

Automatic PPA variant classification from transcripts and diffusion magnetic resonance imaging

Laura Lumaca,¹ Silvia Basaia,¹ Simona Musicco,¹ Annamaria Riva,¹ Elisa Canu,^{1,2} Beatrice Zorniotti,¹ Alice Corini,¹ Stefano Pisano,¹ Veronica Castelnovo,^{1,2} Edoardo G. Spinelli,^{1,2,3} Giordano Cecchetti,^{1,2,3,4} Alma Ghirelli,^{1,2} Francesca Caso,² Giuseppe Magnani,² Paola Caroppo,⁵ Sara Prioni,⁵ Cristina Villa,⁵ Massimo Filippi,^{1-4,6} Federica Agosta^{1,2,3}

¹ Neuroimaging Research Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy; ² Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy; ³Vita-Salute San Raffaele University, Milan, Italy; ⁴ Neurophysiology Service, IRCCS San Raffaele Scientific Institute, Milan, Italy; ⁵ Unit of Neurology 5, Clinical Neuropathology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy; ⁶ Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy.

INTRODUCTION AND AIMS

Primary Progressive Aphasia (PPA) is a neurodegenerative syndrome characterized by progressive language impairment.

According to current diagnostic criteria, three clinical variants are recognized: nonfluent/agrammatic (nfvPPA), semantic (svPPA), and logopenic (lvPPA). Differential diagnosis relies mainly on neuroimaging and speech evaluation, which is time-consuming, subjective, and difficult to standardize.

This study aimed to support PPA classification by developing an end-to-end pipeline that automatically transcribes spontaneous speech, extracts quantitative linguistic features and integrates them with diffusion MRI measures.

MATERIALS AND METHODS

	Healthy controls (HC)	PPA	nfvPPA	svPPA	lvPPA
N	91	94	38	36	20
Age [y]	43/48	39/55	14/24	16/20	9/11
Sex [M/W]	67.47±5.45 (56.25–81.41)	68.88±8.46 (42.06–83.93)	67.91±8.59 (51.58–83.93)	66.23±8.49 (42.06–81.63)	72.09±6.93 [^] (56.34–81.32)
Education [y]	11.80±4.17 (5–22)	11.04±4.70 (3–22)	9.11±4.83 [*] (3–22)	12.42±4.42 [§] (5–18)	12.25±3.77 (5–17)
Disease Duration [m]	-	36.71±19.11 (3.02–126.59)	28.55±13.15 (3.02–68.01)	45.76±21.90 [§] (11.33–126.59)	35.91±16.70 (5.59–62.06)

Table 1. Demographic features of the subjects enrolled in the study. ^{*} significant difference from HC; [§] significant difference from nfvPPA; [^] significant difference from svPPA

All participants completed a spontaneous speech task (*Picnic Scene*), which was audio-recorded for subsequent analysis.

- Recordings were manually transcribed using traditional methods (**Figure 1**).

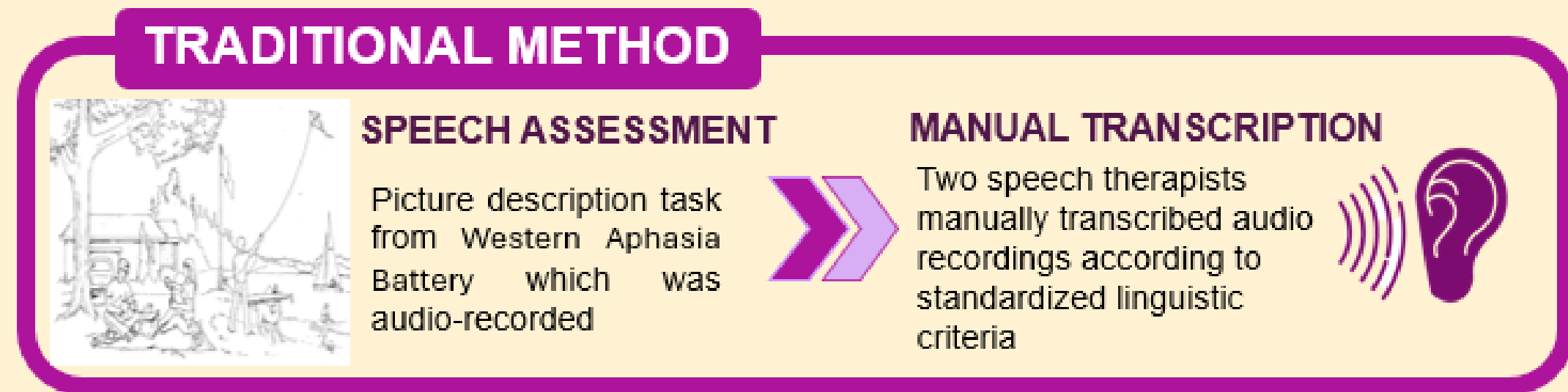


Figure 1. Schematic workflow illustrating the steps of manual transcription from audio recordings

The automatic speech recognition pipeline was built as follows:

All recordings were converted to .wav format, and clinician interventions were removed.

We tested two preprocessing pipelines for the audio:

- Method 1 – resampling to 16 kHz;
- Method 2 – resampling to 16 kHz and loudness normalization.

Speech was automatically transcribed using Microsoft Azure, and filled and empty pauses were detected using a voice activity detection model (**Figure 2**).

From diffusion-weighted images, fractional anisotropy maps were obtained for each subject.

Group comparisons were performed using TBSS, and mean FA values were extracted from the 13 regions shown here for diagnostic classification (**Figure3**).

PPA classification was performed using both unimodal approaches, based on linguistic or imaging features alone, and a multimodal approach combining both feature sets. (**Figure 4**)

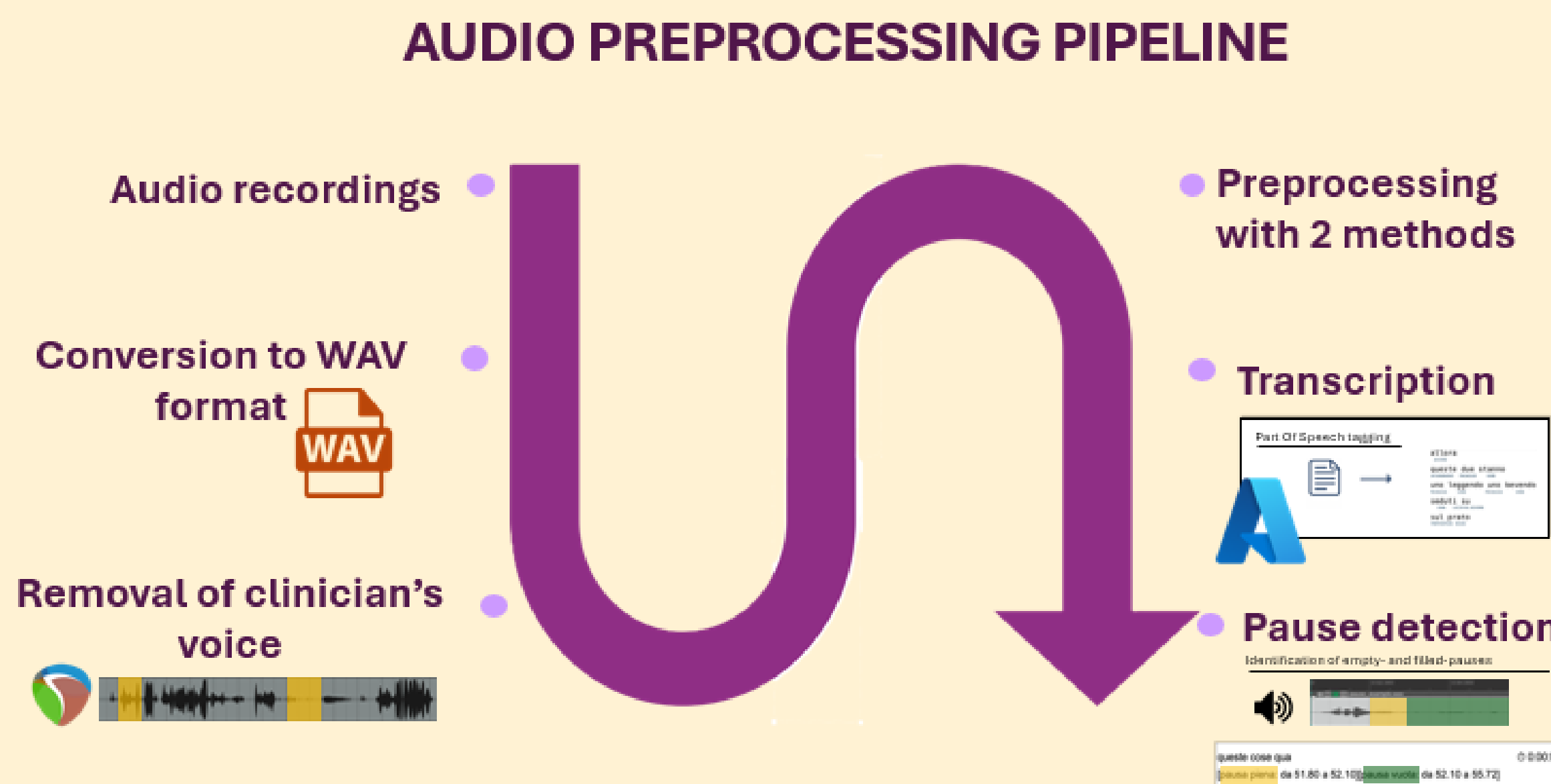


Figure 2. Automatic speech recognition (ASR) pipeline

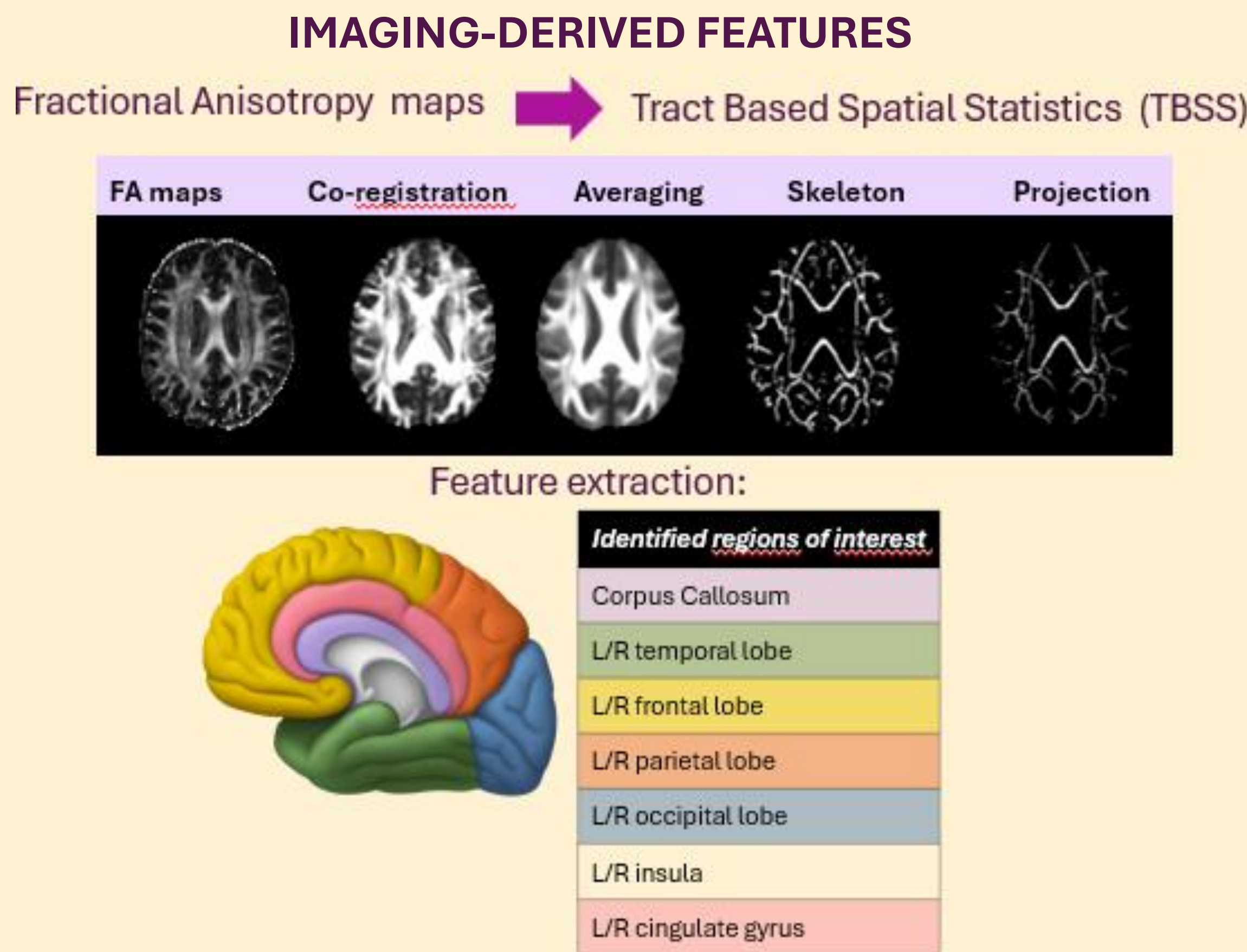


Figure 3. Imaging feature extraction pipeline

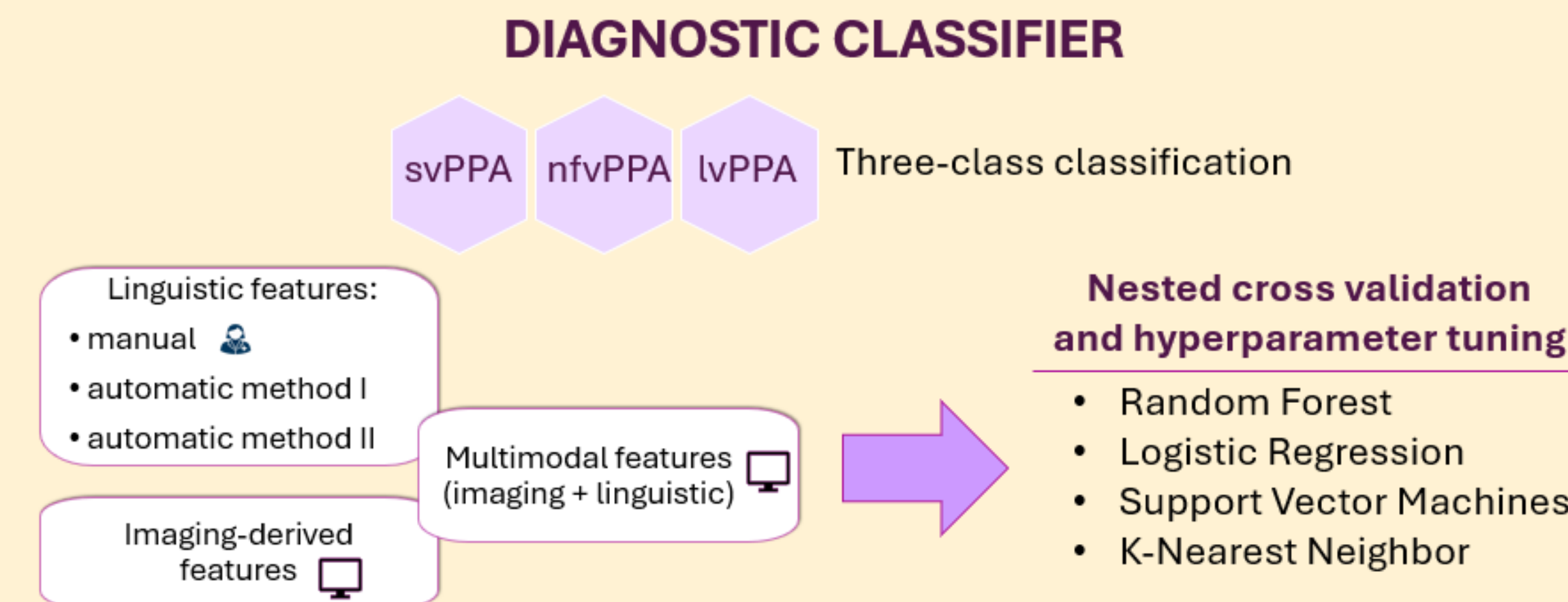


Figure 4. Diagnostic classifier pipeline

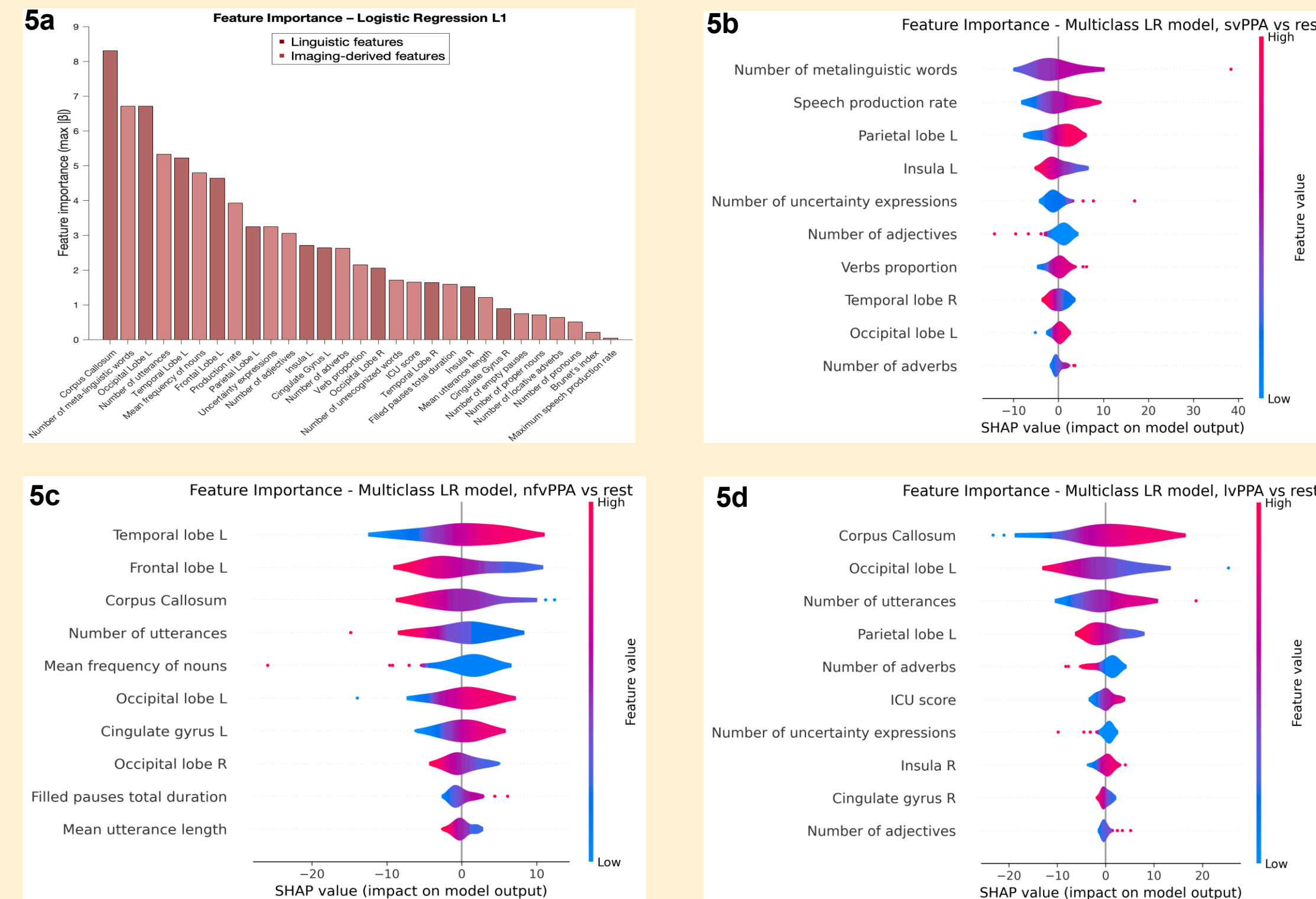
RESULTS

Linguistic unimodal classifiers based on automatically extracted features achieved higher mean balanced accuracy than those using manual features.

Overall, combining linguistic and imaging features led to improved performance, reaching an accuracy of 87% in the three-way classification of PPA variants. The best-performing model was then selected for further analyses (**Table 2**).

	Feature set	Best classifier	Mean Balanced Accuracy across validation sets (%)	Mean balanced accuracy across test sets (%)
Monomodal	Linguistic - Manual	SVM	77.3±3.0	76.5±10.4
	Linguistic - Automatic method I	SVM	81.4±7.6	81.0±9.7
	Linguistic - Automatic method II	LR	78.0±4.4	80.2±3.4
	Imaging	SVM	75.5±3.6	74.1±7.4
Multimodal	Imaging+Automatic - method I	LR	81.4±3.1	87.0±6.9
	Imaging+Automatic - method II	SVM	83.6±6.9	85.1±11.5

Table2. Classification performance of unimodal and multimodal feature sets across validation and test sets. SVM=Support Vector Machine, LR=Logistic Regression



In the multimodal classifier (**Figure 5a**), the corpus callosum was the top-ranked predictor overall. Among linguistic features, the number of meta-linguistic words, number of utterances, and mean frequency of nouns emerged as particularly important.

Figure 5b-d show the variant-specific top predictors distinguishing each PPA variant from the other two (svPPA, nfvPPA, and lvPPA, respectively).

CONCLUSIONS

- Automated speech transcription and extraction of linguistic and brain imaging features support PPA diagnosis.
- Combining linguistic and imaging data improve diagnostic accuracy.
- Automatic methods reduce analysis time and subjectivity compared with manual methods.