JOHNS HOPKINS
MEDICINE

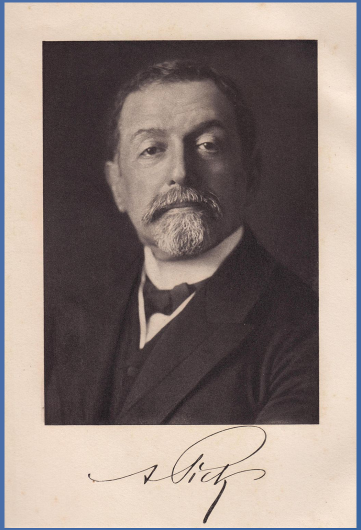
Acoustic markers of PPA variants using machine learning

Charalambos Themistocleous, Bronte Ficek, Kimberly Webster, Dirk-Bart den Ouden, Argye E. Hillis and Kyrana Tsapkini

October 21, 2018

Primary Progressive Aphasia (PPA)

JOHNS HOPKINS MEDICINE



- 1892 - Arnold Pick (1851 – 1924)
- 1982 - Marsel Mesulam’s seminal paper in 1982;
- 2011 - Gorno-Tempini, M. et al. consensus criteria;

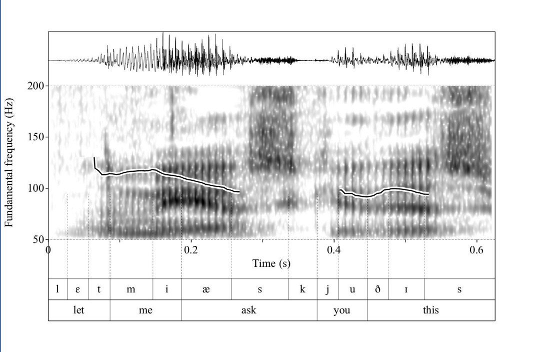
- **Non fluent PPA variant (nfvPPA)**
  - impaired speech articulation or agrammatic speech
  - impaired comprehension of syntactically complex sentences
  - diminished production of verbs and fewer syntactically complex sentences
- **Semantic PPA variant (svPPA)**
  - difficulties in confrontation naming and single word comprehension
  - impaired semantic memory of familiar objects
  - ‘empty speech’ in verbal production
- **Logopenic PPA variant (lvPPA)**
  - difficulties in word retrieval
  - difficulties in repetition of long words and phrases
  - phonological errors in speech production

PPA Variant Classification

JOHNS HOPKINS MEDICINE

- Information from imaging, language and cognitive testing as well as expert neurologist opinion.
- All these take a lot of time and cost a lot.

Speech production



Speech production conveys significant information about the speaker and the linguistic message.

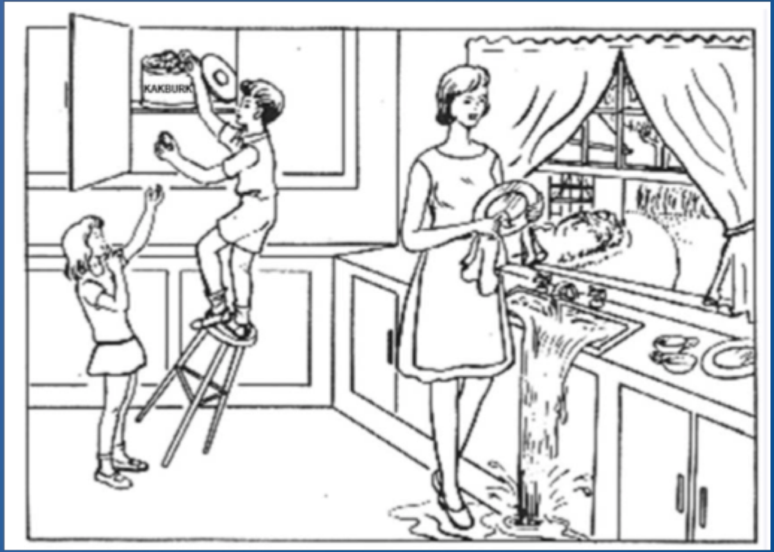
Can we classify PPA variants quickly and accurately from speech using machine learning?

Aims

1. to provide a machine learning (ML) model that can automatically subtype PPA variants and offer diagnosis tailored to specific individuals using information from speech and language
2. to understand speech and language characteristics of PPA variants

Acoustic features

JOHNS HOPKINS
MEDICINE



Connected speech can convey a stunning amount of information that can be used as a biomarker for PPA variant classification.

Speech acoustics machine learning

JOHNS HOPKINS MEDICINE

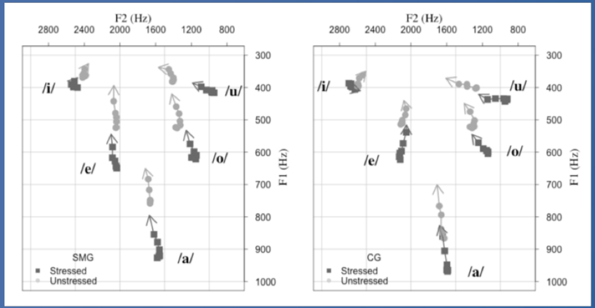
Machine learning and acoustic features to classify dialects and language varieties, e.g.,

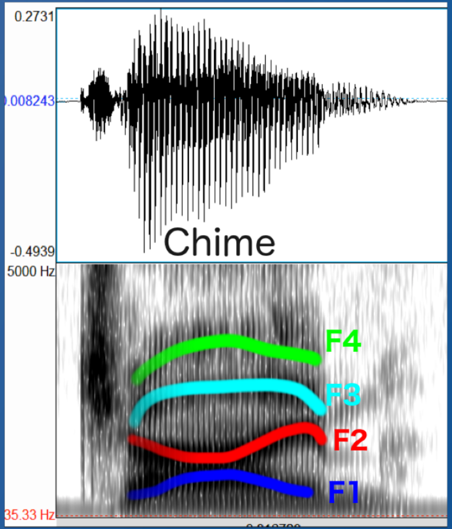
- Prosody: Intonation (Wright, Saxena, Sheppard, & Hillis 2018, Themistocleous, 2011, 2016), Final Lengthening (Themistocleous 2014)
- Vowel spectra, vowel formants, and formant dynamics (e.g., den Ouden, Galkina, Basilakos, & Fridriksson, 2018; Themistocleous, 2017a, 2017b)
- VOT spectra of stops /p t c k/ (Themistocleous 2016)
- Fricative spectra / f v θ ð s z ç ʝ ɣ x ʃ ʒ / (Themistocleous 2017)
- Sonorants, e.g., nasals /m n/, rhotics /r/ and laterals /l/ (Themistocleous, Fyndanis, Tsapkini, submitted)

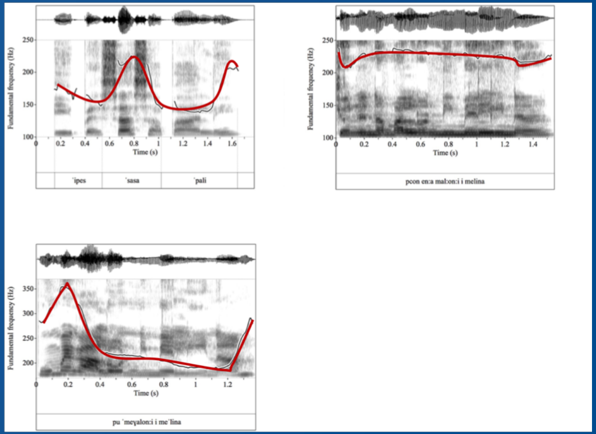
Supervised ML methods: Artificial Neural Networks, SVMs, Random Forests, C5.0, etc. for classification tasks.

Speech acoustics earlier research

JOHNS HOPKINS
MEDICINE

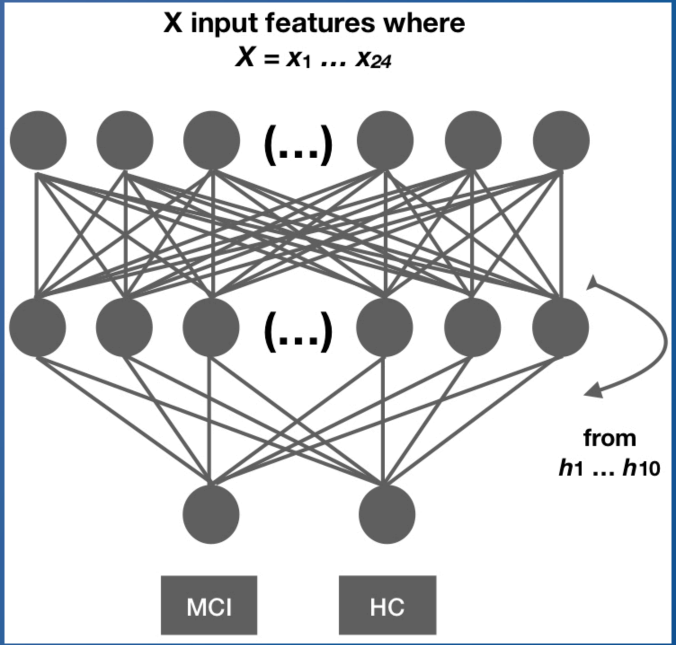






Speech acoustics - MCI vs. HC

Classification of Mild Cognitive Impairment vs. Healthy Controls: **75% classification accuracy.**



Themistocleous, Eckerström, and Kokkinakis (submitted).

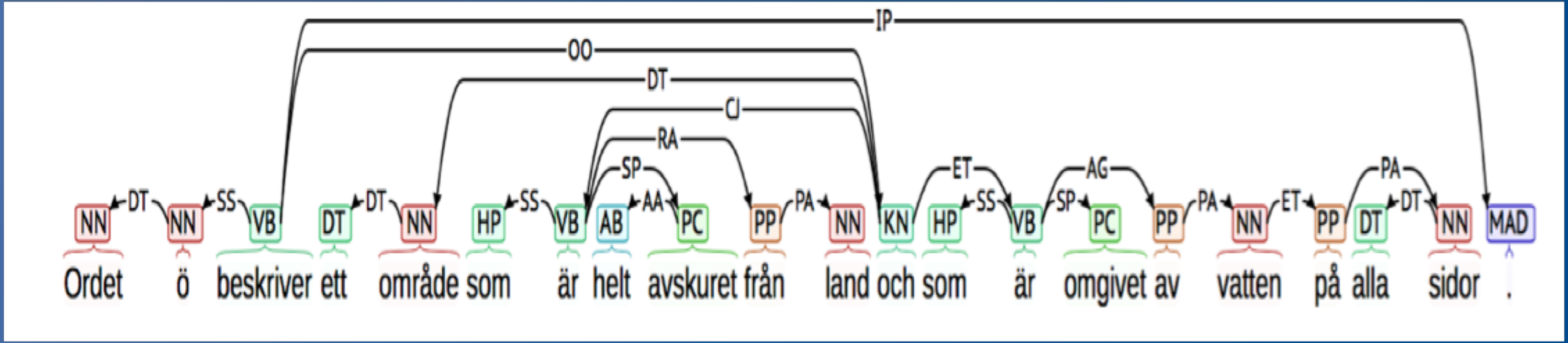
Acoustic markers

- **Vowel formants:** We measured the first five formant frequencies (F1...F5).
- **Formant dynamics:** measurements of F1...F5 formant frequencies in steps of 5 from the onset of the vowel (time = 1) to the offset of the vowel (time 100): i.e., 1, 5, 10 . . . 100
- **Vowel Duration:** the duration of vowels.
- **Pause duration:** (Mack et al. 2015)
- **Intonation:** Fundamental frequency. (F0). We calculated the mean F0, minimum F0 and maximum F0 for each vowel production.
- **Voice Quality:** H1–H2, H1–A1, H1–A2, H1–A3. Harmonic and spectral amplitudes measures of voice quality were extracted from the vowels.

Computational grammars

JOHNS HOPKINS
MEDICINE

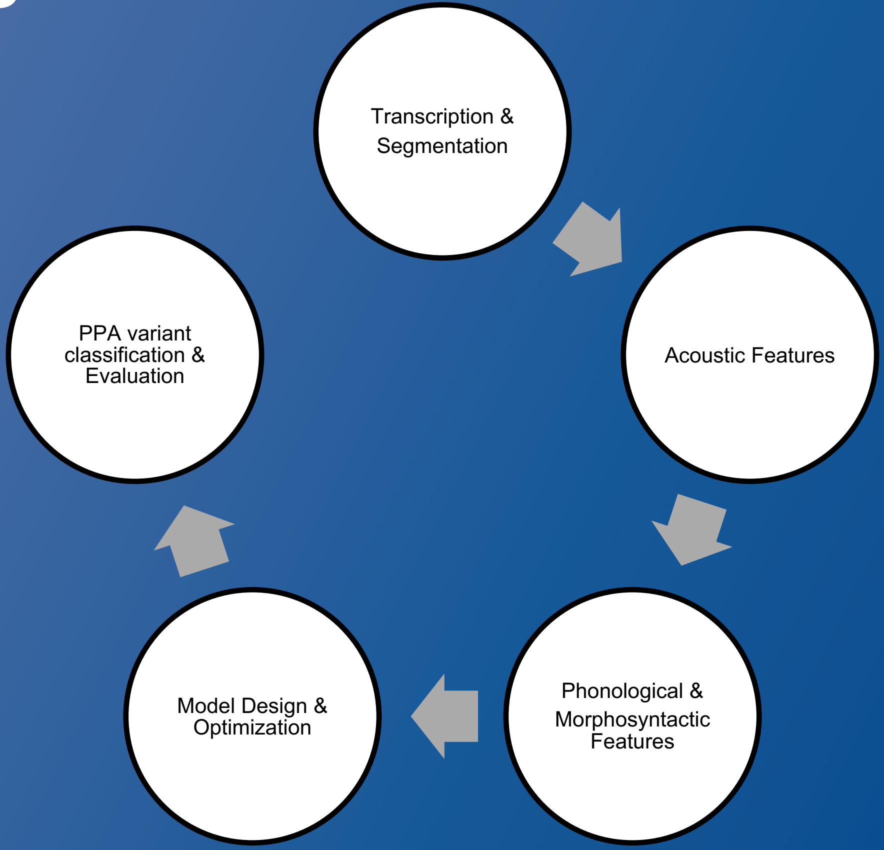
Predictors	Description
Phoneme-to-word ratio	The number of phonemes to number of words for each speaker.
Part of Speech ratio	**noun-verb ratio, noun-adjective ratio, noun-adverb ratio, noun- pronoun ratio, verb-adjective ratio, verb-adverb ratio, verb-pronoun ratio, adjective-adverb ratio, adjective-pronoun ratio,** and **adverb-pronoun**.
Distribution of POS	The proportion of nouns, pronouns, verbs, adjectives, and adverbs with respect to the total number of words per participant. This is a measure of the overall usage of each POS per speaker.



Phonological and morphological markers

- **Phonemes-to-word ratio:** e.g., Do speakers prefer long or short words?
- **Content words:** e.g., Nouns, verbs, adjectives, and adverbs.
- **Function Words:** e.g., Conjunctions, e.g., and, or, and but; Prepositions, e.g., de, in, pre and of; articles, the and a/an; Pronouns such as he/she/it.
- **Part of Speech Ratio:** noun/vowels: preference for vowels or nouns, pronouns/nouns, etc.

Methods

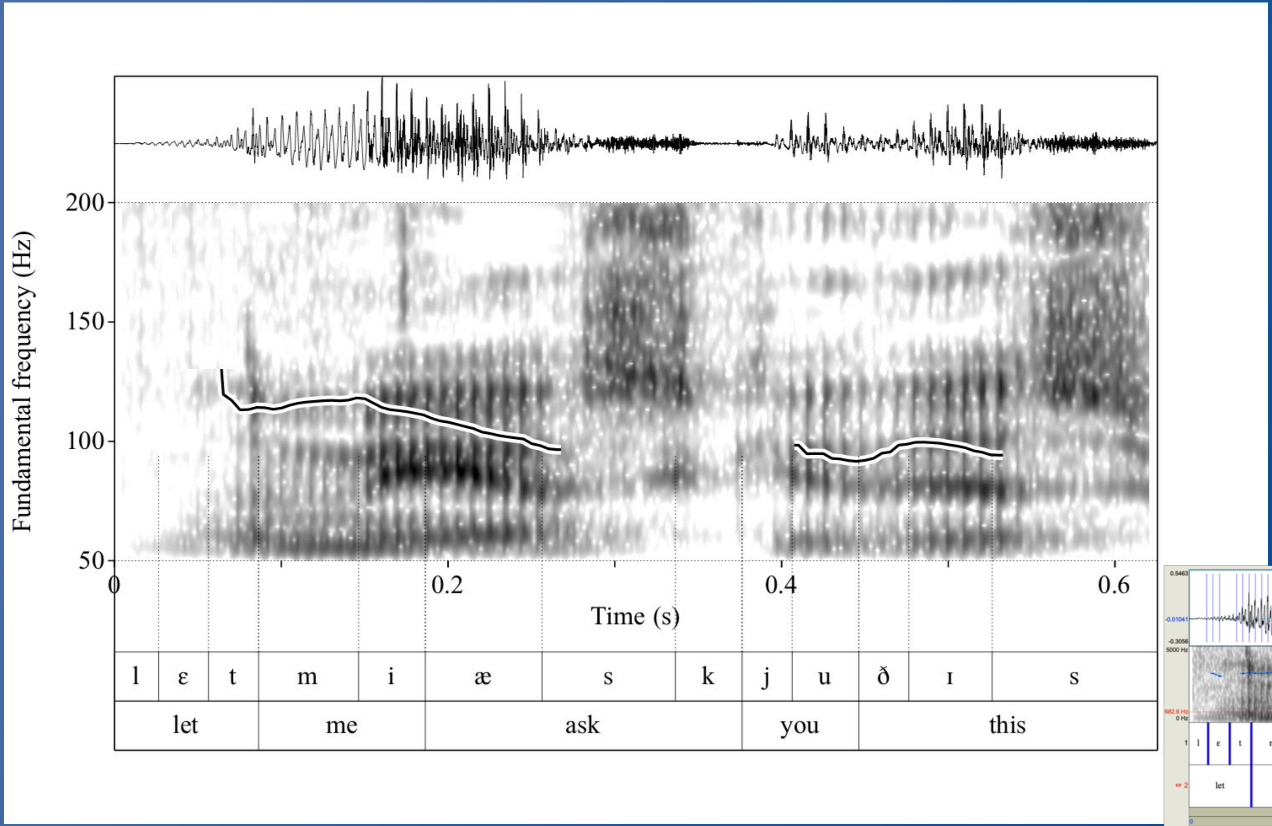

Transcription & Segmentation
Acoustic Features
Phonological & Morphosyntactic Features
Model Design & Optimization
PPA variant classification & Evaluation

Participants

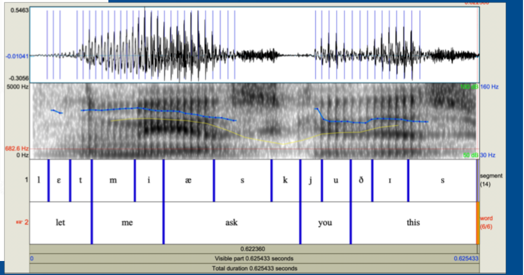
Demographic information of the **36 participants** (for age, variant, education, years post onset of the condition, language severity, and total severity the mean and the standard deviation in parenthesis is provided).

Variant	svPPA	lvPPA	nfvPPA
Female	5	8	6
Male	4	8	5
Total	9	16	11
Age	67 (6)	68 (8)	69 (6)
Education	16 (2)	17 (2)	16 (1)
Language severity	2.27 (0.6)	1.39 (0.8)	2.77 (0.5)

svPPA



Fundamental frequency (Hz)
200
150
100
50
0
0.2
0.4
0.6
Time (s)
l ɛ t m i æ s k j u ð ɪ s
let
me
ask
you
this



l ɛ t m i æ s k j u ð ɪ s
let
me
ask
you
this

Acoustic Data | Textual Data

Feature Engineering: POS, POS ratio,

Transformations: centering, scaling

Feature Selection

Machine Learning

DNN | RF | SVM | DT

Model Optimization

Cross-validation

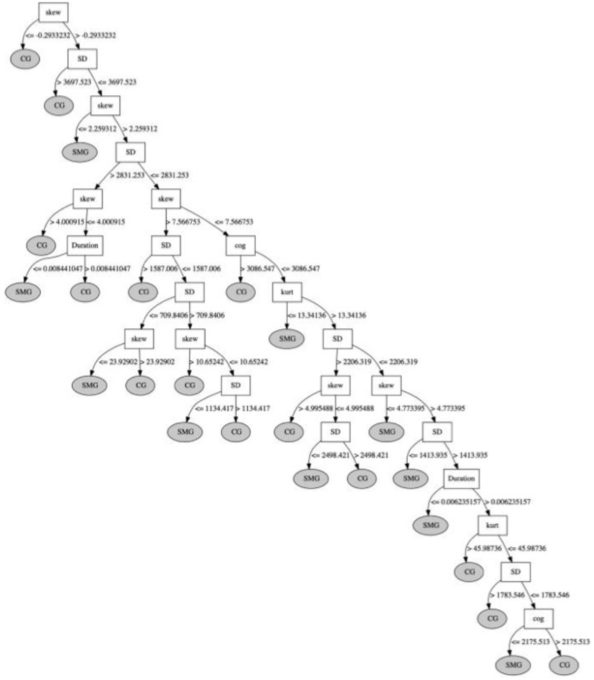
80% train data | 20% test data

Best Model Selection

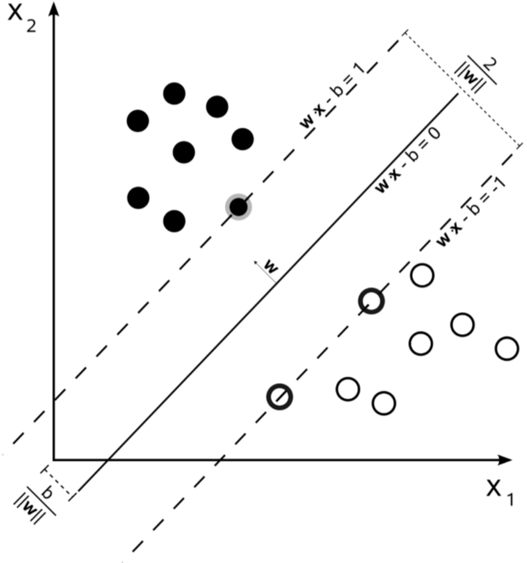
Models

- Decision Trees
- Support Vector Machines
- Random Forests

Decision Trees, Random Forests, SVMs



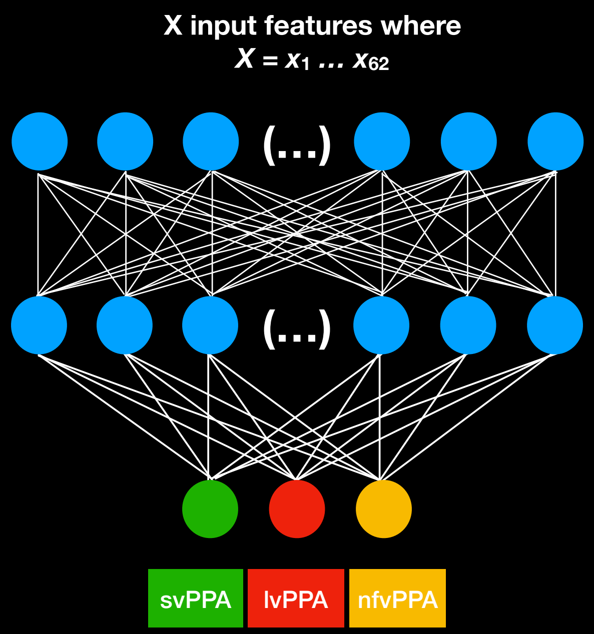
Example of a Decision Tree, using C5.0 (Themistocleous, 2016)



Example of an SVM hyperplane.

Model optimization - DNN

JOHNS HOPKINS MEDICINE



	Layer No	Layer type	Units	Activation
Input	1	Dense	150	ReLu
Hidden	17	Dense	150	ReLu
Output	1	Dense	3	Softmax

DNN: We run several neural network architectures optimizing for the number of hidden layers, dropout, optimizer, and batch size. We run each network for 100 epochs based on the validation loss.

The final DNN consists of 17 hidden layers in addition to the input and output layers.

Input and hidden layers: densely-connected with 150 units - Rectified Linear Unit (ReLu) activation

output layer had 3 units - softmax activation

Model optimization

Models were optimized:

- **DT:** DT models are provided here as a comparison model and their output is reported without optimizations.
- **SVM:** SVMs models were evaluated with both linear and non-linear kernels and optimized for the number of kernels by running the SVM models with 1 - 300 kernels (final: 14 Linear Kernels).
- **RF:** RF models were evaluated by optimizing for the number of trees from 1 - 300 trees (final: 14 trees).

Evaluation and optimization

JOHNS HOPKINS MEDICINE

- **Cross-validation:** We employed a grouped three fold cross-validation: In a 3-fold cross-validation, the data are randomized and split into three different folds and the network is trained three times. In each training setting, a different part of the available data is hold out as a test set.
- **Validation Split 80%-20%:** The 80% of the data serves as a training set and the 20% as an evaluation set.

Evaluation

Predictions of the machine learning model	**vs.**	**Expert classification from clinicians (clinical evaluation from imaging, etc).**

- *Note:* If you want to release a model as a product, you may want retrain the model on the whole data set (without splitting into training and test set).
- *Note:* If you collect more data or want to add more predictors, then you main want retrain the model again.

Evaluation Metrics

	Condition positive	Condition negative
Predicted condition positive	true positive	false positive
Predicted condition negative	false negative	true negative

Table: Confusion matrix.

$$\text{Accuracy} = \frac{\textit{true positive} + \textit{true negative}}{\textit{true positive} + \textit{true negative} + \textit{false positive} + \textit{false negative}} \tag{1}$$

$$\text{Precision} = \frac{\textit{true positive}}{\textit{true positive} + \textit{false positive}} \tag{2}$$

$$\text{Recall} = \frac{\textit{true positive}}{\textit{true positive} + \textit{false negative}} \tag{3}$$

$$F_1 \ \textit{score} = 2 \times \frac{(\textit{precision} \times \textit{recall})}{(\textit{precision} + \textit{recall})} \tag{4}$$

Results from the 3-fold cross-validation

Model	Mean	SD
DNN	77%	5%
SVM	47%	16%
RF	42%	12%
DT	26%	22%

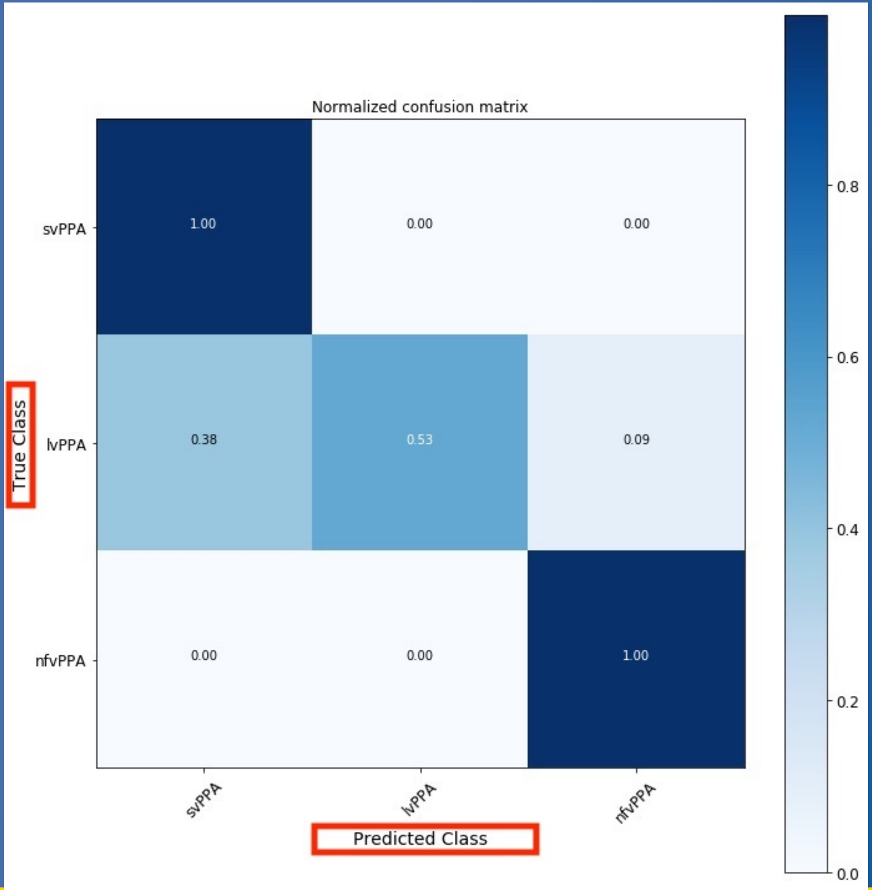
Results from the Validation split: train on the 80% and evaluate on the 20%

JOHNS HOPKINS MEDICINE

Model	Accuracy	Precision	Recall	f1 score
DNN	75%	85%	75%	74%
SVM	65%	81%	65%	61%
RF	67%	78%	67%	66%
DT	36%	51%	36%	36%

Results: Confusion matrix

JOHNS HOPKINS MEDICINE



- DNN predicted correctly all (100%) nfvPPA and svPPA variants.
- Participants with lvPPA were often misclassified as svPPA (38%) and as nfvPPA (9%), which had a cost on the overall accuracy of the DNN model.
- Probably using different markers to capture spelling, syntax, and semantics as well.
- Earlier research also provides problems with the subtyping of logopenic, which points to intrinsic variability of lvPPA.

Discussion: Importance of acoustic and linguistic markers

- Acoustic and Morphophonological predictors encompass aspects of speech production that enable PPA variant identification
- They are in agreement with the current consensus criteria for PPA subtyping

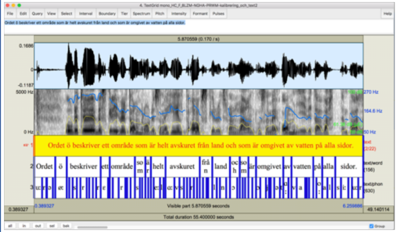
Conclusions

JOHNS HOPKINS MEDICINE

- DNN architectures provided higher classification accuracy than all other automatic classification methods employed, namely random forests, support vector machines, and decision trees.
- A machine learning model has the advantage of offering a diagnosis with error rates that can be estimated from the model.
- **Everlasting effects!** The machine learning model is flexible and its learning is non-fixed: it has the potential to improve when trained with more data, which is an advantage of machine learning models over fixed rule-based descriptions.

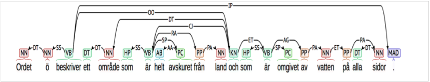
Future research

Audio Recordings



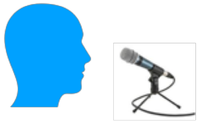
Automatic Speech Analysis; Acoustic Feature Extraction; Audio Transcription

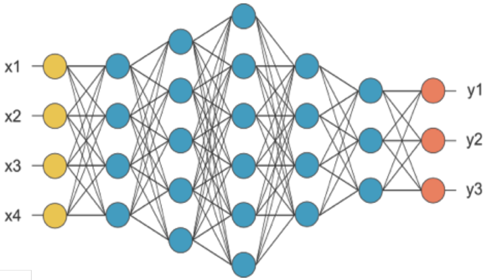
Linguistic Analysis of Transcripts



Morphology (e.g., verbs, nouns, adjectives); syntax: (e.g., Noun Phrases, Verb Phrases, encapsulation); Semantics, Topics, etc.

Linguistic Profile



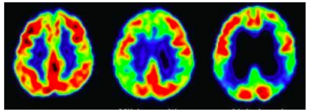


lvPPA | svPPA | nfPPA

Social Information

Sociolinguistic Properties of Speakers

Medical Information



Vital signs; psychological examinations; biomarkers; brain imaging, etc.

Thank you

Acknowledgments

- We are grateful to our patients for their unfailing participation and interest in our study,
- the Science of Learning Institute at Johns Hopkins University to KT and NIH/NIDCD R01 DC014475.
- the Riksbankens Jubileumsfond – The Swedish Foundation for Humanities & Social Sciences (NHS 14-1761:1), and
- to our sponsors, previous participants and their families, for donating their resources for the advancement of science in the pursuit of better treatments for PPA.