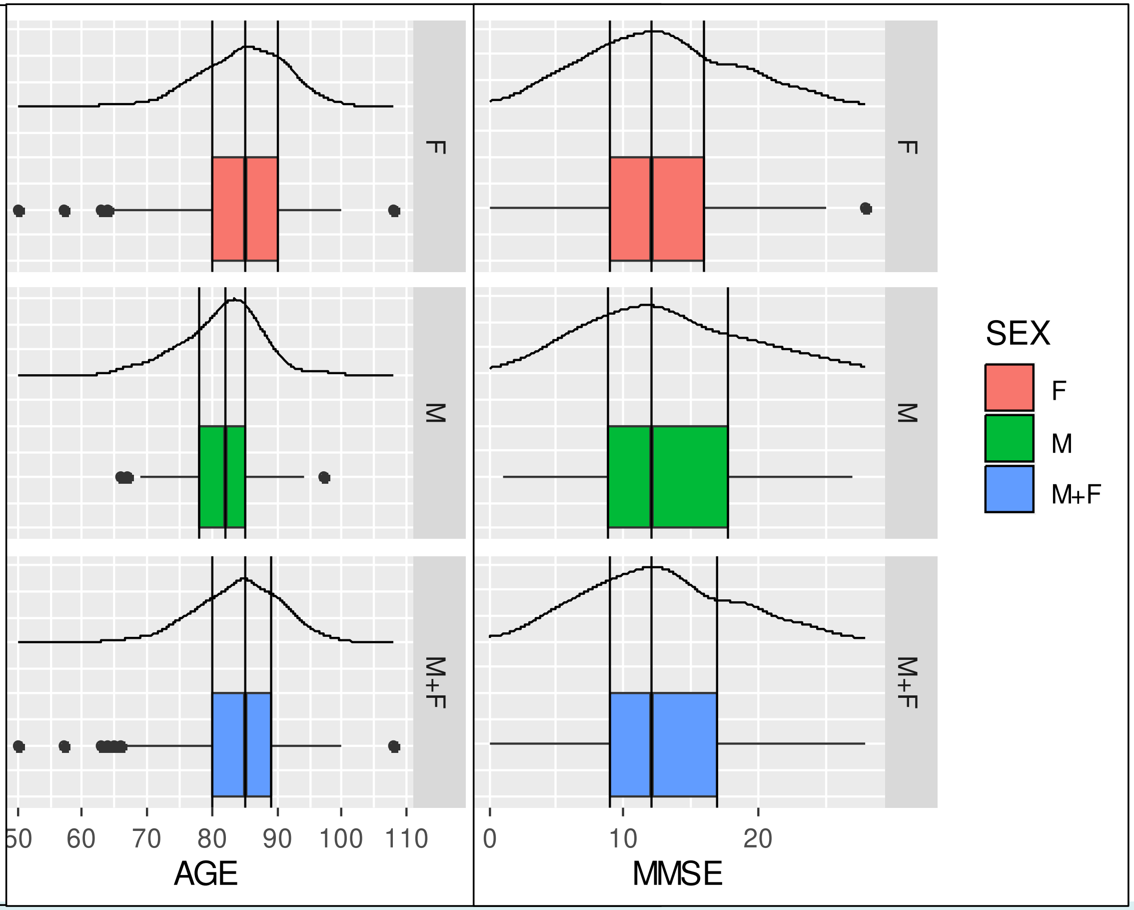
- Duration: 5-10 minutes
- Manual transcription
- Metadata : language | age | sex | MMSE score
- Elimination of written comments by HPs, such as "[Touch the recorder]", "[Silence]", "[Laughs]”

Macro-description of the sample

Unselected series of people diagnosed with **dementia**

«most patients are affected by AD»
(Benvenuti et al., 2020)





Numerosity of patients (and turns) with Age and MMSE information

	Num conversations	Num turns	Turns / Conversations		
			Avg	Std	Std/Avg
PAT w MMSE	238	7596	31.9	20.9	0.656
PAT w MMSE Male	54	1539	28.5	18.0	0.632
PAT w MMSE Female	184	6057	32.9	21.6	0.656
PAT w AGE	355	11227	31.6	22.2	0.701
P					37
P					71
PAT w MMSE and AGE	218	6953	31.9	20.3	0.637
PAT w MMSE and AGE Male	50	1482	29.6	18.2	0.612
PAT w MMSE and AGE Female	168	5471	32.6	20.9	0.64

With over 400 patients, “Anchises 2022” is one of the richest corpora of speech transcripts from people with dementia

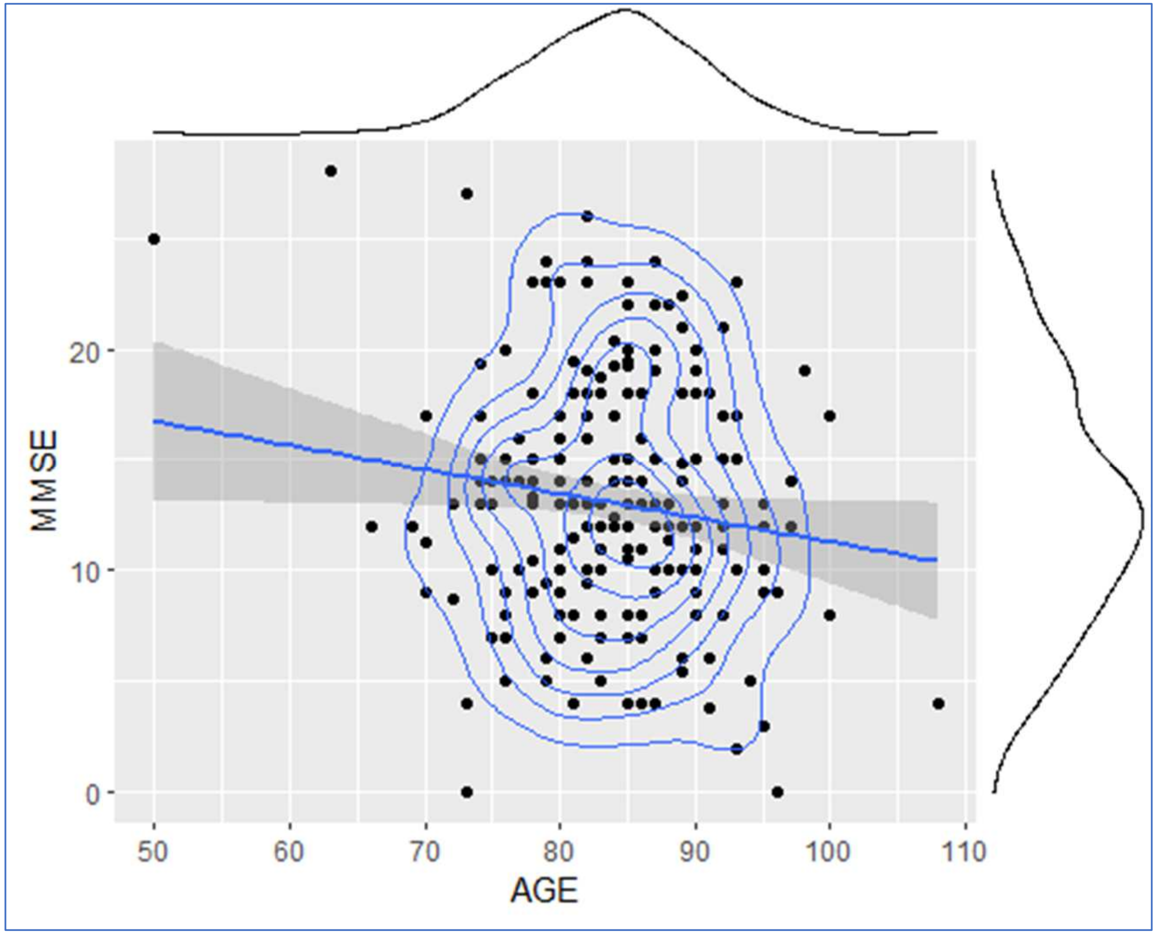
Exploring MMSE vs Age: linear model

`MMSE ~ Age`

The effect of Age is **statistically significant and negative**

- **β = -0.11**, 95% CI [-0.21, -4.26e-03],
- t(217) = -2.05
- **p = 0.041**;
- Std. β = -0.14, 95% CI [-0.27, -5.38e-03]

- model explains a statistically significant **but very weak proportion of variance (R2 = 0.02)**

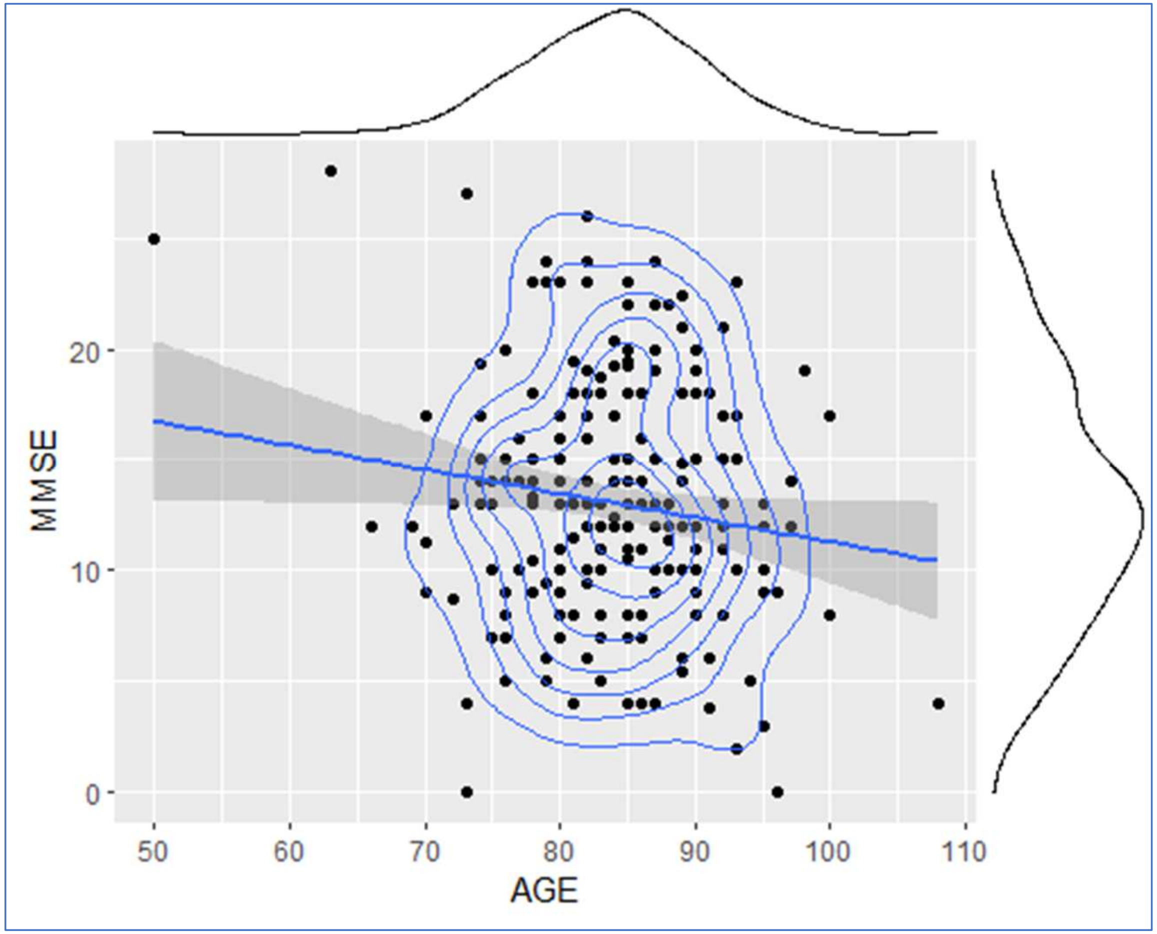


Exploring MMSE vs AGE: Gaussianity

```
mvn(data, mvnTest="hz",
        multivariateOutlierMethod="quan")
```

MVN test	Value	p	MVN
Henze-Zirkler	0.8632	0.148	YES

Univariate test		Value	p	Gauss
Anderson-Darling	AGE	0.7203	0.0594	YES
Anderson-Darling	MMSE	0.9142	0.0197	NO



MMSE vs Linguistic features: some possible approaches

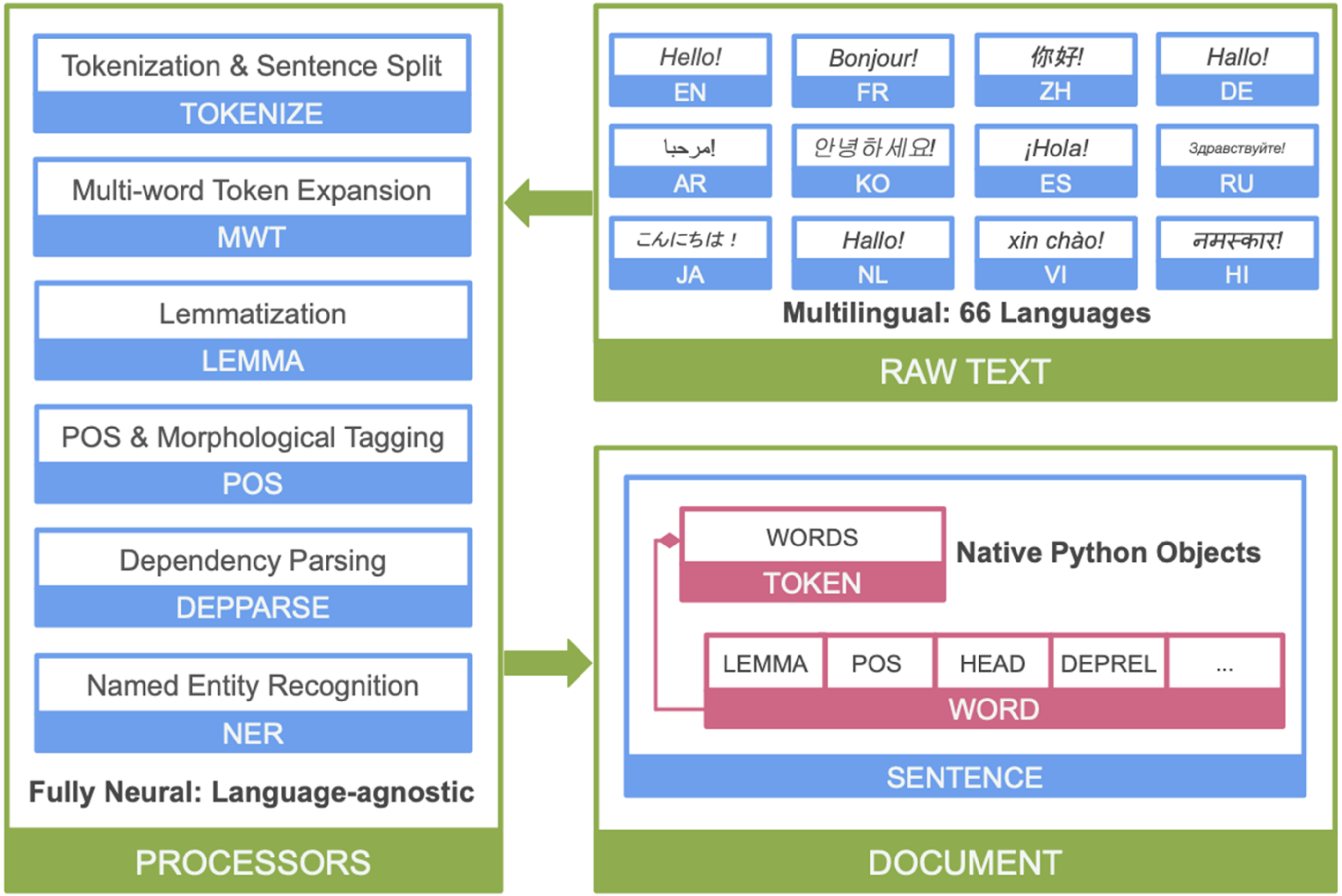
1. Exploring **correlations** between MMSE and the single linguistic feature $\mathtt{FEA}_i$
   - linear fit of the type: $\mathtt{MMSE} \sim \mathtt{FEA}_i \quad i \in 1,2,..,N_{FEA}$
   - Pearson's correlation
   - Spearman's Correlation

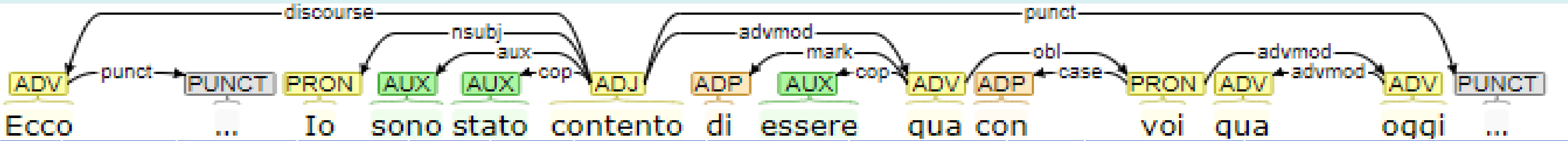
2. Splitting the whole MMSE range [0-30] into smaller subsets (**groups of patients**), based on **cutoff** values
   - Then statistical comparison between groups
   - Classification

Feature extraction : NLP of the Corpus

NLP based on Stanza (Qi et al., 2020)

«Starting from raw text to syntactic analysis and entity recognition, Stanza brings state-of-the-art NLP models to languages of your choosing»





CONLL-U format (https://universaldependencies.org/format.html)

id	form	lemma	upos	xpos	head	deprel	Start char	End char	ner	Multi ner	feats
1	Ecco	ecco	INTJ	I	6	discourse	0	4	O	('O',)	
2	…	…	PUNCT	FS	1	punct	5	6	O	('O',)	
3	Io	io	PRON	PE	6	nsubj	8	10	O	('O',)	Number=Sing| Person=1| PronType=Prs
4	sono	essere	AUX	VA	6	aux	11	15	O	('O',)	Mood=Ind| Number=Sing| Person=1| Tense=Pres| VerbForm=Fin
5	stato	essere	AUX	V	6	cop	16	21	O	('O',)	Gender=Masc| Number=Sing| Tense=Past| VerbForm=Part
6	contento	contento	ADJ	A	0	root	22	30	O	('O',)	Gender=Masc| Number=Sing
7	di	di	ADP	E	9	mark	31	33	O	('O',)	
8	essere	essere	AUX	V	9	cop	34	40	O	('O',)	VerbForm=Inf
9	qua	qua	ADV	B	6	advmod	41	44	O	('O',)	
10	,	,	PUNCT	FF	6	punct	44	45	O	('O',)	
11	con	con	ADP	E	12	case	46	49	O	('O',)	
12	voi	voi	PRON	PE	6	obl	50	53	O	('O',)	Number=Plur| Person=2| PronType=Prs
13	qua	qua	ADV	B	14	advmod	54	57	O	('O',)	
14	oggi	oggi	ADV	B	12	advmod	58	62	O	('O',)	
15	…	…	PUNCT	FS	6	punct	63	64	O	('O',)	

Collected features for each conversation

we explored some of the most used features in similar studies

- *BASIC*: num_turns (+ Age, + MMSE)

Lexical

- content density*
- UPOS rate (e.g. ADV.rate)
- nouns_to_verbs
  - (aka.: reference rate to reality)
- tokens_to_turns
- num_hapax_legomena
- num_words
- num_interrog
- num_ellipsis («…»)

Lexical richness

- num_types
- types_tokens_ratio (TTR)
- Brunet indexes
  - *a* in {0.134, 0.172, 0.185}
- Honore's statistic

Syntactic

- dependency distance(Roark et al., (2007), (2011)
  - mean, std.dev., max
- Maximum structure depth
  - mean, std.dev., max
- Complexity as (Szmrecsanyi, 2004)

$$\frac{2*conj + 2*pron+nouns+verbs}{num\_words}$$

Verbal analysis: Forms

- finite verb (Fin)
- Infinitive (Inf)
- Participle (Part)
- Gerund (Ger)

Verbal analysis: Mood

- Indicative (Ind),
- Subjunctive / Conjunctive (Sub)
- Imperative (Imp),
- Conditional (Cnd)

Verbal analysis: Tense

- Present (Pres)
- Past (Past)
- Imperfect (Imp)
- Future (Fut)

+ 1st person verbs

Exploring linear fit and correlations

	$\mathrm{MMSE} \sim \mathrm{FEA}_i$					Pearson	Spearman	
Feature	beta	t	CI	p.value	R2	r	ρ	p.value
Age	-0.1108	t(214)=-2.0715	[-0.2163, -0.0054]	p=0.04 *	0,0197	-0,1402	-0,0767	p=0,2619
num_turns	0.0109	t(234)=0.622	[	35	0,0017	0,0406	0,0232	p=0,723
num_words	0.0019	t(234)=1.905	[		0,0153	0,1236	0,1393	p=0,0324 *
num_interrog	-0.0249	t(234)=-0.4297	[	58	0,0008	-0,0281	-0,0125	p=0,8484
num_ellipsis	0.0009	t(234)=0.0874	[	3	0	0,0057	-0,002	p=0,9756
num_hapax	0.0166	t(234)=2.8995	[0.0053, 0.0279]	p=0.004 **	0,0347	0,1862	0,1687	p=0,0094 **
TTR	-37.136	t(234)=-1.2692	[-9.4783, 2.0511]	p=0.206	0,0068	-0,0827	-0,0714	p=0,2747
tokens_to_turns	0.0376	t(234)=2.2627	[0.0049, 0.0703]	p=0.025 *	0,0214	0,1463	0,1301	p=0,0458 *
num_types	0.0074	t(234)=2.4883	[0.0015, 0.0133]	p=0.014 *	0,0258	0,1606	0,1526	p=0,019 *
Density*	-32.127	t(234)=-2.5708	[-5.6748, -0.7506]	p=0.011 *	0,0275	-0,1657	-0,1128	p=0,0839 .
Nouns_to_verbs	-0.6627	t(234)=-0.4697	[-3.4422, 2.1169]	p=0.639	0,0009	-0,0307	0,0478	p=0,4647
p1verbs_ratio	-10.012	t(234)=-0.3135	[-7.2931, 5.2907]	p=0.754	0,0004	-0,0205	0,0387	p=0,554
Honore	0.0015	t(234)=1.414	[-6e-04, 0.0037]	p=0.159	0,0085	0,092	0,0986	p=0,131
brunet_.134	0.2072	t(234)=1.887	[-0.0091, 0.4234]	p=0.06 .	0,015	0,1224	0,1129	p=0,0836 .
brunet_.172	0.4335	t(234)=1.3526	[-0.1979, 1.0649]	p=0.177	0,0078	0,0881	0,0725	p=0,267
brunet_.185	0.4651	t(234)=1.0142	[-0.4384, 1.3686]	p=0.312	0,0044	0,0662	0,0545	p=0,4045

Weak portion of variance

Exploring linear fit and correlations

	MMSE ~ FEA_i					Pearson	Spearman	
Feature	beta	t	CI	p.value	R2	r	ρ	p.value
upos.ADJ.perc	-0.4705	t(229)=-2.594	[-0.8279, -0.1131]	p=0.01 *	0,0285	-0,1689	-0,1066	p=0,1061
upos.ADP.perc	0.3033	t(232)=1.8899	[-0.0129, 0.6194]	p=0.06 .	0,0152	0,1231	0,1194	p=0,0682 .
upos.ADV.perc	0.0265	t(234)=0.274	[-0.1638, 0.2168]	p=0.784	0,0003	0,0179	0,0102	p=0,876
upos.AUX.perc	0.1966	t(233)=1.2089	[-0.1238, 0.5169]	p=0.228	0,0062	0,079	0,1062	p=0,1044
upos.CCONJ.perc	-0.0503	t(228)=-0.1981	[-0.5508, 0.4502]	p=0.843	0,0002	-0,0131	0,0048	p=0,942
upos.DET.perc	0.0252	t(234)=0.1862	[-0.242, 0.2925]	p=0.852	0,0001	0,0122	0,0351	p=0,5912
upos.INTJ.perc	-0.2055	t(232)=-3.3036	[-0.328, -0.0829]	p=0.001 **	0,0449	-0,212	-0,1611	p=0,0136 *
upos.NOUN.perc	0.1651	t(234)=1.4948	[-0.0525, 0.3827]	p=0.136	0,0095	0,0973	0,1349	p=0,0383 *
upos.NUM.perc	0.1033	t(170)=0.3312	[-0.5121, 0.7186]	p=0.741	0,0006	0,0254	0,064	p=0,4041
upos.PRON.perc	0.0296	t(234)=0.284	[-0.1755, 0.2346]	p=0.777	0,0003	0,0186	0,0209	p=0,7497
upos.PROPN.perc	-0.0519	t(182)=-0.2406	[-0.4776, 0.3738]	p=0.81	0,0003	-0,0178	-0,0334	p=0,6529
upos.SCONJ.perc	0.2825	t(224)=1.0846	[-0.2308, 0.7958]	p=0.279	0,0052	0,0723	0,0632	p=0,3446
upos.VERB.perc	0.2423	t(234)=2.2232	[0.0276, 0.4571]	p=0.027 *	0,0207	0,1438	0,0814	p=0,2127
va.VerbForm.Fin.perc	-0.0594	t(234)=-1.9415	[-0.1197, 9e-04]	p=0.053 .	0,0159	-0,1259	-0,1245	p=0,0561 .
va.VerbForm.Inf.perc	0.0281	t(234)=0.7152	[-0.0493, 0.1054]	p=0.475	0,0022	0,0467	0,0714	p=0,2749
va.VerbForm.Part.perc	0.0607	t(234)=1.6683	[-0.011, 0.1324]	p=0.097 .	0,0118	0,1084	0,0962	p=0,1405
va.VerbForm.Ger.perc	-0.0792	t(234)=-0.2829	[-0.6309, 0.4725]	p=0.778	0,0003	-0,0185	0,0314	p=0,6316

Exploring linear fit and correlations

	$MMSE \sim FEA_i$					Pearson	Spearman	
Feature	beta	t	CI	p.value	R2	r	ρ	p.value
va.Mood.Ind.perc	-0.0346	t(234)=-0.8505	[-0.1147, 0.0455]	p=0.396	0,0031	-0,0555	-0,0251	p=0,7008
va.Mood.Sub.perc	0.2462	t(234)=2.6905	[0.0659, 0.4265]	p=0.008 **	0,0300	0,1732	0,1999	p=0,002 **
va.Mood.Imp.perc	-0.0364	t(234)=-0.5757	[-0.1611, 0.0883]	p=0.565	0,0014	-0,0376	-0,0776	p=0,2348
va.Mood.Cnd.perc	0.1470	t(234)=0.7415	[-0.2435, 0.5375]	p=0.459	0,0023	0,0484	0,0821	p=0,209
va.Tense.Pres.perc	-0.0323	t(234)=-1.5927	[-0.0722, 0.0076]	p=0.113	0,0107	-0,1036	-0,1048	p=0,1083
va.Tense.Past.perc	0.0540	t(234)=1.7388	[-0.0072, 0.1152]	p=0.083 .	0,0128	0,1129	0,112	p=0,0859 .
va.Tense.Imp.perc	0.0194	t(234)=0.6268	[-0.0415, 0.0802]	p=0.531	0,0017	0,0409	0,0521	p=0,426
va.Tense.Fut.perc	0.1239	t(234)=0.7522	[-0.2007, 0.4486]	p=0.453	0,0024	0,0491	0,1149	p=0,0782 .
dep_dist.avg	23.928	t(234)=2.6202	[0.5936, 4.1921]	p=0.009 **	0,0285	0,1688	0,1368	p=0,0356 *
dep_dist.std	-0.8194	t(234)=-0.3268	[-5.7582, 4.1195]	p=0.744	0,0005	-0,0214	-0,0459	p=0,4826
dep_dist.max	0.9942	t(234)=1.6291	[-0.2082, 2.1966]	p=0.105	0,0112	0,1059	0,0984	p=0,1318
max_depth.avg	10.365	t(234)=2.3305	[0.1603, 1.9127]	p=0.021 *	0,0227	0,1506	0,1318	p=0,0431 *
max_depth.std	0.5240	t(234)=0.8234	[-0.7298, 1.7779]	p=0.411	0,0029	0,0537	0,0641	p=0,3269
max_depth.max	0.2090	t(234)=1.7368	[-0.0281, 0.446]	p=0.084 .	0,0127	0,1128	0,1074	p=0,0997 .
szmrecsanyi	64.628	t(234)=1.7688	[-0.7357, 13.6614]	p=0.078 .	0,0132	0,1149	0,1007	p=0,1229

Exploring MMSE-based groups: preliminary considerations about cut-off values for dementia stages

<table>
<tr><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th><th>19</th><th>20</th><th>21</th><th>22</th><th>23</th><th>24</th><th>25</th><th>26</th><th>27</th><th>28</th><th>29</th><th>30</th><th></th></tr>
<tr><td colspan="6">severe</td><td colspan="4">moderate</td><td colspan="15">mild dementia</td><td colspan="6">normal</td><td>[1]</td></tr>
<tr><td colspan="24"></td><td colspan="7">normal</td><td>[2]</td></tr>
<tr><td colspan="10">severe cognitive decline</td><td colspan="11">moderate cognitive decline</td><td colspan="4">MCI</td><td colspan="6">normal</td><td>[3]</td></tr>
<tr><td colspan="18">severe cognitive impairment</td><td colspan="6">MCI</td><td colspan="7">no impairment</td><td>[4]</td></tr>
<tr><td colspan="14">severe impairment</td><td colspan="4">moderate imp.</td><td colspan="6">Mild impairment</td><td colspan="3">bordeline</td><td colspan="4">normal</td><td>[5]</td></tr>
<tr><td colspan="18"></td><td colspan="6">MCI or early dementia</td><td colspan="7">healthy control (normal)</td><td>[6]</td></tr>
<tr><td colspan="10">severe AD</td><td colspan="5">moderate-severe AD</td><td colspan="6">moderate AD</td><td colspan="6">mild AD</td><td colspan="4"></td><td>[7]</td></tr>
</table>

[1] Vigorelli, Corpus Anchise, 2022

[2] Regione Emilia Romagna, *Strumenti per la valutazione del paziente con demenza, ottobre 2000:*

- «The cutoff score is 23-24, and most non-demented older people rarely score below 24», p. 14
- «A corrected score above 24 is considered normal», p. 15

[3] De Stefano et al., 2021

[4] Tombaugh and McIntyre, 1992

[5] Chopra et al., 2007

[6] Beltrami et al., 2018

[7] National Institute For Health And Clinical Excellence . *Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (Review of TA 111), UK, November 2009, p. 2*

Preliminary considerations about cut-off values for dementia stages

Cutoff values for MMSE ranges are not well established

(not surprising)

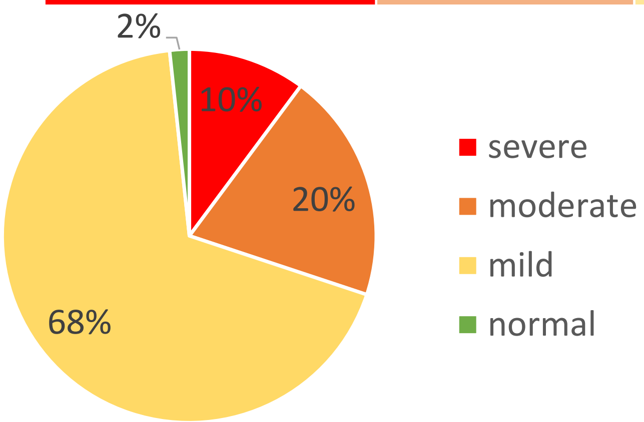
Study	MMSE score	Descriptor used in study
Folstein et al. (1975)	20 or less	Likely dementia
	24–30	Normal
Tombaugh and McIntyre (1992)	24–30	No impairment
	18–23	Mild cognitive impairment
	0–17	Severe cognitive impairment
Chopra et al. (2007)	27–30	Normal
	24–26	Borderline impairment
	18–23	Mild impairment
	14–17	Moderate impairment
	0–13	Severe impairment
Ferrell et al. (2000)	19 or less	Cognitive impairment
Kaasalainen et al. (1998)	23 or less	Cognitive impairment
Scherder and Bouma (2000)	18	Serious cognitive disturbance
Krulewitch et al. (2000)	24	Mild cognitive impairment
Tsai et al. (2008)	9 or less	Severe cognitive impairment
	25–30	No cognitive impairment
Radbruch et al. (2000)	20/21	Impaired
Chopra et al. (2008)	14 or less	Severe cognitive impairment
Shega et al. (2008)	24 or greater	Cognitively intact
	9 or less	Severe cognitive impairment

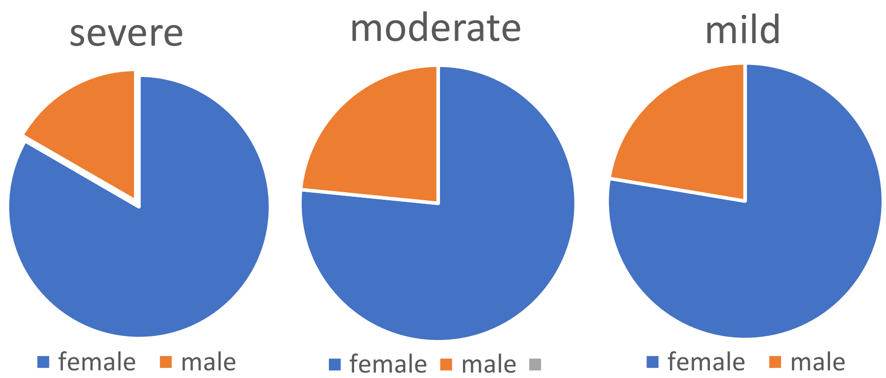
Examples of inconsistent MMSE cut-off scores reported in the literature (adapted from Monroe and Carter, 2012)

Group analysis: preamble

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

name	Severe (24)	Moderate (47)	mild dementia	normal
n.total	24	47	161	4
n. female	20	36	125	
n. male	4	11	36	
age-fem.	87.4 (8.7)	84.5 (7.3)	84.7 (6.1)	
age-male	79 (5.3)	82.0 (4.5)	81.6 (6.2)	





Statistical comparison between groups: Kolmogorov–Smirnov test

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Severe (24)						Moderate (47)				mild dementia (161)															normal (4)					

	severe_vs_moderate
num_words	(24,47) KS=0.452, p=0.002 **
ttr	(24,47) KS=0.452, p=0.002 **
tokens_to_turns_ratio	(24,47) KS=0.496, p<0.001 ***
num_types	(24,47) KS=0.388, p=0.012 *
brunet_.134	(24,47) KS=0.495, p<0.001 ***
brunet_.172	(24,47) KS=0.473, p<0.001 ***
brunet_.185	(24,47) KS=0.433, p=0.003 **
upos.ADJ.perc	(24,46) KS=0.408, p=0.007 **
upos.DET.perc	(24,47) KS=0.371, p=0.017 *
upos.INTJ.perc	(24,46) KS=0.37, p=0.019 *
upos.SCONJ.perc	(22,45) KS=0.484, p=0.001 **
dep_dist.avg	(24,47) KS=0.516, p<0.001 ***
max_depth.avg	(24,47) KS=0.434, p=0.003 **
max_depth.std	(24,47) KS=0.371, p=0.017 *

	severe_vs_mild
num_words	(24,161) KS=0.408, p=0.001 **
num_hapax	(24,161) KS=0.339, p=0.011 *
Ttr	(24,161) KS=0.39, p=0.002 **
tokens_to_turns_ratio	(24,161) KS=0.415, p<0.001 ***
num_types	(24,161) KS=0.39, p=0.002 **
Density*	(24,161) KS=0.306, p=0.031 *
nouns_to_verbs_ratio	(24,161) KS=0.319, p=0.022 *
brunet_.134	(24,161) KS=0.454, p<0.001 ***
brunet_.172	(24,161) KS=0.387, p=0.003 **
brunet_.185	(24,161) KS=0.375, p=0.004 **
upos.ADJ.perc	(24,157) KS=0.402, p=0.001 **
upos.ADP.perc	(23,160) KS=0.321, p=0.024 *
upos.DET.perc	(24,161) KS=0.416, p<0.001 ***
upos.INTJ.perc	(24,160) KS=0.377, p=0.004 **
upos.NOUN.perc	(24,161) KS=0.364, p=0.005 **
upos.SCONJ.perc	(22,155) KS=0.342, p=0.015 *
va.Mood.Sub.perc	(24,161) KS=0.442, p<0.001 ***
dep_dist.avg	(24,161) KS=0.489, p<0.001 ***
max_depth.avg	(24,161) KS=0.4, p=0.002 **
Szmrecsanyi	(24,161) KS=0.301, p=0.035 *

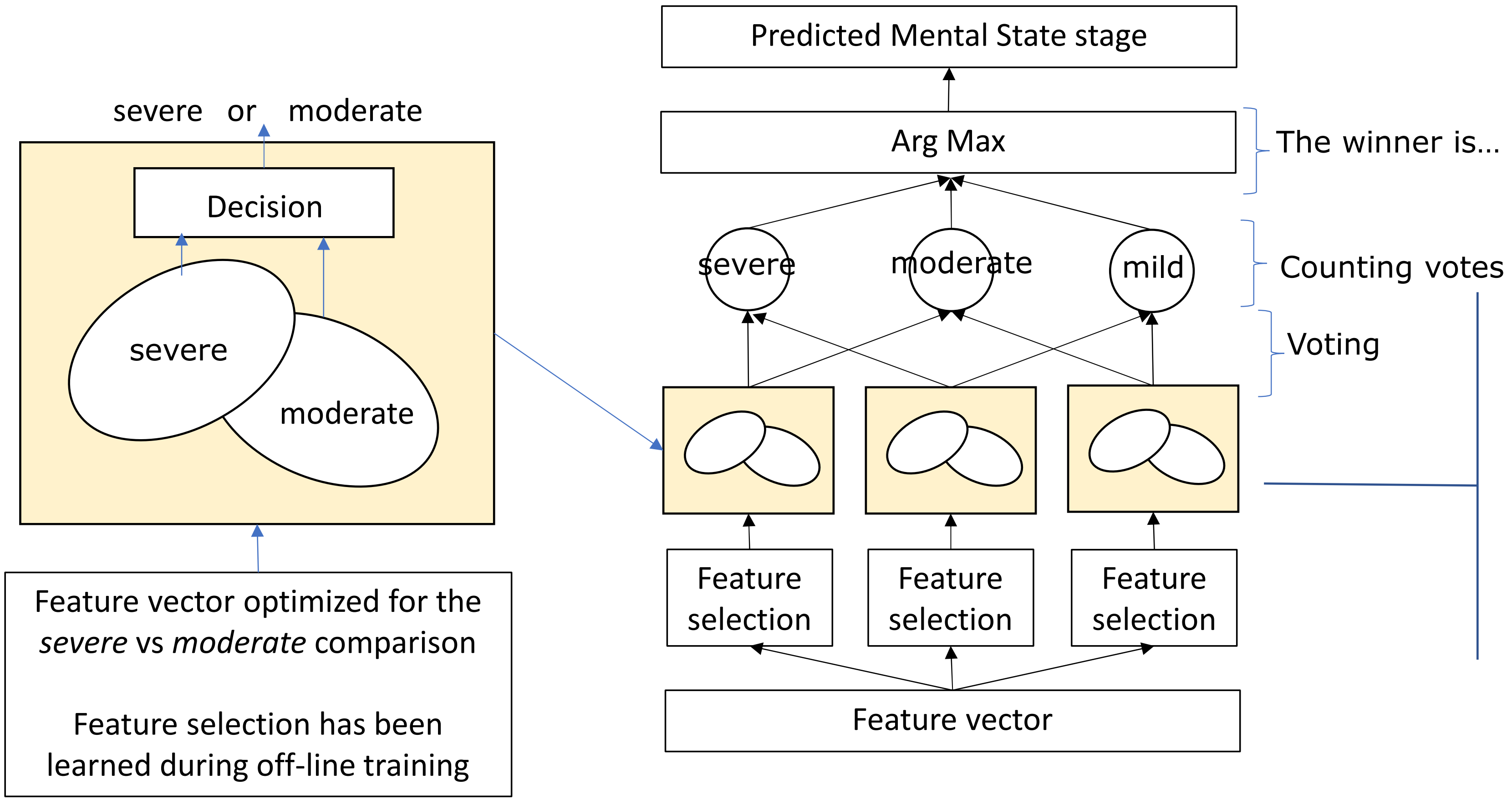
moderate_vs_mild
No statistical significant difference in any feauture between moderate and mild

inadequacy cutoff values?

Exploring classification: preamble

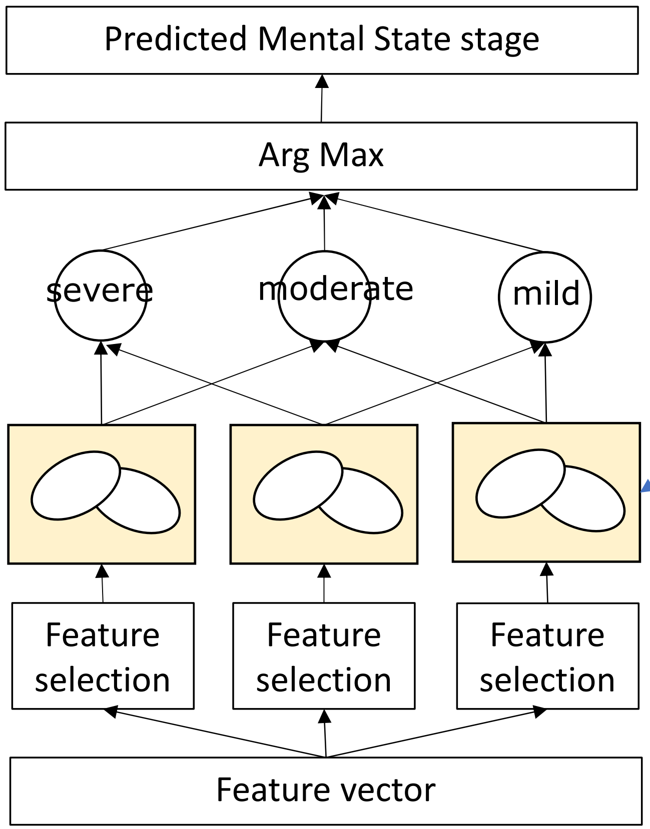
- We deal with a 3-classes classification problem
  - severe, moderate, mild stages
- Classes are quite unbalanced (24, 47, 161), therefore:
  - Accuracy (Recall) is not suitable
  - **F1 is better**
  - $\boldsymbol{F1 = \frac{2 * Recall * Precision}{Recall + Precision}}$**, i.e. the harmonic mean of Recall and Precision,** where:
    - $Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$
    - $Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$

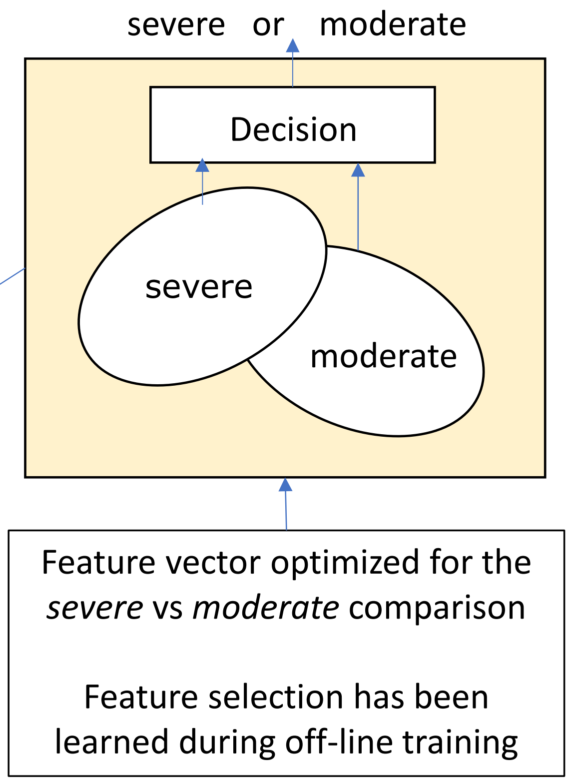
Exploring classification



The 3-classes classifier (for severe, moderate, mild stages) is composed by 3 binary classifier (one for each couple of groups

Exploring classification

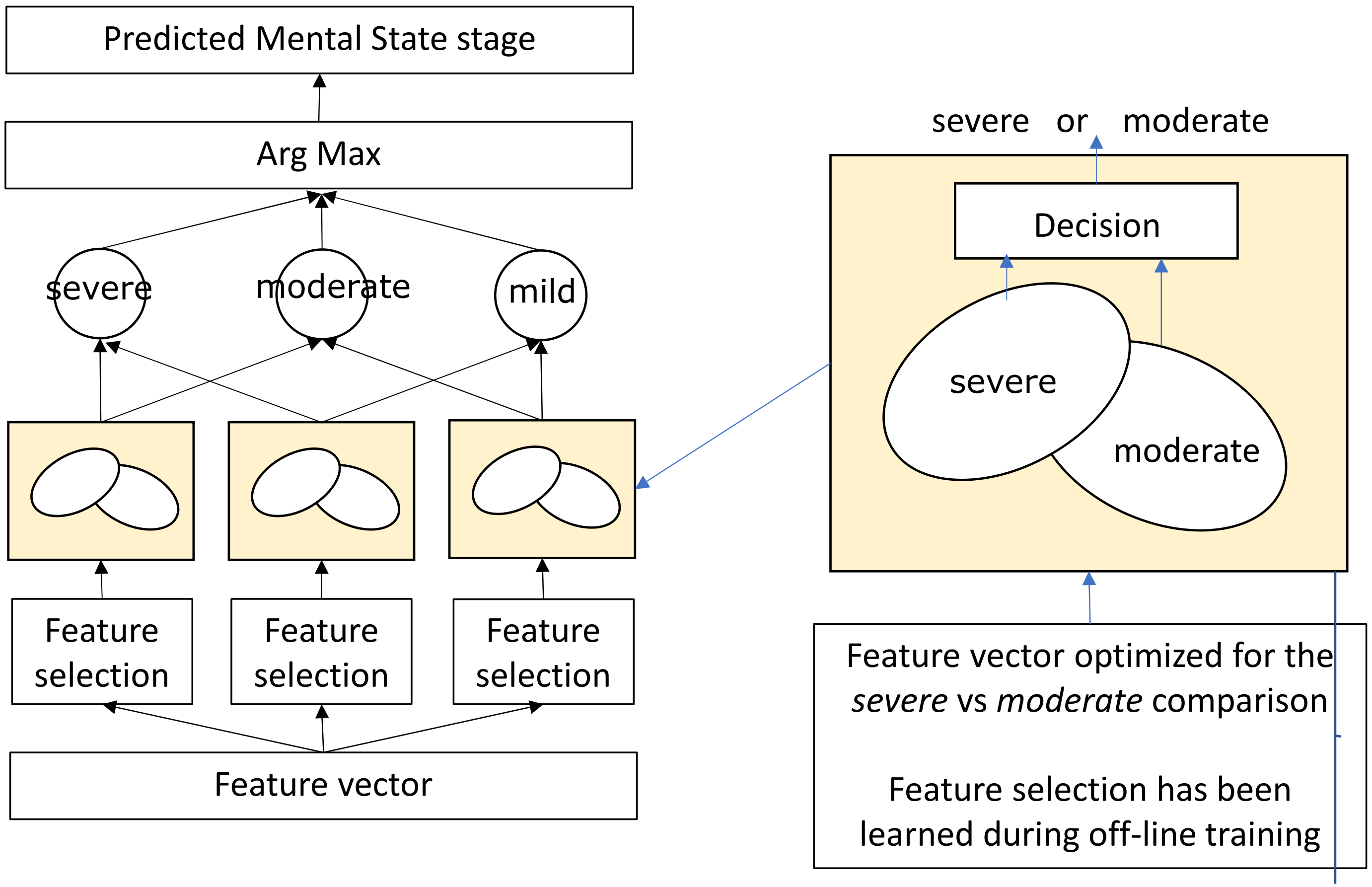




Each binary classifier outputs the most likely label

Gaussian Mixture Model likelihood, trained with features coming from the couple of groups associated to the binary classifier

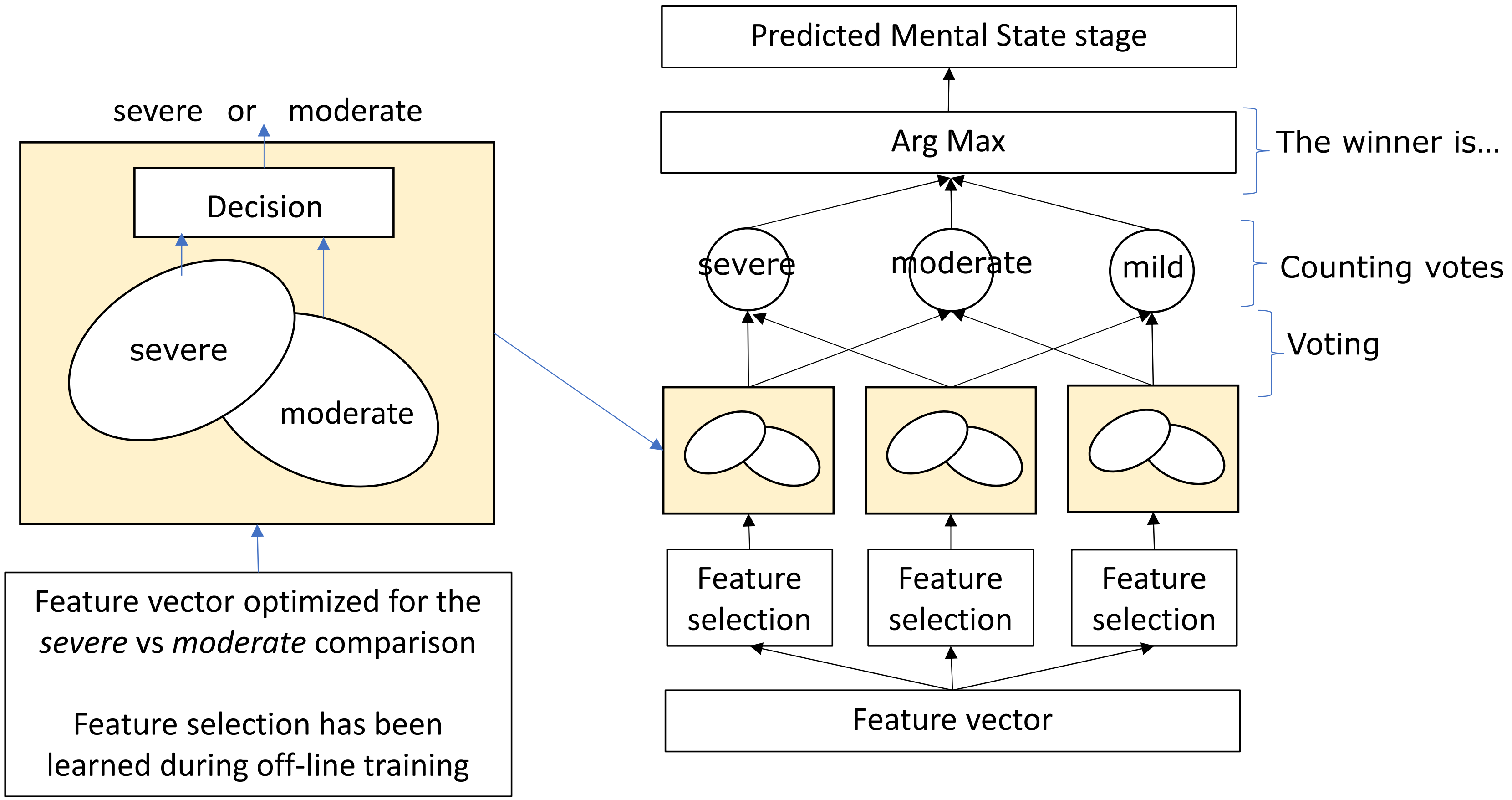
Exploring classification



Training of each binary classifier includes automatic feature selection:

- The (minimum number of) features that produce the highest target metric are selected
- Chosen target metric is F1

Exploring classification



Final classification is done by voting

Exploring classification: best results

Using…	3 classes	Severe vs Moderate	Severe vs Mild	Moderate vs Mild
ALL the explored features	**0.5134**	**0.7441**	**0.6366**	**0.6497**
Lexical features only	0.4996	0.7399	**0.6366**	0.6057
Only set of features (1)	0.4267	0.5886	0.5835	0.5640
Syntactic features only	0.4125	**0.7441**	0.6168	0.5282

Severe vs Mild and Moderate vs Mild perform roughly the same …. inadequacy cutoff values?

(1) {*num_turns, num_words, num_interrog, num_ellipsis, tokens_to_turns_ratio, num_types, p1verbs_ratio*}

Exploring classification: best results

Using ALL the explored features	F1	Selected features	Were they significant?
3 classes	**0.5134**		
Severe vs Moderate	**0.7441**	**dep_dist.avg**	(24,47) KS=0.516, p<0.001 ***
Severe vs Mild	**0.6366**	**nouns_to_verbs_ratio°** **INTJ.rate**	(24,161) KS=0.319, p=0.022 * (24,160) KS=0.377, p=0.004 **
Moderate vs Mild	**0.6497**	**num_words** **AGE** **TTR** **ADP.rate** **VERB.rate** **max_depth_std**	*no single feature was statistically significant*

° aka, reference rate to reality

Discussion: Exploring linear fit and correlations

- Several features are correlated to the MMSE scores
  - num_hapax
  - tokens_to_turns
  - num_types
  - density*
  - ADJ, INTJ, VERB rates, SUBJ rate
  - dep_dist.avg, max_depth.avg
- but each one explains a weak portion of the MMSE variance
- it is possible that other factors influence the trend of the MMSE as a function of the single feature
- It is possible that other features (not yet explored in this study) may work better

Discussion: comparison between Mental-State groups

- Several features seems to be sensitive to the *severe* vs *moderate* and *severe* vs *mild* dementia stages…
- … In terms of statistically significant difference of the distributions in the two groups (KS test)
- none of the explored features is distributed differently in the groups *moderate* vs *mild* dementia stages
  - It is possible that the mild group is too large (MMSE in [10 - 24])
  - In other words: the chosen MMSE cutoff values are not effective
  - It is possible that moderate and mild affects other features, not yet explored in this study

Discussion:
Speech-based Mental-State classification

- 2-classes classification task (F1 from 0.6366 to 0.7441) is always easier than 3-classes task (F1=0.5134)
- Achieved performance are comparable to some recent findings for 2-class classification (MCI vs Healthy Control)
- <u>Although no statistically significant difference was found in any feature alone between *moderate* and *mild* groups, a 2-class classification is still possible</u>
  - <u>Best F1 = 0.6497</u>
  - <u>Probably due to interactions among predictors</u>

Means (μ) and standard deviations (σ) of the automatic classifiers results (macro-averaged F1-score) over 10 runs for the different feature families considering the SVC technique.

Corpus	Feature set	All Tasks μ	All Tasks σ
AAC	Acoustic	0.5972	0.0366
	Demographic	0.3888	0.0239
	Readability	0.3577	0.0273
	Rhythmic	0.5228	0.0355
	Lexical	0.4960	0.0628
	Syntactic	0.6014	0.0319
	ALL	**0.7045**	0.0185
AAC	Significant	0.6662	0.0391
MCC	Acoustic	0.5847	0.0392
	Demographic	0.3888	0.0239
	Readability	0.4968	0.0456
	Rhythmic	0.5713	0.0555
	Lexical	0.4570	0.0437
	Syntactic	0.5990	0.0607
	ALL	**0.7445**	0.0164
MCC	Significant	0.7126	0.0150

Adapted from Calzà et al., 2021

Future directions: corpus augmentation

- the corpus is constantly growing

Future directions: corpus analysis

- Exploration of many other text-based linguistic features

- Exploration of different grouping criteria based on MMSE:
  - Trying different proposed cutoff values
  - Computing the cutoff values that better distinguish one group from another with respect to linguistic features

Future directions: beyond Anchise

- We have achieved nice results with transcriptions of ecological speech
- We want to test a richer set of text-based features on other voice tasks as well
- We want to test other features starting from audio recordings
  - We are confident to be able to achieve even better results
  - (project ongoing)
- We are ready to collaborate with academic and business partners

Grazie per l’attenzione!

Thanks for your attention!

A COMPUTATIONAL ANALYSIS OF SPEECH PATTERNS IN DEMENTIA: THE “ANCHISE 2022” CORPUS

Il parlato in ambito medico: analisi linguistica, applicazioni tecnologiche e strumenti clinici - Università del Salento Lecce - 15-17 febbraio 2023

BACKUP SLIDES

References

- Benedict RH , Brandt J . Limitation of the Mini-Mental State Examination for the detection of amnesia. J Geriatr Psychiatry Neurol. 1992 ; 5 (4): 233 – 237
- Benvenuti N., Bolioli A., Mazzei A, Vigorelli P., Bosca A. The "Corpus Anchise 320" and the Analysis of Conversations between Healthcare Workers and People with Dementia, Proceedings of Italian Conference on Computational Linguistics, 2020
- Bolioli A., Vigorelli P., benvenuti N., Mazzei A. Alzheimer, demenza, disturbi cognitive e psico-affettivi. Atti del 65° Congresso Nazionale SIGG (Società Italiana di Gerontologia e Geriatria), 2020, 71 s.
- Bueno-Cayo AM, Del Rio Carmona M, Castell-Enguix R, Iborra-Marmolejo I, Murphy M, Irigaray TQ, Cervera JF, Moret-Tatay C. Predicting Scores on the Mini-Mental State Examination (MMSE) from Spontaneous Speech. Behav Sci (Basel). 2022 Sep 16;12(9):339. doi: 10.3390/bs12090339Folstein MF , Folstein SE , McHugh PR . " Mini-mental state " . A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975 ; 12 (3): 189 – 198
- De Stefano A, Di Giovanni P, Kulamarva G, Di Fonzo F, Massaro T, Contini A, Dispenza F, Cazzato C. Changes in Speech Range Profile Are Associated with Cognitive Impairment. Dement Neurocogn Disord. 2021 Oct;20(4):89-98. doi: 10.12779/dnd.2021.20.4.89.
- Ferris, S.H.; Farlow, M. Language Impairment in Alzheimer's Disease and Benefits of Acetylcholinesterase Inhibitors. Clin. Interv.Aging 2013, 8, 1007–1014.
- Grut M , Fratiglioni L , Viitanen M , Winblad B . Accuracy of the Mini-Mental Status Examination as a screening test for dementia in a Swedish elderly population . *Acta Neurol Scand*. 1993 ; 87 (4): 312 – 317
- Harvan JR , Cotter V . An evaluation of dementia screening in the primary care setting . *J Am Acad Nurse Pract*. 2006 ; 18 (8): 351 – 360
- Monroe T, Carter M. Using the Folstein Mini Mental State Exam (MMSE) to explore methodological issues in cognitive aging research. Eur J Ageing. 2012 Jun 15;9(3):265-274. doi: 10.1007/s10433-012-0234-8.

- Mungas D , Marshall SC , Weldon M , Haan M , Reed BR . Age and education correction of Mini-Mental State Examination for English and Spanish-speaking elderly . *Neurology*. 1996 ; 46 (3): 700 – 706
- Nys GM , van Zandvoort MJ , de Kort PL , et al . Restrictions of the Mini-Mental State Examination in acute stroke . Arch Clin Neuropsychol. 2005 ; 20 (5): 623 – 629.
- Ostrosky-Solis F , Lopez-Arango G , Ardila A . Sensitivity and specificity of the Mini-Mental State Examination in a Spanish-speaking population . *Appl Neuropsychol*. 2000 ; 7 (1): 25 – 31 Hobson V , Spering C , Humphreys Clark J , O'Bryant SE . An evaluation of the age- and education-adjusted MMSE scores among rural dwelling Mexican American elders: the Cochran County Aging Study . *Texas Public Health J*. 2009 ; 61 (1): 22 – 24.
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In Association for Computational Linguistics (ACL) System Demonstrations. 2020
- Tombaugh TN , McIntyre NJ . The mini-mental state examination: a comprehensive review [see comment] . J Am Geriatr Soc. 1992 ; 40 (9): 922 – 935.
- Vigo I, Coelho L, Reis S. Speech- and Language-Based Classification of Alzheimer's Disease: A Systematic Review. Bioengineering (Basel). 2022 Jan 11;9(1):27. doi: 10.3390/bioengineering9010027

Influence of Age and Education on MMSE scores

- MMSE performance is moderated by **education** and **age.**
- Correction formulas have been developed to adjust the effects of age and education on MMSE scores

		Age interval				
		65-69	**70-74**	**75-79**	**80-84**	**85-89**
Scolarity	**0-4 years**	+0.4	+0.7	+1.0	+1.5	+2.2
	5-7 years	-1.1	-0.7	-0.3	+0.4	+1.4
	8-12 years	-2.0	-1.6	-1.0	-0.3	+0.8
	13-17 years	-2.8	-2.3	-1.7	-0.9	+0.3

Correction coefficients for Italian population (Magni et al, 1996)

Exploring MMSE-based groups: preliminary considerations about cut-off values for dementia stages

SLIDE NASCOSTA DI SERVIZIO

<table>
<tr><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th><th>17</th><th>18</th><th>19</th><th>20</th><th>21</th><th>22</th><th>23</th><th>24</th><th>25</th><th>26</th><th>27</th><th>28</th><th>29</th><th>30</th><th></th></tr>
<tr><td colspan="6">severe</td><td colspan="4">moderate</td><td colspan="15">mild dementia</td><td colspan="6">normal</td><td>[1]</td></tr>
<tr><td colspan="24"></td><td colspan="7">normal</td><td>[2]</td></tr>
<tr><td colspan="10">severe cognitive decline</td><td colspan="11">moderate cognitive decline</td><td colspan="4">MCI</td><td colspan="6">normal</td><td>[3]</td></tr>
<tr><td colspan="18">severe cognitive impairment</td><td colspan="6">MCI</td><td colspan="7">no impairment</td><td>[4]</td></tr>
<tr><td colspan="13">severe impairment</td><td colspan="5">moderate imp.</td><td colspan="6">Mild impairment</td><td colspan="3">bordeline</td><td colspan="4">normal</td><td>[5]</td></tr>
<tr><td colspan="18"></td><td colspan="6">MCI or early dementia</td><td colspan="7">healthy control (normal)</td><td>[6]</td></tr>
<tr><td colspan="31"></td><td></td></tr>
<tr><td colspan="10">severe AD</td><td colspan="5">soderately severe AD</td><td colspan="6">moderate AD</td><td colspan="6">mild AD</td><td colspan="4"></td><td>[7]</td></tr>
</table>

[1] Vigorelli, Corpus Anchise, 2022

[2] Regione Emilia Romagna, *Strumenti per la valutazione del paziente con demenza, ottobre 2000:*

- «The cutoff score is 23-24, and most non-demented older people rarely score below 24», p. 14
- «A corrected score above 24 is considered normal», p. 15

[3] De Stefano et al., 2021

[4] Tombaugh and McIntyre, 1992

[5] Chopra et al., 2007

[6] Beltrami et al., 2018

[7] National Institute For Health And Clinical Excellence . *Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease (Review of TA 111), UK, November 2009, p. 2*

Some notes about the Mini-Mental State Examination (MMSE)

MMSE (Folstein et al., 1975) is a commonly administered measure of global cognitive functioning that is used for:

- tracking changes
- screening
- measuring outcome in clinical trials.

MMSE comprises a series of questions and the performance of some actions which can be classified into five components based on **verbal questions** (Orientation, Registration, Attention and Calculation, Recall, and Language)

It is therefore to be expected that there is a relationship between MMSE scores and language components (Bueno-Cayo et al., 2022)