Automatic Detection of Linguistic Indicators As a Means of Early Prediction of Alzheimer's and of Related Dementias: A Cross Linguistic Analysis

iASiS Big Data for Precision Medicine

Vassiliki Rentoumi*, PhD[1], George Paliouras, PhD[1], Konstantinos Bougiatiotis, MSc[1], Dimitra Arfani[2], Katerina Fragkopoulou[2], Spyridoula Varlokosta PhD[2] and Peter Garrard, PhD, FRCP[3]
[1]NCSR 'Demokritos', Athens, Greece, [2]National and Kapodistrian University of Athens (UoA), ATHENS, Greece, [3]St George's, University of London , London, United Kingdom, *vassiliki.rentoumi@gmail.com

Background

- Alzheimer's Disease (AD) and other types of dementia are associated with **changes in spoken language**
- Language Evaluation is **time consuming** and in most cases **subjective**

Aims of the study

- To adopt a **computational approach** based on **machine learning** (ML) to analyze **language samples** from native speakers of **English** and **Greek** in order to **automatically detect early indicators of AD**
- To identify **AD-induced language characteristics** that are either **cross-linguistic** or **language-specific**

Materials and Methods

Speech samples were taken from **three sources**: Two from archived (English) language resources and one (Greek) collected for this project. Participants were shown the "cookie theft" picture[1] and were asked to describe what they could see happening.

Analytical Approach

Feature Extraction: Bag of Words assumption (BoW), Part of Speech (PoS) tags, Lexical Variation (LV) and Syntactic Complexity (SC) measures
Feature Selection: **(1) Common top ranked Information Gain (IG)** words across the three data sets; **(2) Commonly distinctive** PoS, LV, SC between AD & NC groups (p-value < 0.05) across data sets
Classification of spoken samples

Data sets	DEMENTIA (US)[2]	OPTIMA (UK)[3]	GREEK
Class: Samples (Avg. MMSE)	AD: 309 (19.0)	AD: 180 (21.1)	AD: 17 (20.0)
	NC: 246 (27.5)	NC: 248 (27.0)	NC: 14 (28.6)

Cross Linguistic Analysis Results

(1) Correlations/overlap of common top-100 ranked (IG) words

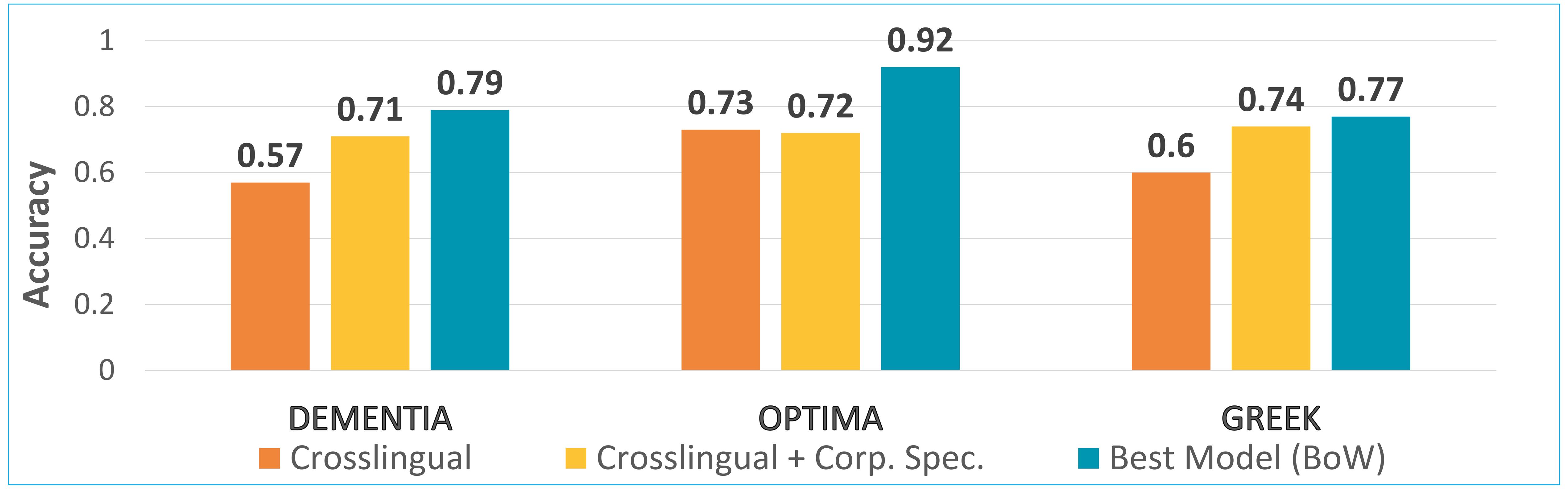
	DEMENTIA-GREEK	DEMENTIA-OPTIMA	OPTIMA-GREEK
Pearson's Correlation	0.3	0.79	0.52
Spearman's Rank Correlation	0.26	0.62	0.39
% Common words in top-100	0.49	0.36	0.41

(2) Commonly distinctive (t-Test, p<0.05) PoS, LV, SC features between AD & NC groups across data sets

Crosslingual (i.e. all samples)	US English & Greek
*CCW (AD >NC)	Adverb Freq. (AD>NC)
**CWR (AD>NC)	Nouns/Tokens (NC>AD)
Nouns Freq. (AD>NC)	Pronouns/Nouns (AD>NC)
	Mean Length Sentence (NC>AD)

*Closed Class Words Count (conjunctions, determiners, prepositions, pronouns)
**Closed Class/Open Class Words Ratio (Open Class: nouns, adjectives, verbs, adverbs)

Classification Results



- **Crosslingual**: CCW, CWR, Nouns Freq.
- **Crosslingual + Corp. Spec**: CCW, CWR, Nouns Freq., along with corpus specific statist. significant features
- **Best Model (BoW):** Tf-idf scores of lemmatized unigrams and bigrams

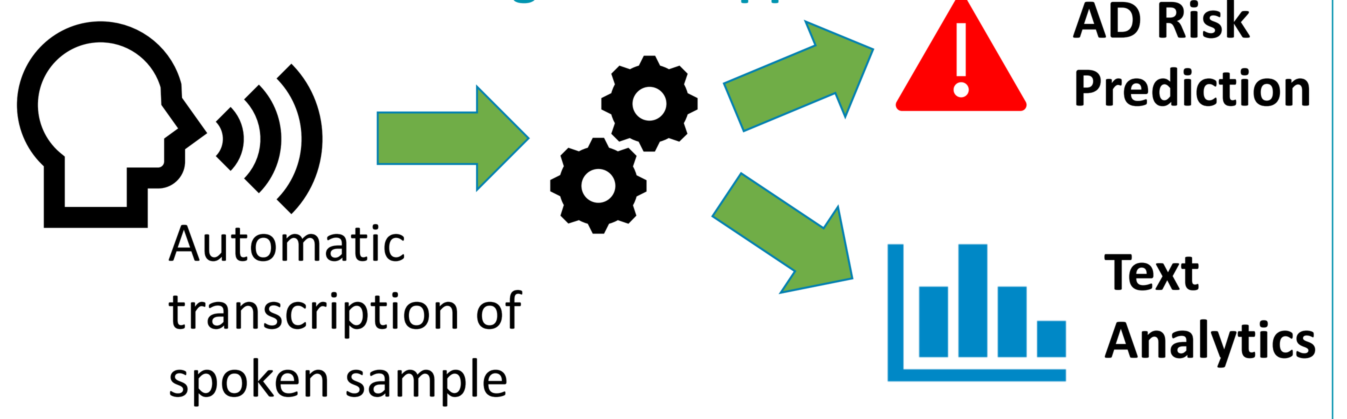
Conclusions

- ✓ **Discriminative** power of LV, SC and PoS features **verified across languages**
- ✓ **State-of-the art accuracy** of deployed system
- ✓ Development of **CogAware prototype**

Future Work

- ❖ **Fusion** of crosslingual features with best model (**complementarity of errors**)
- ❖ Preliminary findings indicate **enhanced performance**

Additional Work



- ➢ **Research prototype** for automatic language-based assessment of **AD risk factor**
- ➢ Android application and Google Drive Add-on
- ➢ **Quick** and **accurate screening** of patients for AD
- ➢ **Pilot deployed** in day-care centers (GR)

Alzheimer's Research UK The Power to Defeat Dementia · ATC Athens Technology Center · CRG Centre for Genomic Regulation · SaludMadrid · UMIACS University of Maryland Institute for Advanced Computer Studies · Leibniz Universität Hannover · Demokritos · St George's University of London · Politécnica · GECP Grupo Español de Cáncer de Pulmón Spanish Lung Cancer Group

Funded by the European Union
www http://project-iasis.eu/
@Project_IASIS

alzheimer's association AAIC 18

References

[1]Goodglass, Harold. Boston diagnostic aphasia examination: Short form record booklet. Lippincott Williams & Wilkins, 2000.
[2]Becker, J. T., Boller, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology, 51(6)*, 585-594.
[3]https://www.ndcn.ox.ac.uk/research/centre-prevention-stroke-dementia/resources/optima-oxford-project-to-investigate-memory-and-ageing