Politecnico di Torino

Master’s Degree in Biomedical Engineering
A.Y. 2024/2025
Graduation Session December 2025

Machine Learning-based multimodal classification of Primary Progressive Aphasia (PPA) variants from automatically transcribed speech and diffusion MRI microstructure

Supervisor
Prof. Filippo MOLINARI
Co-advisors
Prof. Salvi MASSIMO
Prof. Massimo FILIPPI
Prof.ssa Federica AGOSTA
Dott.ssa Silvia BASAIA

Candidate
Beatrice ZORNIOTTI

Table of contents

List of figures

List of tables

List of abbreviations

Abbreviation	**Definition**
AD	Alzheimer's Disease
ADC	Apparent Diffusion Coefficient
ADL	Activities of Daily Living
AI	Artificial Intelligence
ANN	Artificial Neural Network
AQ	Aphasia Quotient
ASR	Automatic Speech Recognition
AUC	Area Under the Curve
BDAE	Boston Diagnostic Aphasia Examination
BDI	Beck Depression Inventory
BET	Brain Extraction Tool
BOLD	Blood Oxygen Level-Dependent
bvFTD	behavioral variant Fronto-Temporal Dementia
CADe	Computer-Aided Detection
CADx	Computer-Aided Diagnosis
CDR	Clinical Dementia Rating
CNN	Convolutional Neural Network
CSF	Cerebrospinal Fluid
DL	Deep Learning
dMRI	diffusion Magnetic Resonance Imaging
DT	Decision Tree
DTI	Diffusion Tensor Imaging
DWI	Diffusion Weighted Imaging
DWT	Discrete Wavelet Transform
EA	Evolutionary Algorithms
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
FAB	Frontal Assessment Battery
FFT	Fast Fourier Transform
FID	Free Induction Decay
FLAIR	Fluid-Attenuated Inversion Recovery
fMRI	Functional Magnetic Resonance Imaging
FN	False Negative
FOV	Field Of View
FP	False Positive
FPR	False Positive Rate
FSE/TSE	Fast Spin-Echo/Turbo Spin-Echo
FSL	FMRIB Software Library

FTD	Fronto-Temporal Dementia
FTLD	Fronto-Temporal Lobar Degeneration
GLM	General Linear Model
GM	Gray Matter
GMM	Gaussian Mixture Model
GRE	Gradient echo
HC	Healthy Control
HMM	Hidden Markov Model
IADL	Instrumental Activities of Daily Living
ICU	Information Content Unit
IFOF	Inferior Fronto Occipital Fasciculus
ILF	Inferior Longitudinal Fasciculus
IR	Inversion Recovery
kNN	K Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LOSO	Leave-One-Subject-Out
LPC	Logistic Regression
LR	Long Short-Term Memory
LSTM	Linear Predictive Coding
LUFS	Loudness Units Full Scale
lvPPA	logopenic variant of Primary Progressive Aphasia
MAE	Mean Absolute Error
MCI	Mild Cognitive Impairment
MD	Mean Diffusivity
MDEFT	Modified Driven Equilibrium Fourier Transform
MFCC	Mel-Frequency Cepstral Coefficient
ML	Machine Learning
MMSE	Mini-Mental State Examination
MPRAGE	Magnetization-Prepared Rapid Gradient Echo
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
mRMR	Minimum Redundancy Maximum Relevance
nfvPPA	Non-fluent variant of Primary Progressive Aphasia
NLP	Natural Language Processing
NMR	Nuclear Magnetic Resonance
PALS	Progressive Aphasia Language Scale
PASS	Progressive Aphasia Severity Scale
POS	Part-Of-Speech
PPA	Primary Progressive Aphasia
PSP	Progressive Supranuclear Palsy
RAVLT	Rey Auditory Verbal Learning Test
RBF	Radial Basis Function
RF	Random Forest
RFr	Radio Frequency

RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
Rs-fMRI	Resting-state functional Magnetic Resonance Imaging
RTF	Real-Time Factor
SE	Spin-Echo
SE-EPI	Spin-Echo Echo-Planar Imaging
SF	Sampling Frequency
SGD	Stochastic Gradient Descent
SHAP	Shapley Additive Explanation
SLF	Superior Longitudinal Fasciculus
SMC	Subjective Memory Complaints
STIR	Short Tau Inversion Recovery
SVM	Support Vector Machine
svPPA	semantic variant Primary Progressive Aphasia
SydBat	Sydney language Battery
TBSS	Tract Based Spatial Statistics
tDCS	transcranial Direct Current Stimulation
TE	Echo Time
TFCE	Threshold-Free Cluster Enhancement
TI	Inversion Time
TMT	Trail Making Test
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
TR	Repetition time
TTR	Type-Token Ratio
UD	Universal Dependency
UF	Uncinate Fasciculus
VAD	Voice Activity Detection
VOSP	Visual Object and Space Perception battery
WAB-R	Western Aphasia Battery – Revised
WER	Word Error Rate
WM	White Matter
WRR	Word Recognition Rate

Abstract

Primary Progressive Aphasia (PPA) is a neurodegenerative syndrome characterized by the gradual decline of language abilities, with three main variants, nonfluent/agrammatic (nfvPPA), semantic (svPPA), and logopenic (lvPPA) that differ in their linguistic profiles and underlying pathology. In clinical practice, differential diagnosis still depends mainly on clinician, which can be time-consuming, subjective, and difficult to standardize. Automatic analysis of speech, with quantitative extraction of linguistic features, offers a way to objectively characterize patients' eloquence. In this context, artificial intelligence (AI)–based models trained on speech- and imaging-derived features may assist in the differential diagnosis of PPA, complementing traditional clinical and neuroimaging assessments.

This study aimed to automatize the diagnostic process of PPA variants. Specifically, the first objective was to define and develop a pipeline for automatically extracting linguistic features from automatic transcriptions. The second objective was to integrate these features with imaging-derived measures to build a multimodal diagnostic classifier for variant differentiation. Finally, the third objective was to compare the automatic process with the manual gold standard and to merge both approaches into a web application designed to support clinical classification.

Data were collected at IRCCS San Raffaele Scientific Institute and included 91 healthy controls and 94 PPA patients (38 nfvPPA, 36 svPPA and 20 lvPPA). Of these, 80 healthy controls and 85 patients underwent diffusion tensor (DT) MRI, and 38 controls and 81 patients completed the Picnic picture description task from the Western Aphasia Battery, designed to evaluate connected speech production.

Audio and imaging data were preprocessed to ensure normalization and quality consistency. Speech samples were automatically transcribed using Microsoft Azure, and linguistic features were extracted in Python through existing libraries and customized models and pipelines developed for this purpose. Fractional Anisotropy (FA) maps were obtained from DTI data, and Tract-Based Spatial Statistics (TBSS) was applied to derive imaging features. A multimodal classifier was then trained and cross-validated to discriminate among PPA variants. Transcription accuracy and feature extraction errors were computed by comparing the automatic system with the manual gold standard. Finally, a prototype web application was developed to integrate audio preprocessing, automatic transcription with a supervised revision step—allowing clinicians to correct transcription

errors—and subsequent automatic feature extraction, ensuring that the linguistic measures accurately reflect the original speech.

The automatic transcription system achieved a mean Word Error Rate (WER) of 12 ± 7% in HC and 19 ± 10% in PPA, with no significant differences across variants. The mean relative error in linguistic feature extraction was 12 ± 8%. The multimodal diagnostic classifier reached a balanced accuracy of 87 ± 7% in the three-class task. The prototype web application achieved a mean WER of 9 ± 3% across the two operators and the test set. Inter-rater variability was still present, but the two operators reached a mean per-subject agreement of 91 ± 4%.

The automatic speech-processing pipeline showed good agreement with clinicians, and the prototype web application enabled clinicians to rapidly refine automatic transcriptions, effectively bridging manual precision and automated efficiency. Future developments will focus on incorporating the full multimodal workflow and improving the treatment of pathological speech patterns, with the goal of delivering a scalable, clinician-friendly tool for diagnosis and longitudinal monitoring in PPA.

1 Primary Progressive Aphasia

1.1 Definition

Primary Progressive Aphasia (PPA) is a clinical neurodegenerative syndrome primarily characterized by the gradual onset and progression of language impairments, with minimal initial involvement of other cognitive or behavioral domains. The condition usually manifests as a slow deterioration in the ability to communicate, which reflects the underlying language network atrophy in the left frontal, temporal and/or parietal regions of the left hemisphere. The onset is often insidious, with symptoms emerging progressively over time; at the beginning, patients may seem cognitively intact except for their linguistic deficits. [1], [2]

The term "Primary Progressive Aphasia" (PPA) was coined in 1992 by Mesulam and Weintraub to give a more accurate clinical and terminological representation of the syndrome. The term "primary" highlights the specific involvement of cortical areas directly related to language function, "progressive" reflects the gradual worsening of the disorder and "aphasia" indicates that the primary impairment is in the language domain. [3], [4]

However, the concept of PPA had already been introduced by Mesulam earlier, in 1982, when he referred to a group of six patients, aged between 17 and 69 years, as having "slowly progressive aphasia without generalized dementia". These patients exhibited a gradual onset of aphasia, without any notable cognitive or behavioral alterations, thus differentiating the condition from Alzheimer's disease (AD) and Pick's disease (now known as fronto-temporal dementia, FTD). All patients showed involvement of the left perisylvian region of the brain, which is essential for language processing, and they were monitored over time to see how their condition developed. [5]

Going back in time, the first documented cases of progressive language impairment date back to the late 19th century. In 1892, Pick described a case of a man with progressive language impairment, who also displayed memory problems and behavioral changes, such as threatening his wife with a knife. [6] The following year, Serieux reported a case of progressive language fluency decline in a patient, but without social, memory or visuospatial impairments. [7] These early descriptions laid the groundwork for understanding the clinical manifestation of language impairment as a distinct neurodegenerative syndrome.

1.2 Clinical Variants

Patients with PPA exhibit heterogeneous speech and language profiles, reflecting the existence of a spectrum of related clinical syndromes rather than a single disorder.[8]

Already in the early 1990s, two distinct forms of progressive language impairment had been described: semantic dementia[9] and progressive non-fluent aphasia [10]. The first formal diagnostic criteria for PPA were proposed in 2001. [2] At that time, only these two variants were recognized, and patients were generally categorized accordingly, or broadly grouped under the "fluent" versus "non-fluent" distinction. However, several cases could not be accommodated within this binary classification. [11] To capture this additional linguistic profile, a new variant was introduced in 2004, termed logopenic progressive aphasia, which led to the establishment of the current three-variant model.[12]

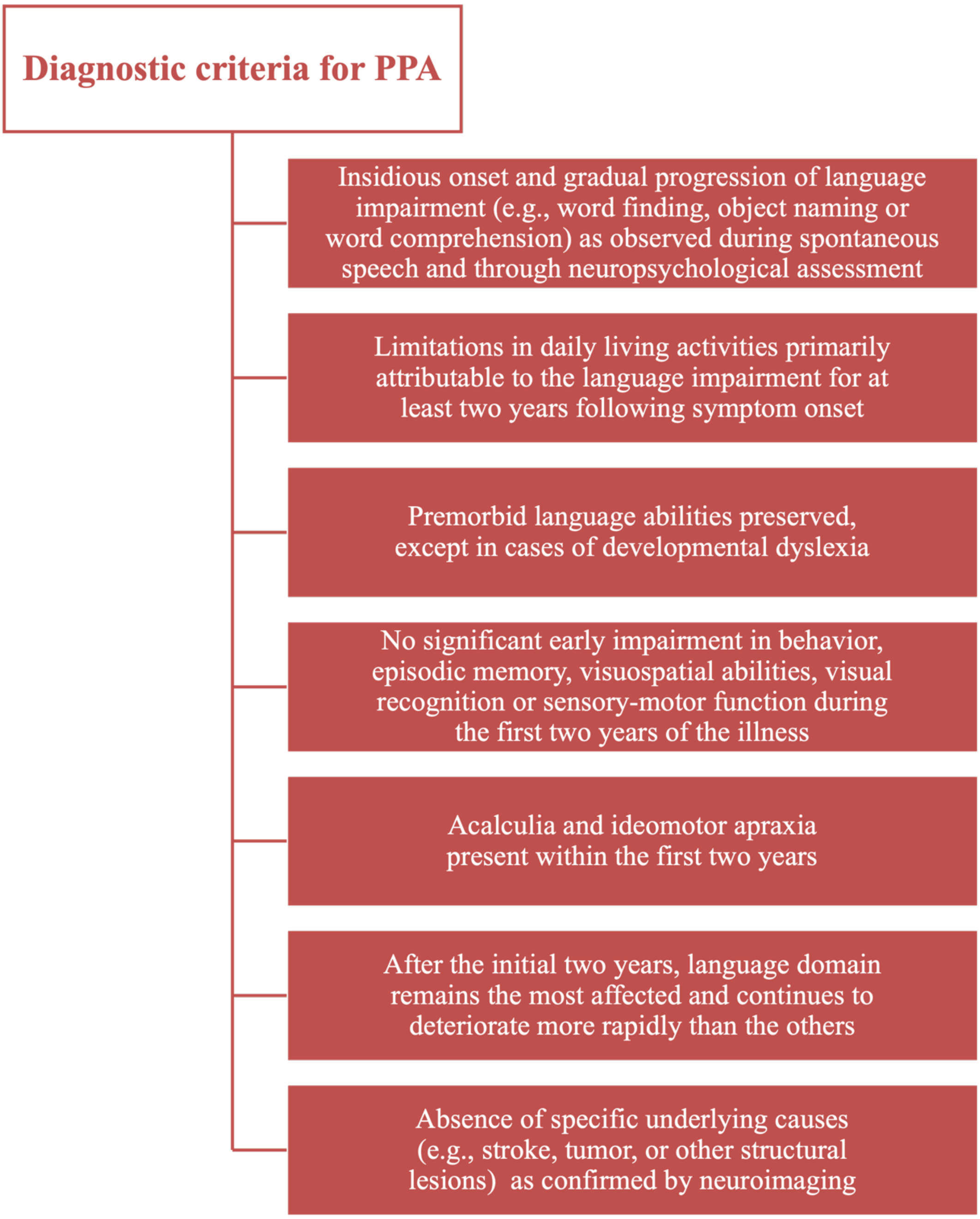
Although this tripartite classification represented a major conceptual advance, a standardized and universally accepted framework was still lacking. To address this, an international consortium of experts met repeatedly between 2006 and 2009 to develop consensus diagnostic criteria for PPA subtypes, particularly applicable to early disease stages. During this period, significant advances in neuroimaging and molecular research revealed that the distribution of cortical atrophy within the language network largely determines the clinical phenotype. [8] Consequently, neuroimaging evidence was incorporated into the evolving classification framework.

A key milestone came in 2011, when updated international consensus criteria for three variants were published. [8] These criteria aimed to standardize clinical and research procedures by defining PPA variants according to characteristic patterns of language impairment, neuroanatomical atrophy and underlying pathology. The names adopted for the three variants: nonfluent/agrammatic (nfvPPA), semantic (svPPA), logopenic (lvPPA), were carefully chosen to preserve consistency with previous literature while aligning with the nomenclature used for behavioral variant frontotemporal dementia (bvFTD), which reflects the inclusion of PPA in the larger spectrum of frontotemporal lobar degeneration (FTLD). [8]

1.2.1 Diagnostic criteria

As mentioned in the previous paragraphs, the first formal diagnostic criteria for PPA were proposed by Mesulam in 2001, outlining the fundamental clinical features of the disorder and providing a framework that preceded the later variant-based classifications. Figure 1-1

After the initial phase, other cognitive abilities may deteriorate; however, language must continue to be the most impaired domain throughout the course of the disease, deteriorating more quickly than other affected functions. Finally, neuroimaging should exclude non-degenerative causes, such as stroke, tumor or other structural brain lesions, to confirm the progressive and primary nature of the aphasia. [2]



Diagnostic criteria for PPA

Insidious onset and gradual progression of language impairment (e.g., word finding, object naming or word comprehension) as observed during spontaneous speech and through neuropsychological assessment

Limitations in daily living activities primarily attributable to the language impairment for at least two years following symptom onset

Premorbid language abilities preserved, except in cases of developmental dyslexia

No significant early impairment in behavior, episodic memory, visuospatial abilities, visual recognition or sensory-motor function during the first two years of the illness

Acalculia and ideomotor apraxia present within the first two years

After the initial two years, language domain remains the most affected and continues to deteriorate more rapidly than the others

Absence of specific underlying causes (e.g., stroke, tumor, or other structural lesions) as confirmed by neuroimaging

Figure 1-1 Diagnostic criteria for PPA. Adapted from [2]

Building upon Mesulam's original formulation, the 2011 international consensus criteria proposed by Gorno-Tempini et al. introduced a more structured and hierarchical diagnostic framework for PPA. [8] Establishing a clinical diagnosis involves a two-step process. In the first phase, patients must fulfill the general diagnostic requirements for PPA, which include three inclusion and four exclusion criteria. [8] [Figure 1-2 Figure 1-3]

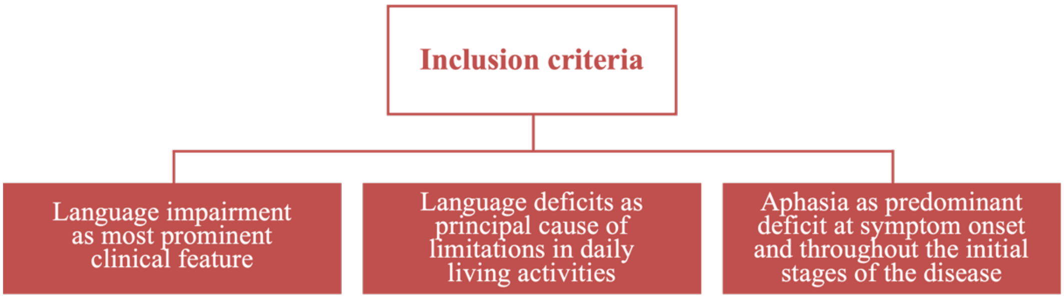

Inclusion criteria
Language impairment as most prominent clinical feature
Language deficits as principal cause of limitations in daily living activities
Aphasia as predominant deficit at symptom onset and throughout the initial stages of the disease

Figure 1-2 Inclusion criteria for the diagnosis of PPA. Adapted from [8]

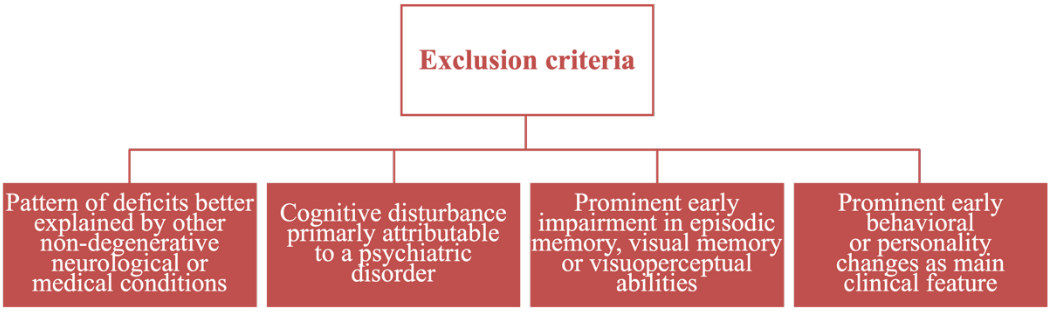

Exclusion criteria
Pattern of deficits better explained by other non-degenerative neurological or medical conditions
Cognitive disturbance primarily attributable to a psychiatric disorder
Prominent early impairment in episodic memory, visual memory or visuoperceptual abilities
Prominent early behavioral or personality changes as main clinical feature

Figure 1-3 Exclusion criteria for the diagnosis of PPA. Adapted from Gorno-Tempini et al., Neurology, 2011 [8]

Once these general conditions are met and a diagnosis of PPA is established, the next step involves classifying patients into one of the clinical variants according to three hierarchical levels: clinical, imaging-supported and definite pathologic diagnosis. At the clinical level, classification is based on the presence of specific speech and language features characteristic of each variant. The imaging-supported diagnosis requires that the clinical criteria are met, and that neuroimaging evidence shows a pattern of structural or functional alterations consistent with the expected distribution for that variant. Lastly, a definite pathological diagnosis is made when the clinical phenotype is accompanied by neuropathological or genetic findings that are consistent with a particular underlying disease, like FTLD, AD, or other defined etiologies. Importantly, the identification of a distinct pathology does not necessarily mean that the clinical syndrome is more defined, but rather that it has been connected to a known biological substrate. In the study, the authors suggest that this evaluation be performed through a short bedside language screening, although a comprehensive assessment by a speech and language pathologist is still the most reliable method to guarantee diagnostic accuracy. [8]

1.2.1.1 Non-fluent/agrammatic variant

NfvPPA is a rare neurodegenerative syndrome characterized by agrammatism in language production or effortful, halting speech. [8] Agrammatism is usually characterized by the use of short, syntactically simple phrases and the omission or substitution of grammatical morphemes, while effortful speech is characterized by slow, labored articulation and frequent pauses, which reflect an underlying motor speech planning deficit. [13]

To confirm the clinical diagnosis of nfvPPA, at least two additional supportive features should also be present. [8] These include difficulties in comprehending syntactically complex sentences, such as passives or relative clauses [13], while single-word comprehension and object knowledge remain relatively preserved. The co-occurrence of grammatical and motor speech impairments, as well as spared lexical and semantic processing, sets nfvPPA apart from the other subtypes.

For an imaging-supported diagnosis, structural or functional neuroimaging must show atrophy and/or hypoperfusion or hypometabolism primarily affecting the left posterior fronto-insular region. At the level of definite pathology, most cases are associated with frontotemporal lobar degeneration with tau inclusions (FTLD-tau), while a smaller percentage shows FTLD-TDP pathology. [8]

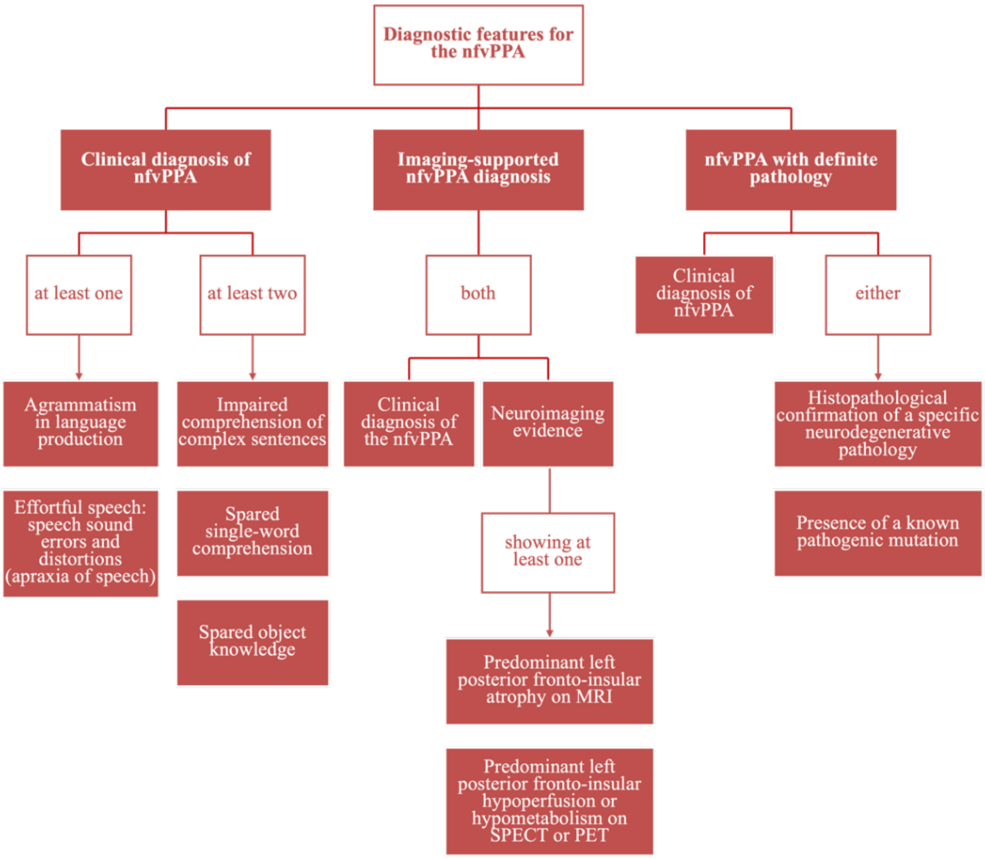

Diagnostic features for the nfvPPA
Clinical diagnosis of nfvPPA
Imaging-supported nfvPPA diagnosis
nfvPPA with definite pathology
at least one
at least two
both
Clinical diagnosis of nfvPPA
either
Agrammatism in language production
Impaired comprehension of complex sentences
Clinical diagnosis of the nfvPPA
Neuroimaging evidence
Histopathological confirmation of a specific neurodegenerative pathology
Effortful speech: speech sound errors and distortions (apraxia of speech)
Spared single-word comprehension
showing at least one
Presence of a known pathogenic mutation
Spared object knowledge
Predominant left posterior fronto-insular atrophy on MRI
Predominant left posterior fronto-insular hypoperfusion or hypometabolism on SPECT or PET

Figure 1-4 Diagnostic features for the nfvPPA. Adapted from [8]

1.2.1.2 Semantic variant

Unlike nfvPPA, the clinical diagnosis of svPPA requires the presence of both core features: impaired confrontation naming (anomia) and impaired single-word comprehension. The deficit in single-word comprehension is usually more pronounced for low-frequency or unfamiliar items, as a result of a breakdown in semantic memory. Furthermore, at least three supportive features should be noted: reduced object knowledge, surface dyslexia or dysgraphia (i.e., difficulties in reading or writing irregular words), and preserved repetition and speech production, which remain grammatically correct and fluent.

For an imaging-supported diagnosis, neuroimaging must show bilateral atrophy and/or hypoperfusion or hypometabolism involving the ventral and lateral parts of the anterior temporal lobes.

At the level of definite pathology, FTLD-TDP is typically observed. [8], [14]

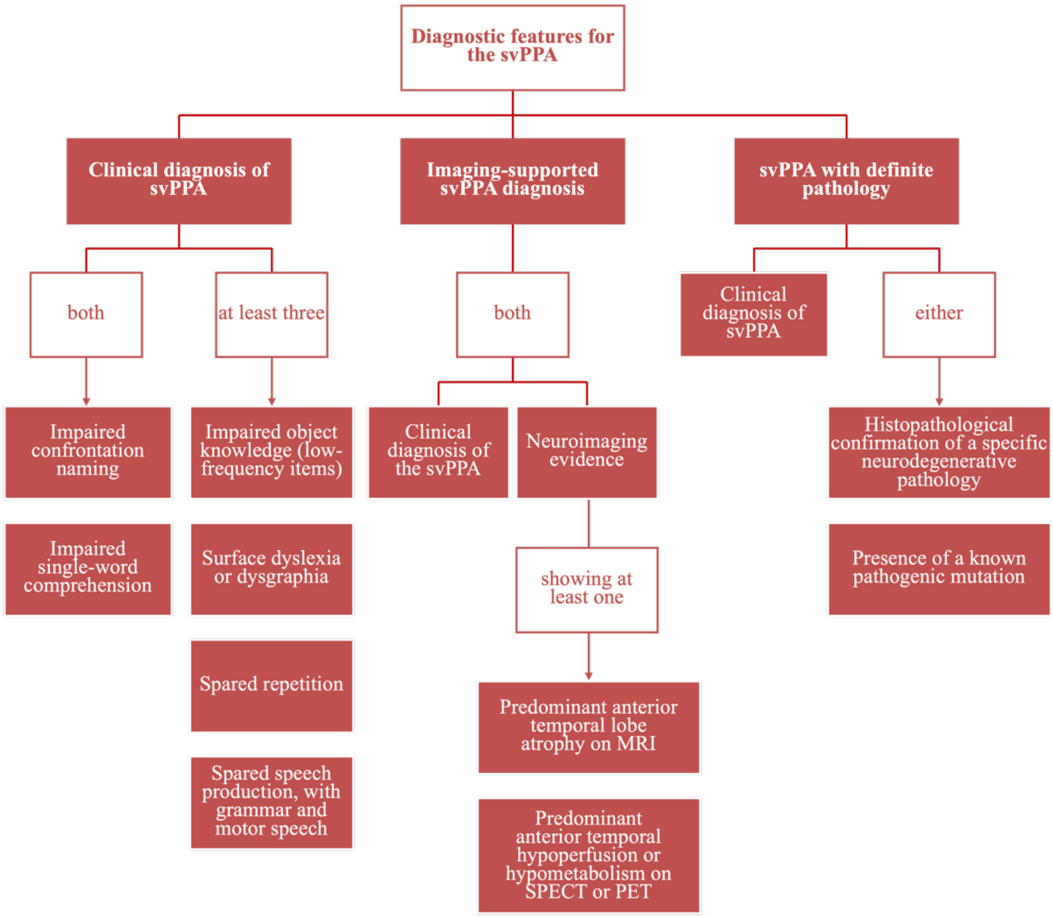

Diagnostic features for the svPPA
Clinical diagnosis of svPPA
Imaging-supported svPPA diagnosis
svPPA with definite pathology
both
at least three
both
Clinical diagnosis of svPPA
either
Impaired confrontation naming
Impaired object knowledge (low-frequency items)
Clinical diagnosis of the svPPA
Neuroimaging evidence
Histopathological confirmation of a specific neurodegenerative pathology
Impaired single-word comprehension
Surface dyslexia or dysgraphia
showing at least one
Presence of a known pathogenic mutation
Spared repetition
Predominant anterior temporal lobe atrophy on MRI
Spared speech production, with grammar and motor speech
Predominant anterior temporal hypoperfusion or hypometabolism on SPECT or PET

Figure 1-5 Diagnostic features for the svPPA. Adapted from [8]

1.2.1.3 Logopenic variant

LvPPA is characterized by impaired single-word retrieval in spontaneous speech or naming and impaired repetition of sentences and phrases. These symptoms are typically accompanied by phonological errors and frequent pauses due to word-finding difficulties, resulting in a slowed speech rate.

At least three additional features should also be observed, including relatively preserved single-word comprehension and object knowledge, intact motor speech, and absence of agrammatism. With the progression of the disease, episodic memory, verbal comprehension, visuospatial skills and executive functions become more impaired. It has been reported that the decline in general cognition in lvPPA patients is about twice as rapid as in the other variants [15] and that anxiety, irritability, apathy and agitation may appear over time, even if they are uncommon at the onset of the disease. [16] For an imaging-supported diagnosis, neuroimaging should reveal atrophy and/or hypoperfusion or hypometabolism predominantly involving the left posterior perisylvian or parietal regions, consistent with the language network affected in this variant. At the definite pathological level, AD pathology represents the most common underlying substrate. [8]

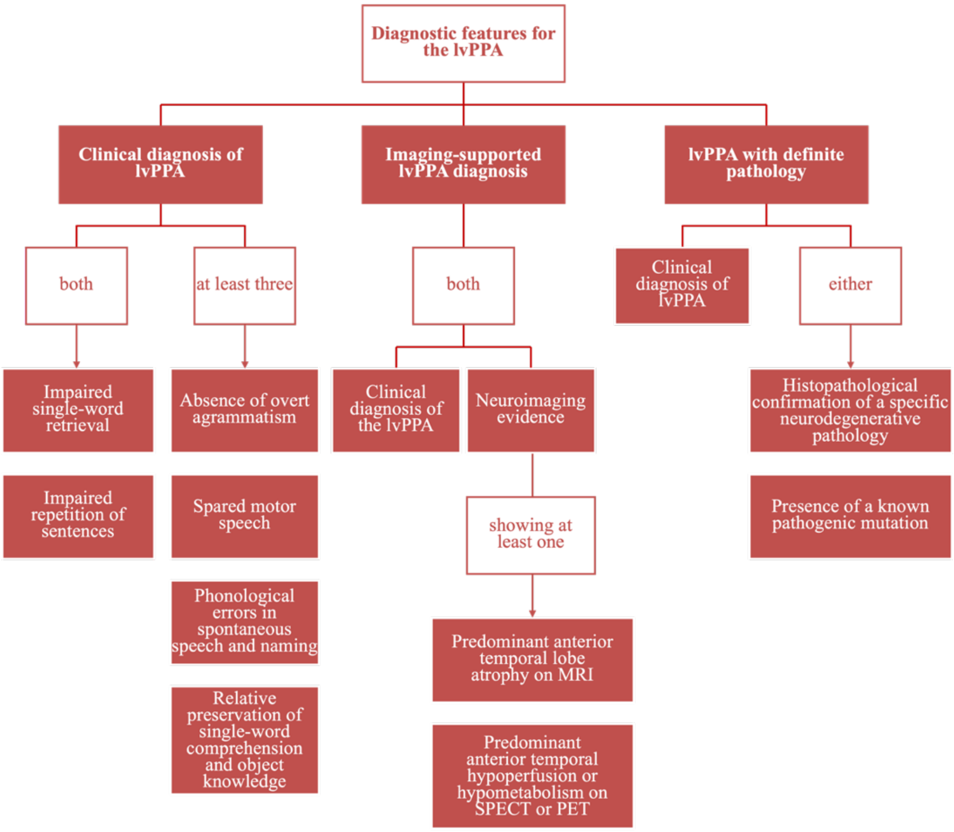

Diagnostic features for the lvPPA
Clinical diagnosis of lvPPA
Imaging-supported lvPPA diagnosis
lvPPA with definite pathology
both
at least three
both
Clinical diagnosis of lvPPA
either
Impaired single-word retrieval
Impaired repetition of sentences
Absence of overt agrammatism
Spared motor speech
Phonological errors in spontaneous speech and naming
Relative preservation of single-word comprehension and object knowledge
Clinical diagnosis of the lvPPA
Neuroimaging evidence
showing at least one
Predominant anterior temporal lobe atrophy on MRI
Predominant anterior temporal hypoperfusion or hypometabolism on SPECT or PET
Histopathological confirmation of a specific neurodegenerative pathology
Presence of a known pathogenic mutation

Figure 1-6 Diagnostic features for the lvPPA. Adapted from [8]

1.3 Epidemiology

The Genetic and Rare Diseases Information Center of the National Institutes of Health classifies PPA as a rare disease, affecting fewer than 50,000 individuals in the United States. The overall prevalence of PPA is estimated to be 3 to 4 cases per 100,000 population. [17]

According to a recent French national registry study, the incidence of PPA was estimated to be approximately 1.14 per 100,000 person-years, compared with 35.7 per 100,000 person-years for AD. Patients with PPA were found to be significantly younger at symptom onset than those with AD, with a median age at diagnosis of 73.7 years versus 81.4 years in AD. The study also reported a more balanced sex distribution in PPA, with 55.7% of females, compared with 69.6% in AD, and noted that the level of education was generally higher among PPA patients. In particular, a larger proportion of individuals with PPA had completed more than six years of schooling, suggesting potential differences in demographic and cognitive profiles between the two populations. [18]

The prevalence of the nfvPPA has been estimated to range between 0.5 to 3.9 per 100,000 people. Approximately 56% to 86% of individuals with PPA present coexisting speech apraxia. [19] Moreover, a recent study showed that bvFTD is more common in men, whereas PPA is more common in women ($p<0.001$). This finding suggests that sex may be a potential factor in determining FTD phenotype, although it does not appear to influence survival. [20]

In a cohort study conducted by Spinelli et al., 69 patients with PPA were analyzed, classified into 29 svPPA, 25 nfvPPA, 11 lvPPA and 4 unclassifiable/mixed cases. The study reported differences among the variants in terms of age at onset, diagnostic delay, sex distribution and survival time.

The svPPA showed the earliest onset, with a mean age at symptom onset of 60 years (59.6 ± 7.2) and a mean age at first evaluation of 64.7 ± 6.7 years, indicating a diagnostic delay of roughly five years. Men were slightly more affected than women (52%). This variant also exhibited the longest survival, averaging 12 years (11.6 ± 4.3) from symptom onset.

The nfvPPA was predominantly observed in women (72%) and presented with a mean age at onset of 64 years (64.4 ± 7.5). The mean age at first evaluation was 68.6 ± 7.6 years, reflecting a shorter diagnostic delay compared with the other variants, possibly due to the more evident speech production difficulties. The average survival for this group was 8 years (8.0 ± 2.5), representing the shortest prognosis among the variants. Statistical analysis confirmed that the survival time for nfvPPA was significantly shorter than for svPPA ($p<0.05$).

The lvPPA was characterized by a mean age at onset of 63 years (63.0 ± 7.9) and an average age at first evaluation of 66.8 ± 8.6 years, corresponding to a diagnostic delay of about four years. Females represented 55% of this group and the mean

survival had an average of 11.0 ± 4.1 years. The lvPPA group also demonstrated a significantly longer survival compared with nfvPPA ($p<0.05$).
Taken together, these findings indicate that, in this sample, svPPA tends to manifest earlier and progress more slowly, lvPPA shows intermediate onset and survival, and nfvPPA presents later but progresses more rapidly, consistent with its more pronounced speech and motor involvement. [21]

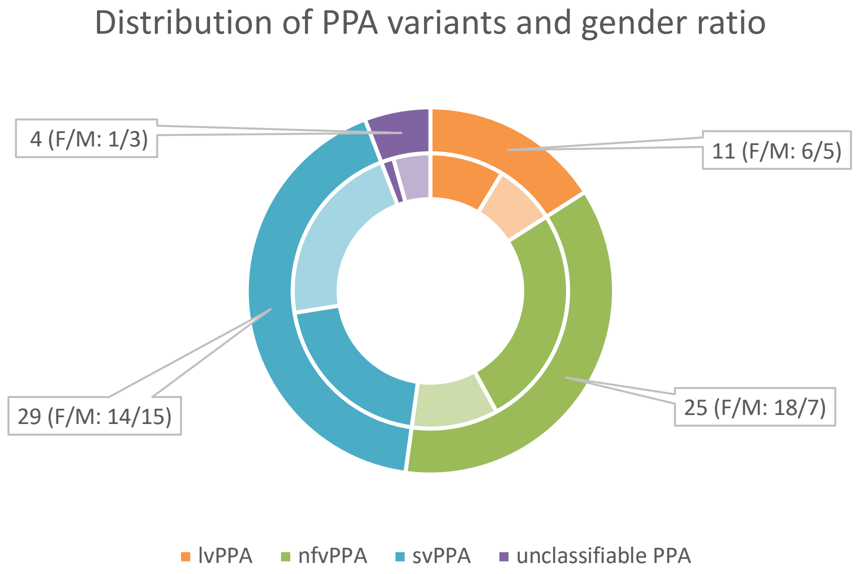


Figure 1-7 Distribution of PPA variants and gender ratio. Adapted from [21]

PPA variant	***Age at onset***	***Age at first evaluation***	***Survival***
lvPPA	63.0±7.9	66.8±8.6	11.0±4.1
nfvPPA	64.4±7.5	68.6±7.6	8.0±2.5
svPPA	59.6±7.2	64.7±6.7	11.6±4.3

Table 1-1 Demographic features of PPA patients. Adapted from [21]

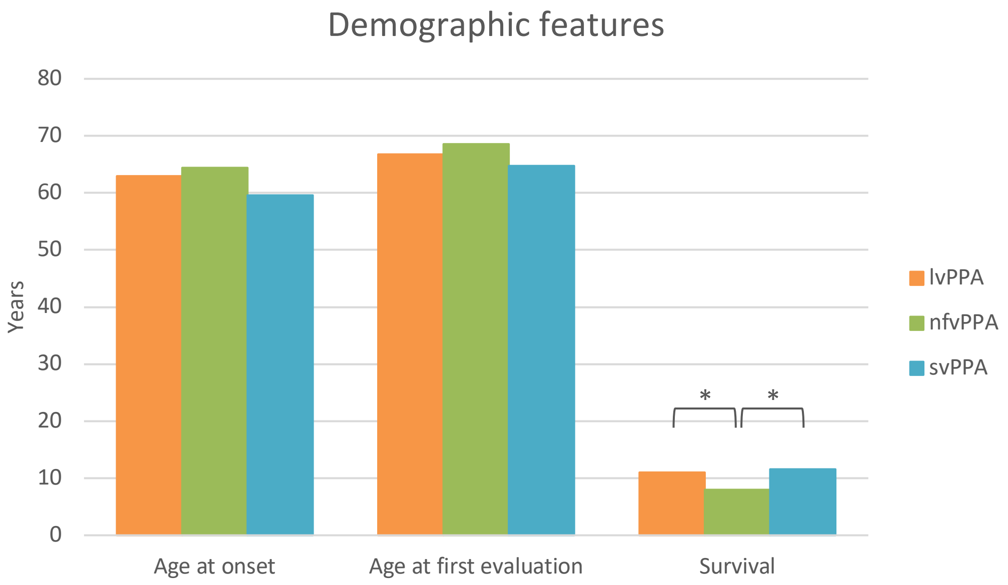


Figure 1-8 Demographic Features of with PPA patients[1]*. Adapted from [21]*

1.4 Etiology

PPA is considered a subtype within the FTD spectrum, together with the more common bvFTD. [19], [22]

Although all forms share progressive language impairment as their main clinical feature, each variant is associated with distinct patterns of neurodegeneration involving specific cortical networks.

Overall, neuropathological studies report that approximately 40-45% of PPA cases are associated with TDP-43 inclusions, a similar proportion with tau pathology, while AD-related pathology accounts for roughly 20% of cases.[21]

The semantic and nonfluent variants are most frequently linked to FTD-type pathologies (TDP-43 or tau), whereas the lvPPA is more commonly associated with AD. [19]

In particular, for the semantic variant, neurodegeneration is most consistently associated with FTLD characterized by TDP-43 type C inclusions, which account for approximately 75-100% of cases, whereas a minority presents FTLD-tau. [19]

Rarely, svPPA has been linked to mutations in the TARDBP gene, which encodes the TDP-43 protein, though such associations remain uncommon.[23]

Regarding nfvPPA, the underlying pathology most often corresponds to FTLD with tau inclusions, particularly the 4-repeat tau form.[19] A smaller proportion of cases

[1] * indicates a statistically significant difference ($p < 0.05$)

shows TDP-43 type A pathology and only rarely AD related changes. From a genetic perspective, mutations in the GRN or C9orf72 genes have been identified in some nfvPPA cases, though they are rare. Together with MAPT, these genes represent the three most frequent genetic causes of FTLD.[19] Notably, GRN mutations located on chromosome 17q21.31 lead to reduced progranulin production and have been identified in two families with PPA [24], although such mutations are more typically associated with the bvFTD. The C9orf72 repeat expansion, which encodes TDP-43 type A aggregates, has been reported, supporting a molecular overlap between nfvPPA and other FTLD subtypes. [25]

By contrast, lvPPA is most frequently associated with AD pathology, observed in up to 95% of cases.[26] However, alternative pathological profiles such as Lewy body disease, TDP-43 or tauopathies have also been reported, although these remain uncommon. [19], [27]

At the genetic level, mutations in classical AD-related genes such as PSEN1, PSEN2 and APP are rare, accounting for less than 2% of lvPPA cases. Conversely, a subset of patients presents GRN or MAPT mutations, suggesting that some cases may instead lie on the FTD spectrum. Furthermore, about 40% of lvPPA patients carry the APOE ε4 allele, while nearly half exhibit the MAPT H1H1 haplotype, both considered potential susceptibility factors for neurodegeneration. [28]

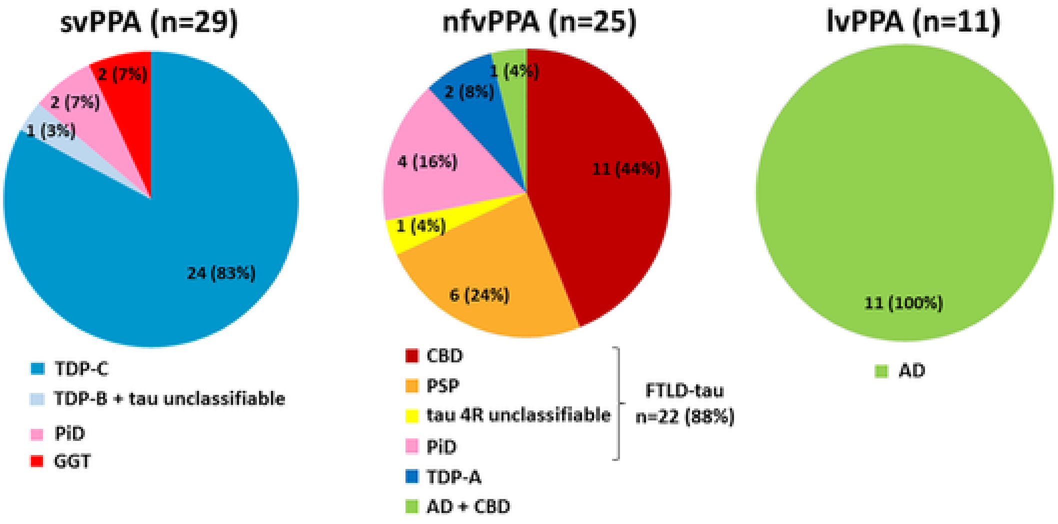


Figure 1-9 Primary neuropathological diagnosis in PPA patients. From [22]

1.5 Pathophysiology

The pathophysiological mechanisms of PPA reflect the progressive degeneration of the cortical networks responsible for language, that results from the accumulation of pathological proteins such as tau, TDP-43, or β-amyloid. Although all variants share a gradual deterioration of linguistic abilities, each form exhibits a characteristic pattern of cortical atrophy, metabolic alteration and functional disconnection.

1.5.1 Cortical degeneration and molecular pathology

In svPPA, degeneration primarily affects the anterior temporal lobes, especially the temporal poles, with a marked left-hemispheric predominance. This pattern accounts for the progressive loss of conceptual and lexical-semantic knowledge. [29] At the molecular level, TDP-43 type C inclusions represent the main pathological hallmark, observed in roughly 80% of svPPA cases. [21] This places svPPA within the FTLD-TDP spectrum, sharing molecular similarities with frontotemporal dementia and amyotrophic lateral sclerosis. [19]

LvPPA is most frequently associated with β-amyloid deposition and hyperphosphorylated tau accumulation, reflecting its close relationship to AD. Atrophy is typically asymmetric, involving the left posterior perisylvian and parietal regions. These areas are responsible for the characteristic deficits in phonological retrieval and sentence repetition that define the syndrome. Over the course of the disease, many patients develop memory and cognitive symptoms similar to those seen in AD, supporting the notion that lvPPA represents a language-dominant presentation of that spectrum. [19]

NfvPPA shows great pathological and clinical heterogeneity. Neuroimaging consistently reveals atrophy in the dominant inferior frontal gyrus (Broca's area) and insular cortex, corresponding to regions involved in syntactic processing and motor speech planning.[30] [29] Histopathological studies indicate that most nfvPPA cases are associated with FTLD-tau pathology, while a minority display TDP-43 or AD-related changes. This heterogeneity suggests that nfvPPA reflects a clinicopathological continuum rather than a single disease entity. [19]

1.5.2 Language-related brain regions

During the 19th century, the studies of Broca and Wernicke led to the formulation of the classical language model, which identifies two main areas: Broca's area, responsible for language production situated in the left frontal lobe, and Wernicke's area, in the left posterior temporal lobe, involved in comprehension. [31][32]

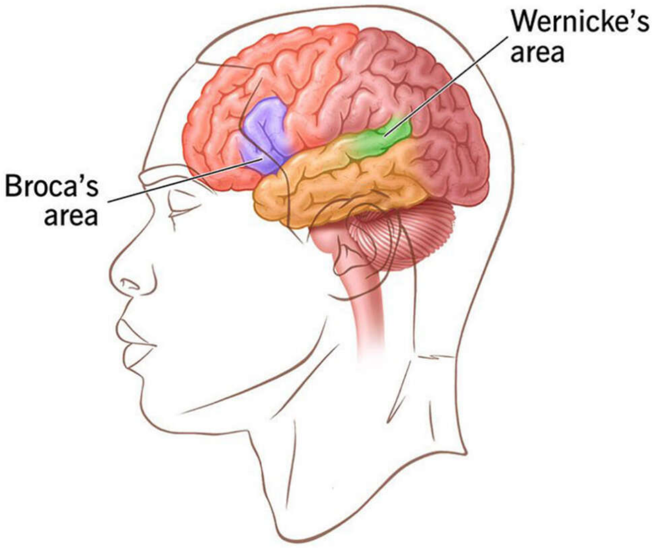


Figure 1-10 Broca and Wernicke brain regions. From [33]

However, recent studies based on modern neuroimaging techniques revealed a wider and more complex organization of the language network.

Language comprehension also involves temporo-parietal regions of the left hemisphere, such as superior, middle and inferior temporal gyri, the angular gyrus and the fusiform gyrus. These areas cooperate in the analysis of speech sounds and in lexical and semantic processing. Similarly, language production is not limited to Broca's area, but extends to lateral and medial prefrontal cortical regions, which play an executive role, coordinating linguistic and semantic information coming from posterior areas and adapting verbal production to communicative goals. Some regions of the middle frontal gyrus, such as Brodmann area 9, showed no activation during language tasks, suggesting that this areas may be related to other cognitive functions. Linguistic processes also involve the cerebellum, which supports neural computations necessary for language planning and fluency. Retrosplenial cortex, located near the splenium of the corpus callosum, is involved in memory functions, contributing to contextual comprehension and lexical retrieval. [34]

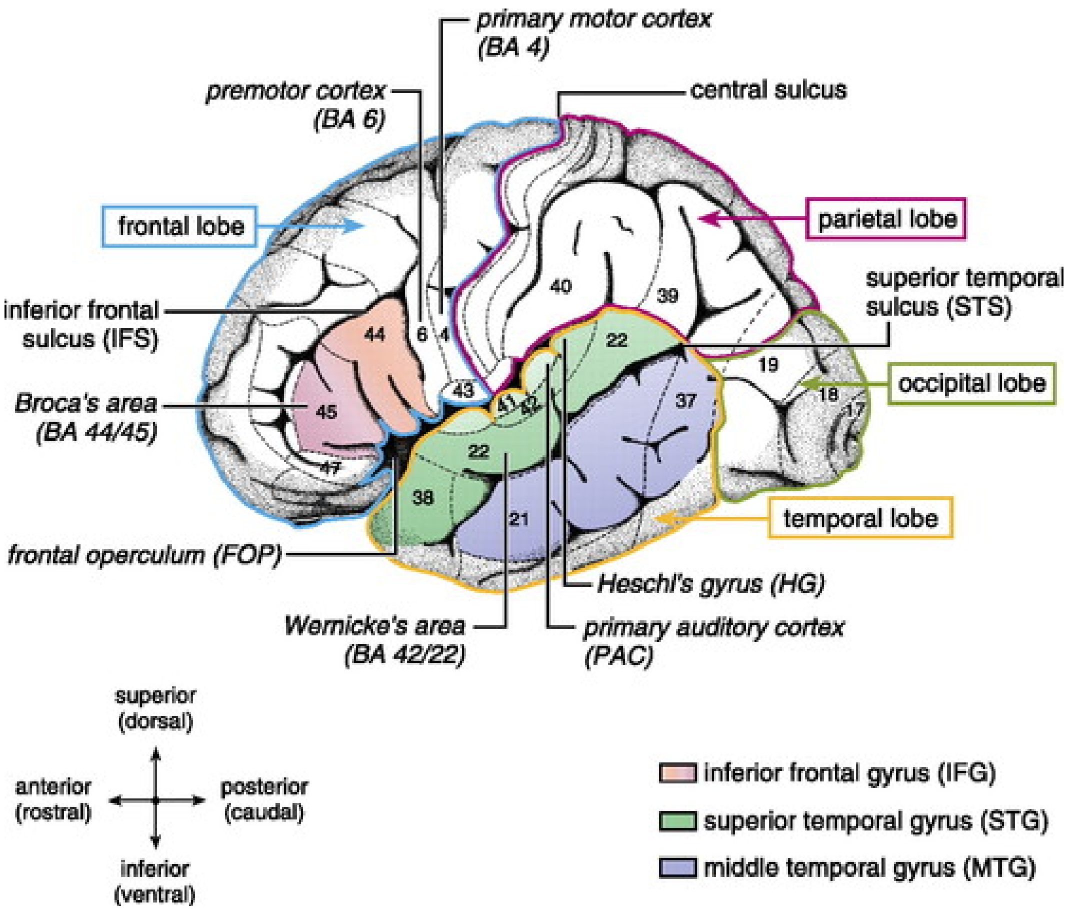


Figure 1-11 Anatomical representation of the left hemisphere. From [35]

1.6 Treatments

Speech and language intervention has been shown to improve language outcomes in individuals with PPA. Behavioral treatments that stimulate residual linguistic knowledge across the semantic, phonological, and orthographic domains have demonstrated cross-domain improvements in communication abilities [36]. Different approaches may be employed to target specific linguistic functions, such as naming impairment (anomia), a core feature across all PPA variants. Effective strategies include repeated practice of names paired with pictures or verbal descriptions of the target items, repetition-based training, in which participants repeat picture labels following auditory presentation and reading-based interventions involving multisyllabic words, which have yielded lasting and generalizable benefits [37], [38]

In addition to behavioral therapy, transcranial direct current stimulation (tDCS) has emerged as a promising adjunct treatment. Given its low cost, noninvasive nature and high safety profile, tDCS represents a promising technique with the potential to enhance behavioral intervention and slow the rate of language decline in PPA. When combined with speech–language therapy, tDCS has been shown to augment generalization and maintenance of positive language outcomes. [38] Notably, a recent study demonstrated that bi-hemispheric tDCS can induce measurable neuroplastic changes in the right hemisphere. Specifically, significant improvements in fluency were associated with increased fractional anisotropy in the right uncinate fasciculus and reduced mean diffusivity in the right frontal aslant tract, suggesting structural reorganization of white matter pathways. Moreover, a positive correlation between changes in the right uncinate fasciculus and fluency improvement supports the hypothesis that right-hemispheric tracts may contribute to compensatory mechanisms underlying aphasia recovery. [39]

Although evidence remains limited, case studies have also reported improvements in language abilities following steroid treatment [40] and Omentum Transposition Therapy [41], suggesting potential alternative or adjunctive therapeutic approaches.

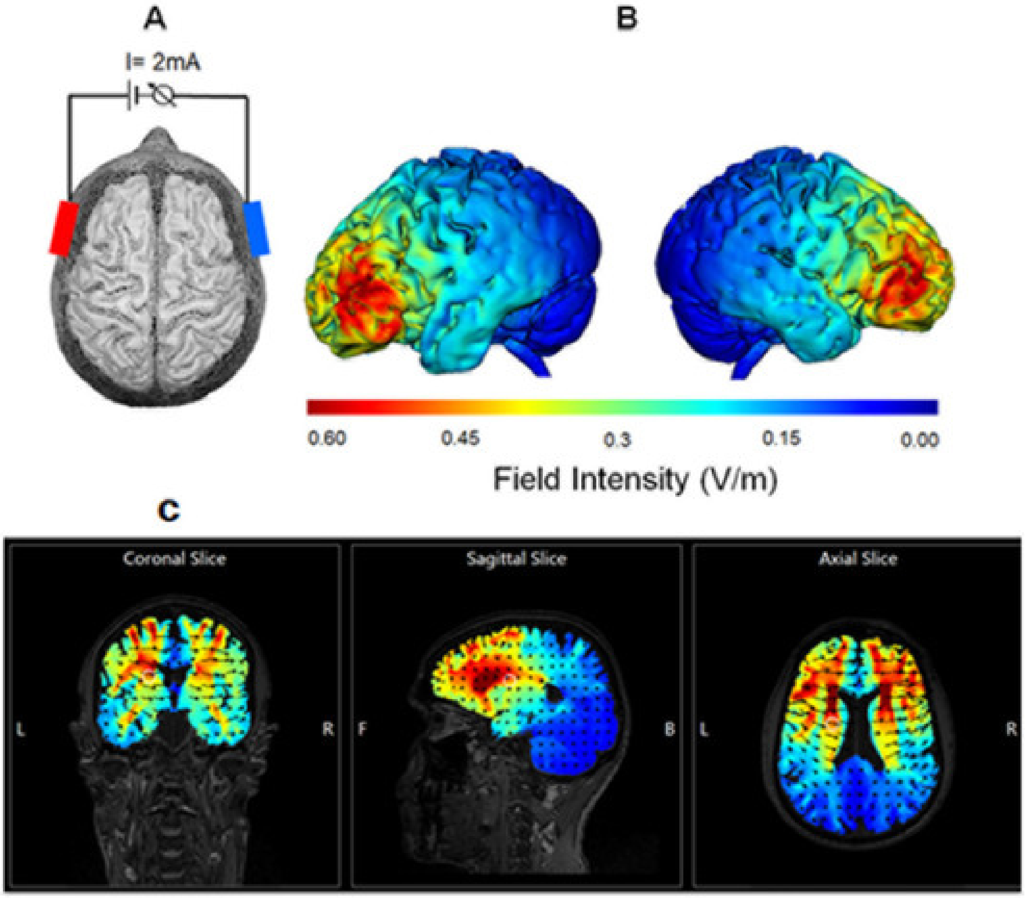


Figure 1-12 Transcranial direct current stimulation in an aphasic patient. Electrode placement over Broca's area and its right hemisphere homologue. Adapted from [39]

1.7 Assessment methods for the diagnosis of PPA variants

The first step in distinguishing PPA variants involves collecting a detailed case history and conducting a clinical interview to gather information about the initial symptoms, the development of additional deficits over time and the extent of linguistic, cognitive, or motor impairment at the time of assessment. [42]

1.7.1 Speech and language assessment

A comprehensive speech–language evaluation should be conducted to characterize both impaired and preserved linguistic abilities in individuals with suspected PPA. The assessment should cover the main language domains typically affected in the three PPA variants.

Individuals with nfvPPA usually show impairments in grammar and motor speech. These include effortful and halting speech, agrammatism, apraxia of speech and dysarthria, as well as difficulties in sentence comprehension, particularly for grammatically complex structures. In contrast, those with svPPA exhibit deficits in confrontation naming due to semantic errors, single-word comprehension and knowledge of objects or people. Impairments in reading and spelling are also common, often characterized by regularization errors. Finally, patients with lvPPA typically produce phonemic errors in naming tasks, show impaired sentence comprehension driven by length and frequency effects, and demonstrate phonological errors in reading and spelling. [8]

To observe these functional impairments, clinicians can choose from a range of tasks assessing both expressive and receptive language abilities. These may include: picture description (an example is the Picnic Scene description task) or story-telling tasks to evaluate grammar and fluency; repetition of multisyllabic words to assess motor speech planning; sentence-to-picture matching, word-to-picture or word-to-definition matching, and gesture-to-object or sound-to-picture association tasks to evaluate sentence comprehension, object and person knowledge, and single-word understanding. [8]

Beyond qualitative characterization, speech and language assessment also provides a quantitative index of aphasia severity, which can serve as a baseline measure to monitor disease progression and treatment outcomes. Standard aphasia batteries, originally developed for stroke-induced aphasia, are widely used in both clinical and research settings involving PPA. [42] These batteries consist of multiple tasks evaluating different language domains, allowing a comprehensive profile of language abilities to be established. Among these, the Western Aphasia Battery – Revised (WAB-R) 11/26/25 11:47:00 AM and the Boston Diagnostic Aphasia Examination (BDAE) [43] are widely employed to describe overall language profiles and quantify aphasia severity. [42] Importantly, the WAB-R has been shown to assist in distinguishing between PPA subtypes. [44]

In addition to these general measures, several assessments have been developed or adapted specifically for differential diagnosis and progression tracking in PPA. [42] The Sydney Language Battery (SydBat) includes picture naming, word comprehension, semantic association and repetition tasks, and is designed to differentiate between PPA variants.[45] The Repeat and Point Test is a short assessment test to distinguish between semantic and nonfluent variants by requiring repetition of multisyllabic words followed by pointing to the correct target among phonological and semantic distractors. [46] The Progressive Aphasia Severity Scale (PASS), allows clinicians to rate symptom severity and monitor longitudinal progression in PPA.[47] The Progressive Aphasia Language Scale (PALS) provides structured clinician ratings of speech and language features based on performance across a standardized set of tasks. [48]

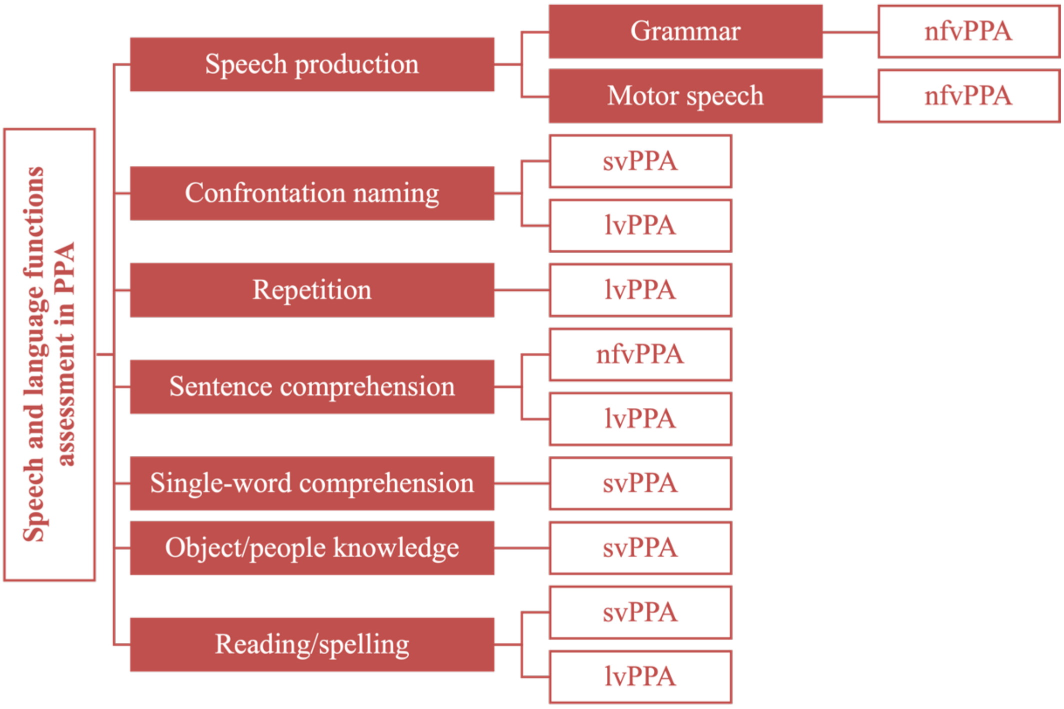


Figure 1-13 Tasks that may be used for speech and language functions assessment in PPA. Adapted from [8]

1.7.2 Neuropsychological and neurologic assessment

It was reported that in a particular sample of individuals with PPA, approximately 31% could not be classified via the quantitative application of the 2011 consensus diagnostic criteria, and that significant neurocognitive differences persisted among participants with logopenic, semantic, and agrammatic PPA even after controlling for aphasia severity (WAB-AQ). These findings suggest that neuropsychological testing may provide additional diagnostic value in distinguishing clinical subtypes. [49]

For this reason, neuropsychological assessment is routinely performed to evaluate cognitive domains beyond language, helping to identify the broader neurocognitive profile of impairment. Memory functions can be explored using the Three Words–Three Shapes Test, where poor performance often reflects left-hemisphere language dysfunction rather than medial temporal memory network impairment. [50] Processing speed is commonly assessed with the Trail Making Test (TMT) Part A [51], which measures visual scanning and visuomotor tracking abilities, while executive functions are typically examined with the TMT Part B, assessing divided attention and cognitive flexibility. [51] Visuospatial abilities can be evaluated through subtests of the Visual Object and Space Perception Battery (VOSP), such as the Cube and Incomplete Letters subtests [52], in which participants are required to recognize degraded letters or shapes presented with only about 30% of their original form. Finally, a comprehensive neurologic examination, including general cognitive testing, is usually conducted by a behavioral neurologist to provide an integrated overview of the patient's cognitive and motor status. [49]

1.8 Speech and language profiles in PPA variants

The three PPA variants present with different brain lesion localization. These different lesions are associated with specific cortical degeneration pattern and impaired white matter (WM) pathways, thus leading to language domain alterations.

The nfvPPA patients present with impaired grammar, speech production and comprehension. [53] These deficits are caused by lesions and cortical atrophy in the left frontal regions which are involved in complex grammatical processes. Their agrammatic patterns are similar to those observed in post-stroke agrammatism, but they differ from lvPPA and svPPA patterns. [54] This variant is also characterized by effortful speech, meaning slow and labored speech production that reflects a speech motor planning deficit. [93] In fact, they may have difficulties coordinating the movements of the tongue and lips, often making speech sound errors, such as distortions, deletions, substitutions and insertions. [53] [93] Prosody is also affected. They tend to produce short and simple sentences, omitting grammatical morphemes. Regarding language production they can omit articles or use wrong morphological endings. The use of verbs is reduced compared with healthy controls. [55]

Patients with the semantic variant show a progressive loss of semantic knowledge, gradually losing the meaning of words. Their speech is fluent but less informative than the other variants; during spontaneous speech, severe anomia is often evident. Naming tasks can reveal semantic paraphasias. These patients usually replace low-frequency words with the ones that are more familiar to them. It is usual they substitute a word with its superordinate one. [55] This struggle with less frequent words is also present in single word comprehension, which is impaired. [53] This means they tend to replace the less frequent words with more familiar ones, using for example a substitute from the superordinate category. The disease can progress until object recognition becomes affected and the impairment extends to all their sensory modalities. [55] Patients may struggle to find words to express their thoughts or to recall facts. [53]

Finally, logopenic patients present with impaired sentence repetition and word finding difficulties. [55] Because of this, they make frequent pauses, searching for words, resulting in slow speech. They usually preserve single words repetition, but a short-term memory deficit reduces their sentence comprehension; this impairment is mainly phonological. [53][55] They may produce phonological paraphasias when struggling in confrontation naming. Phonemic speech sound errors can occur too. Difficulties with calculation are also common. [55]

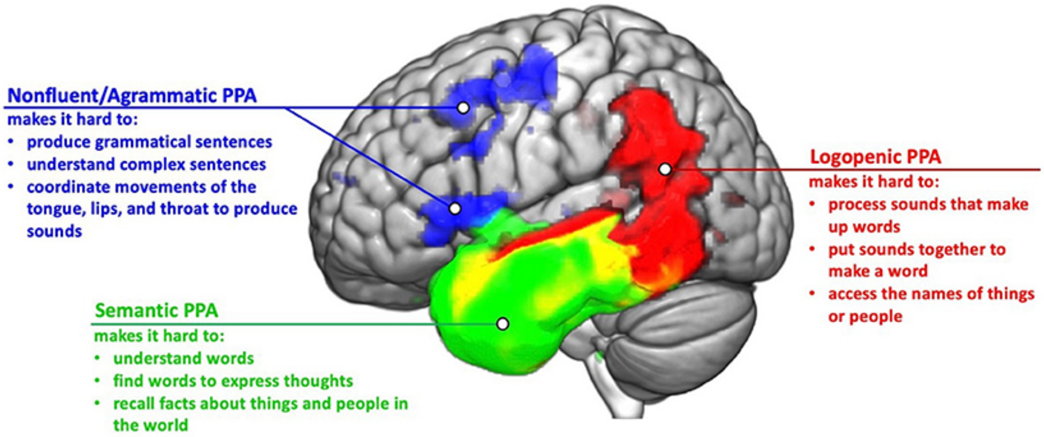

Nonfluent/Agrammatic PPA
makes it hard to:
• produce grammatical sentences
• understand complex sentences
• coordinate movements of the tongue, lips, and throat to produce sounds
Logopenic PPA
makes it hard to:
• process sounds that make up words
• put sounds together to make a word
• access the names of things or people
Semantic PPA
makes it hard to:
• understand words
• find words to express thoughts
• recall facts about things and people in the world

Figure 1-14 Main damaged areas from each PPA variant. Adapted from [53]

2 Magnetic Resonance

Magnetic resonance imaging (MRI) is a noninvasive diagnostic technique that uses strong magnetic fields and low-energy radiofrequency waves, thereby avoiding exposure to ionizing radiation [56]. As the technology becomes more affordable and accessible, its use in clinical medicine continues to expand.

In 1946, Edward Purcell and Felix Bloch independently discovered the magnetic properties of atomic nuclei, forming the basis for nuclear magnetic resonance (NMR). This finding earned them the 1952 Nobel Prize in Physics. [57], [58] Initially regarded as a topic of fundamental physics, NMR soon attracted biomedical interest. In 1971, Raymond Damadian demonstrated that biological tissues differ in their relaxation properties (T_1 and T_2), which allowed for the distinction between healthy and cancerous tissues [59]. This discovery marked the transition of NMR from a theoretical construct to a potential diagnostic tool.

A decisive breakthrough followed in 1973, when Paul Lauterbur, Peter Mansfield, and Richard Grannell introduced the use of magnetic field gradients to encode spatial information, transforming NMR spectroscopy into an imaging modality [59]. Their innovation laid the groundwork for MRI, fundamentally changing the way internal anatomical structures could be visualized noninvasively.

The transition of MRI from experimental research to clinical application occurred in 1980, when the first diagnostic images were successfully obtained in Nottingham and Aberdeen [59]. Since then, the technology has evolved rapidly. Contemporary MRI systems make use of superconducting magnets cooled to cryogenic temperatures, producing magnetic fields that usually range from 0.5 to 1.5 Tesla (T). [57]

Although scanners operating at 1.5 T have long represented the clinical standard, more powerful 3 T systems are now widely implemented, particularly within research and advanced diagnostic environments. Their stronger magnetic fields yield considerable gains in image quality, including an increase in SNR and up to a 96% enhancement in contrast to noise ratio [57]. These improvements enable more precise visualization of anatomical details and greater measurement consistency. Nonetheless, high-field imaging also presents technical challenges: susceptibility effects and geometric distortions become more pronounced, requiring meticulous optimization of acquisition protocols and post-processing methods [56].

2.1 Basic principles

NMR describes the interaction between the magnetic moments of certain atomic nuclei and an externally applied magnetic field [57], [58]. The magnetic moment of a nucleus originates from a property known as nuclear spin, representing an intrinsic angular momentum determined by the configuration of protons and neutrons within the nucleus. The value of this spin is defined by the spin quantum number (I), which varies according to the nuclear composition.

Nuclei containing even numbers of both protons and neutrons have no net spin and, consequently, no magnetic moment. In contrast, nuclei with an odd number of protons or neutrons exhibit a non-zero spin and generate a measurable magnetic moment [46][47]. Isotopes such as ^{1}H, ^{31}P, ^{13}C, and ^{15}N, each with I = ½, are particularly significant for magnetic resonance applications due to their favorable spin properties and their abundance in biological tissues [58]. Among these, hydrogen plays a central role in MRI because of its high concentration in water and fat, the main constituents of the human body. [56][47]

The term "magnetic" in MRI refers to the application of a strong static external magnetic field (B_0), while "resonance" describes the condition in which an oscillating electromagnetic field matches the precessional frequency of the spinning nuclei [59]. In classical terms, a nucleus with a non-zero spin can be visualized as a minute bar magnet that spins about its own axis, creating a local magnetic field with distinct north and south poles [47]. Under normal conditions, these magnetic dipoles are randomly oriented, and their magnetic effects cancel out, resulting in no net magnetization.

When the sample is placed in the static magnetic field B_0, each nucleus behaves like a small magnetic dipole attempting to align with the field. The nuclear spins can adopt one of two possible orientations: parallel to B_0 (the lower-energy configuration) or antiparallel to B_0 (the higher-energy configuration), corresponding to magnetic quantum numbers of +½ and −½ [56] [58] [60]

The small energy difference between these two states forms the physical basis for magnetic resonance.

At absolute zero, all nuclear spins would occupy the parallel, low-energy state. However, at physiological temperatures, thermal energy disrupts perfect alignment, resulting in a Boltzmann distribution of spins between the two energy levels [49]. Although the vast majority of spins remain randomly oriented, a slight excess of nuclei align parallel to B_0. This small population difference gives rise to a net macroscopic magnetization vector (M_0) that points along the direction of the external magnetic field [46]. This equilibrium magnetization constitutes the initial state of the system and is essential for generating the detectable MRI signal.

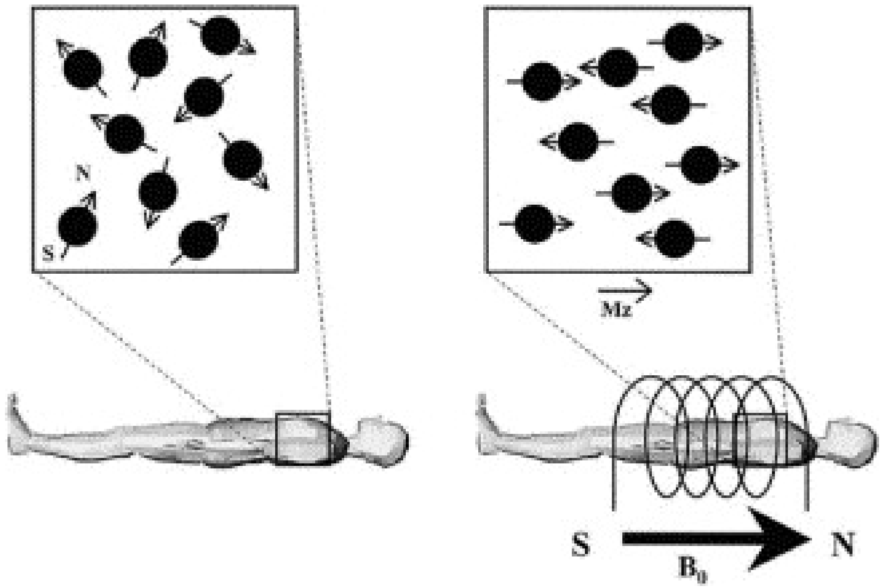


Figure 2-1 Nuclei orientation in a static magnetic field. From [56]

2.1.1 Precession and Larmor frequency

When the net magnetization vector (M_0) is perfectly aligned with the static external magnetic field (B_0), no rotational motion occurs, since the torque acting on the magnetization is zero and therefore $M \times B_0 = 0$. In this equilibrium condition, the magnetization remains static. However, if M is displaced from the longitudinal axis, the torque generated by the magnetic field induces a rotational motion around the direction of B_0 , a phenomenon known as precession. [56] This motion can be described by the Larmor equation, which defines the angular frequency of precession as:

$$\omega_0 = \gamma B_0$$

where ω_0 is the Larmor frequency, γ is the gyromagnetic ratio (a constant specific to each nucleus), and B_0 is the strength of the static magnetic field. [57], [58]

The Larmor frequency represents the rate at which the magnetic moment of a nucleus precesses around the external field and is a key parameter in magnetic resonance phenomena.

It is important to note that individual nuclei do not align perfectly with the direction of B_0, but rather wobble or precess around it. Each nucleus maintains its own phase of precession, meaning that while the angular frequency (ω_0) is identical for all nuclei of the same species in a uniform magnetic field, their instantaneous orientations are phase-shifted relative to one another [56].

In human MRI applications, magnetic field strengths typically range between 0.1 and 4 Tesla (T), corresponding to radiofrequency (RFr) resonance frequencies up to approximately 170 MHz [61]. For example, in a 1.5 T MRI system, hydrogen nuclei precess at a frequency of about 63.75 MHz [56]. More advanced scanners operating at 7 T achieve Larmor frequencies approaching 300 MHz, offering improved spatial resolution and signal-to-noise ratio, although these benefits are accompanied by increased technical challenges [58].

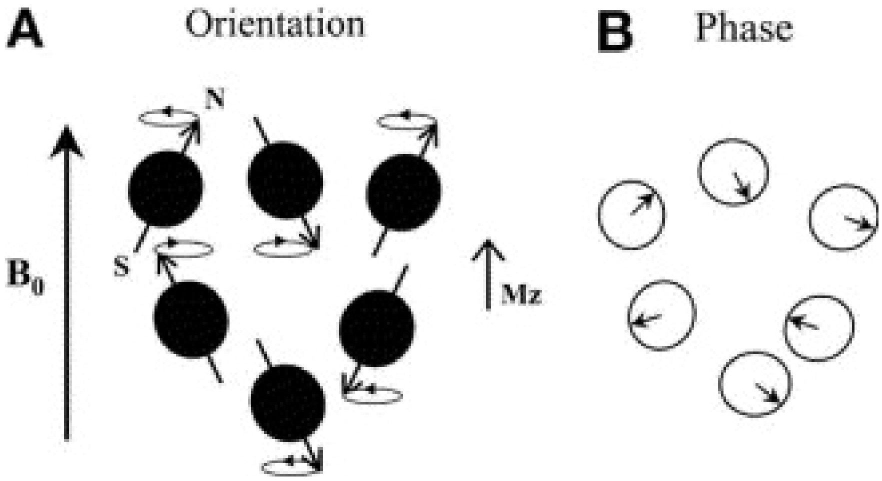


Figure 2-2 Nuclei orientation and phase with an external field. From [56]

2.1.2 Radiofrequency excitation

At equilibrium, the net magnetization vector (M_0) of the nuclei within the magnetic field (B_0) is aligned along the longitudinal axis. In this configuration, the magnetization is static and does not generate a measurable signal, since a time-varying magnetic flux, required by Faraday's law of induction, is absent [56] [58]. To obtain an observable signal, the orientation of the magnetization must be perturbed from its equilibrium alignment.

Directly altering B_0 is technically impractical; instead, a radiofrequency (RFr) magnetic field, denoted B_1, is applied perpendicular to B_0. This secondary field oscillates at the Larmor frequency, the specific resonance frequency of the nuclei. When B_1 is applied at this frequency, resonance occurs: in a rotating reference frame, the oscillating RFr field behaves as if it were a static field capable of rotating the net magnetization vector by a defined flip angle. [57], [58]

The RFr energy is typically delivered in short pulses lasting from microseconds to a few milliseconds, known as RFr pulses. The magnitude and duration of these pulses determine the resulting flip angle. A 90° pulse tips the magnetization into the transverse plane, producing the excitation pulse, while a 180° pulse inverts the magnetization, generating an inversion pulse [58].

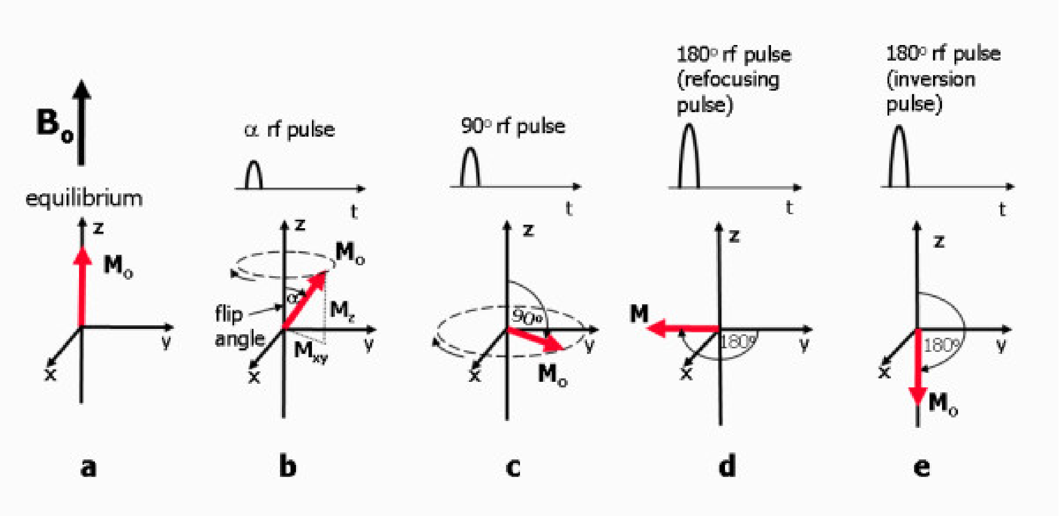


Figure 2-3 Magnetization vector with different RFr pulses. Adapted from [62]

From a quantum-mechanical perspective, the absorption of RFr energy induces transitions between nuclear spin energy levels, driving the spins to precess in phase within the transverse plane. [57] Consequently, the net magnetization vector now possesses both a longitudinal component (M_z), parallel to B_0, and a transverse component (M_{xy}), perpendicular to it. At the end of the RFr excitation, immediately following a 90° pulse, the magnetization lies entirely in the transverse plane, where M_{xy} precesses around B_0 at the Larmor frequency [59].

This precessional motion induces a time-varying magnetic flux in the receiver coil, generating an alternating current according to Faraday's law of induction. The resulting voltage signal, known as the free induction decay (FID), represents the measurable nuclear magnetic resonance response of the system [58], [59]. The FID exhibits a sinusoidal oscillation whose amplitude decays over time as the spins lose phase coherence and return to equilibrium. Through mathematical transformation, typically via the Fourier transform, this time-domain signal can be expressed in the frequency or spatial domain, forming the foundation for MR image reconstruction [57], [58].

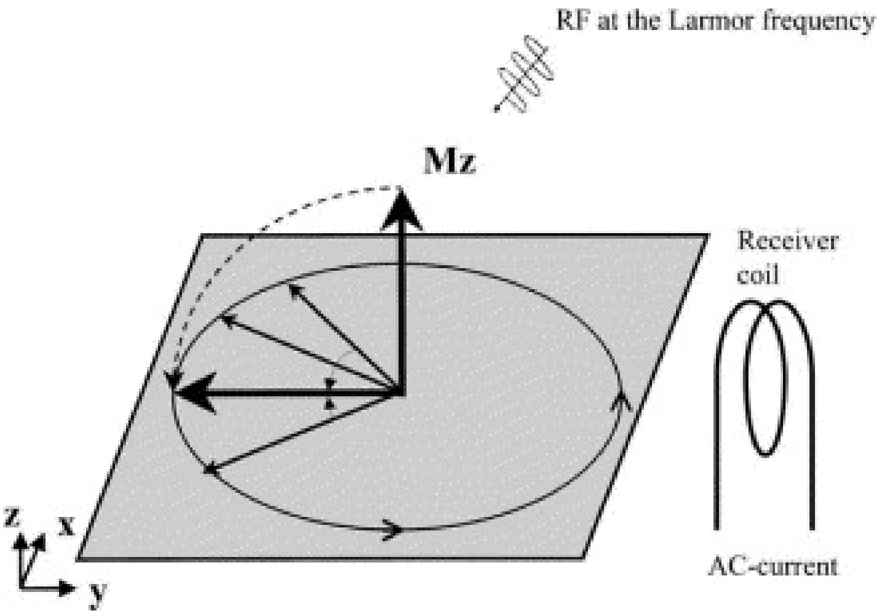


Figure 2-4 Radiofrequency excitation. From [56]

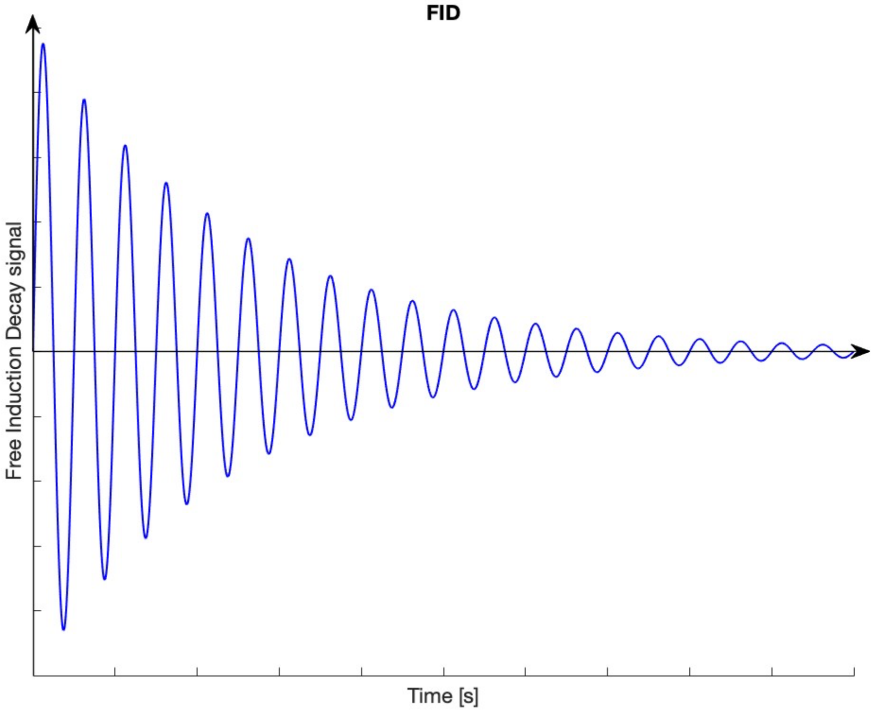


Figure 2-5 Free Induction Decay signal

2.1.3 Relaxation mechanisms

Following RFr excitation, the net magnetization vector returns to its equilibrium configuration through two independent relaxation processes: longitudinal (spin–lattice) relaxation and transverse (spin–spin) relaxation. These mechanisms describe how the magnetization components recover and decay over time, ultimately restoring the equilibrium state of the system. The characteristic time constants associated with each process, T_1 and T_2, are intrinsic properties of the tissue and play a fundamental role in determining MRI contrast. [56], [58]

- Longitudinal relaxation (T_1)

The longitudinal relaxation process governs the recovery of the magnetization component aligned with the main magnetic field (B_0). After an RFr pulse, the longitudinal magnetization (M_z) is reduced from its equilibrium value (M_0) and gradually realigns with B_0 as energy is transferred from the excited nuclear spins back to their surrounding molecular environment, known as the "lattice." This exchange occurs because of interactions between nuclear spins and fluctuating local magnetic fields, typically arising from neighboring dipoles [58].

The time constant T_1 represents the time required for the longitudinal magnetization to recover to approximately 63% of its equilibrium value following a 90° RFr pulse. [56] Since spontaneous transitions between spin states are extremely rare, this recovery primarily results from stimulated interactions with local field fluctuations. In biological tissues, the efficiency of these interactions depends on molecular mobility: fluids with rapid molecular motion exhibit long T_1 times, while more rigid tissues, such as fat or muscle, display shorter T_1 values [57], [58].

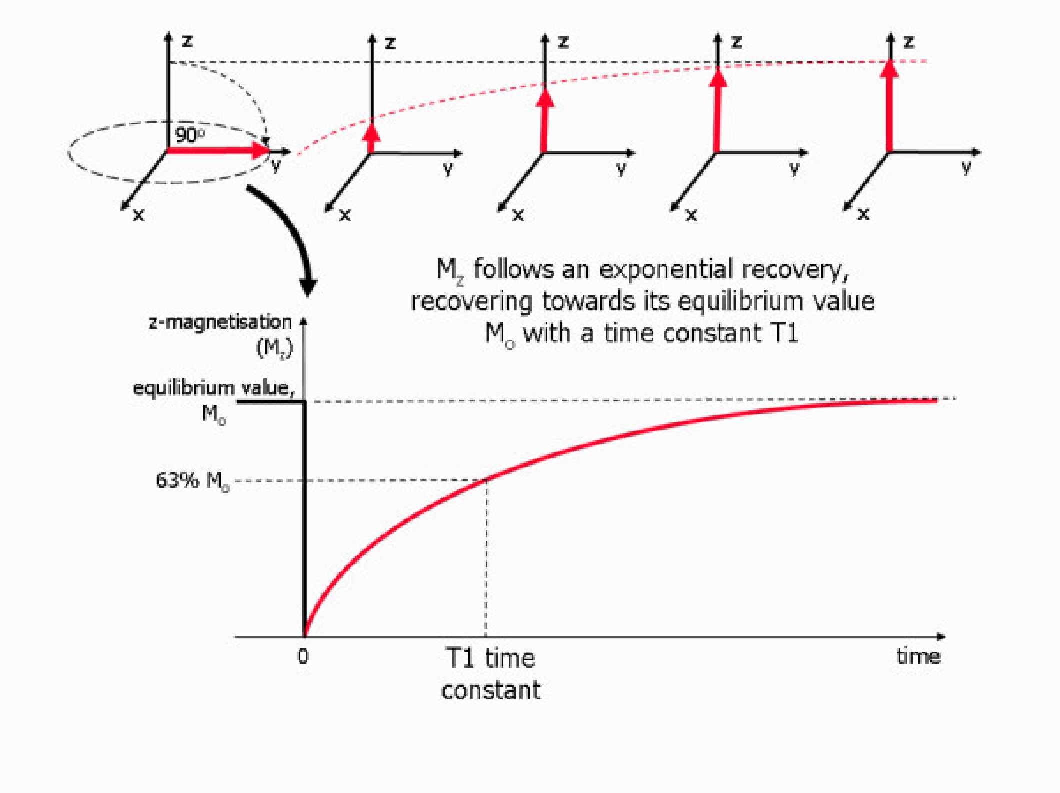

z
y
x
90°
M_z follows an exponential recovery, recovering towards its equilibrium value M_o with a time constant T1
z-magnetisation (M_z)
equilibrium value, M_o
63% M_o
0
T1 time constant
time

Figure 2-6 From [63]

- Transverse relaxation (T_2)

While longitudinal relaxation restores magnetization along B_0, the transverse component (M_{xy}) decays due to the loss of phase coherence among individual spins precessing in the transverse plane. Immediately after excitation, all spins precess synchronously, generating a strong coherent signal. Over time, slight variations in local magnetic fields cause individual spins to precess at slightly different frequencies, leading to progressive dephasing and signal decay [58], [59].

This process, driven by mutual interactions between neighboring spins (hence the term *spin–spin relaxation*), is characterized by the T_2 time constant, defined as the time required for the transverse magnetization to decay to 37% of its initial value. [56], [58]. Even in an ideally homogeneous magnetic field, random molecular motion and local dipolar interactions cause irreversible dephasing and energy exchange between spins.

In practical MRI systems, additional inhomogeneities in the external field (B_0) accelerate signal decay, producing an apparent relaxation time called ${T_2}^*$, which is shorter than the intrinsic T_2. Unlike T_2, this dephasing is not random and can be partially refocused using spin-echo techniques. [59]

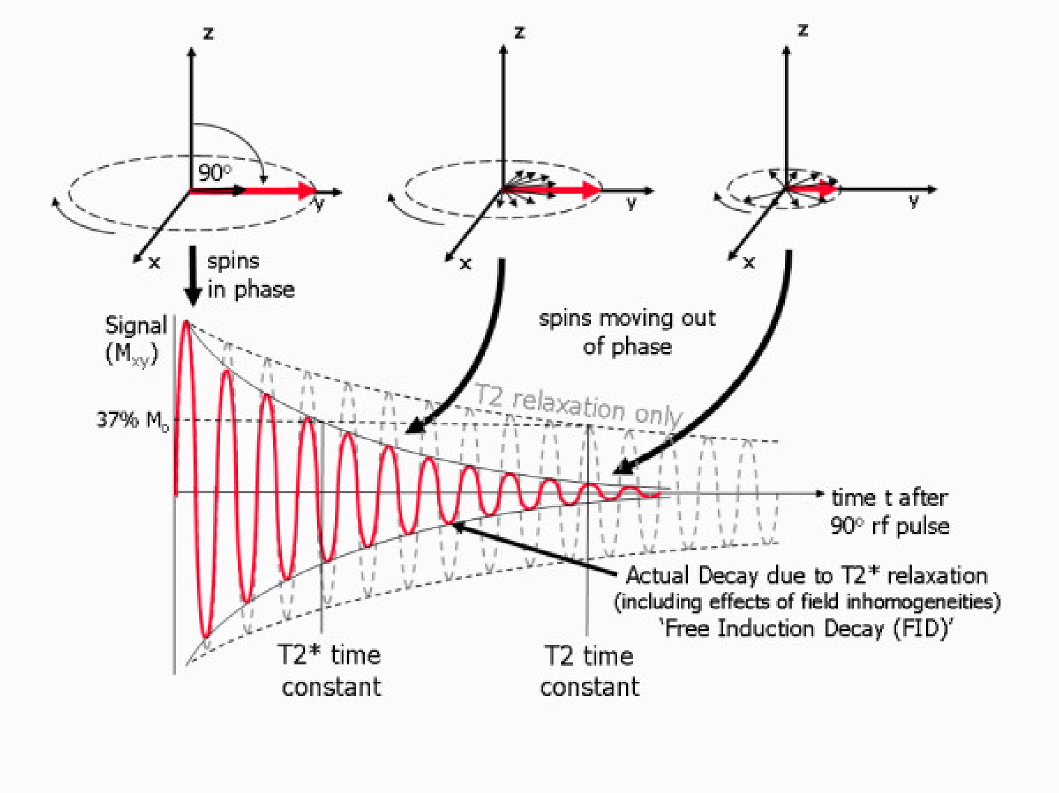

z
z
z
90°
y
y
y
x
x
x
spins in phase
spins moving out of phase
Signal (M$_{xy}$)
37% M$_0$
T2 relaxation only
time t after 90° rf pulse
Actual Decay due to T2* relaxation (including effects of field inhomogeneities) 'Free Induction Decay (FID)'
T2* time constant
T2 time constant

Figure 2-7 Transverse relaxation: T_2 and $T_2{}^*$ processes and time constants. Adapted from [64]

T_1, T_2 and $T_2{}^*$ relaxation times vary widely across biological tissues. For example, water and cerebrospinal fluid (CSF) exhibit long T_1 values (3000–5000 ms) and long T_2 times, whereas fat has a short T_1 (~260 ms) and shorter T_2, making it appear bright on T_1-weighted images and relatively dark on T_2-weighted ones [57], [58]. In general, T_1 relaxation reflects how efficiently spins exchange energy with their surroundings, while T_2 relaxation reflects how quickly they lose coherence with each other.

These relaxation mechanisms are central to MRI because the different recovery and decay rates of tissues provide the intrinsic contrast that allows the visualization and differentiation of anatomical and pathological structures. [57]

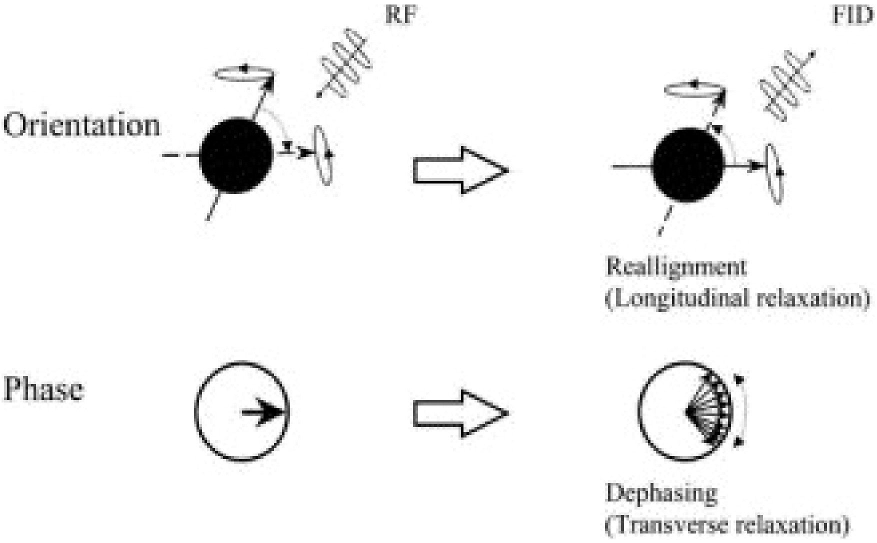


Figure 2-8 Relaxation processes: realignment and dephasing. From [56]

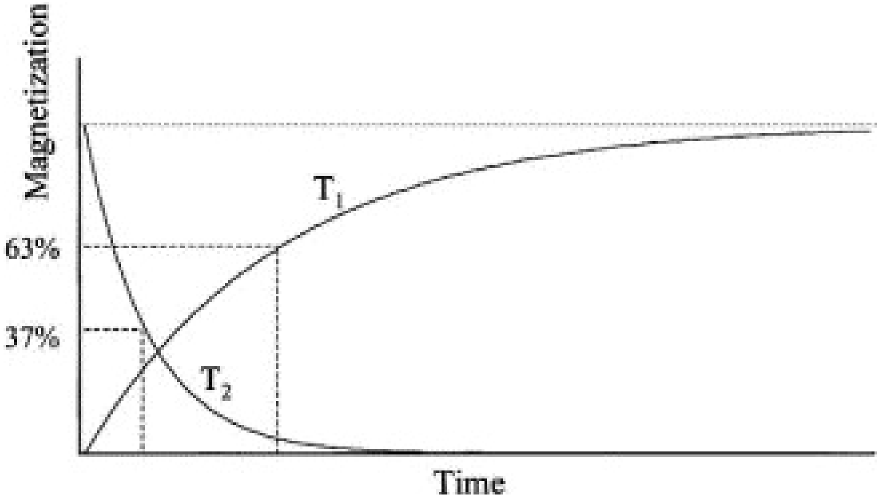


Figure 2-9 Longitudinal and transverse relaxation constants. From [56]

Tissue	*T_1 (ms)*	*T_2 (ms)*
Gray matter	1260.8±178.4	109.4±52.9
White matter	999.9±443.3	112.3±82.9
Cerebrospinal fluid	4627.3±788.1	-
Blood	1414.8±103.6	308.5±26.2

Table 2-1 Longitudinal and transversal relaxations times for human tissues (1.5 T), expressed as mean±standard deviation. Adapted from [65]

2.1.4 Pulse sequences

In routine imaging, the FID is seldom acquired directly because it decays too quickly to permit complete spatial encoding and robust contrast manipulation. Instead, clinical MRI relies on pulse sequences: structured combinations of RFr pulses, gradient events, sampling windows, and timing delays, that rephase or dephase spins in controlled ways to generate measurable echoes and to shape image contrast. [56], [57] Among these, spin-echo (SE) and gradient-echo (GRE) families are foundational.

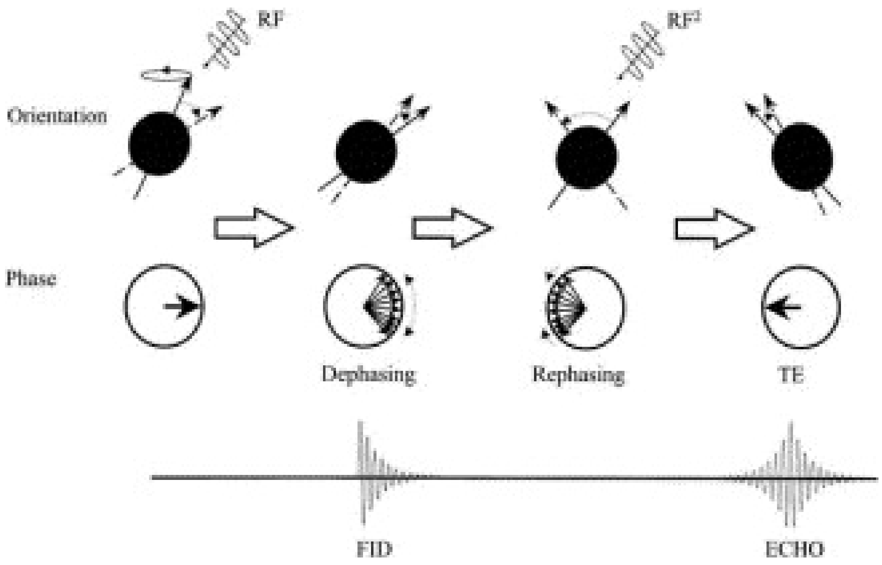


Figure 2-10 Representation of FID and echo signals. From [56]

A canonical SE sequence begins with a 90° RFr pulse that tips the net magnetization into the transverse plane. After excitation, a time delay is introduced before signal measurement. This delay, known as the echo time (TE), corresponds to the interval between the initial excitation pulse and the point at which the echo signal reaches its maximum amplitude. To refocus the dephasing spins and generate this echo, a

180° RFr pulse is applied midway through the interval, at $\frac{\mathrm{TE}}{2}$, to invert the accumulated phase dispersion caused by static B_0 inhomogeneities. As a result, the transverse magnetization rephases, producing a measurable signal at time TE. [56] This refocusing compensates for field inhomogeneities but not for intrinsic T_2 relaxation, so SE signals retain T_2 weighting.

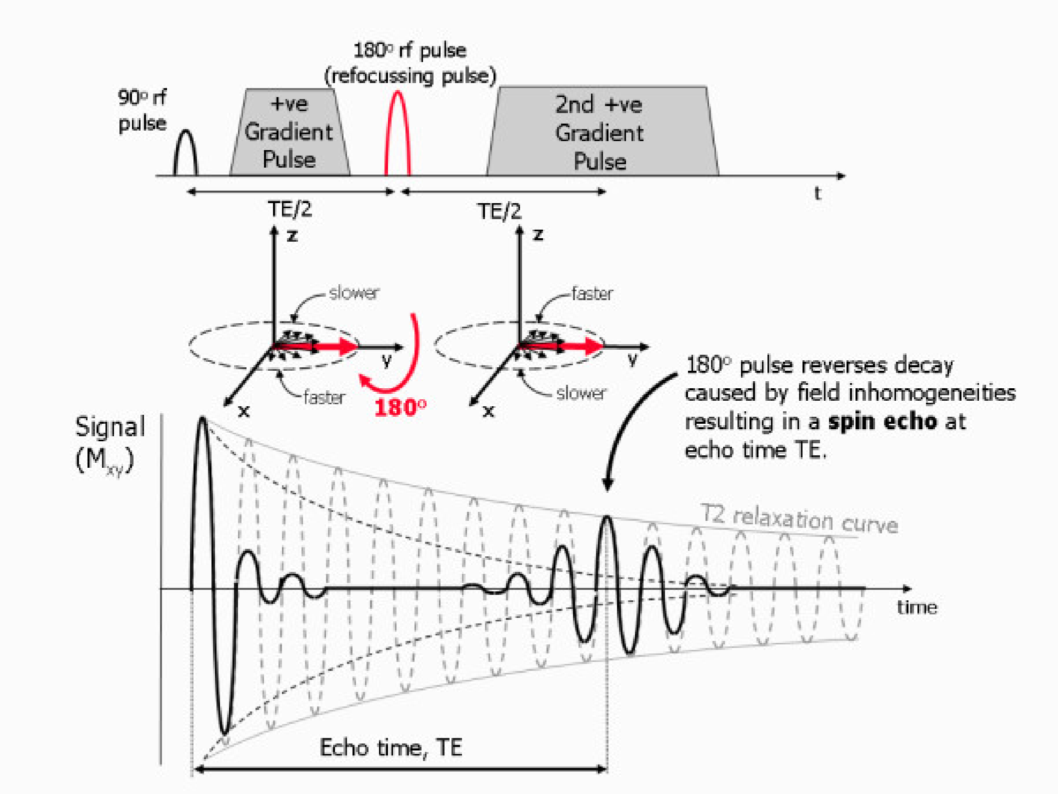


Figure 2-11 Spin echo. From [66]

The timing between successive excitation pulses, defined as the repetition time (TR), together with TE and the tissue-specific relaxation properties (T_1 and T_2), determines the predominant image contrast.

- Short TR / short TE → T_1-weighted, accentuating differences in longitudinal recovery (fat bright, water dark).
- Long TR / long TE → T_2-weighted, emphasizing transverse decay (fluids bright).
- Long TR / short TE → proton-density weighting (ρ), minimizing relaxation contrast [56], [67].

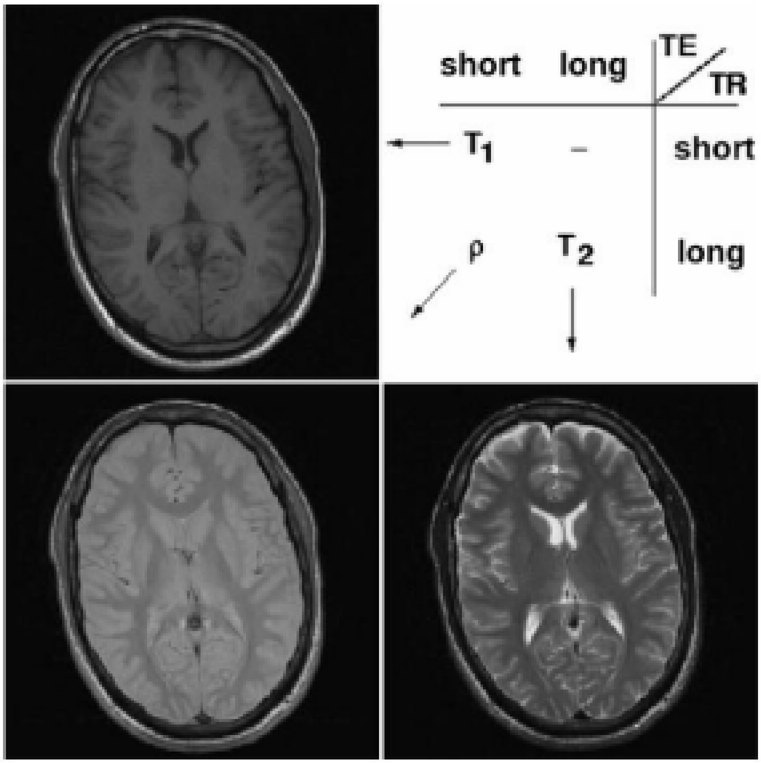


Figure 2-12 Brain slice images T_1, T_2 and ρ weighted. From [67]

The SE refocusing concept extends to multi-echo trains. In the Carr–Purcell–Meiboom–Gill (CPMG) scheme, a 90° pulse is followed by a series of evenly spaced 180° pulses, yielding multiple echoes whose amplitudes decay with T_2. Building on this, Fast/Turbo Spin Echo (FSE/TSE) collects many phase-encoded echoes per TR (echo-train length), drastically reducing scan time while preserving T_2-type contrast; TE within the train governs the dominant contrast of the reconstructed image. [61] GRE sequences achieve signal refocusing through controlled gradient polarity reversals instead of 180° RFr pulses. After excitation, a dephasing gradient is applied, then an opposite gradient rephases the spins to form a gradient echo. Because static field inhomogeneities are not refocused by RFr, GRE signals are T_2^*-sensitive and enable rapid acquisitions with flexible contrast (via flip angle, TR, and TE). GRE underpins many fast 2D/3D protocols and specialized applications [59], [61].

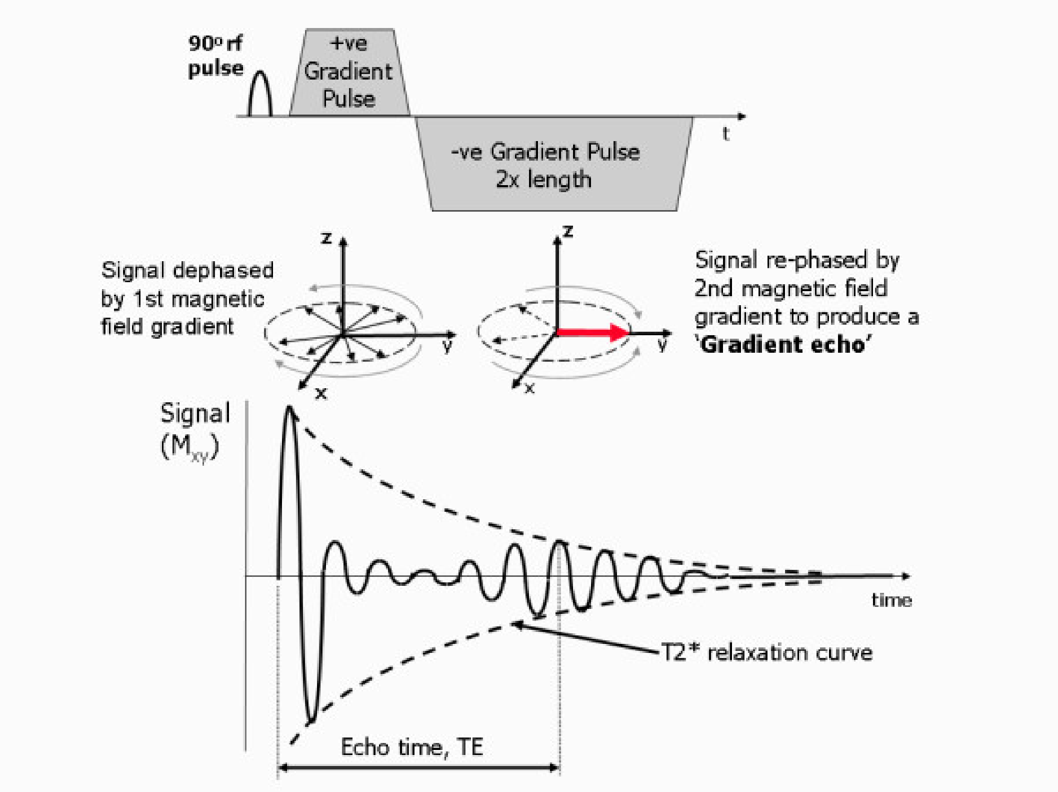

90º rf pulse
+ve Gradient Pulse
t
-ve Gradient Pulse 2x length
Signal dephased by 1st magnetic field gradient
z
y
x
Signal re-phased by 2nd magnetic field gradient to produce a 'Gradient echo'
Signal (M_{xy})
time
T2* relaxation curve
Echo time, TE

Figure 2-13 Gradient echo. From [68]

Inversion recovery (IR) sequences introduce a 180° inversion pulse prior to excitation to manipulate longitudinal magnetization recovery. By choosing an appropriate inversion time (TI), that is the delay between inversion and the subsequent 90° pulse, specific tissues can be nulled (e.g., fat in STIR, CSF in FLAIR), enabling powerful contrast manipulation beyond TR/TE alone. [59] In the simplest experiment, a single 90° pulse generates a transverse magnetization that yields an FID, but one must wait several T_1 times before repeating the measurement, and the signal typically decays before spatial encoding can be completed. Pulse sequences overcome these limitations by rephasing signal components (as in SE and GRE) and by systematically encoding contrast through the manipulation of TR, TE, TI and flip angle parameters. [56]

Because acquisition time is determined by the number of phase-encoding steps, slices, and TR intervals, several acceleration strategies have been developed to improve efficiency. These include reducing the number of phase encodes (at the cost of resolution), interleaved multi-slice imaging, and multi-echo trains such as those used in FSE/TSE sequences, which can shorten total acquisition times from hours to minutes. [61]

Echo-Planar Imaging (EPI) represents one of the fastest approaches: with a single excitation, rapid gradient switching traverses k-space to form one or multiple echoes, enabling single-shot EPI acquisitions on the order of a tenth of a second, crucial for fMRI, diffusion, and angiographic applications. Multi-shot EPI variants mitigate some artifacts while remaining fast. [61]

For high-resolution anatomical imaging, especially at high magnetic fields, Magnetization-Prepared Rapid Gradient Echo (MPRAGE) has become a standard T_1-weighted sequence. It combines a magnetization-preparation module with a rapid GRE readout to achieve strong T_1 contrast within practical times [69], [70]. The MP2RAGE sequence refines this concept by acquiring two GRE blocks at different inversion times and flip angles within a single preparation period, combining them to enhance signal uniformity and tissue contrast, thereby enabling precise delineation of cortical and subcortical structures at high fields strenghts. [70]

2.1.5 Spatial encoding

Spatial localization in MRI is achieved through the controlled application of magnetic field gradients, which systematically vary the strength of the main magnetic field (B_0) across different spatial directions. These gradients, generated by three orthogonal gradient coils (G_x, G_y and G_z), enable the encoding of spatial information in the MR signal by making the resonance frequency and phase of nuclear spins position dependent. [56], [57] The process involves three main steps: slice selection, phase encoding, and frequency encoding.

During slice selection, a magnetic field gradient, typically ranging between 5 and 40 mT/m, is applied along one spatial direction, commonly the longitudinal (z) axis. Because the Larmor frequency is directly proportional to the local magnetic field strength, this gradient introduces a spatially dependent variation in precessional frequency. An RFr pulse with a narrow bandwidth is then applied; only spins whose resonance frequency falls within this specific range are excited, resulting in the selection of a thin axial slice. The slice thickness (d) can be approximated by the relationship:

$$d \approx \gamma \cdot G \cdot \tau$$

where γ is the gyromagnetic ratio, G the gradient amplitude, and τ the duration of the RFr pulse. By adjusting the RFr bandwidth or the gradient strength, slices of different thicknesses can be selected, while modifying the RFr carrier frequency or the gradient orientation allows for the acquisition of slices in arbitrary planes. [56]

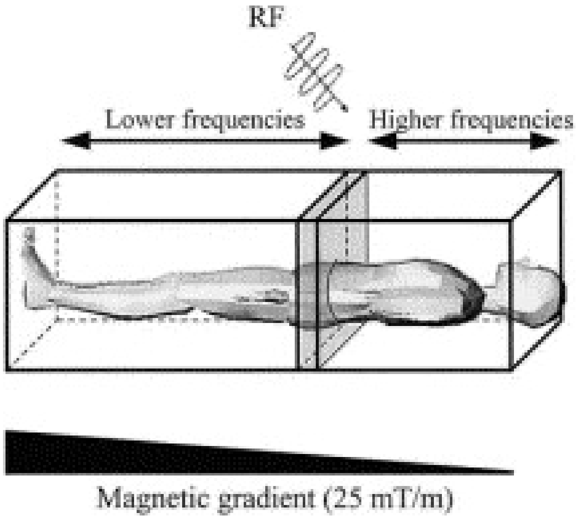


Figure 2-14 MRI spatial encoding. From [56]

Once a slice has been selected, spatial information within that slice must be encoded in two dimensions. This is accomplished through phase encoding and frequency encoding. In phase encoding, a short gradient pulse is applied along a second axis (commonly y) after the RFr excitation but before signal readout. During this brief period, spins precess at different frequencies depending on their spatial position within the gradient field. When the gradient is switched off, all spins return to the same frequency, but they retain the phase shift accumulated during the gradient application. The degree of phase shift is determined by both the amplitude and the duration of the phase-encoding gradient. [59] Repeating the acquisition multiple times while systematically varying this gradient generates signals with different phase distributions across the slice. Typically, 128 or 256 distinct phase-encoding steps are used, corresponding to the number of lines in the final image matrix. [61]

Frequency encoding, also known as the readout gradient, is applied during signal detection along the third orthogonal axis (typically x). This gradient imposes a linear change in magnetic field strength along the readout direction, causing spins at different positions to precess at distinct frequencies. As a result, each frequency in the detected signal corresponds to a specific spatial location. In gradient-echo (GRE) sequences, the readout gradient is first applied with one polarity to dephase the spins and then reversed to rephase them, creating the echo signal. [59]

Combining phase and frequency information enables each pixel within the slice to be uniquely identified by its specific combination of phase and frequency values. The set of acquired data points fills a matrix in frequency–phase space, known as k-space. The signal measured in k-space represents amplitude as a function of time; through the application of a Fast Fourier Transform (FFT), these data are converted into amplitude as a function of spatial frequency, which corresponds to image intensity in the final reconstruction. [56]

Accurate image formation requires that the number of acquired data points in k-space be at least equal to the number of pixels in the final image. To achieve this, the pulse sequence must be repeated for each phase-encoding gradient amplitude until the entire k-space is filled. While frequency encoding contributes minimally to the total acquisition time, since it occurs during a brief readout period on the order of milliseconds, phase encoding significantly affects scan duration, as it requires one RFr excitation per phase step and depends on the repetition time (TR). [59] Consequently, optimizing the balance between image resolution, acquisition time, and contrast is a central consideration in MRI protocol design.

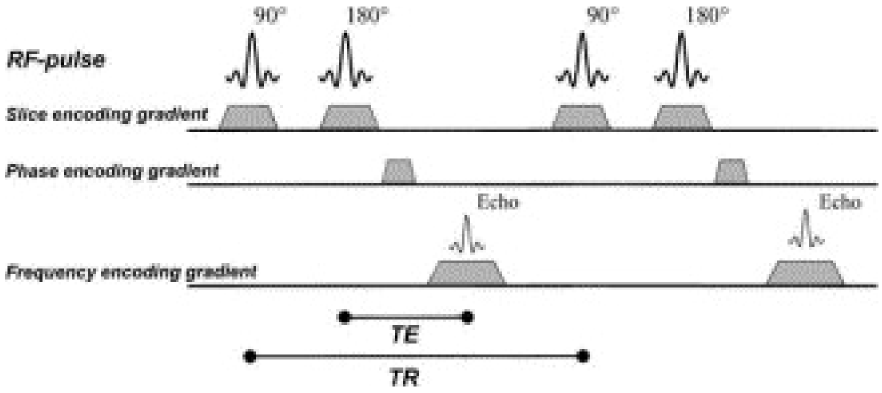


Figure 2-15 Example of a spin-echo sequence. From [56]

2.1.6 Contrast

The contrast in a MR image represents the difference in signal intensity between distinct regions or voxels, reflecting variations in the physical and biochemical properties of tissues. An MR image is said to be weighted toward a particular parameter when the observed contrast primarily depends on differences in that property across tissues. [71] By appropriately tuning the acquisition parameters, MRI can emphasize the influence of specific relaxation mechanisms, most notably T_1, T_2, and ρ, thereby enhancing diagnostic sensitivity to pathological changes.

Among the numerous sequence parameters, two are particularly critical for image contrast: the TR and the TE. As discussed earlier, TR defines the interval between consecutive excitation pulses, while TE denotes the time between the initial RFr pulse and the peak of the measured signal. Adjusting these parameters selectively enhances the contribution of longitudinal or transverse relaxation effects, thus determining whether an image is T_1-weighted, T_2-weighted, or ρ –weighted.

T_1-weighted images primarily reflect differences in longitudinal relaxation times among tissues. They are obtained using short TR values (typically ≤200 ms) and short TE values (≤25 ms). Under these conditions, tissues with shorter T_1, such as fat, recover their longitudinal magnetization more rapidly between successive excitations and therefore appear bright, whereas tissues with longer T_1, such as water or CSF, remain partially saturated and appear darker. This weighting provides strong anatomical detail and is particularly useful for visualizing structures rich in lipids or assessing post-contrast enhancement. [59]

Conversely, T_2-weighted images emphasize differences in transverse relaxation times. A long TE (usually >80 ms) allows transverse magnetization to decay, accentuating tissues that retain signal longer, such as fluids, which thus appear bright. To minimize T_1-related effects, a long TR (greater than 2 s) is also applied. This combination enhances contrast based on the rate of transverse relaxation, making T_2-weighted imaging especially sensitive to edema, inflammation, and lesions with high water content [59], [61].

When both TR and TE are long, the effects of relaxation are minimized and the contrast instead depends mainly on the local proton density. These ρ-weighted images exhibit a high signal-to-noise ratio but relatively low intrinsic contrast, making them useful for anatomical delineation and for sequences requiring uniform signal intensity across tissues. [59]

Additional contrast manipulation can be achieved using inversion recovery techniques, which introduce a 180° inversion pulse before excitation to selectively null the signal of specific tissues. The key parameter in these sequences is the inversion time (TI): the delay between the inversion and the excitation pulse. Short TI values (around 150 ms) suppress signals from fat, forming the basis of Short Tau Inversion Recovery (STIR) sequences, whereas longer TIs (approximately 1–2 s) suppress signals from fluids such as CSF, as in Fluid-Attenuated Inversion Recovery (FLAIR) imaging. [61] These methods allow selective attenuation of unwanted tissue signals, improving lesion visibility and diagnostic specificity, particularly in neuroimaging.

2.2 Structural Magnetic Resonance Imaging

MRI represents the gold standard for the noninvasive assessment of intracranial anatomy, offering excellent spatial resolution and soft-tissue contrast for the identification of structural abnormalities.[72] The human brain is composed primarily of two distinct tissue types, gray matter (GM) and white matter (WM), whose differing microstructural properties give rise to characteristic MRI signal patterns. White matter is predominantly composed of myelinated axonal tracts that facilitate long-range neuronal communication, whereas gray matter consists mainly of neuronal cell bodies, dendrites, and glial cells.[73]

The principal determinant of MRI contrast between these two tissue classes is the presence of myelin, the multilamellar lipid membrane that ensheathes axons. Myelin influences MRI contrast by shortening both T_1 and T_2 relaxation times and slightly reducing proton density. This occurs because the highly organized lipid–protein structure of the myelin sheath restricts the mobility of surrounding water molecules and limits their interaction with macromolecules.[74] Consequently, white matter exhibits a shorter T_1 and T_2 compared to gray matter, which underlies the brightness differences typically observed in conventional anatomical images.

Structural MRI encompasses imaging sequences that capture the morphology and organization of brain tissues with high spatial detail. While the term is often used synonymously with T_1-weighted imaging, it also includes T_2-weighted and advanced contrast mechanisms such as FLAIR. These techniques provide complementary information on tissue composition, water content, and pathological alterations, forming the foundation of both clinical diagnosis and morphometric research [73], [74].

High-resolution three-dimensional T_1-weighted sequences, such as MPRAGE and MDEFT (Modified Driven Equilibrium Fourier Transform), are widely used for detailed anatomical mapping. Both are designed to enhance the contrast between gray matter, white matter, and cerebrospinal fluid (CSF). MPRAGE achieves this by introducing a magnetization preparation phase followed by rapid gradient-echo acquisition, whereas MDEFT modifies the equilibrium magnetization to improve contrast uniformity and spatial resolution. [74]

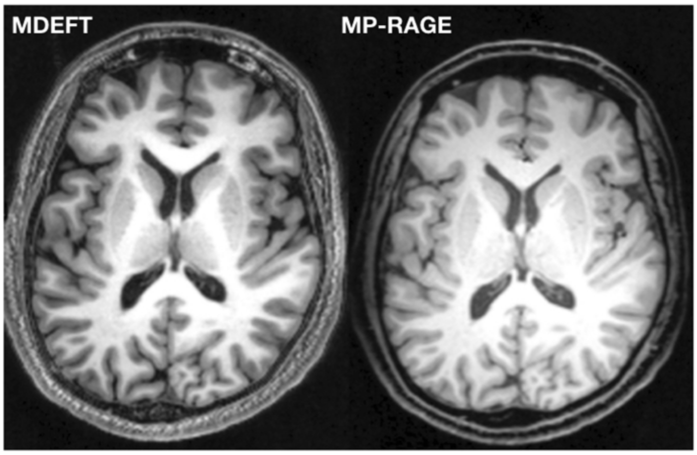


Figure 2-16 MDEFT and MPRAGE MR sequences. From [75]

In T_1-weighted images, tissues rich in lipids, such as myelinated white matter, appear brighter due to their short T_1 relaxation times, while fluids like CSF, with longer T_1, appear dark. The resulting images show white matter as light gray, gray matter as darker gray, and CSF as black. These contrasts enable the precise delineation of cortical and subcortical boundaries, though the distinction between gray matter and CSF can sometimes be limited in pure T_1-weighted acquisitions.

In contrast, T_2-weighted imaging provides an inverted contrast pattern: fluids appear bright, while white matter appears darker than gray matter. Although this contrast is less effective for differentiating white and gray matter, it enhances the visibility of fluid-filled spaces such as ventricles and the subarachnoid space, making it valuable for detecting edema, inflammation, and other pathological fluid accumulations. [73]

A further refinement of T_2-weighted imaging is the Fluid-Attenuated Inversion Recovery (FLAIR) sequence, which suppresses the high signal from CSF to improve the visualization of periventricular and cortical lesions.[76] FLAIR has become an essential complement to conventional T_1 imaging, particularly in neurological disorders such as multiple sclerosis, where lesion detection near CSF boundaries is crucial.

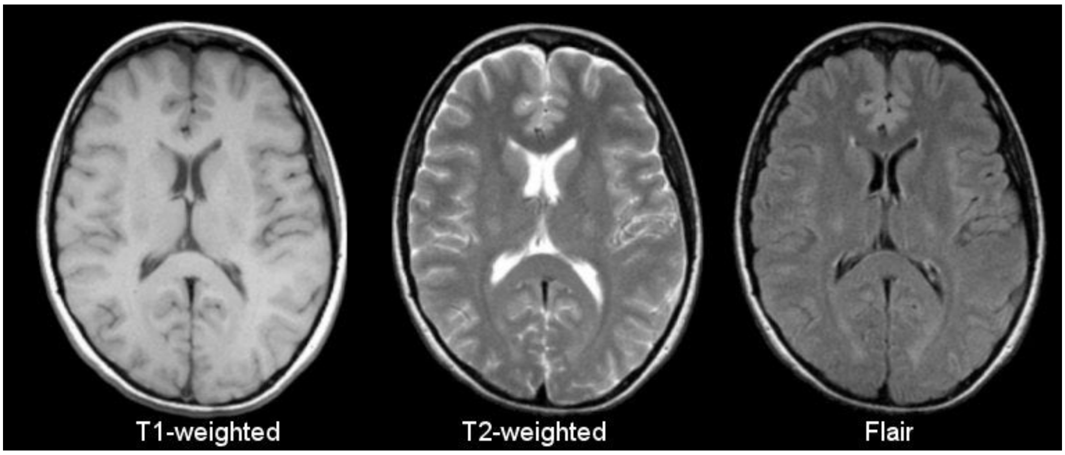


Figure 2-17 T_1-weighted, T_2-weighted and FLAIR brain images. From [77]

2.2.1 Diffusion weighted Magnetic Resonance Imaging

The concept of diffusion originates from Fick's law, which describes the flux of particles from regions of high to low concentration. Einstein later established the relationship between the mean-squared displacement of particles undergoing

Brownian motion and the diffusion coefficient, which depends on temperature, pressure, and the properties of the medium. [78], [79]

In biological tissues, diffusion magnetic resonance imaging (dMRI) exploits the sensitivity of MR signals to the random motion of water molecules, providing a non-invasive probe of tissue microstructure. The development of diffusion-weighted imaging (DWI) in 1985 marked a major breakthrough in neuroimaging, followed by further advances from Le Bihan in 1989 [80] and the introduction of diffusion tensor imaging (DTI) and fiber tractography by Basser [81] and Le Bihan [82] in the early 1990s. [78], [83]

In DWI, magnetic field gradients are applied before and after the refocusing pulse in a SE or gradient-echo sequence. These gradients encode the displacement of water protons into the phase of the MR signal, leading to signal attenuation that follows a mono-exponential decay as a function of the b-value, a parameter determined by the gradient strength, duration and diffusion time. The apparent diffusion coefficient (ADC) quantifies the average diffusivity of water molecules within a voxel. [78]

Although extensions of the DWI model allow separation of diffusion and perfusion effects, the most widely used and conceptually straightforward model for biological tissues is diffusion tensor imaging, which assumes anisotropic but Gaussian diffusion. [78]

Because axonal membranes and myelin sheaths constrain water movement primarily along the fiber axis, diffusion in white matter (WM) is inherently anisotropic. [83] Thus, DWI provides indirect yet powerful information about WM microstructure.

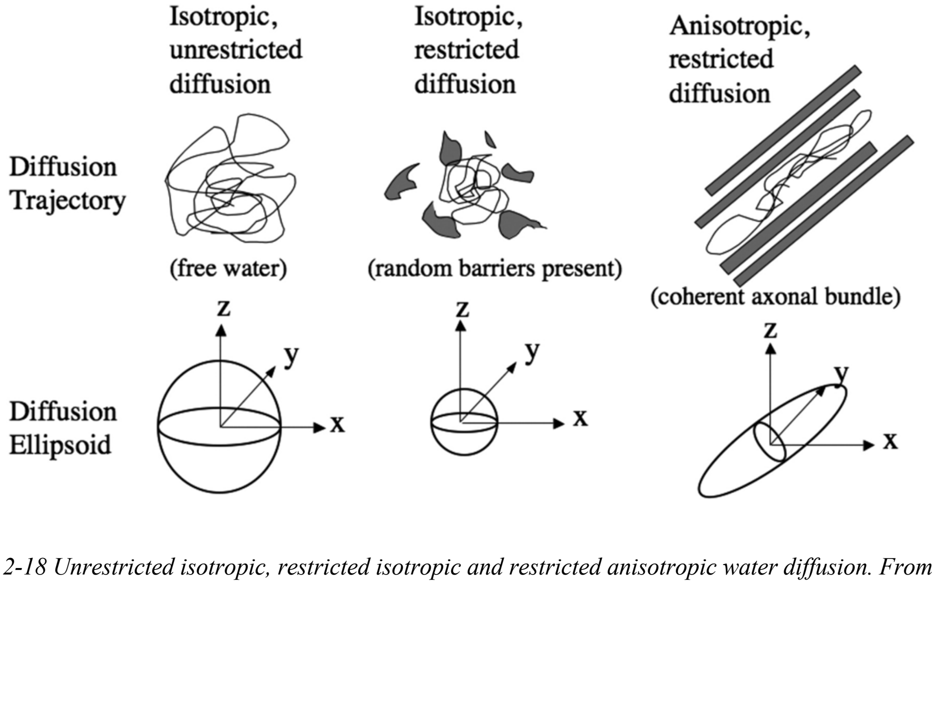

Isotropic, unrestricted diffusion
Isotropic, restricted diffusion
Anisotropic, restricted diffusion
Diffusion Trajectory
(free water)
(random barriers present)
(coherent axonal bundle)
Diffusion Ellipsoid
z
y
x
z
y
x
z
y
x

Figure 2-18 Unrestricted isotropic, restricted isotropic and restricted anisotropic water diffusion. From [84]

2.2.1.1 Diffusion tensor imaging

To model anisotropic diffusion, DTI represents diffusion as a 3×3 symmetric tensor, in which each element corresponds to the diffusion coefficient along or across a principal direction. [78] The tensor can be diagonalized to yield eigenvalues (λ_1, λ_2, λ_3) and eigenvectors, which describe the magnitude and orientation of diffusion along the main axes of the diffusion ellipsoid.

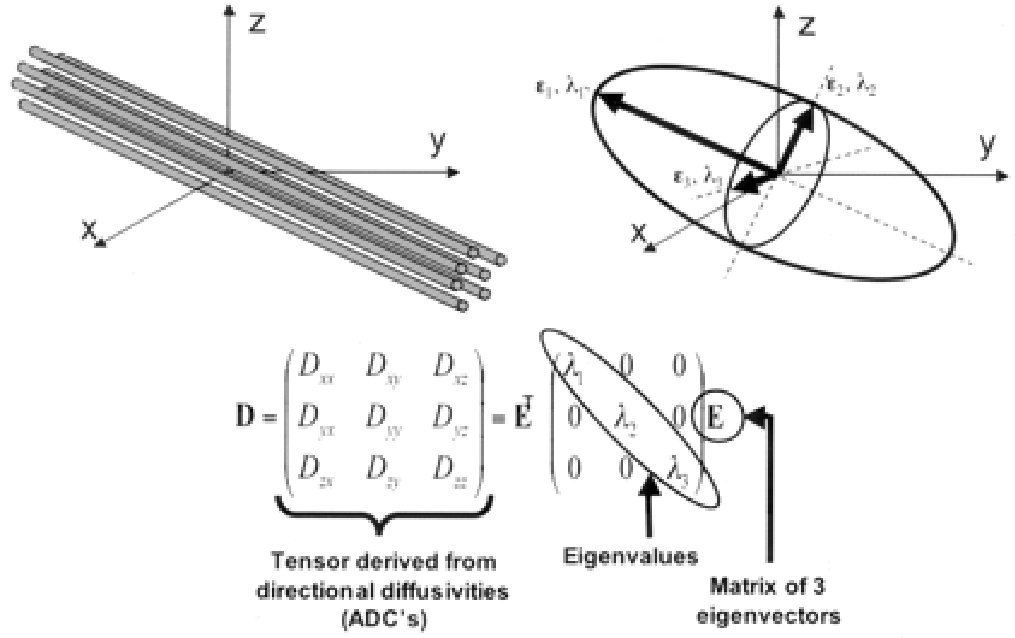

z
y
x
z
y
x
ε₁, λ₁
ε₂, λ₂
ε₃, λ₃
D = (Dxx Dxy Dxz; Dyx Dyy Dyz; Dzx Dzy Dzz) = E^T (λ₁ 0 0; 0 λ₂ 0; 0 0 λ₃) E
Tensor derived from directional diffusivities (ADC's)
Eigenvalues
Matrix of 3 eigenvectors

Figure 2-19 Diffusion tensor. From [85]

From these eigenvalues, two scalar indices can be derived: the mean diffusivity (MD) and the fractional anisotropy (FA).

MD represents the average molecular mobility and is defined as:

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}$$

FA quantifies the degree of directional dependence of water diffusion and is mathematically defined as the normalized variance of the eigenvalues of the diffusion tensor.

$$FA = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_1 - \lambda_3)^2}{2 \cdot ({\lambda_1}^2 + {\lambda_2}^2 + {\lambda_3}^2)}}$$

FA values range from 0 (isotropic diffusion) to 1 (maximally anisotropic diffusion). [78], [86] In a completely isotropic medium, where $\lambda_1 = \lambda_2 = \lambda_3$, the diffusion ellipsoid assumes a spherical shape, as observed in cerebrospinal fluid, and FA equals 0. Conversely, in tissues with highly organized fiber architecture, the eigenvalues differ substantially, the diffusion ellipsoid becomes elongated and FA approoaches 1, as seen in coherent WM tracts. [78], [86], [87]

Higher FA values are observed in regions containing densely packed, coherently oriented and heavily myelinated fibers, such as the corpus callosum or internal capsule. In contrast, FA decreases in tissues with reduced myelination, fiber crossings, axonal loss or increased extracellular space. [83], [86]
These DTI-derived metrics serve as quantitative indicators of WM integrity and are widely used to investigate both healthy and pathological brain microstructure.
By combining local estimates of diffusion tensors across voxels, fiber tractography can reconstruct the major WM pathways of the brain. Deterministic and probabilistic approaches are employed to generate whole-brain structural connectomes, providing valuable insight into large-scale network organization and connectivity. [83]

2.2.1.2 White matter microstructure

Diffusion MRI provides multi-scale insights about WM architecture. Local, voxel-wise measures such as FA, MD and their derived indices (axial and radial diffusivity) reflect microstructural characteristics including axon density, fiber coherence, membrane integrity and myelin content. [83] [86]
At the macroscopic level, tractography-based analyses enable the reconstruction of fiber bundles and whole-brain connectomes, revealing the structural organization of major pathways and the inter-individual variability of WM networks. [83]
Alterations in these microstructural properties can influence physiological parameters such as conduction velocity, signal synchronization and functional connectivity, which in turn may impact behavioral and cognitive outcomes. [83]

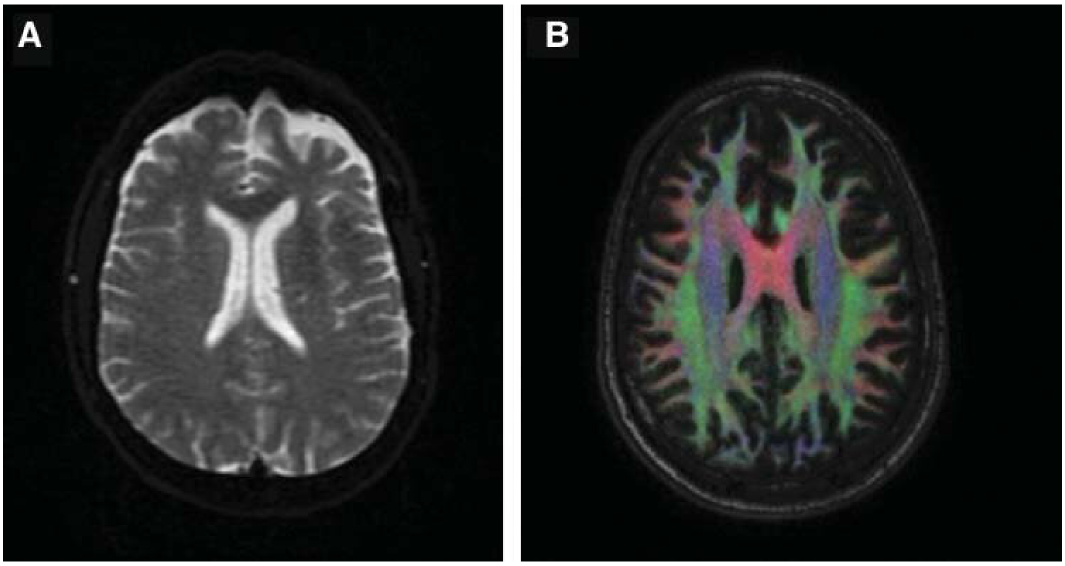


Figure 2-20 Diffusion weighted image (A) and diffusion tensor image (B) of the brain. From [88]

2.3 Functional Magnetic Resonance Imaging

Conventional MRI provides high-resolution anatomical images based on the spatial distribution and relaxation properties of mobile hydrogen nuclei. Its exceptional soft-tissue contrast and submillimetric spatial resolution make it indispensable in clinical diagnostics. However, beyond structural characterization, magnetic resonance techniques can also measure physiological processes occurring in real time. Among these, functional MRI (fMRI) has emerged as a major tool for investigating brain activity and metabolism [43].

fMRI employs the same hardware as conventional MRI and is available on standard clinical scanners, typically operating at 1.5 T or higher. It is noninvasive, does not require radioactive tracers or contrast agents, and provides relatively high spatial resolution at comparatively low cost. The method detects regional, time-dependent changes in cerebral metabolism, reflecting neural activation induced either by cognitive tasks or by spontaneous fluctuations in the resting brain [44].

The signal of interest in fMRI is the blood oxygen level–dependent (BOLD) contrast, which arises from changes in the magnetic properties of hemoglobin depending on its oxygenation state. Neuronal activity increases local energy demand, stimulating oxidative metabolism and leading to a rise in cerebral blood flow and oxygen delivery. Intriguingly, the inflow of oxygenated blood often exceeds the actual metabolic requirement, resulting in a transient reduction in deoxyhemoglobin concentration within active regions. Since deoxyhemoglobin is paramagnetic, while oxyhemoglobin is not, this shift alters local magnetic susceptibility and reduces spin dephasing, producing a modest increase (typically around 1%) in MR signal intensity [43][46].

The dynamics of this hemodynamic response are relatively slow compared with the underlying neuronal activity. The BOLD signal begins approximately 2 seconds after neural activation, peaks around 5–8 seconds, and subsequently falls below baseline in a post-stimulus undershoot that can last up to 10 seconds. An initial negative dip is sometimes observed within the first seconds, likely reflecting transient oxygen consumption before the vascular response compensates [46].

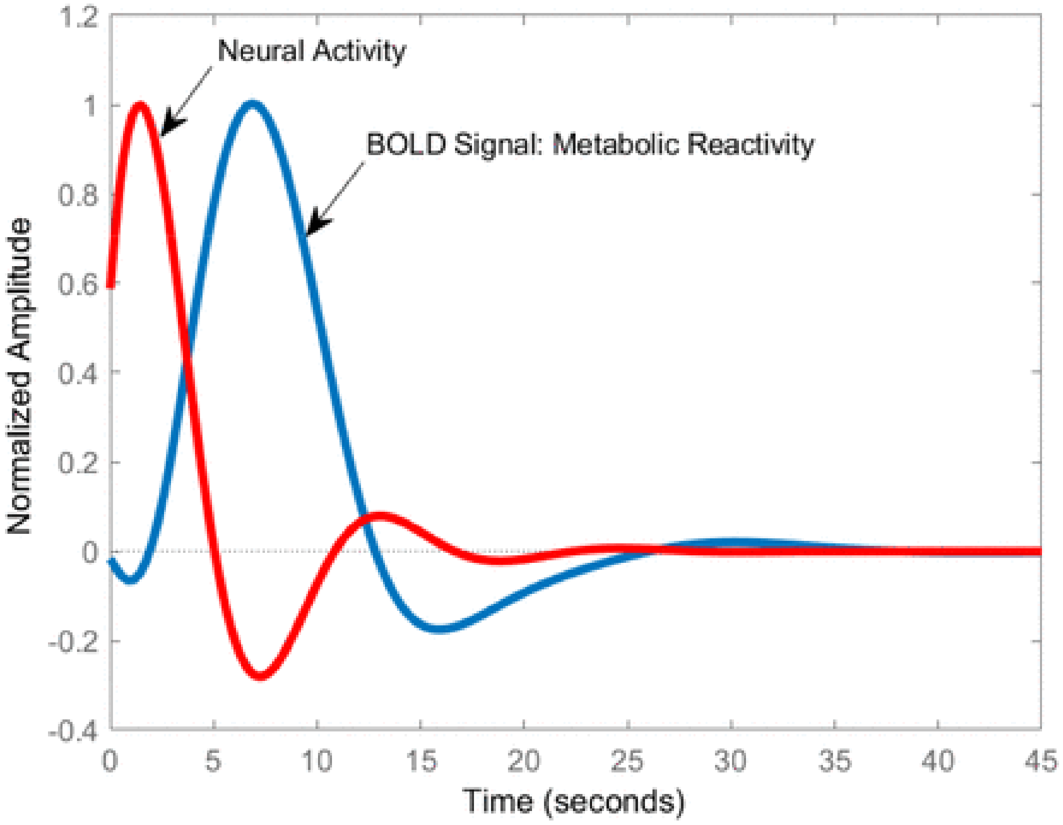


Figure 2-21 BOLD signal and neural activity over time. From [89]

Although powerful, the fMRI signal is affected by numerous non-neuronal sources of variability. Random thermal noise, arising from both the scanner electronics and tissue conductivity, is uniformly distributed and typically mitigated by averaging. Scanner drift causes slow, low-frequency intensity fluctuations, while physiological noise, stemming from respiration, cardiac pulsation, and subtle head motion, induces structured artifacts that vary across voxels [45]. These sources of noise can obscure the true BOLD fluctuations if not properly addressed during data preprocessing.

To mitigate such confounds, fMRI data undergo extensive preprocessing. This includes motion correction, removal of non-neuronal signals from cerebrospinal fluid and white matter, and spatial registration to a standard anatomical template to ensure intersubject comparability [46][48]. Temporal filtering, typically restricting the frequency range to below 0.1 Hz, and spatial smoothing further improve the signal-to-noise ratio. In resting-state fMRI (rs-fMRI), it is common to discard the initial volumes of the acquisition to eliminate transient, non–steady-state effects and enhance data stability.

Despite its versatility, fMRI presents inherent limitations in spatial and temporal resolution. High-resolution structural MRI can achieve voxel sizes below 1 mm^3, but such acquisitions are too slow for dynamic imaging. Typical fMRI scans achieve a spatial resolution of approximately 3 × 3 × 5 mm^3, with whole-brain coverage obtained every 2 seconds. This represents a balance between spatial detail and temporal sampling: shorter TR improve temporal resolution but reduce signal intensity and image quality. Moreover, variability in brain anatomy across individuals necessitates spatial normalization to a common atlas, a process that can introduce minor distortions or blurring of activation maps [45].

2.4 MRI evidence of cortical involvement in PPA

The nfvPPA patients brain is characterized by lesions in the left posterior fronto-insular region. [8] Among these regions, the inferior frontal gyrus is severely impacted in this PPA variant: the left inferior frontal area has been identified as the epicenter specific to this variant and shows severe atrophy, while the posterior inferior part presents both structural and functional impairments. [55] Atrophy is also present in the insula, premotor cortex and supplementary motor areas. [8]
Regarding WM fibers, the most affected in nfvPPA are those connecting frontal, subcortical and parietal regions, particularly along the dorsal pathway, the superior longitudinal fasciculus. Functional connectivity is impaired too, mainly involving the posterior middle temporal gyrus and the inferior frontal gyrus of the left hemisphere. [55]

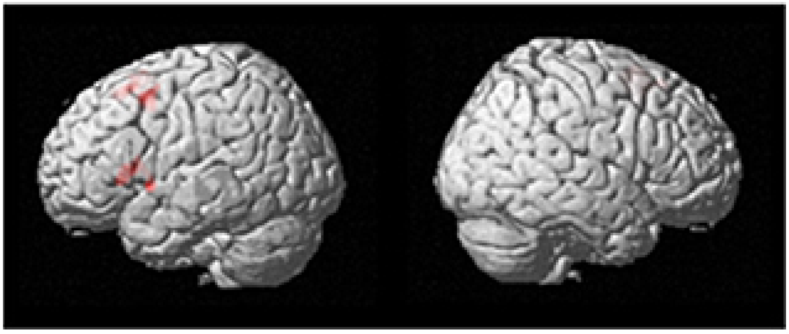


Figure 2-22 Regions showing significant hypometabolism in nfvPPA, with respect to healthy controls. Statistical maps thresholded at p < 0.05, family-wise error corrected at the peak level. Adapted from [90]

The epicenter of the semantic variant is the anterior temporal lobes. [55] These regions show atrophy in the ventral and lateral parts of both hemispheres, usually with a worse involvement on the left side at disease onset. [8] Atrophy also involves the right hemisphere, particularly the temporal regions, and the hippocampus, particularly in its anterior portion. With progressive deterioration, even regions such as the ventral and lateral temporal cortices and contralateral temporal and frontal regions become affected. Regarding WM, lesions can be seen between the temporal and the occipital lobes and between the temporal lobe and the orbitofrontal cortex, with more severe damage on the left side; the damage extends to the temporal segment of the dorsal frontoparietal pathway. The right side is affected as well, with damage of the WM bundles. Functional connectivity is reduced at the temporal pole and between the anterior temporal lobe and early sensory and modality-specific cortices. [55]

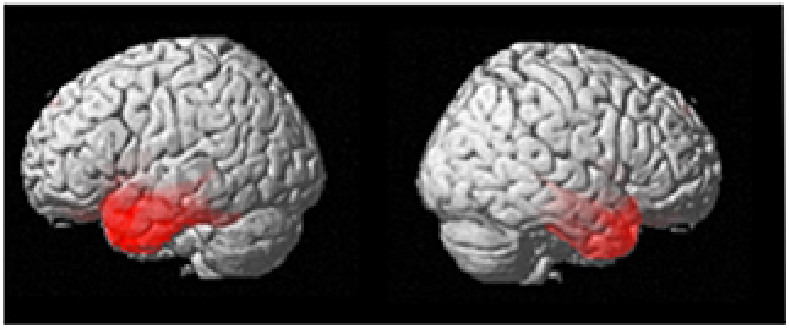


Figure 2-23 Regions showing significant hypometabolism in svPPA, with respect to healthy controls. Statistical maps thresholded at p < 0.05, family-wise error corrected at the peak level. Adapted from [90]

The left posterior superior and middle temporal gyri are the epicenter of lvPPA. Atrophy involves the left posterior perisylvian region and the inferior parietal lobule; [8] [55] damage can extend to the ipsilateral parietal and frontal lobes as well as the contralateral temporal lobe. WM fibers are damaged on the left side, particularly those connecting the parietal with frontal and posterior temporal areas, and thoose within the left posterior inferior longitudinal fasciculus. In more advanced patients, lesions may also affect the anterior inferior longitudinal fasciculus, the uncinate fasciculus and the superior longitudinal fasciculus. Functional connectivity is impaired in regions such as the inferior parietal lobule, superior and middle temporal gyri and frontal regions. Alterations may be present in both the posterior superior temporal gyrus and the inferior frontal lobe. [55]

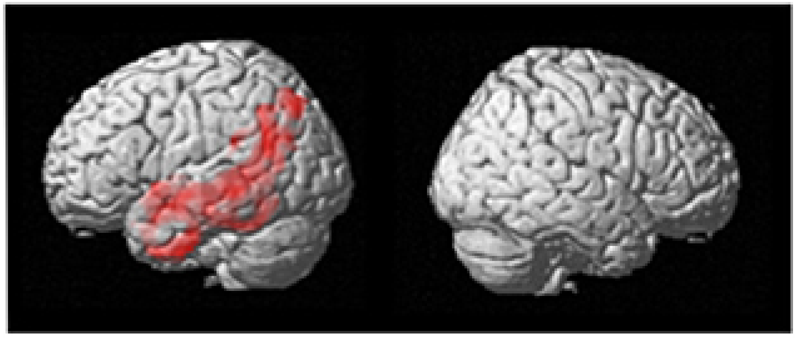


Figure 2-24 Regions showing significant hypometabolism in lvPPA with respect to healthy controls. Statistical maps thresholded at $p < 0.05$, family-wise error corrected at the peak level. Adapted from [90]

2.5 MRI findings in neurodegenerative diseases

MRI has become an essential tool for investigating neurodegenerative diseases, allowing researchers and clinicians to delineate both structural and functional alterations associated with these conditions. Beyond traditional visual assessment, MRI now encompasses a broad spectrum of advanced techniques, ranging from automated volumetric analysis and cortical thickness estimation to diffusion tensor and functional MRI, that enable the in vivo characterization of multiple aspects of neurodegeneration. [91]

These approaches facilitate the quantitative evaluation of cortical thinning, subcortical atrophy, and white matter disruption, providing objective biomarkers that support differential diagnosis and longitudinal monitoring. [92]

Importantly, the ability of MRI to identify disease-specific structural and connectivity patterns holds significant promise for the development of personalized, disease-modifying therapies, which will increasingly depend on early and precise characterization of neural degeneration. [91]

Table 2-2-4 Summary of MRI studies investigating neurodegenerative diseases.

Authors-Year	***Neurodegenerative disease***	***Imaging technique***	***Main features/metrics extracted***	***Focus***
Canu et al. - 2025	PPA	Structural MRI	Gray matter volumes	Variant classification
Bârlescu et al. - 2025	Progressive supranuclear palsy (PSP)	DTI	Whole brain BSS	In vivo analysis of regions involved in PSP
Rodriguez-Vieitez et al. - 2025	Genetic frontotemporal dementia	T_1 and DW MRI	Cortical MD, cortical thickness	Investigation of the association between cortical microstructure and disease severity and clinical progression
Owens-Walton et al. - 2024	PD	DWI	FA and MD maps	Evaluation of WM microstructure
Qin et al. - 2024	Amyotrophic lateral sclerosis	DTI	MD and FA	Correlation between DTI parameters and neurofilaments; Classification
Zhang et al. - 2023	AD	DTI	TBSS and multi-metric (FA, MD and radial diffusivity)	Classification
Neophytou et al. - 2023	PPA	DTI	tractography	Investigation of right hemisphere role in language processes
Spotorno et al. - 2023	AD	Structural and diffusion MRI	Cortical thickness values, Mean Squared Displacement and Return-To-Origin Probability voxel-wise maps	Investigation of the sensitivity of diffusion MRI markers to cortical microstructural changes

Illán-Gala et al. - 2022	PPA	Structural and DW MRI	Cortical thickness and MD	Study of microstructural changes using cortical MD
Lavrador et al. - 2022	Huntington's Disease	T_1 and DW MRI	GM voxel-based measure and FA values from all brain and from subcortical regions of interest	Disease stage classification
Adanyeguh et al. - 2021	Huntington's Disease	DW Magnetic Resonance Spectroscopy	Fixel based analysis and MD	Identification of microstructural imaging markers of WM degeneration
Bouchard et al. - 2019	svPPA	DTI	TBSS and GM volumetric data	Characterization of WM damage
Wen t al. - 2016	PD	DTI	TBSS and diffusion connectometry	White matter microstructural characteristics study
Agosta et al. - 2015	PPA	DTI	Cortical thickness and white matter damage metrics	Variant classification (nfvPPA and svPPA)
Schwindt et al. - 2011	PPA	DTI	TBSS, voxel-based morphometry	Investigation on svPPA and nfvPPA differences

Recent research has shown that combining structural and diffusion-based MRI metrics offers a more sensitive and specific assessment of neurodegenerative pathology than either modality alone. This multimodal approach is particularly relevant in disorders such as PPA, where distinct cortical and white matter patterns reflect the clinical heterogeneity of the syndrome.[92]

A study applying this framework investigated whether the integration of T_1-weighted structural measures and DTI indices could discriminate between the nfvPPA and the svPPA at the individual level. Thirteen patients with each PPA subtype and twenty-three healthy controls underwent high-resolution MRI scans. Cortical thickness and diffusion parameters from major associative and interhemispheric tracts were extracted and analyzed using a random forest (RF) classifier to identify the features most predictive of each clinical phenotype. [92]

The results indicated that the strongest discriminative markers were increased diffusivity in the left inferior longitudinal fasciculus and uncinate fasciculus, and

reduced cortical thickness in the left temporal pole and inferior frontal gyrus. When both cortical and diffusion-derived measures were combined, the classification achieved an area under the curve (AUC) of 0.91, accuracy of 0.89, sensitivity of 0.92, and specificity of 0.85. The performance of this combined model surpassed that of models relying solely on gray matter or white matter features (accuracies of 0.73 and 0.68, respectively). Cross-validation through leave-one-out analysis confirmed the robustness of the multimodal approach. [92]

Cortical surface reconstructions revealed distinct topographical patterns of atrophy across the two variants. nfvPPA exhibited predominant thinning in the left inferior frontal and insular regions, while svPPA showed marked degeneration of the anterior temporal cortex, particularly on the left. These spatial distributions closely paralleled the linguistic and semantic deficits typical of each form of PPA. [92]

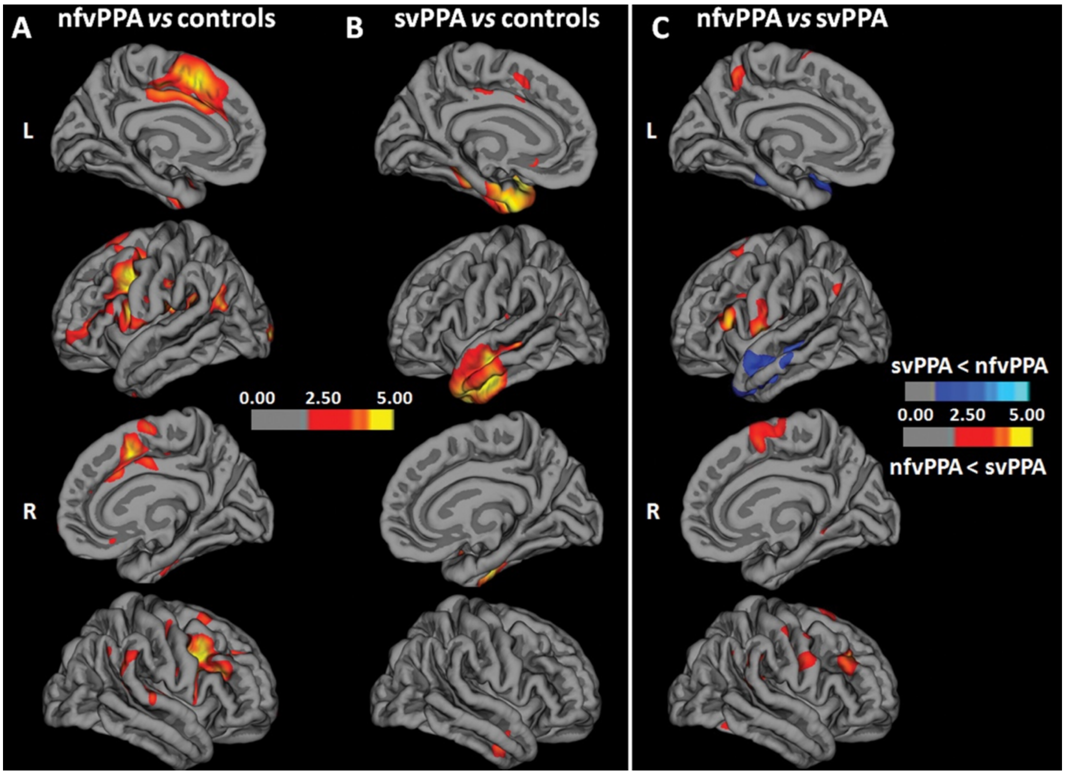


Figure 2-25 Three-dimensional reconstructed MRI maps of PPA variants. From [92]

2.5.1 Tract-based spatial statistics

Tract-Based Spatial Statistics (TBSS) is an advanced voxel-wise analysis method developed to improve the sensitivity and interpretability of diffusion tensor imaging (DTI) studies. [93], [94] TBSS introduced the concept of projecting diffusion data onto a WM skeleton, representing the core of major fiber tracts that are common across subjects. [94] The pipeline operates on FA maps, since FA is independent of local fiber orientation and provides a robust and comparable metric across individuals, making it suitable for group-level analyses. [93] It consists of several stages, designed to ensure accurate alignment and group-level comparability of diffusion data.

First, a preprocessing phase is performed to correct for motion and eddy current distortions, typically through affine registration of all diffusion-weighted images to the non–diffusion-weighted image. From these prealigned data, the diffusion tensor is estimated voxel-wise, allowing the computation of eigenvalues and eigenvectors from which fractional anisotropy (FA) and other diffusion metrics are derived. [93]

This is followed by a nonlinear registration step, in which each subject's FA map is aligned to a common reference space. The target is chosen as the "most representative" subject (i.e., the one requiring minimal warping for all other subjects to align). All FA maps are then transformed into a standard space. Importantly, perfect alignment is not required, as subsequent projection onto the white matter skeleton compensates for residual misregistrations. [93]

Next, the mean FA skeleton is generated by averaging all registered FA maps and extracting the central trajectories of the white matter tracts common to all subjects. This skeletonization process involves identifying the local perpendicular direction to each tract and applying non-maximum suppression to retain only the tract centers, where FA values are highest. A threshold is then applied to exclude regions of gray matter, cerebrospinal fluid, or poorly aligned tracts [93]

Finally, each subject's FA data are projected onto the mean FA skeleton by searching, at each skeleton voxel, along the perpendicular direction for the maximum FA value and assigning this value to the skeleton voxel. This step effectively aligns homologous tracts across subjects, reducing the impact of partial volume effects. The resulting skeletonized FA data can then be used for voxel-wise statistical analyses across subjects, commonly employing the general linear model (GLM) to test for group differences or correlations with behavioral or clinical measures. [93], [94]

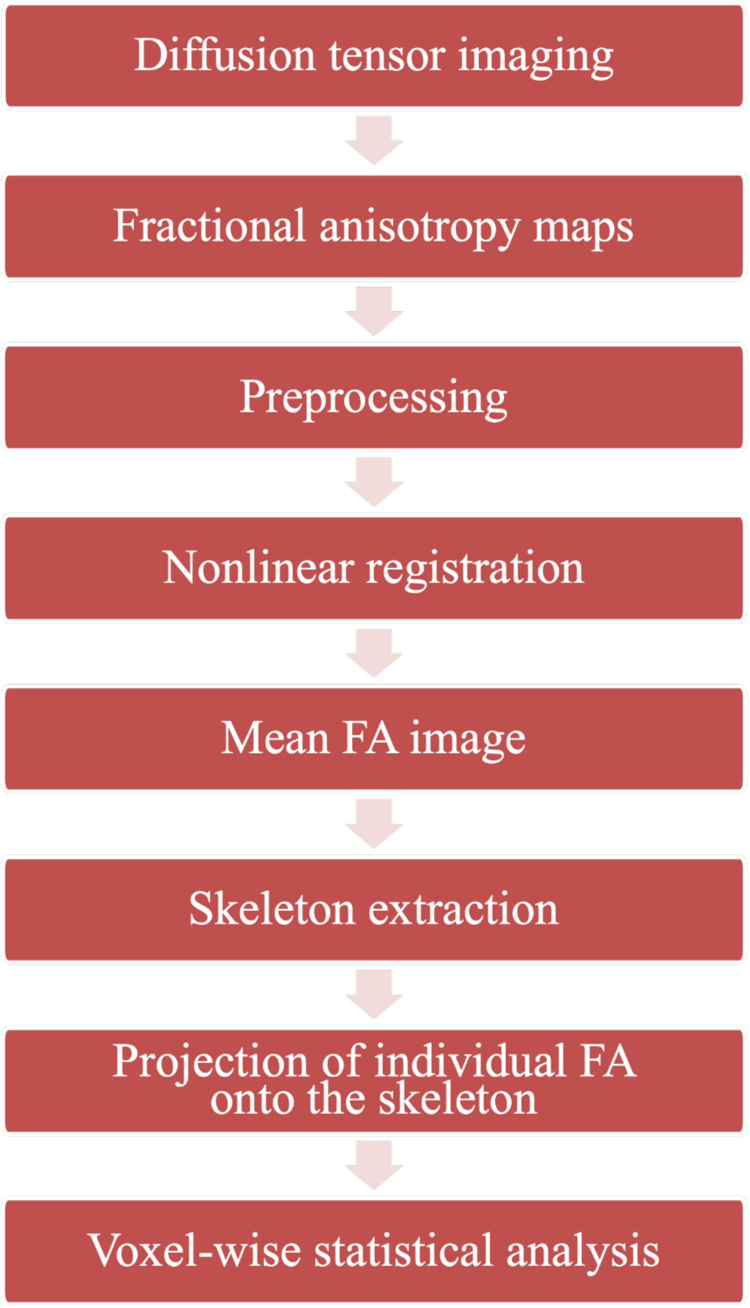

Diffusion tensor imaging
Fractional anisotropy maps
Preprocessing
Nonlinear registration
Mean FA image
Skeleton extraction
Projection of individual FA onto the skeleton
Voxel-wise statistical analysis

Figure 2-26 TBSS workflow.

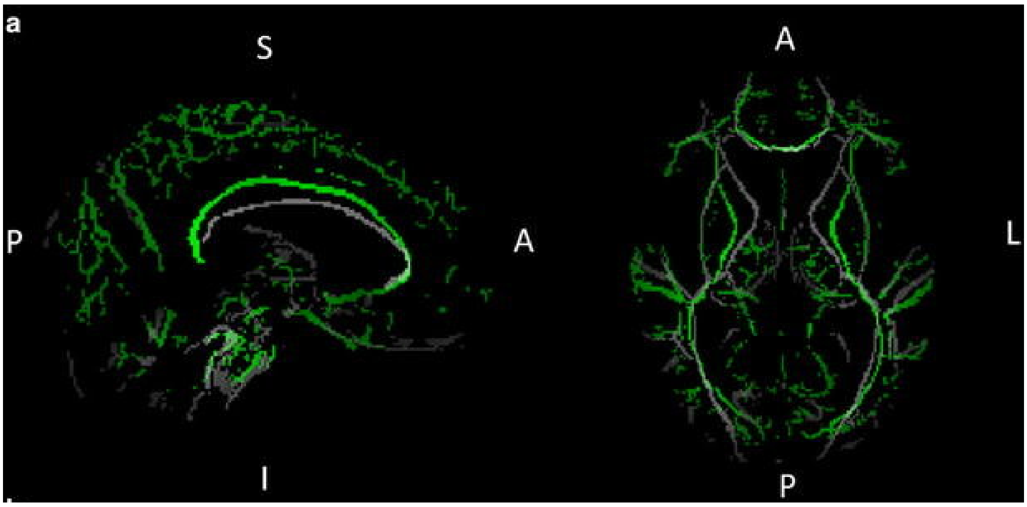


Figure 2-27 Comparison of FMRIB58 FA standard skeleton (grey) with the mean FA skeleton derived from the study [95] (green).

2.5.2 Application of TBSS in clinical research

TBSS has become a widely used method for assessing white matter integrity across groups, particularly in neurodegenerative and neurocognitive disorders. The following examples illustrate typical methodological parameters and applications adopted in clinical diffusion studies.

In one study, individual FA maps were projected onto a common white matter skeleton derived from the FMRIB58_FA template and thresholded at FA > 0.2 to exclude peripheral and non–white matter regions. Voxel-wise group comparisons were then performed using the *randomise* tool in FSL, typically employing 5000 permutations and significance thresholds of $p < 0.05$ or $p < 0.01$. Multiple-comparison correction was applied using the threshold-free cluster enhancement (TFCE) method, which increases sensitivity without the need for arbitrary cluster-size thresholds. Results were visualized with *FSLEyes*, overlaying statistical maps on the mean FA skeleton, while the FSL white matter atlas was used to identify tracts showing significant group differences. Additional processing with PANDA enabled the conversion of significant TBSS clusters into tract-based tables for anatomical interpretation. This study on AD revealed widespread reductions in FA within major association and projection fibers, particularly in the anterior corona radiata, as well as in frontoparietal and callosal tracts. These alterations reflect axonal disorganization and myelin loss consistent with AD-related white matter degeneration. [96]

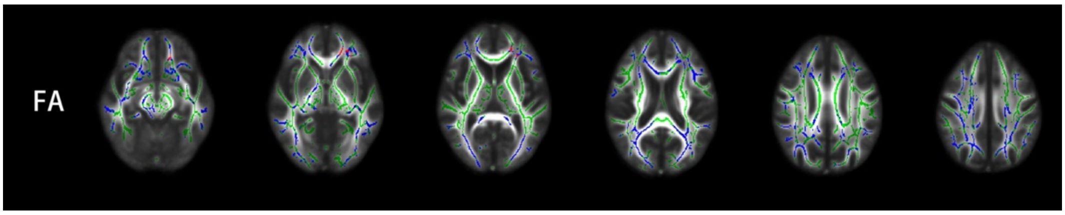


Figure 2-28 TBSS images of FA in horizontal slices of brain at different Z-axis values. [96]

TBSS has also proven valuable for distinguishing between variants of PPA. In svPPA, FA reductions were predominantly left-lateralized, involving ventral pathways such as the uncinate fasciculus (UF), inferior fronto-occipital fasciculus (IFOF), and inferior longitudinal fasciculus (ILF), extending into the temporal pole and ventral frontal areas. Dorsal tracts such as the forceps minor, genu of the corpus callosum, and superior longitudinal fasciculus (SLF) also showed reduced FA, whereas right-hemisphere changes were limited.

Conversely, in nfvPPA, microstructural damage was more focal, with FA reductions confined to the left IFOF and UF, within the external capsule and ventrolateral prefrontal white matter. Additional involvement extended into the superior temporal and prefrontal regions, paralleling the asymmetrical cortical atrophy typically observed in these patients on voxel-based morphometry. [97]

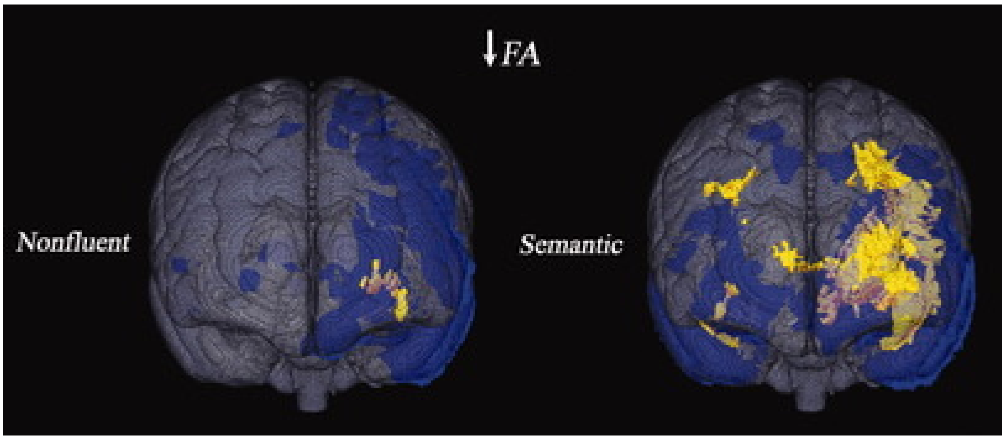


Figure 2-29 Nonfluent and semantic FA findings. Adapted from [97]

3 Artificial Intelligence

The idea of artificial intelligence (AI) can be traced back to Alan Turing's work in the 1930s when he defined the principles of computation through the concept of a universal machine capable of performing any operation that could be described. Turing and Emil Post demonstrated that mathematical problems could be solved by abstract machines manipulating symbols according to a finite set of rules. [98]
However, one of the first publications related to AI was the 1943 work of McCulloch and Pitts; they developed a computational model inspired by the functioning of biological neurons. Their model produced binary outputs and relied on fixed threshold and weight parameters, providing the prototype for later artificial neural networks. [99][Figure 3-1]

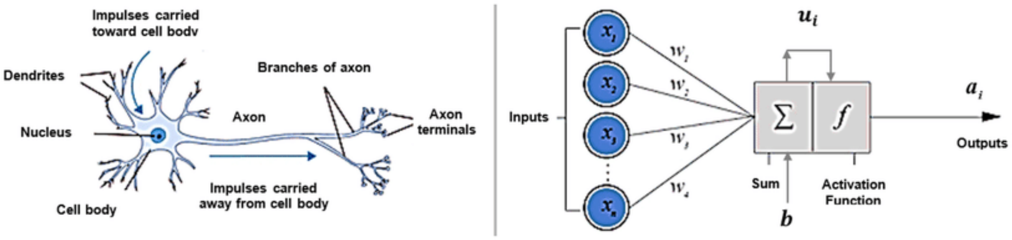


Figure 3-1 Biological neuron and McCulloch and Pitts model based on it. From [100]

Building on these foundations, Turing focused in the following years on the question "Can machines think?", later proposing the Turing test to empirically assess machine intelligence. [101]
During the 1950s, creating a thinking machine became a major scientific challenge, and John McCarthy introduced the term Artificial Intelligence, referring to a machine capable of thinking, giving birth to AI as a scientific discipline. [102] This was considered possible provided that every aspect of learning could be precisely described, so that a machine could be made to simulate it. [103], [104]
In 1958, Frank Rosenblatt proposed the perceptron, a probabilistic model made of layers of connected units that learn from data. The connections are weighted and these weights change during the algorithm's learning process. [105]

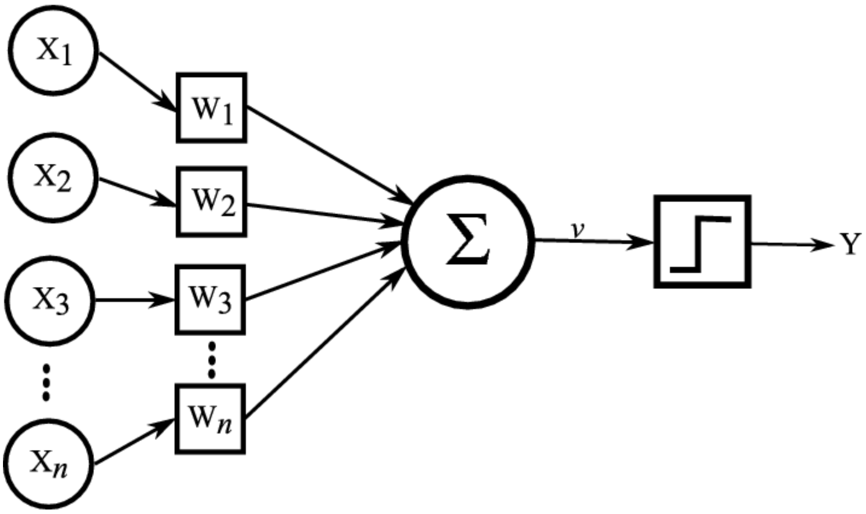


Figure 3-2 The perceptron by Frank Rosenblatt. From [106]

A few years later, in the 1970s, the AI field entered a so called ‘winter’: interest in the topic decreased as disillusionment spread among researchers. This was caused both by hardware limitations and by the complexity of trying to replicate the human brain function. [107]

However, some progress was made; systems based on expert knowledge were developed, though only in specific domains.[108] In 1974, Paul Werbos introduced the backpropagation algorithm in his doctoral thesis. This algorithm computes the contribution of each weight to the output error and then propagates the error backward through the layers, updating the weights to improve the network. [109]

Despite this, in the 1980s a second AI winter occurred. Attention shifted to evidence-based approaches, in which model validation relied on empirical data and experimental results. [110]

With further improvements in hardware, AI rose again, driven by the rapid increase in available data. Modern AI aims to simulate and augment human intelligence through computational algorithms capable of efficiently analyzing complex data. By applying such algorithms, these systems can classify, predict and extract meaningful conclusions. [111]

Within this large field, machine learning (ML) and deep learning (DL) have found their places. ML focuses on algorithms that can learn from data to perform complex tasks or generate predictive models. A critical stage in ML workflows involves the manual design or selection of informative variables from data for model construction, called feature engineering or extraction. In contrast DL, that is a subfield of ML, has revolutionized the discipline by enabling algorithms to automatically learn feature representations directly from the data. DL relies on multilayer neural networks conceptually derived from neurobiological models of cognition; this finally closes the historical gap between symbolic reasoning and neural-inspired learning models. [112]

3.1 Introduction to Machine Learning

In 1959, Arthur Samuel, a researcher at the International Business Machines Corporation, coined the term "machine learning" after demonstrating that a computer could be programmed to learn how to play checkers. [113] ML refers to a group of computational methods that allow algorithms to identify patterns in data and make predictions without being explicitly programmed through fixed rules. In fact, its usual that concepts or objects cannot be easily defined by precise rules, but with ML they can instead be taught to a machine through examples. The computer transforms those examples into useful knowledge using the algorithms designed for machine learning. [114]

Unlike traditional approaches, which are often time-consuming, expensive, and prone to human bias, ML can efficiently manage large, complex, and high-dimensional datasets. By applying statistical and data-driven rules, ML algorithms automatically detect relationships between inputs and outputs and improve these patterns through an iterative learning process. Once trained on sufficient data, the model can generalize what it has learned, accurately predicting or classifying new and unseen data. Machine learning proves particularly effective in tasks such as classification, regression and clustering, which are a great tool for repeatable decision-making. However, achieving high predictive accuracy in complex applications requires large training datasets, since data needs increase exponentially with the number of model parameters. [115]

3.2 Learning Paradigms

ML approaches can be classified into three main categories: supervised, unsupervised and reinforcement learning.

- In supervised learning, algorithms are trained to obtain predictive models that, given certain inputs, produce outputs which can be discrete values if it is a classification task or continuous values if it is a regression task. The available data, consisting in examples and their associated labels, is usually divided into a training set and a validation set. The aim is to approximate a function (f) that maps an input data (x_j) to its corresponding label (y_j):

  $$y_j = f(x_j)$$

  Once this function is learned, it can be applied to unseen data, the test set, which includes examples coming from the same underlying distribution as the training data. This step is usually used to evaluate the model’s ability to generalize. [116][Figure 3-3]

- In unsupervised learning, algorithms aim to discover hidden patterns or structures in the data without using predefined labels. A common approach within this category is clustering, which automatically groups similar data points together based on shared characteristics. The main idea is to form clusters in which elements are alike under specific criteria, while remaining distinct from those in other groups. Similarity between data points is usually measured through a distance metric in a multidimensional space. Other similarity measures exist for variables of different types, such as binary or nominal. In some cases, certain features can be considered more important than others, and this can be accounted for by assigning them greater weight in the similarity calculation. [116][Figure 3-3]
- Reinforcement learning is a learning approach based on interaction with the environment. Through this interaction, the system observes the outcomes of its actions and uses that feedback to achieve specific goals. The objective is to learn a policy, which is a mapping from states to actions, that maximizes the total expected reward over time. Rewards may be received at the end of a task or during intermediate steps. The agent learns by exploring different actions and observing their consequences, improving its behavior through a process of trial and error. Unlike other ML methods, learning and evaluation occur simultaneously, as the agent continually adjusts its decisions based on experience. [116] [Figure 3-4]

Beyond these three main paradigms, other approaches have been developed to address specific learning scenarios. One of the most studied is semi-supervised learning, which is particularly useful when large amounts of data are available but only a small portion is labeled. This situation is common in fields such as image and text analysis, where obtaining labeled examples can be costly and time-consuming because it requires expert human annotation. Semi-supervised methods make use of both labeled and unlabeled data, combining them to build models that are often more accurate than those trained solely on the available labeled examples. [117]

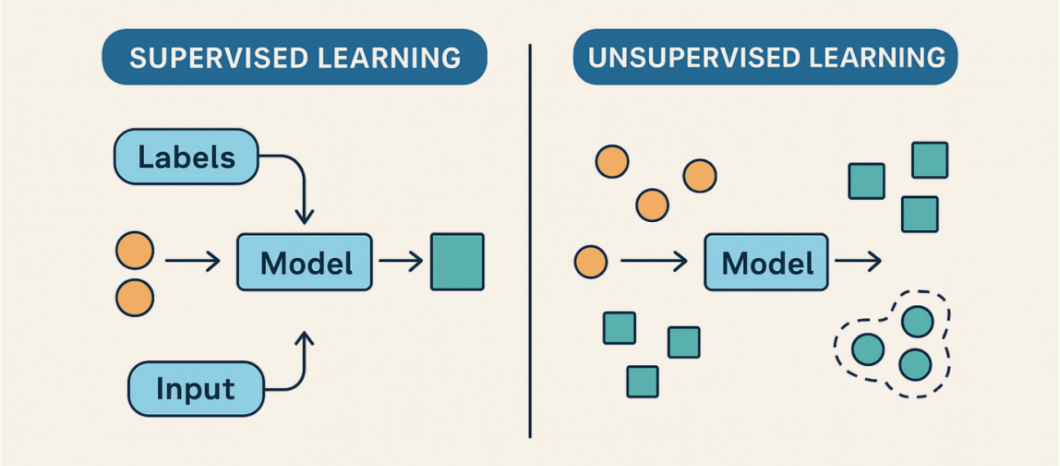


Figure 3-3 Supervised and unsupervised learning as machine learning paradigms. Adapted from [118]

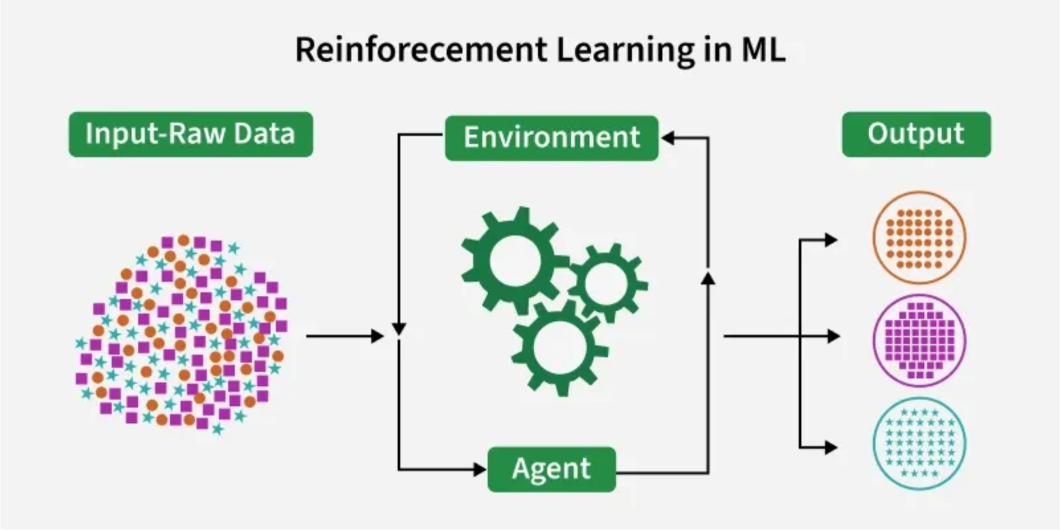


Figure 3-4 Reinforcement learning. From [119]

3.3 Standard Supervised Learning Workflow

The standard supervised learning pipeline consists of several steps:

1. Data collection and preprocessing – ensuring high data quality, consistency and noise reduction to train reliable models
2. Feature extraction – transforming raw input signals (e.g., audio, text, images) into numerical descriptors that captures relevant information for the analysis
3. Feature selection – reducing dimensionality of the input and retaining only the most informative variables to ensure and improve model performance and generalization

4. Model training and parameter tuning – training the model on labeled data to learn decision boundaries or predictive patterns
5. Evaluation and validation – assessing generalization using techniques such as cross-validation and external test data to ensure the robustness of results and prevent overfitting

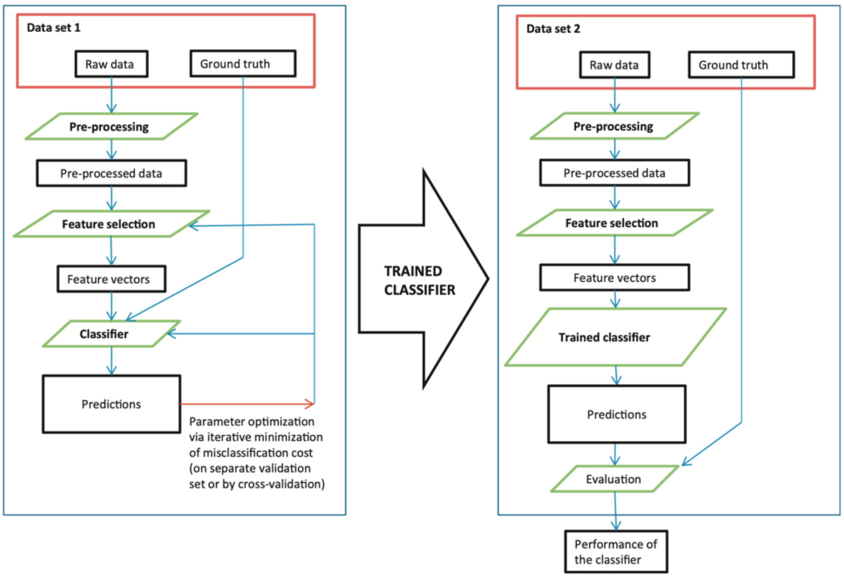


Figure 3-5 Workflow of supervised machine learning studies [120]

3.3.1 Data collection and preprocessing

In ML, the quality and consistency of data are critical factors which could directly influence model performance. Raw data often contain noise, inconsistencies or irrelevant information due to differences in collection procedures or acquisition conditions. For this reason, preprocessing represents a fundamental step in the standard ML workflow. It usually includes operations such as normalization, artifact removal, harmonization across sources and converting measurements into standardized formats. [121]

Once the dataset has been collected and preprocessed, it is usually divided into three subsets: the training set, the validation set and the test set. The training set is used to fit the model, the validation set to tune the hyperparameters of the model in order to prevent overfitting, and the test set to assess the model's ability to generalize to unseen data. [Figure 3-6]

At the end of the training phase, once the model has been optimized and the hyperparameters fixed, it is often beneficial to retrain the model on both the training and validation sets. This allows the model to take advantage of all available data before being tested on the new examples in the test set. [122], [123] [124]

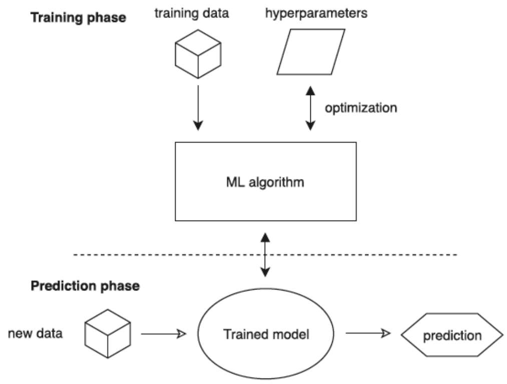


Figure 3-6 Model training phase using training and validation sets, prediction phase using the test set with the new data. From [124]

During the training phase, one of the main challenges in supervised learning is finding the right balance between fitting the data and maintaining generalization. When a model learns the training data too precisely, including its noise and random fluctuations, it may fail to perform well on the other examples, this is a phenomenon known as overfitting. In this case, the model essentially memorizes the training set instead of learning the underlying patterns that describe the data. Conversely, if the model is stopped too early or lacks sufficient complexity, it may fail to capture the essential structure of the data, leading to underfitting. Typically, during training, the model's performance is monitored on both the training and validation sets: while the training error keeps decreasing, an increase in validation error indicates the onset of overfitting. [125]

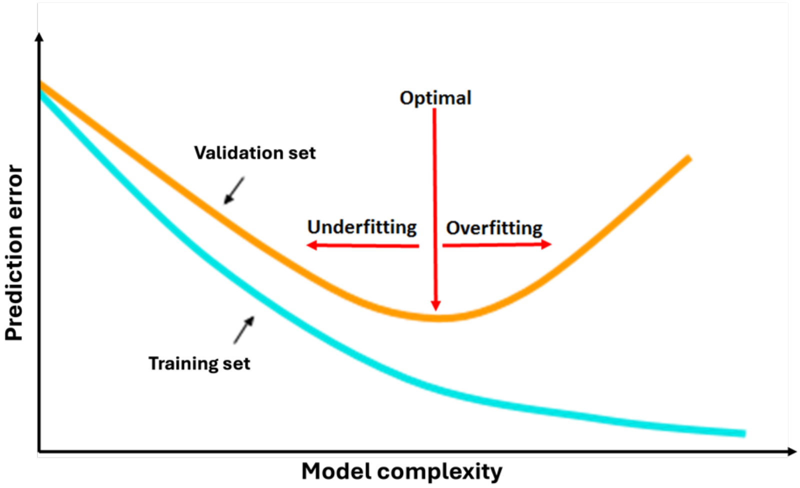


Figure 3-7 Overfitting and underfitting areas in the model's training phase. Adapted from [126]

3.3.2 Feature extraction

Feature extraction refers to the process of transforming raw signals into a simplified set of quantitative descriptors that capture informative characteristics of the signals while reducing noise and dimensionality. [127] Extracted features may describe physical, temporal or statistical properties of the data, depending on the modality and task.

This process may involve dimensionality reduction techniques or the creation of new features derived from existing ones. The resulting features often reveal hidden relationships or patterns that are not directly visible in the raw data. It also facilitates data visualization, since reducing the number of dimensions allows patterns to be explored in two or three dimensions. [128]
Feature extraction methods can be divided into two main categories:

- manual feature extraction
- automated feature extraction

Handcrafted feature extraction relies on domain expertise and predefined algorithms to identify relevant characteristics, offering high interpretability. On the other hand, automated feature extraction uses data-driven models to automatically learn useful representations from large datasets, capturing complex patterns when enough data are available.

Regardless of the method, ensuring consistency and comparability of features across samples, through normalization and standardization procedures, is essential for reliable model performance. [128]

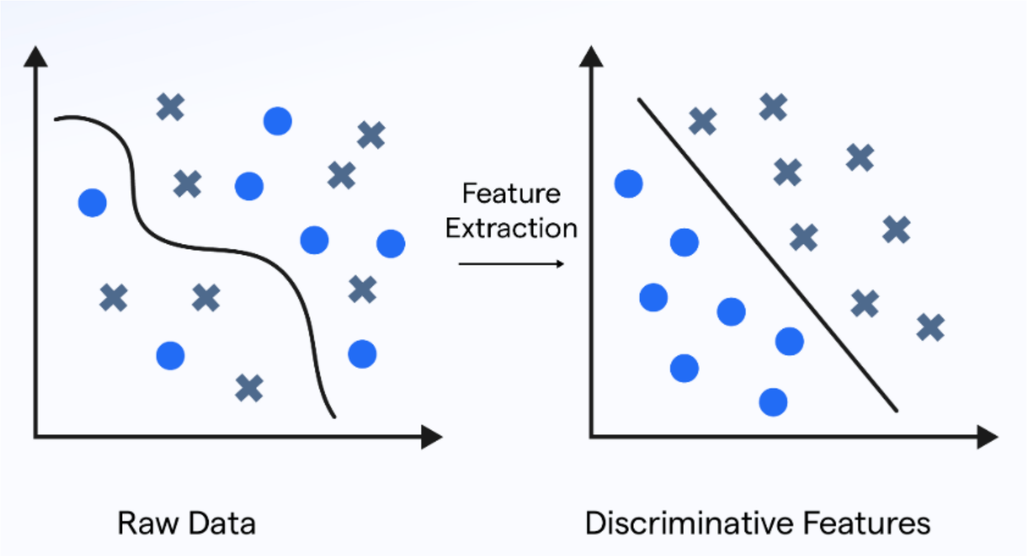


Figure 3-8 Feature extraction from raw data. Adapted from [129]

3.3.3 Feature selection

Feature selection refers to the process of identifying the most relevant variables for a predictive task while removing redundant or irrelevant ones. This step reduces the dimensionality of the data while preserving the information that is most useful for the model. By focusing on informative features, it helps minimize noise and overfitting, resulting in models that are more efficient and easier to interpret. [130]

Reducing the number of features also brings practical advantages: it lowers computational and memory demands, speeds up model training and helps models generalize better. [128] In fact, using too many features (even when all of them contain information about the target variable), can actually degrade prediction performance due to overfitting. Feature selection helps reduce this risk. [131]

Feature selection algorithms aim to identify the subset of predictors that best represent the data, sometimes under specific constraints such as the number of features or the inclusion or exclusion of particular variables. These methods are generally divided into three main categories: filter methods, wrapper methods and embedded methods. [132]

- Filter type feature selection algorithms assess each feature independently of the classifier, using statistical criteria such as correlation, mutual information, or chi-square tests. In this approach, important features are selected and then a model is trained only using the selected features. These methods are computationally efficient but do not account for interactions between features. A widely adopted approach is the minimum redundancy maximum relevance (mRMR) method which identifies features that maximize relevance to the target variable while minimizing inter-feature redundancy. [132][133]
- Wrapper type feature selection algorithms evaluate subsets of features based on model performance, using strategies such as forward or backward selection, recursive feature elimination or exhaustive search. The training phase is repeated and the model improved until the stopping criteria are satisfied. These methods are model-dependent and can yield high accuracy but are computationally expensive and prone to overfitting. [132][133]
- Embedded type feature selection algorithms perform feature selection as part of the model learning process itself – for example, L1 regularization in linear models or feature importance in tree-based algorithms. This approach allows the importance of the features to be derived directly from the trained model. Embedded methods offer a good compromise between efficiency and accuracy, even though they are tied to specific model families. [133]

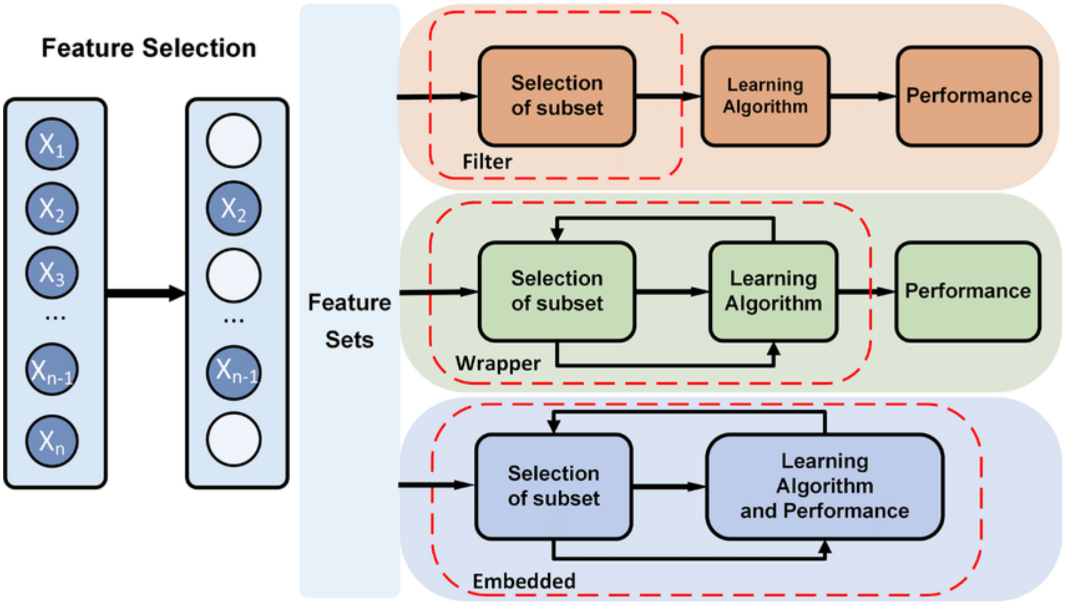


Figure 3-9 Filter, wrapper and embedded feature selection. From [134]

Hybrid or ensemble strategies can also be employed, for instance combining an initial filter step to remove obviously irrelevant features with a wrapper or embedded method for fine-tuning selection.

3.4 Machine Learning Classifiers

Machine learning classifiers are algorithms that automatically assign items to a discrete group or class, based on a specific set of features. In this section, the most common algorithms for supervised machine learning are listed. [116] [135]

3.4.1 Linear Regression

Linear regression is a supervised learning method used to model the relationship between a dependent variable and one or more independent variables through a linear function parameterized by coefficients β. The general form of the model is:

$$Y_i = f(X_i, \beta) + e_i$$

where Y_i is the target (dependent) variable, X_i are the predictors (independent variables), f is the linear function, β are the unknown parameters and e_i are the error terms. [135] This method is suitable when the relationship between the variables can be reasonably approximated by a linear function and the dependent variable is continuous.

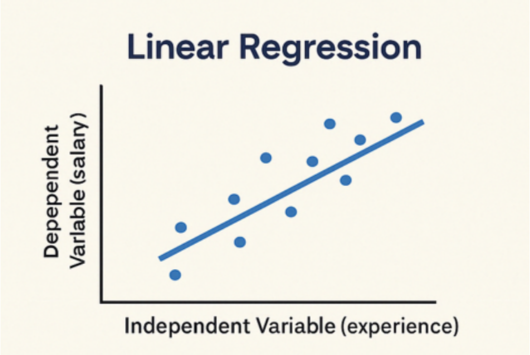


Figure 3-10 Linear regression: relationship between dependent and independent variables. Adapted from [136]

3.4.2 Logistic Regression

Logistic Regression (LR) is a widely used statistical model for both binary and multiclass classification tasks, valued for its simplicity and interpretability. It belongs to a family of statistical models called generalized linear models (GLMs), which relate a combination of predictor variables to the expected value of the outcome through a link function. LR specifically uses the logistic function to model the probability of class membership.

In binomial logistic regression, the model predicts the probability of an observation belonging to one of two possible classes. Multinomial logistic regression extends this approach to problems with more than two categories, modeling the probability of each class relative to a reference category.

The logistic function is defined as:

$$f(Z) = \frac{1}{1 + e^{-Z}}$$

where Z is the linear combination of the input features and their corresponding coefficients. [137] [138]

In the context of binary classification, LR relies on several key assumptions for valid results. These assumptions are properties of the data that the model requires to function correctly; violating them may lead to misleading outcomes. Specifically:

- Observations must be independent.
- There must be no perfect multicollinearity among the independent variables.
- Continuous predictors should have a linear relationship with the logit (a transformed version of the outcome). [137]

To improve generalization and reduce the risk of overfitting, regularization can be applied, although LR is applicable even without it. Regularization is particularly useful when the dataset has high dimensionality or a small number of instances, as it helps reduce noise from less informative features. Depending on the regularization method, the coefficients of less important features can be shrunk toward zero, removed entirely, or both. [138]

An example of a regularization method is the Least Absolute Shrinkage and Selection Operator (LASSO), which shrinks some coefficients while setting others to zero, effectively combining the benefits of subset selection and ridge regression. [138], [139]

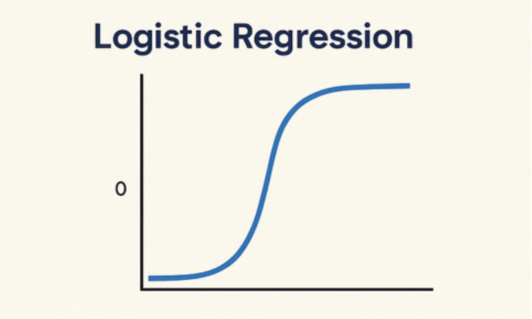


Figure 3-11 Sigmoid curve used in logistic regression to model the probability of class membership. Adapted from [136]

3.4.3 Decision Trees

Decision Trees (DTs) are non-parametric supervised ML models widely used for classification and regression tasks. [140] They were among the first generation of statistical algorithms implemented electronically during the adoption of digital circuitry in the later decades of the 20th century. Over time, DTs have evolved into highly cross-disciplinary, general-purpose, and computationally intensive methods, applied not only in prediction and classification, but also in machine learning, artificial intelligence, knowledge discovery, and data mining.

The main characteristic of DTs is the recursive partitioning of the target variable based on the values of input features or predictors. This process creates a hierarchical, tree-like structure of nodes and leaves, where each internal node represents a decision based on a feature, each branch corresponds to a possible outcome of that decision, and each leaf node assigns a class label. [Figure 3-12]

At each node, a splitting criterion is used to determine how to partition the data. Common criteria include the Gini index and Information Gain, which aim to create child nodes that are as homogeneous as possible with respect to the target variable. [140]

As a result of this recursive subsetting, target values within each node or leaf are progressively more homogeneous (intra-node similarity), while between nodes they become increasingly dissimilar (inter-node dissimilarity). This property allows DTs to model complex interactions between features while remaining interpretable, making them a versatile tool in both traditional statistical applications and modern computational approaches. [141]

However, despite their high interpretability, single trees tend to overfit the training data, especially when grown to full depth. To overcome this limitation, ensemble methods such as Random Forests (RFs) are often preferred. In these approaches, multiple trees are combined, and their decisions are aggregated through mechanisms like majority voting or weighting, resulting in more stable and accurate predictions. [140]

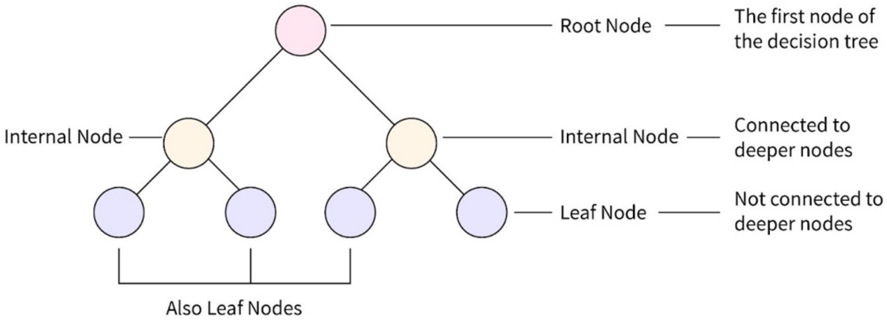

Root Node
The first node of the decision tree
Internal Node
Internal Node
Connected to deeper nodes
Leaf Node
Not connected to deeper nodes
Also Leaf Nodes

Figure 3-12 Basic structure of a decision tree: root, internal and leaf nodes. From [142]

3.4.4 Random Forest

RFs are ensemble learning methods that combine the predictions of multiple DTs to improve stability and accuracy. Each tree is trained on a bootstrap sample of the original dataset, while at every split, a random subset of features is considered. This dual source of randomness, in both samples and features, produces decorrelated trees, whose predictions are aggregated to form the final classification, as illustrated in Figure 3-13. [143]

Two main hyperparameters control the model’s behavior:

- the number of trees, which determines the ensemble’s size and affects its stability;
- the number of features evaluated at each split, which regulates tree diversity and decorrelation.

By averaging across many weakly correlated trees, RFs effectively reduce variance and mitigate the overfitting typical of single decision trees. Moreover, RFs provide estimates of feature importance, offering insight into which variables contribute most to the model's predictions, an aspect particularly valuable in clinical and biomedical research, where interpretability supports hypothesis generation and validation.

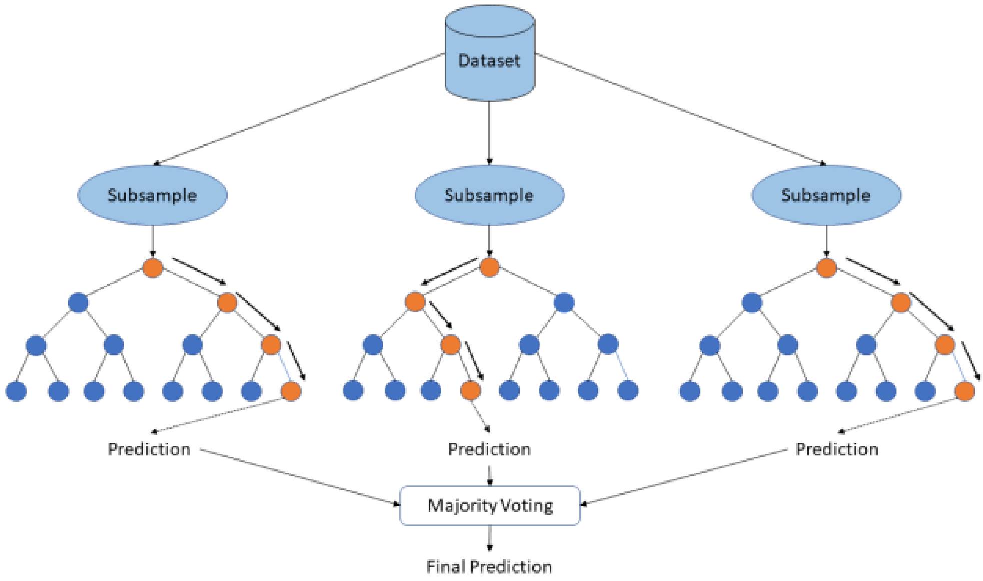


Figure 3-13 Structure of a Random Forest model, combining multiple decision trees trained on data subsamples and aggregating their predictions through majority voting. From [144]

3.4.5 Support Vector Machines

Support Vector Machines (SVMs) are supervised learning algorithms that find the hyperplane that best separates data points from different classes by maximizing the margin between them. In two-dimensional space this boundary is a line, while in three-dimensional spaces it becomes a plane. Only a subset of samples, called support vectors, defines the position of this boundary; all other points have no direct influence on it. [145]

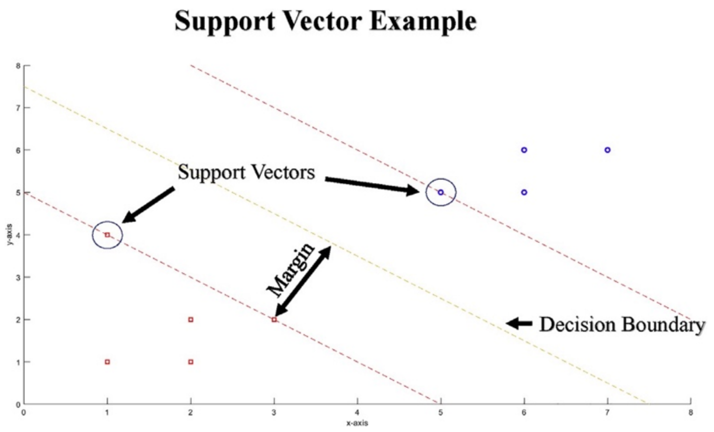


Figure 3-14 Representation of a SVM, showing the decision boundary, support vectors and margins that define the optimal separating hyperplane. From [112]

When classes are linearly separable, the SVM seeks the boundary that maximizes the distance between the closest samples of opposite classes, as illustrated in Figure 3-15. In practice, most datasets are not perfectly separable, so soft-margin SVMs introduce a regularization parameter C that balances the trade-off between maximizing the margin and allowing some misclassifications. Larger values of C emphasize correct classification, while smaller values prioritize a wider margin and better generalization. [146]

By applying kernel functions (e.g., the radial basis function, RBF), SVMs can implicitly project the data into a higher-dimensional feature space and thus model non-linear decision boundaries.

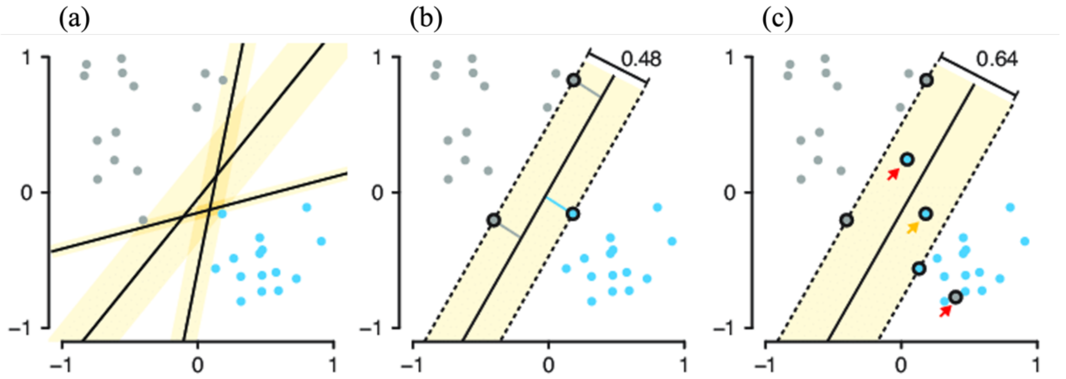


Figure 3-15 Illustration of the SVM optimization process: (a) linearly separable classes, (b) maximization of the margin between classes and (c) margin violations and misclassifications in non-separable classes. Adapted from [146]

3.4.6 k-Nearest Neighbors

k-Nearest Neighbors (kNN) is a non-parametric, instance-based classification algorithm that assigns a label to a new sample according to the majority class among its k nearest training samples, based on a chosen distance metric, such as the Euclidean distance. [146]

The hyperparameter k controls the balance between model complexity and generalization. It is tipically chosen as an odd number to avoid ties. Smaller values of k produce finely textured decision boundaries, that can closely follow the training data, but are sensitive to noise and outliers, leading to high variance. In contrast, larger values of k generates more rigid boundaries, effectively averaging the influence of distant neighbors, which reduces variance but increases bias. [Figure 3-16]

While kNN can capture highly non-linear class boundaries, it is computationally demanding at prediction time, since classification requires computing distances from the new sample to all training points. [146]

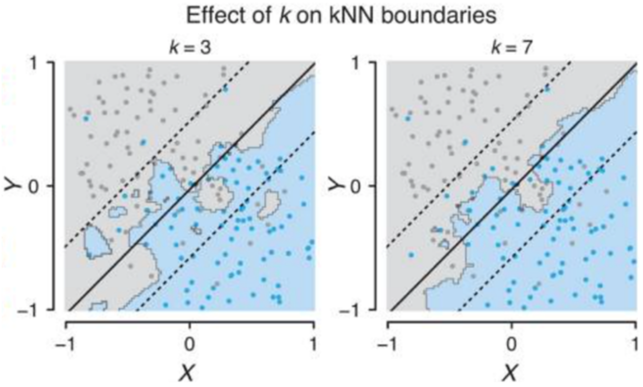


Figure 3-16 Effect of the parameter k on kNN decision boundaries. Adapted from [146]

3.4.7 Naïve Bayes

Naïve Bayers classifiers are probabilistic models based on Bayes' Theorem, which is used to estimate the probability of a class y given a set of features $x_1, x_2, \ldots, x_n$:

$$P(y|x_1, x_2, \ldots, x_n) = \frac{P(x_1, x_2, \ldots, x_n|y)P(y)}{P(x_1, x_2, \ldots, x_n)}$$

This represents the probability that an observation belongs to class y, after considering the evidence provided by its features. [135]

In practice, the denominator is constant across classes, so classification is based on the proportional relationship:

$$P(y|x_1,\ x_2, \dots , x_n) \propto P(y)P(x_1,\ x_2, \dots , x_n|y)$$

The model makes the assumption that all features are conditionally independent given the class. [135] Under this assumption, the joint likelihood can be decomposed as:

$$P(x_1,\ x_2, \dots , x_n|y) = P(y)\prod_{i=1}^{n} P(x_i|y)$$

Which lead to the final form:

$$P(y|x_1,\ x_2, \dots , x_n) \propto P(y)\prod_{i=1}^{n} P(x_i|y)$$

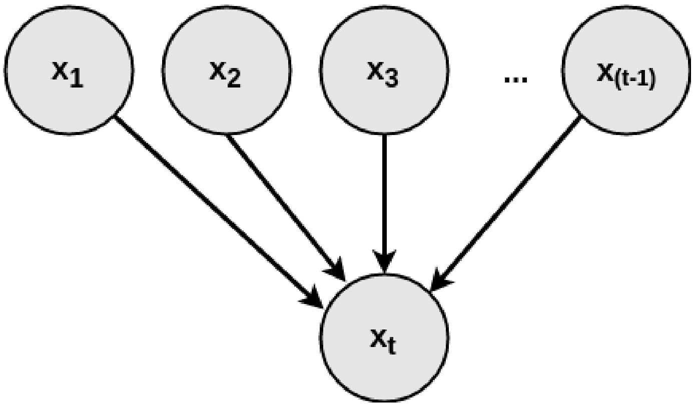


Figure 3-17 Graphical representation of the Naive Bayes classifier, based on the assumption that features are conditionally independent given the class variable (x_t). Adapted from [147]

3.5 Model evaluation and optimization strategies

Machine learning algorithms improve their performance through iterative training procedures, where model parameters are adjusted to minimize the difference between predicted and actual values. This process is guided by a learning method, which defines how the model updates its internal parameters during training.

Among the most common learning algorithms are Stochastic Gradient Descent (SGD) and Evolutionary Algorithms (EA). SGD iteratively updates model parameters by computing the gradient of the loss function with respect to each parameter and adjusting them in the direction that minimizes the loss. This method is computationally efficient and widely used in deep learning. This process is illustrated in Figure 3-18 where the algorithm follows the slope of the cost function to move higher values toward a global minimum. In contrast, EA are population-based optimization methods inspired by natural selection. They explore the parameter space through mutation and recombination operations, selecting the best-performing solutions over multiple generations. [148] Figure 3-19 represents the workflow of a Genetic Algorithm, a subclass of EA, showing how the population evolves through successive steps of initialization, evaluation, selection, crossover and mutation until convergence. [149]

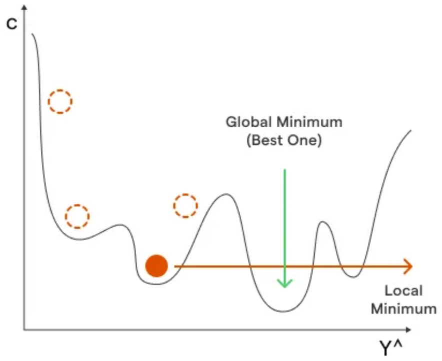


Figure 3-18 Illustration of the SGD optimization process. Adapted from [150]

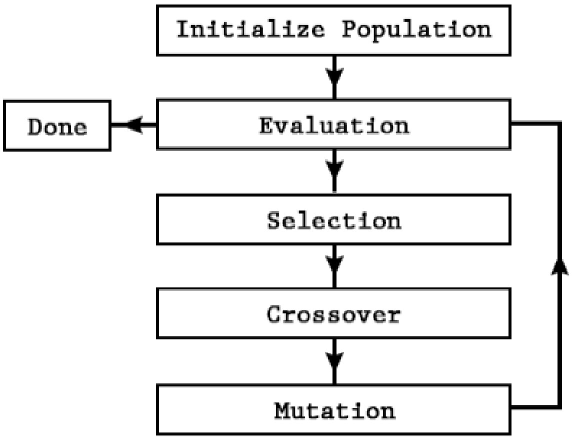


Figure 3-19 Flowchart of a Genetic Algorithm, representing the main evolutionary steps of the optimization process. Adapted from [149]

Beyond learning the model parameters, many machine learning methods require the tuning of hyperparameters, which control aspects such as model complexity, regularization strength, or learning rate. Proper hyperparameter optimization is crucial to achieve good generalization and avoid overfitting or underfitting. To objectively assess how well a learning algorithm generalizes beyond the data it was trained on, cross-validation strategies are commonly employed. [122]

These methods partition the available data into multiple training and validation subsets, allowing the model to be trained and tested across different folds of the dataset. A common approach is k-fold cross-validation, where the data are divided into k equally sized folds. The model is trained k times, each time using $k - 1$ folds for training and the remaining one for validation. The resulting performances scores are averaged, providing a more stable and reliable estimate of model performance while enabling hyperparameter tuning. [151]

In clinical or subject-based datasets, where multiple samples may originate from the same participant, a leave-one-subject-out (LOSO) cross-validation scheme is often preferred. In this case, all data from one subject are excluded during training and used only for testing at each iteration. This design prevents information leakage across folds and provides a more realistic estimate of subject-level generalization. [152]

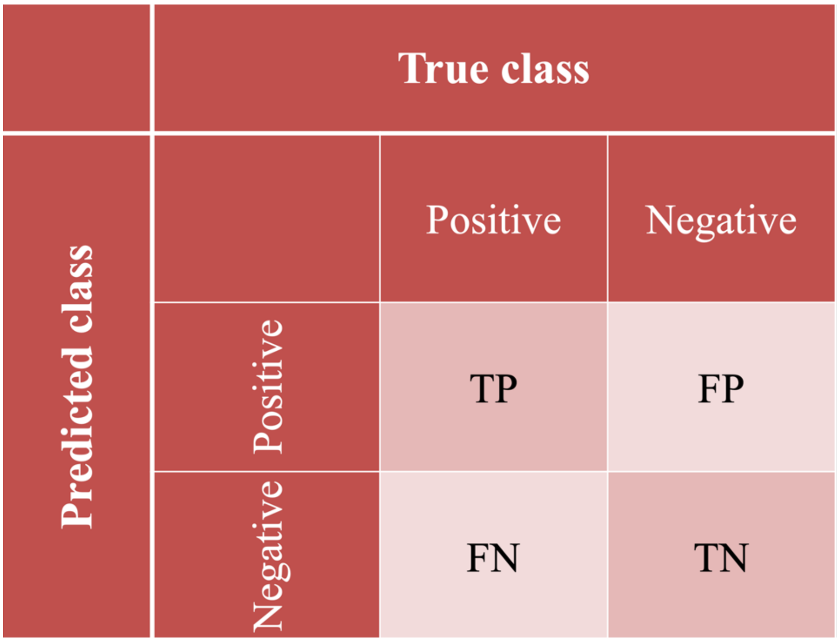
After selecting the optimal model configuration, the final classifier is retrained on the entire construction set and evaluated on the held-out test set, providing an unbiased estimate of its generalization ability.

3.6 Performance metrics

Model evaluation aims to quantify how well a trained model performs when predicting unseen data. In binary classification problems, each sample can belong to one of two classes, usually referred to as *positive* and *negative*. [153] [154]

The predicted and true labels can be combined in a confusion matrix, which summarizes the number of correct and incorrect predictions. [Figure 3-20] The confusion matrix is composed of four elements. [155]

- True positive (TP): the model correctly predicts the positive class
- False positive (FP): the model incorrectly predicts a positive outcome for a negative instance (it is also known as a Type I error)
- True negative (TN): the model correctly predicts the negative class
- False negative (FN): the model incorrectly predicts a negative outcome for a positive instance (Type II error)



		True class	
		Positive	Negative
Predicted class	Positive	TP	FP
	Negative	FN	TN

Figure 3-20 Confusion matrix illustrating the relationship between predicted and true labels. TP=true positive, TN=true negative, FP=false positive, FN=false negative.

The confusion matrix provides the basis for calculating several performance metrics, which describe different aspects of model behavior.

Accuracy

Accuracy represents the proportion of correctly classified instances out of all predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

This measure can be misleading in imbalanced datasets, where one class is much more frequent than the other. In such cases, a model could achieve a high accuracy simply by predicting the majority class and not by actually having a good performance. [155]

Precision

Precision measures how many of the instances predicted as positive are actually positive:

$$Precision = \frac{TP}{TP + FP}$$

A high precision indicates a low number of false positives and is particularly important in tasks where false alarms must be minimized, but gives no information about FN. [155]

Recall

Also called sensitivity or true positive rate (TPR), recall measures the proportion of actual positives that are correctly identified:

$$Recall = \frac{TP}{TP + FN}$$

It is an important measure when missing a positive case is more costly than missing a false positive (false positives are not accounted for in this measure). [155]

Specificity

Specificity, or true negative rate (TNR), measures the proportion of actual negatives correctly identified:

$$Specificity = \frac{TN}{TN + FP}$$

It is a useful measure when false positives must be minimized, complementing sensitivity in assessing overall discrimination ability. [155]

False Positive Rate

The False Positive Rate (FPR) represents the proportion of negative instances that are incorrectly classified as positive:

$$FPR = \frac{FP}{TN + FP}$$

It is the complement of the specificity $FPR = 1 - Specificity$.

A low FPR indicates that the model rarely misclassifies negative instances as positive. [155]

Balanced accuracy

Balanced accuracy is the average of sensitivity and specificity:

$$Balanced\ accuracy = \frac{Sensitivity + Specificity}{2}$$

This metric is particularly suitable for imbalanced datasets, as it provides an unbiased view of model performance across both classes. [154]

F1-score

The F1-score combines precision and recall into a single metric using the harmonic mean:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

It is usually used when both information about false positives and false negatives are important, offering a balance between the two errors. [155]

Receiver Operating Characteristic Curve

The Receiver Operating Characteristic (ROC) curve plots the TPR against the FPR at various classification thresholds, providing a visual representation of the model's discriminative power. [155]

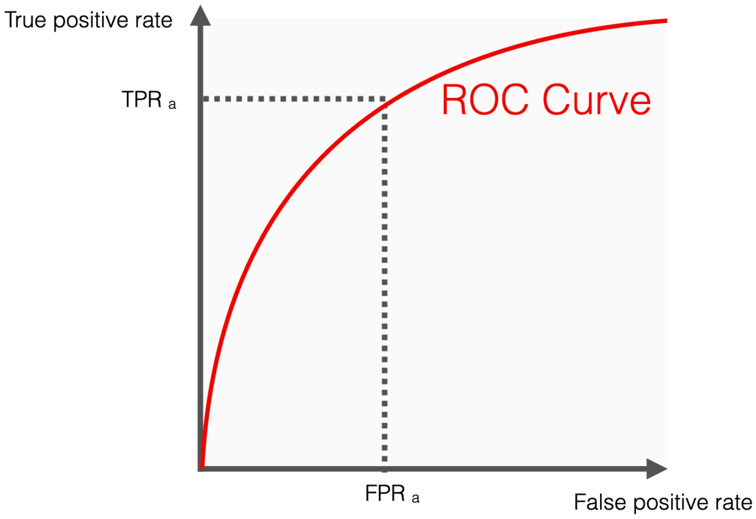


Figure 3-21 Receiver Operating Characteristic curve showing the relationship between the True Positive Rate and the False Positive Rate across different classification thresholds. From [156]

Area Under the ROC Curve

The Area Under the ROC Curve (AUC) quantifies the model's discriminative power as a single value between 0 and 1:

$$AUC = \int_0^1 TPR(FPR)dFPR$$

A value of 1 represents a perfect classifier, while a value of 0 indicates a model that inverts the classes. $AUC = 0.5$ corresponds to a random classifier. The ROC-AUC is threshold independent and can be used to compare models. [155]

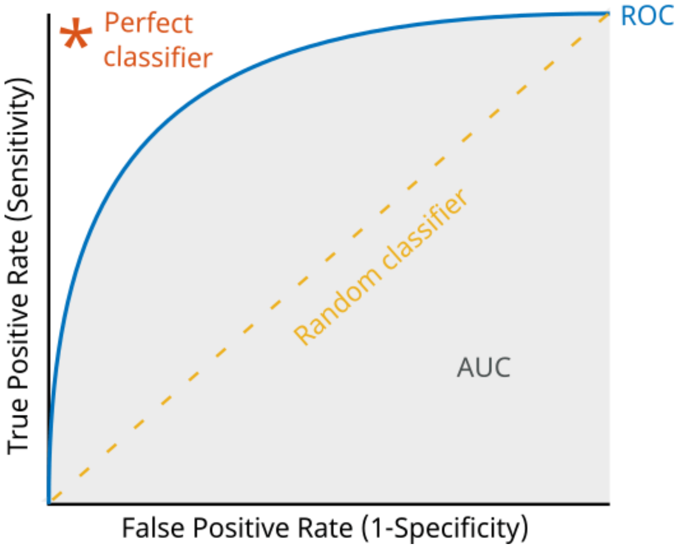


Figure 3-22 Illustration of the Area Under the Curve for a classifier. The plot highlights the behavior of a prefect and a random classifier. From [157]

3.7 Machine learning in biomedical applications

The rapid progress of computational technologies and the growing availability of biomedical data have created new opportunities for applying AI within biomedical research. ML methods are now used across multiple domains, such as risk assessment, early disease detection, diagnosis, outcome prediction and therapy monitoring, offering new ways to extract clinically meaningful patterns from data. [158]

These approaches have been particularly influential in medical analysis and diagnostics, where they assist clinicians in decision-making by providing quantitative and reproducible assessments of complex biological data. [158] [159] ML is being increasingly adopted in a wide range of biomedical applications, such as genomics, neuroimaging, digital pathology and speech analysis, supporting the development of data-driven models of disease and advancing precision medicine. To ensure reliable results, ML systems require high-quality annotated datasets, stable computational infrastructures and rigorous validation procedures. Because biological variations can be subtle, AI tools in healthcare are primarily designed as decision-support systems rather than replacements for human experts. Their purpose is to enhance diagnostic accuracy and efficiency. [159]

Traditional ML classifiers, such as LR, RF and SVM, have proven particularly useful in settings where data availability is limited but interpretability is crucial. These algorithms can handle complex, high-dimensional data such as neuroimaging signals, genomic profiles or acoustic features, while providing insight into which variables most strongly influence classification outcomes. [159]

3.7.1 Machine learning in Speech and language Analysis

Speech is a highly informative and multidimensional signal that conveys information across acoustic, prosodic, lexical, syntactic and semantic levels. These linguistic components can be selectively impaired depending on the neuropathological condition, making speech a valuable source of information for ML-based assessment. [1], [160]

However, speech signals are complex and often noisy, particularly in pathological populations, where recordings may be affected by articulatory deficits or background noise. This variability increases the difficulty of feature extraction and model interpretation, requiring robust analytical pipelines to ensure consistent and reliable results. In clinical speech processing, ML systems are often implemented as modular pipelines, in which the workflow is divided into distinct stages: preprocessing, automatic transcription, feature extraction and classification. Each step can be independently optimized, improving flexibility and interpretability. Features may be derived directly from the audio signal, capturing acoustic or prosodic properties such as pitch, intensity or rhythm, or extracted from transcriptions, focusing on lexical or syntactic information. This modular design allows the integration of linguistic knowledge and supports analysis even when datasets are small or heterogeneous. [161]

Research on ML in speech pathology has so far focused mainly on dysarthria, which represents the most common speech motor disorder. A considerable number of studies have analyzed dysarthric speech in Parkinson's disease, while fewer have addressed other conditions such as aphasia, apraxia, dysphonia or dysphagia. Overall, most datasets currently available are in English, which represents more than half of the total, followed by French, with other languages such as Spanish, Japanese, Italian and Korean less frequently represented. About 67% of published studies in this field have applied ML techniques to speech recognition and classification tasks. [162]

Among the algorithms used, SVM and Logistic Regression LR are the most common. In particular, SVM has demonstrated strong performance in detecting and classifying speech disorders due to its ability to handle nonlinear decision boundaries and high-dimensional feature spaces. This makes it well-suited for modeling complex acoustic characteristics found in disordered speech. In the case of dysarthria, SVM-based models effectively capture subtle variations in articulation and phonation, leading to improved accuracy and generalization across subjects. [162]

3.7.2 Machine learning in medical imaging

Medical imaging represents one of the most dynamic areas for the application of machine learning in healthcare. Imaging data contain detailed spatial and quantitative information on tissue structure and function, enabling ML algorithms to detect subtle and complex patterns that may not be visible by the human eye. These techniques have been successfully applied across various radiological tasks, particularly in lesion detection and disease diagnosis. [158]

Two main paradigms are commonly recognized in this context: computer-aided detection (CADe) and computer-aided diagnosis (CADx). CADe systems focus on localizing suspicious regions within medical images, serving as a second reader to support radiologists during screening procedures such as mammography. In contrast, CADx systems aim to characterize identified regions, whether marked by the radiologist or automatically detected, by estimating their likelihood of being pathological. In both cases, the final diagnostic decision remains the responsibility of the clinician. [163][Figure 3-23]

A particularly active research area within medical imaging is radiomics, which involves the extraction of a large number of quantitative descriptors from medical scans. These features can be related to shape, intensity and texture and capture tissue characteristics beyond visual inspection, acting as potential imaging biomarkers. Radiomics facilitates the identification of image-based phenotypes associated with pathological or molecular profiles and contributes to multi-omics integration, where imaging data are combined with other clinical information for more comprehensive disease modeling. [158]

For ML-based imaging systems to achieve clinical relevance, several prerequisites are necessary. Large, well-annotated datasets are essential for reliable model training and validation, while robust computational infrastructures ensure reproducibility and scalability. Furthermore, standardized preprocessing and feature extraction protocols are crucial to minimize variability across scanners and imaging centers. [158]

Computer-extracted radiomic features are then used as input to machine learning algorithms, which combine them into higher-level representations, such as tumor signatures, that may correspond to disease states or clinical outcomes. Over the years, a variety of ML methods have been applied in this field, including linear discriminant analysis, support vector machines, decision trees, random forests and neural networks, each contributing differently to pattern recognition and predictive modeling. More recently, convolutional neural network (CNNs) have emerged as a widely used model for image-based analysis, due to their ability to automatically learn hierarchical representations directly from imaging data, without the need for a feature extraction phase. [158]

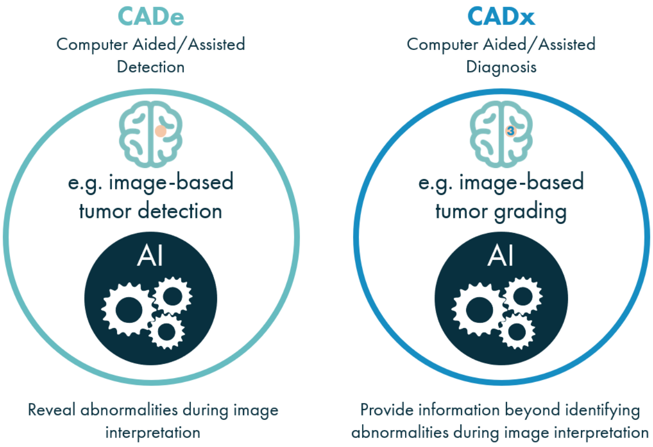

CADe
Computer Aided/Assisted Detection
e.g. image-based tumor detection
AI
Reveal abnormalities during image interpretation
CADx
Computer Aided/Assisted Diagnosis
3
e.g. image-based tumor grading
AI
Provide information beyond identifying abnormalities during image interpretation

Figure 3-23 Representation of Computer Aided Detection (CADe) and Computer Aided Diagnosis (CADx) systems for image-based tumor detection and grading. Adapted from [163]

4 Automatic Speech Recognition

Speech is the most natural and efficient way for humans to communicate. It expresses not only linguistic content but also paralinguistic cues such as emotion, intent and emphasis. Given its intuitiveness, it has also been regarded as an ideal channel for human–machine interaction. For this reason, developing systems capable of interpreting spoken language has long been a central goal in computer science. [164]
To analyze spoken communication, researchers have traditionally relied on manual transcription, which remains the gold standard for linguistic accuracy and detailed annotation. However, manual transcription is extremely time-consuming, requires trained personnel and is difficult to scale. These constraints make it impractical for large speech corpora or real-time applications. Automatic Speech Recognition (ASR) emerged as a solution to this problem: its objective is to convert spoken language into written text automatically, providing a faster and more cost-effective alternative that enables large-scale analysis and interactive speech-based technologies. [165]

An ASR system can be thought of as a machine that perceives an acoustic input, recognizes the spoken words and transforms them into a textual representation that may then serve as input for further processing or action. In practice, achieving this goal remains challenging. Human speech varies widely across speakers, languages, and contexts, for example due to differences in accent, dialect or speaking style, and such variability often prevent ASR models from performing uniformly well across conditions. Moreover, the scarcity of large annotated speech corpora for many languages further limits the generalization of these systems. [164]

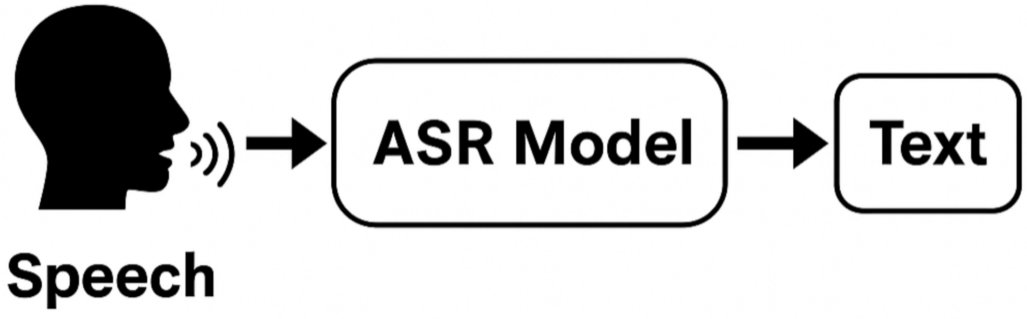


Figure 4-1 Automatic Speech Recognition process.

4.1 Historical development

Attempts to build machines that can recognize or produce speech date back to the 1950s. [166] The Audrey system developed at Bell Laboratories was among the first, capable of identifying spoken digits from a single user. [167] Shortly after, researchers at the MIT Lincoln Laboratory created a system able to discriminate a limited number of phonemes. [168]

In the 1970s, advances in dynamic programming and pattern recognition algorithms made it possible to recognize connected words. The introduction of Hidden Markov Models (HMMs) in the early 1980s represented a decisive step: HMMs provided a probabilistic approach to handle the temporal variability of speech, soon replacing earlier template-matching methods. Later, n-gram language models were developed to represent the statistical relationships between words and improve contextual coherence. [164]

By the 2000s, HMMs were combined with Artificial Neural Networks (ANNs), producing hybrid systems that integrated probabilistic temporal modeling with discriminative learning. The emergence of Recurrent Neural Networks (RNNs) and their variant, Long Short-Term Memory (LSTM) networks, further advanced the field by capturing long-term dependencies in the speech signal. These developments marked the transition from rule-based and statistical systems to modern deep learning architectures, which remain the foundation of state-of-the-art ASR. [164]

4.2 System architecture

Despite the variety of modeling approaches, most ASR systems can be conceptually divided into four functional modules: pre-processing, feature extraction, classification and language modeling. [164]

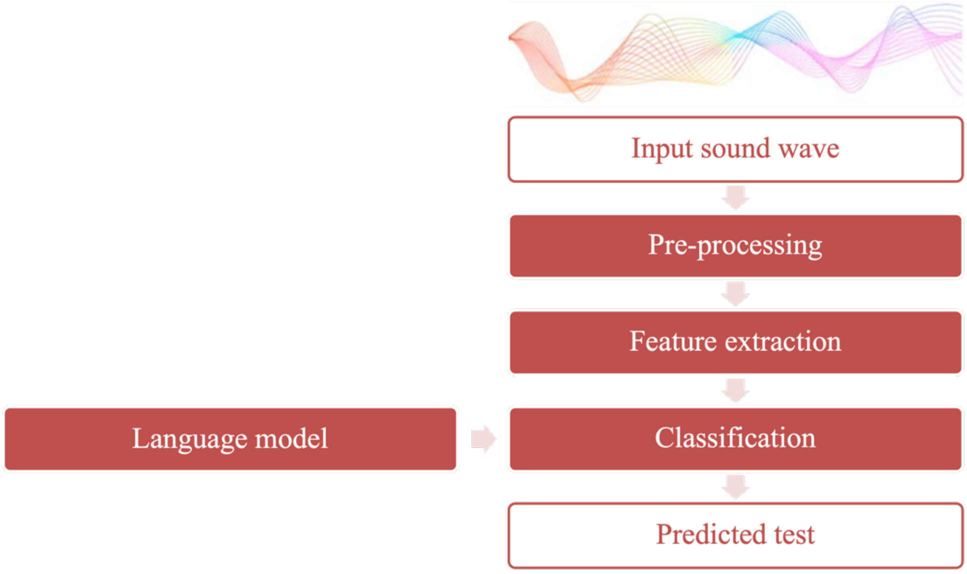


Figure 4-2 Automatic speech recognition structure. Adapted from [164]

The pre-processing stage improves the quality of the incoming audio signal by reducing noise and normalizing amplitude. [169] Common techniques include framing, pre-emphasis, end-point detection and normalization, after which the waveform is ready for analysis.

Feature extraction converts the continuous speech signal into a compact set of numerical descriptors that capture relevant spectral and temporal characteristics. Robust feature representations are critical to achieve reliable recognition under varying conditions. Frequently used methods include Mel-Frequency Cepstral Coefficients (MFCCs), Linear Predictive Coding (LPC) and Discrete Wavelet Transform (DWT) features. [164]

The classification stage maps these features to linguistic units such as phonemes, subwords or words. These models can be generative, such as HMMs and Gaussian Mixture Models (GMMs), or discriminative, like SVMs and Artificial Neural Networks (ANNs). The first ones learn the joint probability distribution of inputs and outputs, while the others directly learn the decision boundaries. Hybrid architectures combine the advantages of both approaches; an example are the HMM-ANN systems. [170]

Finally, the language model applies linguistic constraints and probabilistic rules to form valid word sequences. Traditional systems rely on statistical n-grams, whereas modern architectures often embed the language model within a neural network. Although some end-to-end systems can operate without an explicit language model, incorporating contextual knowledge typically improves accuracy. [164]

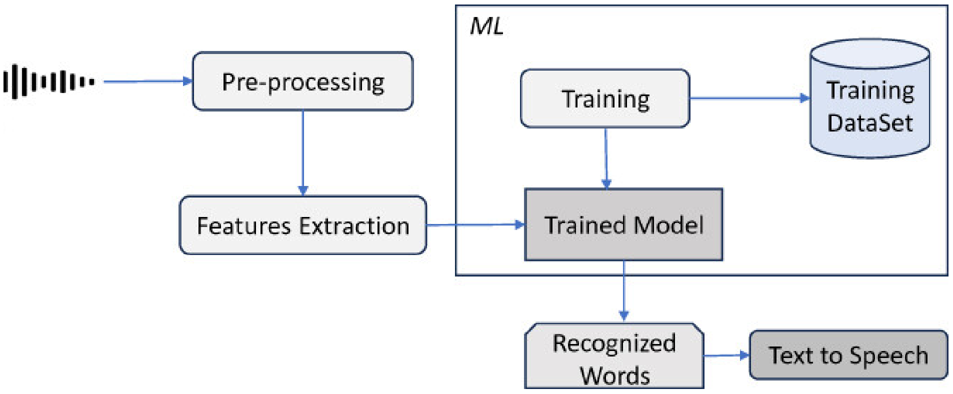


Figure 4-3 Architecture of an ASR system. Adapted from [162]

4.3 Evaluation Metrics

The performance of ASR systems is generally assessed in terms of accuracy and processing speed. [164]

The Real-Time Factor (RTF) evaluates efficiency as the ratio between the system's processing time (P) and the duration of the input signal (I):

$$RTF = \frac{P}{I}$$

An RTF of 1 indicates that the system operates in real time, while values greater than 1 denote slower processing. Since RTF depends strongly on hardware and implementation, it is mainly used to compare computational efficiency across models. [164]

Recognition accuracy is most commonly expressed through the Word Error Rate (WER), which quantifies the discrepancy between the ASR output and a human reference transcription. WER is defined as:

$$WER = \frac{S + D + I}{N}$$

where:

- S = number of substitutions,
- D = number of deletions,
- I = number of insertions,
- N = total number of words in the reference transcription.

A lower WER indicates better performance, meaning higher accuracy, with 0% representing a perfect match. Importantly, WER is sensitive not only to misrecognition of words but also to omissions and insertions. [171]

A complementary measure, the Word Recognition Rate (WRR), expresses the percentage of correctly recognized words and is defined as

$$WRR = \frac{H}{N}$$

where $H = N - (S + D)$. [164]

4.4 Commercial and open-source ASR Platforms

Beyond academic and research prototypes, a variety of commercial and open-source ASR platforms are now available for general use. These systems, ranging from large cloud-based services to freely distributed models, have made speech recognition widely accessible and easily integrable into diverse applications. Among the most commonly used solutions are Microsoft Azure Speech-to-Text, Google Cloud Speech-to-Text and OpenAI Whisper, which represent distinct approaches to large-scale ASR development. [172]

- Microsoft Azure Speech-to-Text: Microsoft Azure is a comprehensive cloud computing platform offering a suite of over 200 services across domains such as data management, AI, and machine learning. Within its Cognitive Services framework, the Speech-to-Text API provides real-time or offline transcription capabilities through pre-trained neural language models, which can be optionally adapted to domain-specific data. It currently supports over 130 locales, including Italian, and integrates features such as pronunciation assessment and custom model refinement [172]. The service can operate both in asynchronous and streaming modes, making it suitable for real-time clinical data collection and analysis.
- Google Cloud Speech-to-Text: The Google Cloud Speech-to-Text API provides automatic transcription for a wide range of applications, from dictation to real-time captioning. It supports more than 380 acoustic and language models, with several variants for major languages, and offers synchronous, asynchronous, and streaming processing modes. The platform allows limited model customization and integration via the Google Cloud Console or directly through browser-based tools. [172]

- OpenAI Whisper: Whisper is an open-source ASR model released by OpenAI, trained on hundreds of thousands of hours of multilingual and multitask speech data. Unlike commercial APIs, Whisper does not require a cloud-based subscription and is available for local or offline use. It employs a transformer-based encoder–decoder architecture capable of robust transcription and language identification across a wide range of domains. However, Whisper currently supports only asynchronous (batch) transcription and not real-time streaming. [172]

Mahmoud et al. evaluated and compared several off-the-shelf ASR platforms using both healthy and pathological speech, specifically including samples of aphasic speech obtained from the naming and repetition subtests of a standardized aphasia battery. [172] Among the commercial systems tested, Microsoft Azure Speech-to-Text and Google Cloud Speech-to-Text were included as general-purpose solutions not specifically trained on the experimental data.

Microsoft Azure consistently achieved lower word error rates than Google Cloud, both on recordings from healthy speakers and on speech samples from individuals with aphasia [Figure 4-4, Figure 4-5]. In particular, when applied to Italian-language speech, Azure outperforms Google. A recent study reports the WER obtained from 10 hours of audio recordings, transcribed in both asynchronous and streaming modes, showing consistently lower error rates for Azure [173]. [Table 4-4-1]

Dataset	***Hours***	***Mode***	***OpenAI***	***Google***	***Azure***
YouTube videos	10	Async	8.41%	16.07%	8.54%
		Streaming	N/A	8.76%	8.54%

Table 4-4-1 WER obtained from 10 hours of audio recordings [173]

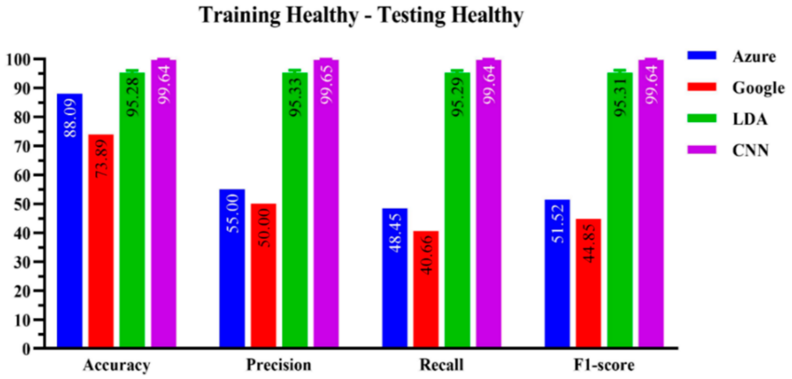


Figure 4-4 Performance evaluation of the ML-based algorithms on healthy dataset [172]

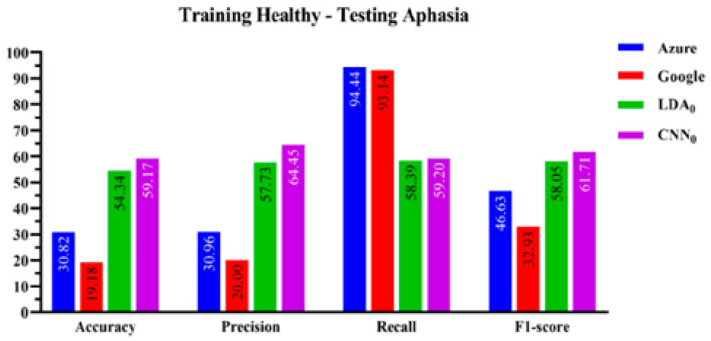


Figure 4-5 Performance evaluation of the ML-based algorithms on aphasic dataset [172]

4.5 Voice activity detection

The quality of the audio input is a key determinant of both ASR accuracy and downstream feature extraction. Recordings collected in different environments are often affected by background noise, variable microphone placement, low recording levels or irregular speaking patterns. Preprocessing is therefore a fundamental step in speech-based pipelines, ensuring greater consistency and reliability across datasets. [174]

Voice Activity Detection (VAD) is a preprocessing technique used to identify and separate speech from non-speech segments within an audio signal. It plays an important role in many speech processing applications that must operate under noisy or degraded conditions. [175] VAD systems are designed to detect when speech is present and to isolate those intervals from silence, environmental noise or other non-verbal sounds. By doing so, they reduce the amount of irrelevant or redundant data that may otherwise interfere with subsequent analysis. [176] Some advanced implementations also include speaker diarization, which determines who is speaking and when in a conversation. [177]

A typical VAD pipeline involves several stages. [Figure 4-6] The audio input, recorded via microphone or retrieved from an existing file, is first fed into the system. During pre-processing, the signal is divided into short overlapping frames, typically 10-30 ms, to capture fine temporal dynamics and improve robustness to noise. [175]

Next, acoustic features are extracted from each frame to distinguish voiced from unvoiced regions; common features include energy, SNR, zero-crossing rate, entropy, correlation and MFCCs. [175]

Finally, a classifier assigns each frame to one of the two classes, speech or non speech. Traditional approaches employ methods such as GMMs, HMMs or SVMs, whereas recent systems increasingly adopt deep neural networks for improved robustness in low SNR conditions. [175]

By detecting when speech occurs, VAD marks the temporal boundaries of speech segments within the audio signal. Removing silent intervals helps reduce insertion errors, which occur when non-speech segments are mistakenly treated as speech, and deletion errors, where speech is not recognized as such and is therefore not transcribed [176]. This process results in cleaner input for subsequent steps (like transcription, feature extraction and classification), improving the overall performance of the ASR system.

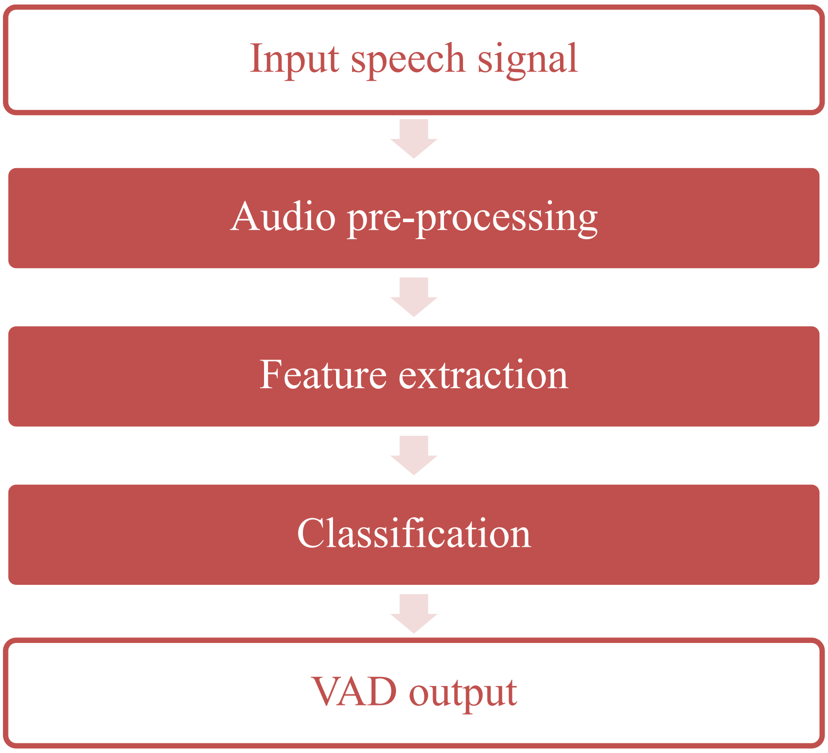


Figure 4-6 VAD pipeline. Adapted from [175]

4.5.1 Commercial VAD Systems

Several VAD systems have been developed to meet the needs of both real-time speech recognition and automatic audio analysis. Some of the leading VAD technologies currently available include:

- Silero-VAD: a system developed by Silero AI Team, based on a pre-trained neural network model for voice activity detection. The model, available as open-source, is implemented in Torch and supports audio with sample rates of 8 kHz and 16 kHz. [177]

- NeMo Deep ASR by NVIDIA: this tool is part of NVIDIA's NeMo toolkit, which provides a suite of deep learning models designed for speech detection, transcription and speaker identification. Its architecture allows the system to handle complex acoustic environments, including multi speaker scenarios. [177]

Both systems are effective in detecting pauses and silence, achieving comparable error rates, but Silero-VAD has demonstrated higher precision and efficiency (in terms of execution time), particularly when applied to clinical context audio recordings, making it well-suited for high variability speech datasets. [177]

4.6 Feature extraction from ASR transcript

Speech provides an accessible window into the cognitive and linguistic functions that are frequently affected in neurodegenerative conditions. When speech is converted into text transcripts, it becomes possible to extract quantitative descriptors that capture both linguistic and temporal aspects of communication. [178], [179]

Natural Language Processing (NLP) is the computational discipline that enables machines to analyze, understand and generate human language. It combines theoretical models of linguistics with algorithmic and statistical methods to represent naturally occurring texts at one or more levels of linguistic analysis. [180] Traditionally considered a subfield of artificial intelligence, NLP has evolved into a vast field of research that now includes computational linguistics, data mining and machine learning. It includes tasks such as language modeling, semantic similarity estimation, translation and speech understanding. [181] In the context of speech research, NLP methods are applied to the textual output of ASR systems to identify linguistic patterns that reflect cognitive processes. NLP enables the extraction of interpretable linguistic features using techniques such as tokenization (splitting text into words), part-of-speech (POS) tagging (labelling words by their grammatical role), syntactic parsing (analyzing sentence structure) and semantic analysis (interpreting meaning and relationships between words). [182]

Recent research has shown that a wide range of lexical, syntactic and acoustic features derived from transcribed speech can effectively discriminate between individuals with neurodegenerative disorders and healthy speakers. [179] The combination of NLP and ML techniques allows these features to be computed automatically from spoken samples, such as narrative picture descriptions or spontaneous conversations, providing an efficient and scalable alternative to traditional hand-crafted analyses. [179]

Table 4-6 Overview of studies on transcription approaches and feature extraction in neurodegenerative disorders.

Authors - Year	*Neurodegenerative condition*	*Transcription method*	*Extraction method*	*Linguistic feature types*	*Study focus*
Peters et al. - 2025	PPA	Whisper	Linguistic Feature Toolkit	Surface features, lexico-semantic, discourse and syntax features	Classification
Crawford - 2025	PD	Whisper	Word2Vec, BERT, XLNet, GPT-2, text-embedding-ada-002, text-embedding-3 (small and large)		Disease detection
Heitz et al. - 2024	AD	Wave2vec2, Whisper, Google Speech "Chirp"	BERT, manual	Syntactic features based on POS tags and grammatical constituents, lexical and repetitiveness features	Classification for AD recognition
Konig et al. - 2024	Subjective Cognitive Decline, MCI, Dementia	Manual, SIGMA (ki:elements ASR pipeline)	SIGMA	Word count, semantic cluster size and switches, word frequencies	Investigation of the agreement between manual and automatic transcription
Huang et al. - 2024	AD, MCI	Feishu	Stanford CoreNLP, manual	POS tagging, type-token ratio (TTR), information	Comparison between acoustic and

				words and units	linguistic features for screening
Lopes da Cunha et al. - 2024	AD, bvFTD	Google Speech to Text, manual revision	FreeLing (POS tagger, morphological tagging module), TELL, FastText	Lexical category, first and third person usage, lexico-semantic features, semantic variability	Analysis of free speech for diagnosis
Cho et al. - 2024	AD, FTLD	Semi-automatic annotation protocol	SpaCy, NLTK	POS tagging, semantic ambiguity, word frequency and familiarity, concreteness, lexical diversity	Classification
Rezaii et al. - 2024	PPA	Microsoft Dictate	Stanza	POS tagging, high and low frequency verbs	Classification
Soroski et al. - 2022	AD, Mild cognitive impairment (MCI), Subjective Memory Complaints (SMC)	Google speech-to-text, manual	Stanford CoreNLP, MRC database, pydub (Python), Syllables (Python)	POS tagging, grammar rules, vocabulary richness, speech rate, pauses	Comparison between manual and ASR transcriptions for classification
Cho et al. - 2021	FTD (bvFTD, svPPA, nfvPPA)	Manual, annotated with Linguistic Data Consortium	SpaCy	POS tagging, lexical diversity, noun frequency, abstractness	Analysis of lexical features
Slegers et al. - 2021	PPA	Manual	SpaCy	POS tagging, named entity recognition, dependency parsing, psycholinguistic, semantic	Classification

				and pragmatic features	

Since neurodegenerative conditions can alter both how people speak (fluency, rhythm, hesitations) and what they say (lexical choice, syntactic structure, semantic content), feature extraction is often organized into two complementary domains: temporal-fluency features and linguistic features. [183]

4.6.1 Temporal and fluency features

Temporal or fluency features describe the dynamic properties of speech, such as timing, rate and pause behavior and provide insight into underlying cognitive processing during language production. Studies have highlighted, for instance, that individuals with cognitive decline tend to produce longer or more frequent hesitations and exhibit reduced speech rate. [183]

Hesitations can be divided into silent pauses and filled pauses (e.g., “ehm”, “um”), both of which can be quantified automatically through ASR systems that output time-aligned phoneme sequences. Typical temporal descriptors include articulation rate (phones per second excluding pauses), speech tempo (phones per second over the entire utterance), utterance length, duration and number of silent or filled pauses, as well as the overall hesitation rate, calculated as the ratio of total pause duration to utterance duration. [183]

Automating this extraction greatly reduces manual workload and enables large-scale analyses using ML classifiers. These markers are particularly informative because they reflect psycholinguistic processes such as lexical retrieval and working-memory load, which are frequently disrupted in neurodegenerative disorders. [183]

4.6.2 Linguistic Features

Linguistic features capture the structure and meaning of spoken language, reflecting how words, sentences, and discourse are organized and used to convey information. They include indices of lexical diversity, syntactic complexity and discourse organization. [184]

These measures provide interpretable indicators of language impairment and are closely aligned with traditional clinical assessments of speech and language function. This category also encompasses semantic embeddings, that are vector representations of word or sentence meaning derived from large-scale NLP models (such as word2vec or BERT), which quantify semantic relationships through data-driven learning rather than predefined linguistic rules. [184]

4.6.2.1 Lexical Features

Lexical measures quantify vocabulary richness, diversity and usage patterns. One of the most widely used metrics is the Type–Token Ratio (TTR), defined as the ratio between the number of unique words (types) and the total number of words (tokens). Because TTR is sensitive to text length, several length-independent indices have been developed, including Brunet's Index (W) and Honoré's Statistic (R), which estimate lexical richness while correcting for sample size. [184]
Brunet's Index is computed as:

$$W = N^{V^{-0.165}}$$

Where N is the total number of words and V is the total number of unique words. Lower values of W indicate a richer vocabulary. [185]
Honoré's Statistic is obtained as:

$$R = 100 \cdot \log\left(\frac{N}{1 - \frac{V_1}{V}}\right)$$

Where V_1 represents words occurring only once, V is the total number of unique words and N is the total text length. Higher values of R correspond to a richer vocabulary. [185]

Beyond lexical diversity, additional descriptors capture word frequency and grammatical category distributions, providing complementary insights into vocabulary use. Mean word frequency (from reference corpora) reflects how common or rare a speaker's word choices are, while ratios such as verb-to-noun or open-class to closed-class words describe the balance between content and function words. [186]
Lexical analyses may also incorporate psycholinguistic variables, including imageability, concreteness, familiarity, and age of acquisition, often computed separately for different parts of speech. These variables provide fine-grained indicators of semantic degradation, word-finding difficulty and conceptual impoverishment, which are characteristic of several neurodegenerative syndromes. [184]

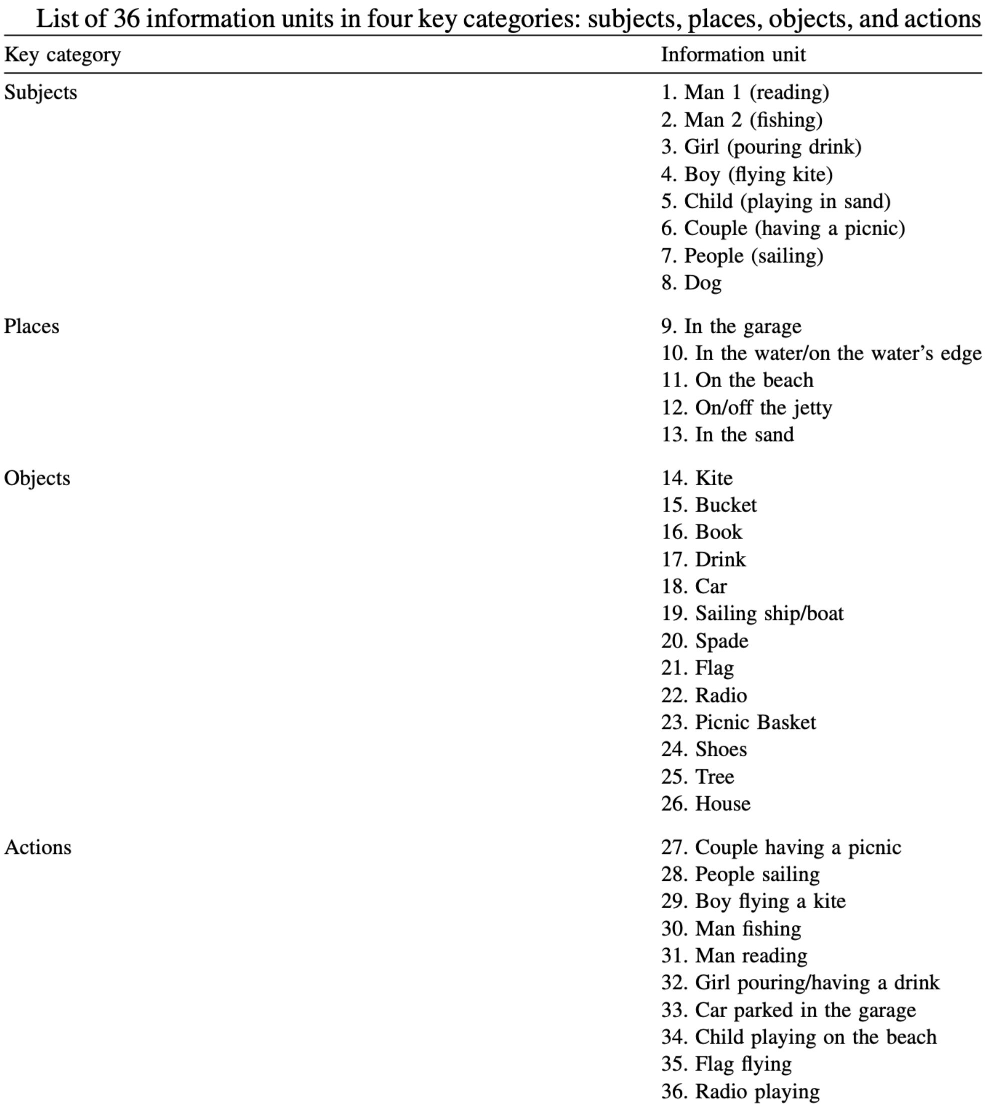
4.6.2.2 Syntactic features

Syntactic features capture how words are combined into phrases and sentences, reflecting grammatical organization and expressive complexity. Using NLP tools such as spaCy, syntactic relations can be derived through Universal Dependency (UD) parsing, which assigns each word a grammatical function (e.g., subject, object, modifier) within the sentence. [184]
From these dependency trees, researchers compute indices such as the average number of dependents per word (syntactic complexity), mean sentence length,

number of noun or verb phrases and the proportion of specific dependency relations (subjects, objects or modifiers) relative to the total number of words. [184]
Simplified sentence structures and reduced grammatical complexity are commonly observed in neurodegenerative speech, particularly in conditions involving frontal or perisylvian network degeneration. [186]

4.6.2.3 Semantic and pragmatic features

Semantic features assess the meaning and informativeness of speech, while pragmatic features capture how language is used in context. A popular semantic measure is the quantification of Information Content Units (ICUs): pieces of meaningful information expressed by the speaker during a narrative description. [187] [Figure 4-7]
Each ICU is encoded as a Boolean variable (mentioned or not mentioned), and the total ICU count provides a measure of overall informativeness. [184]
At a higher level, semantic idea density estimates the amount of information conveyed per word, computed as the average cosine similarity among word embeddings within a sliding window. [184] Lower idea density values indicate reduced informativeness and semantic cohesion, both associated with progressive language deterioration.
Pragmatic descriptors, on the other hand, describe how language is used in context, capturing communicative intent and discourse organization. These measures focus on functional aspects of communication, such as markers of uncertainty ("maybe", "I think"), deictic expressions ("this", "that", "here"), word-finding expressions ("I can't remember") and self-monitoring or repair markers ("I mean, I don't know"). Such indicators can reflect the hesitation and the effort or compensatory strategies employed to maintain communicative effectiveness. [184]
Filled pauses are also examined as a pragmatic features, since they carry information about hesitation type and discourse planning, complementing the quantitative pause measures described above in the temporal domain.



List of 36 information units in four key categories: subjects, places, objects, and actions

Key category	Information unit
Subjects	1. Man 1 (reading) 2. Man 2 (fishing) 3. Girl (pouring drink) 4. Boy (flying kite) 5. Child (playing in sand) 6. Couple (having a picnic) 7. People (sailing) 8. Dog
Places	9. In the garage 10. In the water/on the water's edge 11. On the beach 12. On/off the jetty 13. In the sand
Objects	14. Kite 15. Bucket 16. Book 17. Drink 18. Car 19. Sailing ship/boat 20. Spade 21. Flag 22. Radio 23. Picnic Basket 24. Shoes 25. Tree 26. House
Actions	27. Couple having a picnic 28. People sailing 29. Boy flying a kite 30. Man fishing 31. Man reading 32. Girl pouring/having a drink 33. Car parked in the garage 34. Child playing on the beach 35. Flag flying 36. Radio playing

Figure 4-7 Information content units. From [148]

4.6.3 Linguistic feature extraction in PPA

In the context of PPA, linguistic feature extraction plays a particularly crucial role, as each PPA variant exhibits a distinct combination of linguistic and cognitive impairments. [188]

- The nfvPPA typically presents with agrammatism and motor-speech deficits.
- The svPPA is characterized by impaired single-word comprehension and loss of semantic knowledge.

- The lvPPA is marked by word-finding difficulties, phonological errors and sentence repetition deficits.

To capture these heterogeneous profiles, recent studies have adopted multidimensional feature sets with lexical, syntactic, semantic, and discourse-level descriptors, often combined with acoustic and neuroimaging correlates. [188] For example, ratios of high- to low-frequency verbs or measures of idea density have been shown to differentiate PPA variants, with lvPPA showing selective difficulty retrieving low-frequency verbs and svPPA demonstrating reduced semantic diversity. [184], [186]

4.7 Multimodal Integration

Human interaction with the world is inherently multimodal: we constantly combine information from multiple sensory channels, like vision, hearing, touch and proprioception, to explore our surroundings and perceive new stimuli. [189] Each modality contributes complementary cues; for example, auditory signals convey not only linguistic content but also prosody, speaker identity and contextual information about the environment. [190]

In contrast, human–computer interaction and computational modeling have historically relied on unimodal communication, which is usually limited to a single input or output channel such as text or audio. With the evolution of sensing technologies, signal processing and machine learning, multimodal interfaces have been developed to better capture the complexity of real-word interactions and take advantage of human communication abilities, like speech, gesture, touch and facial expression. [190][191]

In the context of neurodegenerative disease research, multimodal approaches are increasingly adopted to connect behavioral, acoustic and neurobiological levels of analysis. Speech and language data can reflect subtle cognitive and motor alterations, while neuroimaging provides complementary markers of structural and functional brain integrity. Integrating these heterogeneous data sources enables a more comprehensive understanding of disease mechanisms, linking communication deficits to underlying neural alterations. [188]

The extraction of features from speech and neuroimaging data often results in high-dimensional datasets. However, clinical datasets often include a relatively small number of participants, increasing the risk of overfitting when too many features are used simultaneously. This makes feature selection a crucial step in building reliable and interpretable ML models. [192]

Approaches to multimodal data fusion can be categorized as follows:

- Early fusion (feature-level integration): normalized feature sets from multiple modalities are combined into a single input vector before classification. [193], [194]
- Late fusion (decision-level integration): separate models are trained on each modality and their outputs are combined, for example via majority voting or weighted averaging. [193], [194]
- Hybrid methods: intermediate representations, such as embeddings from NLP models and metrics derived from imaging, are merged before classification, offering a compromise between early and late fusion.

Early fusion is optimal when modalities are well-aligned and model assumptions are known, whereas late fusion is often more practical in heterogeneous data and can mitigate issues of dimensionality and allow differential weighting of modalities. [193], [194]

Empirical findings support the benefits of multimodal integration. Agosta et al. demonstrated that combining structural and diffusion MRI metrics improved the discrimination of neurodegenerative phenotypes compared to single-modality analyses. [195] Similarly, studies integrating speech-derived and neuroimaging features have reported higher diagnostic accuracy and provided insights into the brain–behavior relationships underlying the course of disease. [192]

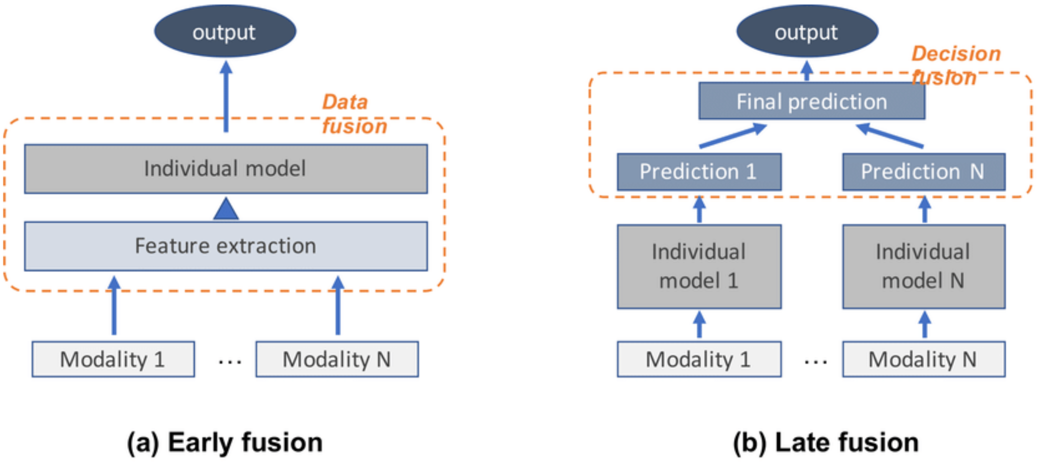


Figure 4-8 The early fusion and late fusion [196]

5 AIMs

PPA is a neurodegenerative syndrome characterized by the progressive deterioration of language abilities, while other cognitive functions remain relatively preserved in the early stages. Accurate diagnosis and differentiation of PPA subtypes—nfvPPA, svPPA, and lvPPA—are critical for clinical management, prognosis, and the development of targeted interventions. Traditional assessment approaches rely heavily on manual transcription and feature extraction, which are time-consuming and require specialized expertise, and integrating multimodal data such as linguistic and neuroimaging measures remains challenging.

Recent advances in machine learning and neuroimaging provide promising opportunities to improve diagnostic precision and efficiency. Automated pipelines combining speech analysis and diffusion tensor imaging can capture subtle structural and linguistic differences between PPA subtypes, offering scalable tools for clinical support. In this context, the present study aims to develop and validate a multimodal framework for PPA diagnosis, focusing on audio processing, feature extraction, and workflow optimization.

Specifically, the goals of this study are:

- Aim 1 – Evaluation of audio preprocessing and automatic transcription methods

  To assess the performance and reliability of automated audio preprocessing pipelines combined with automatic speech transcription. This aim focuses on comparing the outputs of automated transcriptions against manual transcriptions by expert speech therapists, in order to identify methods that provide accurate, reproducible, and clinically feasible linguistic data for downstream analyses.
- Aim 2 – Automatic extraction of linguistic and imaging-derived features and development of a diagnostic classifier

  To define and automatically extract a comprehensive set of linguistic and neuroimaging-derived features from speech recordings and DTI data, and to integrate them into machine learning classifiers for differentiate PPA subtypes. This aim includes developing models that are both accurate and interpretable, allowing the identification of key features associated with each subtype, and enabling insights into the neural and linguistic mechanisms underlying PPA.
- Aim 3 – Comparison between the manual workflow, the fully automated system, and a hybrid approach

  To evaluate and compare three different workflows: the traditional manual approach, a fully automated pipeline, and a hybrid system that integrates automated processing with clinical supervision. This aim focuses on assessing accuracy, efficiency, and standardization, while exploring how hybrid approaches can leverage both computational power and expert knowledge to optimize clinical applicability and reliability.

6 Material and methods

6.1 Participants

A total of 185 subjects participated in the study, including 91 healthy controls (HC) and 94 PPA patients, subdivided into 38 nfvPPA, 36 svPPA and 20 lvPPA. Between 2010 and 2025, patients were prospectively recruited at six referral centers across Lombardy (Italy) and referred to IRCCS San Raffaele Hospital in Milan. All patients were diagnosed as PPA according to the established criteria. [8] Among the total cohort, a sub-sample of 81 patients (34 nfvPPA, 27 svPPA and 20 lvPPA) underwent a comprehensive clinical protocol including a neurological examination, a detailed neuropsychological assessment, and the Picnic Scene description test, from the WAB, recorded via audio recorder. In addition, 85 patients underwent MRI acquisition: 34 nfvPPA, 32 svPPA and 19 lvPPA; for those who also had been recorded, the MRI scan had place within 6 months from the recording.

Healthy control subjects were recruited among non-consanguineous relatives, acquaintances, or selected from available datasets, in order to achieve a balanced and representative sample size. Among them, 80 HC underwent DWI acquisition and a subset of 38 completed the Picnic Scene description test with audio recording.

All participants provided written informed consent before participating in the study.

	HC	***PPA***	***nfvPPA***	***svPPA***	***lvPPA***
Number of subjects	91	94	38	36	20
Picnic scene recording	38	81	34	27	20
MRI acquisition	80	85	34	32	19
Both	27	72	30	23	19

Table 5-1 Sample distribution across subject groups and availability of imaging and audio recording data

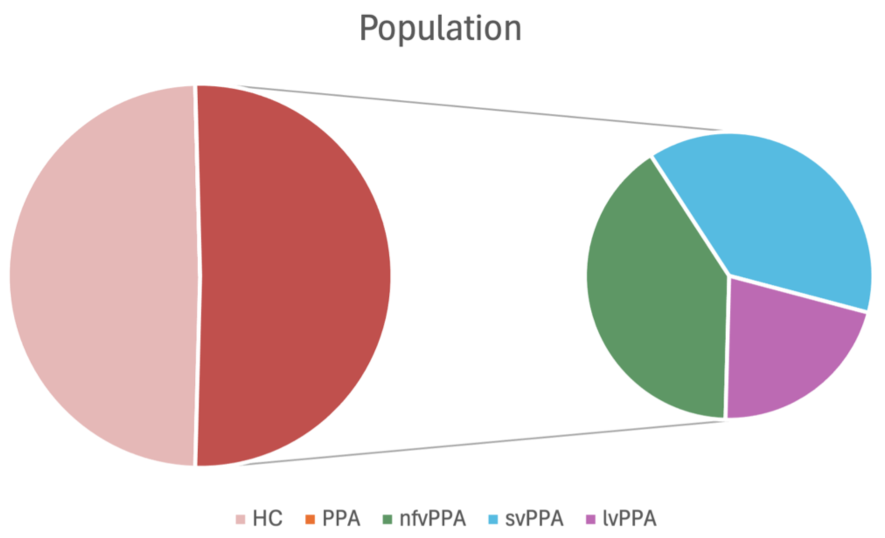


Figure 6-1 Dataset composition in terms of number of HC and PPA patients.

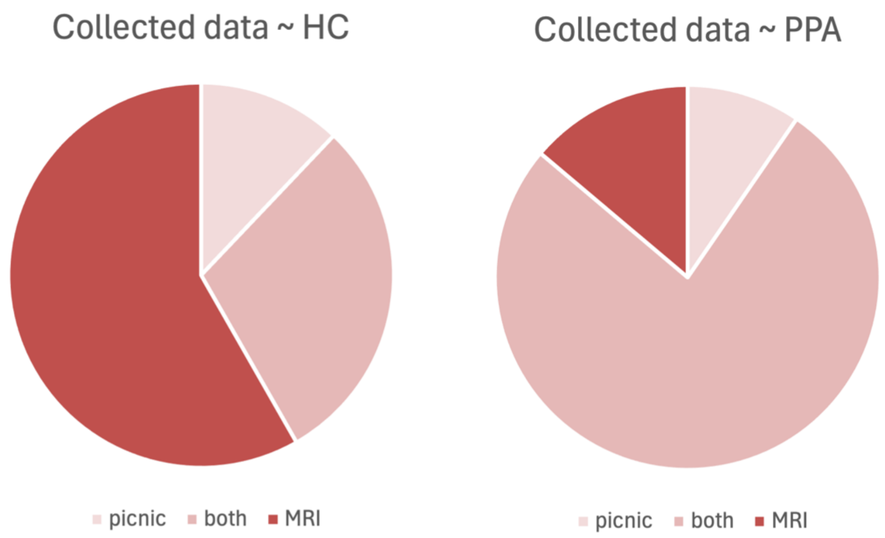


Figure 6-2 Acquired data for HC and PPA participants: Picnic picture description audio recordings and MRI scans.

6.1.1 Inclusion and exclusion criteria

To be included, all patients had to undergo a neurologic examination and a comprehensive neuropsychological assessment, as well as have at least the Picnic test audio recording and/or the MRI scan available.

HC who underwent the Mini-Mental State Examination (MMSE) and Beck Depression Inventory (BDI) had to obtain at least a MMSE score above 27 and a BDI score below 15, to confirm absence of mood disturbances. Subjects with a family history of neurodegenerative diseases were excluded from this category.

Exclusion criteria for both groups included substance abuse, major medical illnesses (neurological, psychiatric or systemic) and any causes of brain damage other than PPA, as these could have interfered with cognitive functioning.

6.2 Neuropsychological assessment

Neuropsychological evaluations were performed by experienced neuropsychologists unaware of the MRI results. In all patients, the neuropsychological assessment investigated the global cognitive functioning with the MMSE [197] ; frontal functioning with the Frontal Assessment Battery (FAB) [198]; verbal memory with the digit span forward [199] and the Rey Auditory Verbal Learning Test (RAVLT) [200]; non-verbal memory with the spatial span forward [199] and the Rey's figure delayed recall [201]; attention and executive functions with the attentive matrices [202], the digit span backward [203], the trail making test [204], the colored progressive matrices [205], and the clock drawing test [206]; visuospatial abilities with the Rey's figure copy [201]; levels of autonomy with Activities of Daily Living (ADL) [207] and Instrumental Activities of Daily Living Scales (IADL) [208]; behavior with the Neuropsychiatric inventory [209] and the Frontal Behavioural inventory [210]; disease severity with the clinical dementia rating scales (CDR, CDR-Sum of boxes and CDR-FTLD) [211],[212]. Furthermore, patients underwent a comprehensive language assessment which evaluated: syntactic comprehension with the Token test [213]; confrontation naming and single word comprehension with the subtests of the CaGi battery [214]; object knowledge with the Pyramids and Palm Tree Test [215]; repetition, reading, and writing with the AAT [216]; fluency with phonemic and semantic fluency tests [217].

6.3 Speech evaluation

The oral version of the Picnic Scene subtest of the WAB was selected for the evaluation of the spontaneous speech of patients and controls. [Figure 6-3] During the evaluation, participants were instructed as follows: "Take a look at this picture, tell me what you see, and try to talk in sentences." Speech samples consisted in the oral description of the Picnic Scene and were audio-recorded using Audacity software (*audacity.sourceforge.net*).

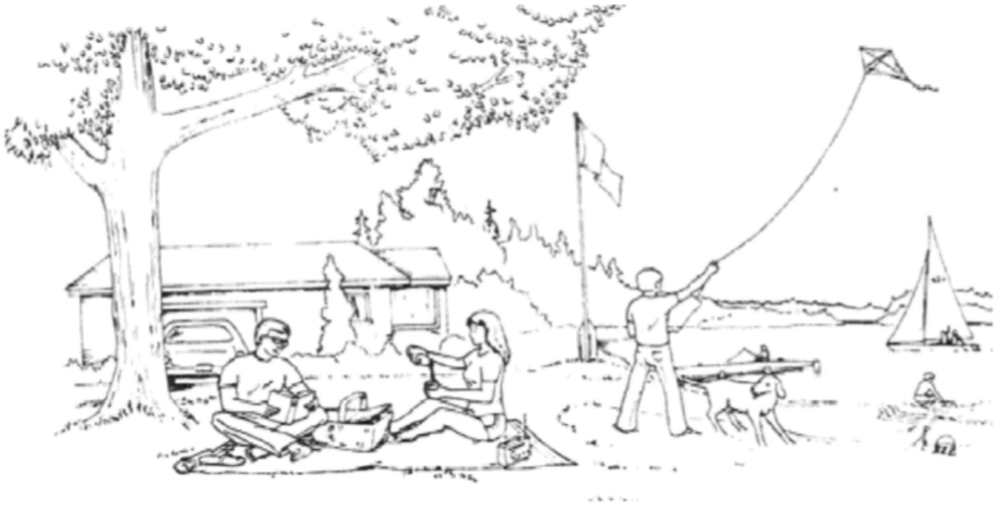


Figure 6-3 Picture description task: the Picnic Scene. From the Western Aphasia Battery.

6.4 MRI acquisition

Brain MRI was performed using two different 3.0 T Philips Medical Systems scanners at IRCCS San Raffaele Hospital between 2010 and 2025.

For 50 patients and 50 HC, diffusion-weighted images (DWI) were acquired on the Ingenia CX scanner with the following parameters: TR=5900 ms, TE=78 ms, a section thickness of 2.3 mm, number of slices=56, matrix size 112x85 and field of view (FOV) = 240x232 mm^2. Diffusion gradients were applied along 6, 30 and 60 directions, with three b-values (700, 1000 and 2855 $\mathrm{s/mm}^2$).

For 36 patients and 26 HC, data were acquired on the Intera scanner using a pulsed-gradient spin-echo echo-planar imaging (SE-EPI) sequence with sensitivity encoding. Acquisition parameters were: TR=8986 ms, TE=80 ms, section thickness=2.5 mm, number of slices=55, matrix size=96x96, FOV=240x240 mm^2, with 32 diffusion gradient directions and a b-value of 1000 $\mathrm{s/mm}^2$).

For the remaining 2 PPA patients and 4 HC, data were acquired on the Intera scanner using a pulsed-gradient SE-EPI sequence with sensitivity encoding and the following parameters: TR=8773 ms, TE=58 ms, 55 contiguous axial slices (2.3 mm thickness), matrix size=112x88, in-plane pixel size=1.87x1.87 mm, FOV=231x240 mm^2, 35 non-collinear diffusion directions, and a b-value of 900 $\mathrm{s/mm}^2$.

6.4.1 Diffusion Weighted MRI preprocessing

MRI preprocessing was performed using the FMRIB Software Library (FSL, version 5.0.5) and relied on four main tools: TOPUP, EDDY, BET and DTIFIT.

TOPUP was applied to correct for susceptibility-induced distortions by estimating the off-resonance field using pairs of b_0 images acquired with opposite phase-encoding directions. The resulting field map was then used by EDDY, which corrected for eddy-current distortions and head motion, identifying and replacing signal outliers and rotating the corresponding diffusion gradient directions (b-vectors). After distortion and motion correction, non-brain tissue was removed using BET (Brain Extraction Tool) to generate a brain mask. Subsequently, diffusion tensor was estimated using DTIFIT, which applies a linear regression model with a multishell approach (b=700, 1000 and 2855 s/mm2). From the resulting tensors, fractional anisotropy (FA) maps were computed.

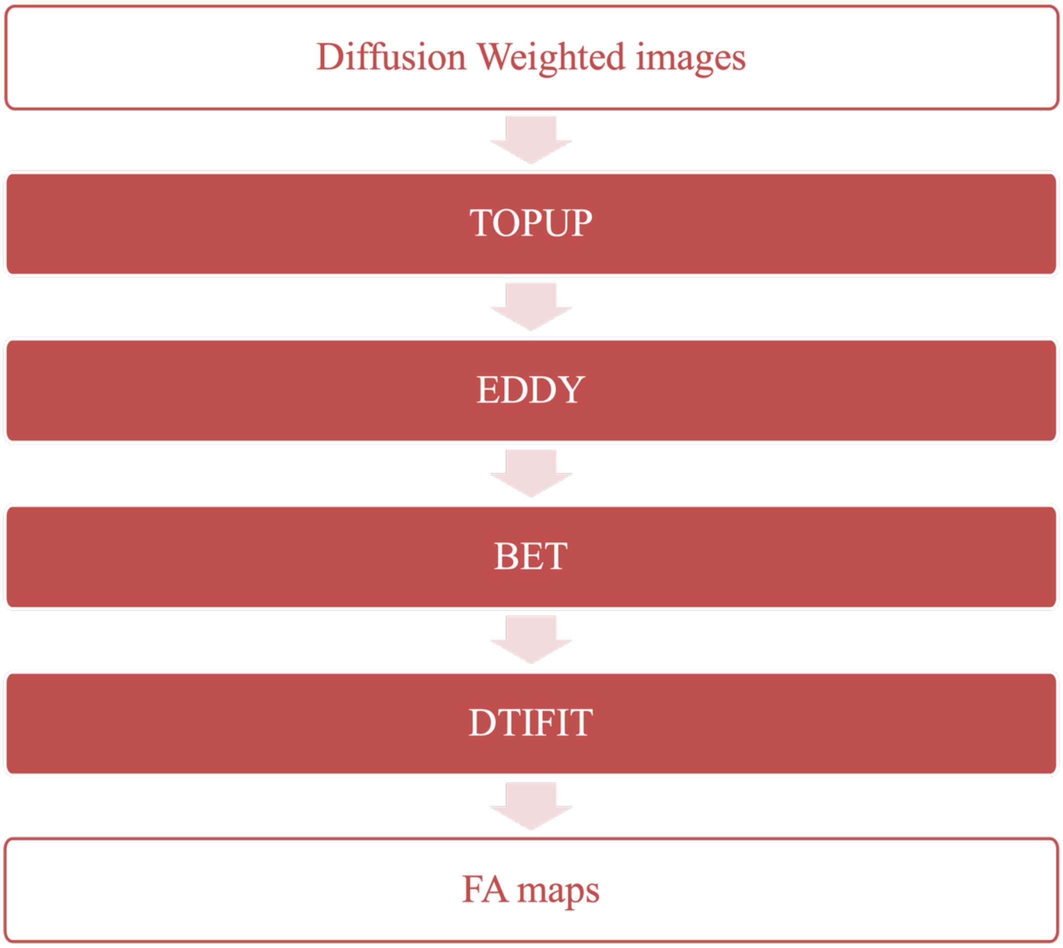


Figure 6-4 Workflow of diffusion weighted MRI preprocessing

6.4.2 TBSS analyses

Voxel-wise diffusion tensor imaging analyses were performed using the Tract-Based Spatial Statistics (TBSS) tool implemented in FSL (version 5.0.9) (http://www.fmrib.ox.ac.uk/fsl/fdt/index.html) focusing on fractional anisotropy maps. The analysis was conducted according to the standard TBSS pipeline.

First, all FA images were preprocessed to remove potential outliers from the diffusion tensor fitting.

Subsequently, all individual FA images were aligned to FMRIB58_FA standard-space template using a nonlinear registration procedure. This template is in MNI152 standard space and represents the mean FA image of 58 healthy adults. Each subject's FA image was non-linearly registered to the target, resampled to 1x1x1 mm MNI space and subsequently used for group-level comparisons.

The aligned FA images were averaged to obtain a mean FA image, from which a mean FA skeleton representing the centers of all major white matter tracts was created. Each participant's aligned FA map was subsequently projected onto this common skeleton, producing skeletonized FA maps for voxel-wise analysis.

Group comparisons were performed using the *randomise* tool in FSL, applying non-parametric permutation testing. Design and contrast matrices were created based on group membership (e.g., controls vs. patients), ensuring the order of entries matched the alphabetical order of the original FA images.

6.4.3 Feature extraction

After performing voxel-wise statistical analyses with *randomise*, the resulting contrast matrices were inspected to identify regions showing significant white-matter alterations (thresholded at $p < 0.05$). From each significant contrast, clusters of contiguous voxels surviving the statistical threshold were extracted and defined as reference Regions Of Interest (ROIs). For each subject, the mean FA value within each ROI was then computed by averaging FA intensities across all significant voxels belonging to that region on the individual skeletonized FA map. This procedure yielded, for every contrast, one FA value per subject per significant ROI. Finally, to obtain a compact set of imaging-derived features, FA values corresponding to ROIs belonging to the same anatomical region were averaged across contrasts, resulting in a final matrix used for subsequent classification.

6.4.4 Feature selection

A redundancy analysis was conducted to identify highly correlated feature pairs that might increase model complexity without contributing useful information. Pairwise Pearson correlations were computed among all features, applying a threshold of 0.97. For any pair exceeding this threshold, one of the two features would be discarded based on methodological considerations.

6.5 Audio processing

All audio samples were converted to .wav format and edited using Reaper [218] (v7.36) to remove examiner interventions during the discourse, keeping only the participant's voice in the track. The resulting cleaned recordings were used for transcription and subsequent analyses.

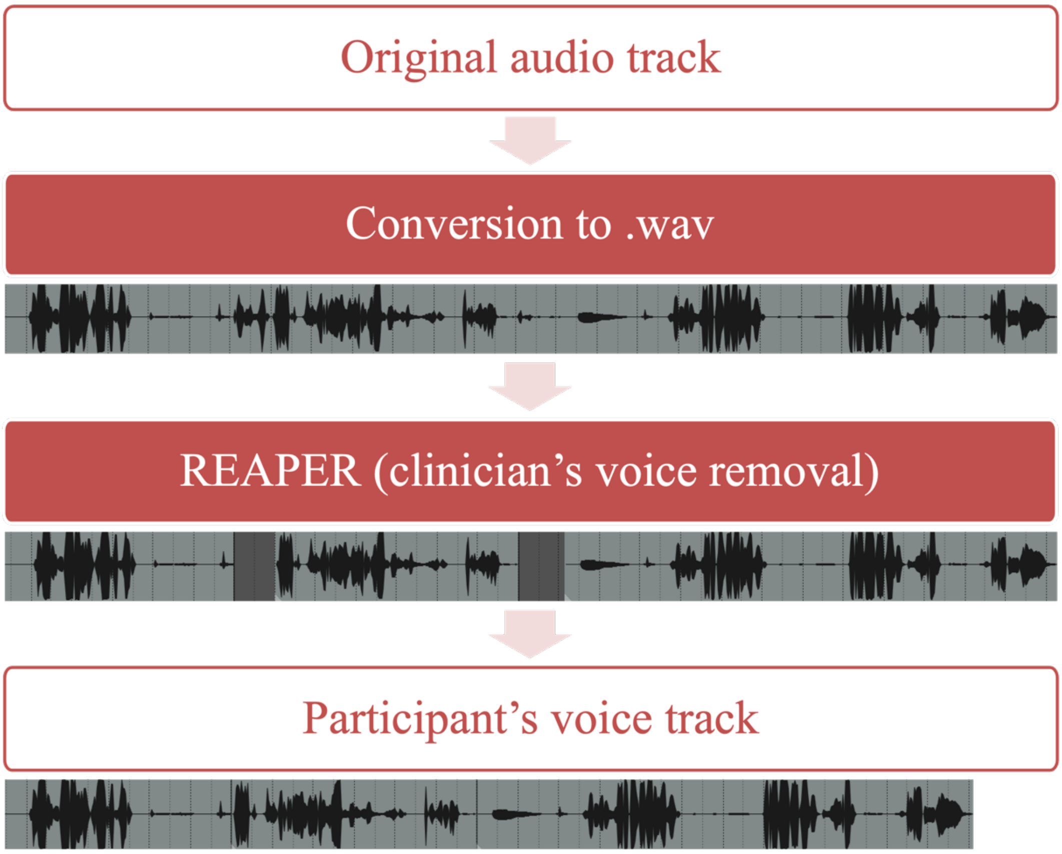
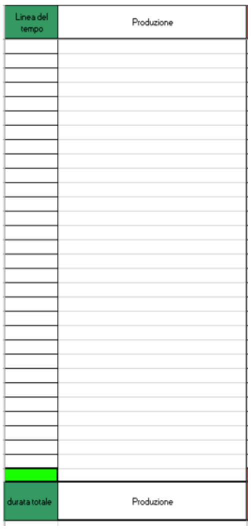


Figure 6-5 Workflow of the audio preprocessing pipeline.

.

6.5.1 Manual transcription

Following audio preprocessing, the recordings were traditionally transcribed manually. Manual transcriptions were performed by speech therapists, using Microsoft Excel for annotations, while reproducing the audio with Audacity. The Excel sheet was organized as shown in Figure 6-6: the first column contained the time line, indicating the start and end points of each utterance of the spontaneous speech during the image's description; the second column included the transcribed text, divided into utterances, sentences and pause periods. Utterances were defined as sequences of words not interrupted by a pause lasting more than two seconds, whose boundaries could be identified based on prosodic cues. An utterance could therefore correspond to a single word, a phrase, a part of a phrase or a full sentence. The duration of silences was represented by a dot (“.”) for each second.



Linea del tempo	Produzione
durata totale	Produzione

Figure 6-6 Excel sheet organization for transcription.

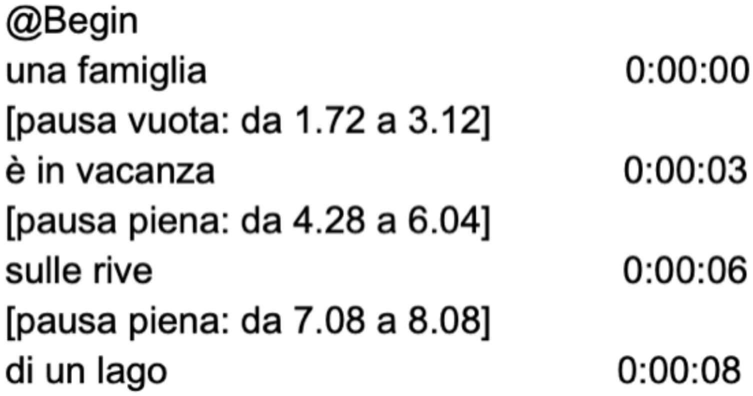
6.5.2 Automatic transcription

An automatic transcription pipeline was also applied to the audio recordings.

Automatic transcription of the audio recordings was performed using the Microsoft Azure Speech to Text service through the asynchronous batch transcription API (version 3.0). [225] Audio files were uploaded to Azure Blob Storage in anonymized form to ensure privacy compliance. A transcription job was created via a POST request to the corresponding endpoint, specifying Italian (“it-IT”) as the target language. The parameters of the request body were configured to produce a detailed output including linguistic metadata, such as the temporal boundaries of each recognized word, with the corresponding offset and duration. Once the transcription process was completed, the resulting file was downloaded in JSON format, which included text, word-level timestamps and confidence scores for each recognized word.

These JSON files were then converted into CHA files to improve readability. Words were arranged sequentially and a line break was inserted whenever two consecutive words were separated by a silent or filled pause. To automatically classify pauses as silent or filled, the Silero VAD model (v5.1.2) was employed. For each audio segment between two words (as defined by the JSON timestamps), the interval was processed by Silero, which labeled it as either speech or non-speech. The model also returned the precise onset and offset times of detected vocal activity, which could represent a subinterval of the analyzed segment. The timestamps of segments labeled as speech were then inserted into the CHA file on a new line and tagged as filled pauses.

Pauses longer than one second that were classified as non-speech were tagged as silent pauses, and their temporal boundaries were recorded on a separate line. Conversely, non-speech segments shorter than one second were discarded as non-pausal events. In addition, the duration of each line of words was annotated within the CHA file to preserve temporal information. [Figure 6-7]



```
@Begin
una famiglia                        0:00:00
[pausa vuota: da 1.72 a 3.12]
è in vacanza                        0:00:03
[pausa piena: da 4.28 a 6.04]
sulle rive                          0:00:06
[pausa piena: da 7.08 a 8.08]
di un lago                          0:00:08
```

Figure 6-7 Example of some lines from a CHA file transcription. Timestamps are reported in seconds.

6.6 AIM 1: evaluation of automatic transcription accuracy

6.6.1 Audio preprocessing

Different audio preprocessing pipelines were tested in order to evaluate their impact on transcription quality. The following preprocessing methods were applied:

I. Raw audio, kept in its original sampling frequency (SF) and channel configuration.
II. Resampling to a 16 kHz SF and conversion to a mono channel. This procedure has been widely adopted in literature [219][220] since the majority of ASR systems are trained at 16 kHz. [221]
III. Resampling audio to 16 kHz mono followed by loudness normalization (parameters: targeted integrated loudness=-16 LUFS (loudness units full scale, 1 LUF= 1 dB), true peak=-1.5 dB and loudness range=11 LU). [222]

IV. A noise reduction performed in Audacity (v 3.7.4) [223]: a 1-2 s silent interval was used to estimate noise profile, which was then subtracted from the entire track (12 dB noise reduction, and both sensitivity and frequency smoothing bands at their default setting of 6). [Figure 6-8] Two variants were tested:

   a. The denoised audio was amplified in Python (gain factor of 1.5, using NumPy library), with amplitude values clipped to the range [−1, 1] to prevent signal overflow, and then resampled to 16 kHz mono.

   b. The audio was converted to 16 kHz mono, followed by loudness normalization (as in method III).

V. Amplitude gain applied directly in Python (NumPy), followed by resampling to 16 kHz mono.

VI. Noise reduction using the *Noisereduce* Python library, with non-stationary noise removal enabled, followed by peak normalization (target peak=0.99). [224]

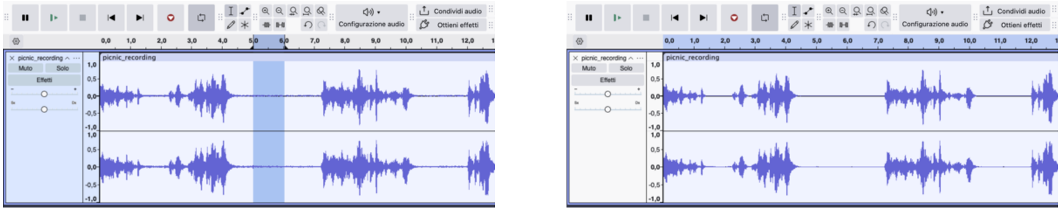


Figure 6-8 Noise reduction using Audacity: a 1- s silent segment is selected to estimate the noise profile. On the right, the cleaned audio track after noise reduction.

6.6.2 Word error rate

The Word Error Rate (WER) was used to compare the automatic transcriptions with the manually produced ones. For the subsample of 40 participants (see AIM 1 in section 5), the WER was first computed across all seven preprocessing and transcription pipelines, in order to identify which method achieved the best transcription accuracy.

To compute the WER, the CHA files produced by the automatic transcription were converted into plain text (TXT) format, containing only the transcribed utterances without timestamps or pause markers. All files, automatic and manual, were cleaned using the same procedure (removal of punctuation, capitalization, and extra spaces) to ensure that the comparison reflected transcription accuracy rather than formatting differences.

For each of the seven pipelines, the mean and standard deviation of WER were calculated separately for the 20 PPA participants and the 20 H. Within the HC group, the Shapiro–Wilk test was applied to assess the normality of the WER distributions. Pairwise comparisons across the seven preprocessing methods were then performed in MATLAB using the Wilcoxon signed-rank test to evaluate whether WER significantly differed between pipelines and to identify the best-performing method. Bonferroni correction was applied to adjust for multiple comparisons.
After selecting the optimal preprocessing method, the final comparison between automatic and manual transcriptions was carried out using the pipeline that achieved the highest accuracy.

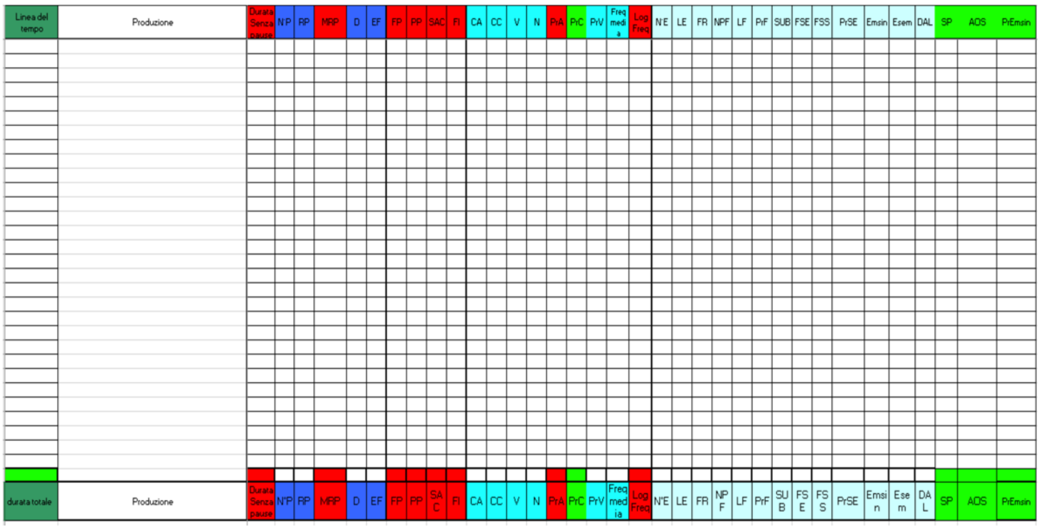
6.7 AIM 2: Multimodal features for classifying PPA variants

Once transcription was completed, linguistic, speech rate or fluency features were extracted from the text and timing data.

6.7.1 Manual feature extraction

From the manual transcription, a quantitative analysis of syntactic, semantic and lexical content was performed, following the reference *Manuale Analisi Eloquio*, a guide developed by neuropsychologists and speech therapists at San Raffaele Hospital. The Excel sheet continued after the first two columns with additional ones dedicated to the extracted features, which can be grouped in the following domains: speech rate and speech sound errors (e.g., total audio duration or speech production rate), features regarding other fluency disruptions, linguistic features related to lexical content and features describing syntactic structure and complexity.

Features related to lexical access difficulties, such as anomia, were excluded from the analysis.



Linea del tempo	Produzione	Durata Senza pause	N°P	RP	MRP	D	EF	FP	PP	SAC	FI	CA	CC	V	N	PrA	PrC	PrV	Freq media	Log Freq	N°E	LE	FR	NPF	LF	PrF	SUB	FSE	FSS	PrSE	Emsin	Esem	DAL	SP	AOS	PrEmsin
durata totale	Produzione	Durata Senza pause	N°P	RP	MRP	D	EF	FP	PP	SAC	FI	CA	CC	V	N	PrA	PrC	PrV	Freq media	Log Freq	N°E	LE	FR	NPF	LF	PrF	SUB	FSE	FSS	PrSE	Emsin	Esem	DAL	SP	AOS	PrEmsin

Figure 6-9 Excel sheet organization for transcription and feature extraction.

To extract the features, a preliminary cleaning phase was necessary. All non-descriptive expressions, such as words directed to the examiner, interjections and meta-linguistic comments, were identified in the transcribed text and excluded from subsequent analysis. Regarding meta-linguistic comments, if their duration was shorter than one second, the corresponding word duration was included in the timing data, although the word itself was excluded from the total word count. These procedures were done by expert speech therapists, who then manually extracted the following features:

- Total duration of the sample: calculated in seconds, from the beginning of the description to the end of the recording. Non-descriptive expression times were excluded from this total
- Duration of pauses: obtained as the sum of unnatural pauses between utterances, both silent and filled.
- Duration of speech without pauses: total duration minus the cumulative duration of pauses
- Total number of words: number of pronounced words, identified by orthographic boundaries. Non-descriptive comments were not considered. Both filled pauses and false starts were excluded. Expressions derived from the verb *esserci* were counted as one word, as were reflexive verb forms (e.g., *si vede*). Articles or elided forms orthographically connected to the following word (e.g., *quell'*) were counted separately. Auxiliary verbs were counted as single words.
- Speech production rate: total number of words divided by total audio duration
- Maximum rhythm of lexical production: total number of words divided by the duration of speech excluding pauses

- Phonotactic distortions: articulatory errors, involving distorted phonemes
- Phonologic errors: omissions, repetitions, transposition or substitution of phonemes, including neologisms. If phonemes were prolonged at the end or within a word as a hesitation strategy, they were considered filled pauses rather than phonological errors
- False starts: words that were only partially pronounced and then abandoned
- Filled pauses: prolonged phonemes occurring at the end or within words. If longer than one second, they were excluded from the time count; if shorter and inseparable from the word, the were included in the duration but also counted as filled pauses
- Repaired or self-corrected sequences: one or more completed words followed by repetitions or reformulations of the same word
- Incomplete sequences: sentences interrupted after the production of the subject or the verb
- Open-class words: nouns, verbs, qualifying adjectives and derived adverbs. Auxiliary verbs *essere* and *avere* were excluded from this class, as well as *stare* when used as a gerund auxiliar
- Closed-class words: prepositions, conjunctions, articles, pronouns, non-derived adverbs and the verbs *essere* and *avere*, both when used as auxiliaries and as main verbs
- Number of nouns: total count of nouns
- Number of verbs: total count of verbs, excluding auxiliaries that belong to the open class. *Essere* and *avere* were counted only when not used as auxiliaries
- Open class proportion: total number of open-class words divided by the number of closed-class words
- Closed class proportion: total number of closed-class words divided by the number of open-class words
- Verb proportion: total number of verbs divided by the sum of verbs and nouns
- Mean frequency of nouns: computed using the lemma frequencies from LIP corpus. Repeated nouns were not considered; words absent from the corpus were assigned a frequency of zero
- Logarithmic frequency of nouns: base-10 logarithm of the mean noun frequency
- Number of utterances: utterances were identified as explained in Section 6.5.1; pauses and meta-linguistic comments were excluded from the count
- Mean utterance length: number of words divided by the number of utterances
- Number of sentences: a sentence was defined as a syntactic structure including a subject (which could be implied) and a predicate (verb plus complement). The complement could be missing if not required by the verb. Complex sentences with subordinate clauses were counted as one sentence,

with the number of subordinates recorded separately. Coordinate clauses were counted as separate sentences
- Number of words in sentences: total number of words, where self-corrected sequences were excluded
- Mean sentence length: number of words in sentences divided by the number of sentences
- Proportion of sentences: number of sentences divided by the number of utterances
- Number of embeddings: number of subordinate clauses
- Sentences with explicit subject: number of sentences where the subject was phonetically realized by a noun or personal pronoun
- Sentences with implicit subject: count of sentences with no phonetic realization of the subject
- Proportion of sentences with explicit subject: number of explicit-subject sentences divided by the total number of explicit and implicit-subject sentences
- Morphosyntactic errors: omissions of articles or prepositions when required, errors in gender or number agreement, incorrect word order, misuse of transitive versus intransitive verbs, or errors in verb inflection
- Proportion of morphosyntactic errors: number of morphosyntactic errors divided by the total words in sentences
- Semantic errors: semantic paraphasias or inappropriate lexical choices
- Syntactic production rate: number of words in sentences divided by total number of words

6.7.2 Automatic feature extraction

The automatic feature extraction aimed to replicate the manual analysis while reducing computational time. In addition, new features were extracted to extend the linguistic and temporal characterization of the speech samples. Not all manually derived features were reproduced, as the automatic procedure was optimized for features that could be consistently and objectively computed from the transcription output. All three files format (JSON, CHA and TXT) were used during feature extraction, depending on whether temporal, pause-related or textual information was required in the process.

Automatically computed features corresponding to the manual ones included:

- Total duration of the recording
- Duration of pauses, subdivided into the duration of silent and filled pauses
- Duration of speech without pauses (both filled and silent ones)
- Total number of words

- Speech production rate
- Maximum speech production rate
- False starts, identified as repetitions of at least three characters in the root of the following word
- Number of filled pauses
- Number of silent pauses
- Open-class words, identified through POS tagging performed by SpaCy Python library (*it_core_news_lg* model)
- Closed-class words (SpaCy POS tagging)
- Number of nouns (SpaCy POS tagging)
- Number of verbs (SpaCy POS tagging)
- Open class proportion
- Verb proportion
- Mean frequency of nouns, computed using the same lexical corpus adopted in the manual analysis
- Logarithmic frequency of nouns
- Number of utterances
- Mean utterance length: calculated as the total number of words divided by the number of utterances
- Number of embeddings, identified as the number of subordinate conjunctions in the text, based on SpaCy POS tagging
- Number of coordinate clauses, corresponding to the number of coordinating conjunctions detected in the text (SpaCy POS tagging)

New features were extracted:

- Number of adjectives (SpaCy POS tagging)
- Number of determiners, including articles, determiner adjectives and pronouns (SpaCy POS tagging)
- Number of auxiliaries (SpaCy POS tagging)
- Number of adverbs (SpaCy POS tagging)
- Number of locative adverbs (SpaCy POS tagging)
- Number of derived adverbs (SpaCy POS tagging)
- Number of prepositions (SpaCy POS tagging)
- Number of pronouns (SpaCy POS tagging)
- Number of stuttering events, identified as the repetition of up to two characters at the beginning of the following word
- Number of proper nouns (SpaCy POS tagging)
- Number of uncertainty expressions (such as *boh* or *ehm*)
- Number of unrecognized words, computed as the number of transcribed words that do not belong to the Italian dictionary

- Total duration of filled pauses, distinguishing between those longer and shorter than one second
- Duration of speech excluding silent pauses
- Type-token ratio, computed using the formula described in section 4.6.2.1
- Brunet’s index, computed using the formula described in section 4.6.2.1
- Honoré’s R, computed using the formula described in section 4.6.2.1
- Percentage of high frequency verbs: computed as the ratio of high-frequency verbs to the total number of verbs
- ICU score: a list of ICU words related to the Picnic picture description task was compiled by expert speech therapists at San Raffaele Hospital [Figure 6-10], following the example of [187]. The number of ICU words appearing in each transcription was then counted.
- Idea density: each word was converted into a vector representation using Spacy’s NLP model. Following [184], the cosine distance between each word vector and both its preceding and following words was computed using the SciPy Python module. The mean of these distances was used as a representative measure of semantic similarity.
- Number of meta-linguistic words: each audio track was processed in two versions: one with the examiner’s voice removed (standard) and one with both examiner speech and meta-linguistic comments removed using Reaper. Both audio versions underwent the same preprocessing pipeline and transcription procedure. The difference in total word count between the two transcriptions was computed.

“*Abitazione, acqua, agnello, albero, alberi, amico, amici, animale, animaletto, aquilone, asse, asta, auto, automobile, bambino, banchina, bandiera, barca, barchetta, battello, bevanda, bevande, bibita, bibite, bicchiere, caffè, cagnolino, campagna, cane, canna, casa, case, casetta, cestino, cesto, cestello, ciabatte, infradito, ciotola, collina, colline, colore, compagnia, coperta, coppia, croce, disegno, donna, famiglia, famigliola, femmina, foglia, foglie, foglio, foglietto, foto, frutta, frutto, garage, gatta, gatto, gioco, laghetto, lago, legno, libro, macchina, mano, mare, marito, maschio, moglie, molo, montagna, monte, moto, musica, nave, navigante, occhiali, paletta, palo, pastore, padrone, pecora, persona, pescatore, pianta, piantina, picnic, ponte, pontile, posto, prato, radio, radiolina, ragazza, ragazzino, ragazzo, riva, sabbia, sandali, scarpa, scarpe, scena, secchiello, signora, signore, signori, spiaggia, tavolino, tavolo, tovaglia, telo, tronco, uomo, uomini, vela, vicinanza, villa, vino, vivanda, vivande, weekend, zattera.”*

Figure 6-10 List of ICU words for the Picnic picture description task, compiled by speech therapists at San Raffaele Hospital.

6.7.2.1 Utterance classifier

To segment the transcription text in a way consistent with speech therapists’ practice, the *Manuale Analisi Eloquio* guidelines for utterance boundary identification were followed, which rely on prosodic cues and pause duration between words. To automatically detect these boundaries, a classifier was developed to label each word as either the final word of an utterance or a non-final word (i.e., occurring at the beginning or within an utterance).

A subsample of 55 subjects was used: 42 participants (21 HC and 21 PPA, balanced by variant) were included in the training and 5-fold cross-validation set, while 13 (9 PPA balanced by variant and 4 HC) formed the test set.

For these subjects, the audio segments corresponding to each transcribed word timestamp were processed to extract temporal features and acoustic features using the Python library *Parselmouth*, yielding numerical descriptors of prosody. The extracted features included:

- duration of the pause between the current word and the next one
- duration of the current word
- pitch values (mean and standard deviation within the window)
- pitch metrics excluding zero-valued frames, which in *Parselmouth* may indicate frames were pitch-estimation failed due to silence, signal instability or background noise (indicated as mean, standard deviation, minimum and maximum within the window)
- Energy of the signal (mean, standard deviation and maximum)
- Pitch slope (mean, standard deviation, maximum and minimum)

Min-max scaling was performed separately for male and female participants to account for physiological differences in vocal frequency not related to word position within the utterance.

The class ratio in the construction set (non-final vs. final words) was balanced to 2:1 by proportionally removing non-final words from each participant. Several combinations of feature selection methods and classifiers (RF, SVM, kNN and LR) were tested. A grid search exploring different number of selected features and various classifier hyperparameters was performed using cross-validation, and the best configuration was selected to maximize the average balanced accuracy across folds.

The final model was trained on the entire construction set and applied to all participants. Utterances were then reconstructed by initiating a new segment right after each word classified as final.

6.7.2.2 High- and low-frequency verbs

To define the distinction between high- and low-frequency Italian verbs, verb frequency data from an Italian corpus commonly used by speech therapists were employed. Before the analysis, all verbs with a frequency lower than 4 were removed to reduce dataset dimensionality. Following [187], the frequencies of *avere* and *essere* were set to equal values. K-means clustering was then applied with $k = 2$ to identify the optimal separation between the two frequency groups. After determining the division point, verbs were sorted by frequency and assigned to the corresponding high- or low-frequency cluster.

6.7.3 Feature selection

An initial feature selection was performed to reduce redundancy and ensure data quality. Features that were null across the entire dataset were analyzed and removed. Then, cross-correlations were computed on the HC linguistic features. Features with Pearson's r values above 0.97 were reviewed by the research team, and one feature from each highly correlated pair was removed based on the relevance to the clinical and linguistic framework.

6.7.4 Multimodal approach for diagnosis

After the feature selection, three approaches were explored to build an automatic diagnostic classifier for distinguishing among the three PPA variants (svPPA, nfvPPA and lvPPA)

- using only linguistic features
- using only imaging features
- a multimodal approach combining both feature types.

The population size differed across the approaches: for the linguistic classifier, 71 PPA patients with the Picnic picture description task recording were included; the imaging approach included all 85 patients with available imaging data; the multimodal approach included 63 PPA patients who had both linguistic and imaging data.

Each dataset was processed in MATLAB as follows:

- A stratified 80-20 % split was performed to create the construction set and the test set. The construction set was normalized via standardization and the resulting parameters were applied to the test set. A 5-fold cross-validation was then performed on the construction set, splitting it into training and validation folds. The model was trained five times, once per fold, and the balanced accuracy was computed on the validation fold. Hyperparameter tuning was carried out at each iteration, and the optimal values were selected as those achieving the highest mean balanced accuracy across the five validation folds. These values were then fixed for subsequent analyses.
- A nested cross-validation scheme was then applied to obtain robust performance estimates. This consisted of an outer 5-fold split between

construction and test sets, and an inner 5-fold split between training and validation sets. This procedure yielded a 25-fold mean balanced accuracy on the validation sets, and a 5-fold mean balanced accuracy and macro F1-score (the mean F1-score across classes) on the test sets.

For methodological consistency, cross-validation splits were kept identical across all classifiers within each dataset by fixing the random seed, ensuring a fair comparison of model performance.

6.7.4.1 Classifiers

Four classifier families were evaluated using the pipeline described in the previous section:

- Logistic regression classifier (one-vs-rest) with L1 regularization. The strength of the L1 penalty and therefore the amount of feature selection performed by the model were controlled by tuning the λ parameter over the following values: (0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5).
- Random forest classifier, tuning both the number of trees (50, 150, 250, 350) and the number of selected features. Features were ranked using RF feature importance (*TreeBagger*, based on the decrease in impurity), and subsets ranging from 1 feature up to the full set were tested.
- Support vector machines classifier with a linear kernel (one-vs- rest). Sequential backward feature selection was applied using an external 5-fold cross-validation. After selecting the feature subset, the regularization parameter C was tuned over (0.01, 0.1, 1, 10, 100).
- kNN classifier, tuning the number of neighbors (3, 5, 7, 9) and the number of selected features. Features were ranked using the *ReliefF* algorithm, and all possible subset sizes (from 1 to the full set) were evaluated.

All these classifiers were evaluated using separately:

a) linguistic features manually extracted from the speech therapists' transcriptions
b) linguistic features automatically extracted using automatic transcriptions
c) features derived from the TBSS analysis
d) multimodal features combining automatic linguistic features and imaging features

6.8 AIM 3: comparative analysis of manual vs automatic processing

To evaluate the performance of the automatic system and identify its strengths and weaknesses, it was compared against the manual system, which served as the ground truth.

6.8.1 Transcription accuracy and pauses detection

The comparison was carried out in terms of transcription accuracy and pause identification ability in a subset of 20 HC and 20 PPA. In particular, transcription performance of the automatic system was evaluated by computing the WER against the manual reference. Statistical analyses were performed to identify significant differences among preprocessing methods within the same group, and between groups using the same method.

Word count error was assessed using the Mean Absolute Error (MAE), computed as the absolute difference between the number of words identified in the automatic and manual annotations.

In addition, two metrics were computed separately for filled pauses and silent pauses:

- Precision, quantifying FP (segments incorrectly identified as filled or silent pauses)
- Recall, quantifying FN (actual filled or silent pauses that were not correctly identified as such by the automatic method)

For this analysis, manual annotations were considered the ground truth for filled-pause identification, whereas silent pauses were verified by a speech therapist and used as the ground truth.

6.8.2 Comparison of linguistic feature extraction

For 20 healthy controls, the same group used for the WER analysis, feature extraction performed manually and automatically was compared while keeping the text constant (manual transcription). The following linguistic features were analyzed: number of words, production rate, number of open-class words, number of closed-class words, number of verbs, number of nouns, proportion of closed-class words, proportion of verbs, and mean noun frequency.

Agreement between the two extraction methods was assessed using the Bland–Altman procedure implemented in MATLAB. [226] This method evaluates the concordance between paired measurements by inspecting the distribution of their differences and determining whether these fall within the expected limits of agreement, thereby identifying the presence of any systematic bias between methods.

In addition, the relative error percentage between manual and automatic measures was computed to quantify the deviation of the automatic extraction from the manual reference.

6.8.3 Development of a web application

To overcome the main limitations of the manual approach, such as its time-consuming nature and limited temporal precision, and to mitigate the weaknesses of the fully automatic system, particularly transcription errors and difficulties in detecting filled pauses, a hybrid approach was developed in the form of a web application designed for use by neuropsychologists and speech therapists.
The application was developed in Python, with integrated HTML components for the user interface, and implemented using the *Streamlit* framework, which enables interactive execution of Python scripts through a web-based environment.

Given an audio input, the application preprocesses the file using the same audio preprocessing pipeline as the automatic transcription system and applies identical Microsoft Azure Speech to Text parameters. The audio is uploaded to Azure Blob Storage and transcribed asynchronously through the Azure Speech API.
The resulting transcription is displayed within the app, where users can review and correct the text directly. All edits can be visualized in revision mode, allowing the operator to easily track modifications throughout the analysis.

The interface includes an audio player to listen to the recording, with navigation controls to move forward and backward through the signal. Additional dedicated tools allow users to play the intervals between words and label them as silent pauses, filled pauses, or non-pausal segments. The system automatically updates the count and cumulative duration for each pause category.
Once the transcription is finalized and validated, the linguistic features are automatically extracted from the corrected text and saved for further analysis.

6.8.4 Testing the WEB-APP on unseen patients

To test the web application, 10 patients recording from the original dataset of 81 audio tracks were used. These recordings were not included in any of the previous analyses, ensuring that they represented unseen data for both the web-app system and the operators. The three PPA variants were equally represented within this subset, comprising 3 nfvPPA, 3 svPPA and 4 lvPPA cases.

6.8.4.1 Fully automatic processing vs automatic processing with operator revision

To quantify the amount of revision required to align the transcription with the patient's actual speech, the WER was computed between the automatically generated transcription and the operator-corrected version for the 10 patients, performed by a speech therapist. The mean WER across patients was then calculated to estimate the overall level of manual correction needed.

6.8.4.2 Inter-operator variability rate

Two speech therapists were asked to use the web application to perform the full analysis on the 10-patient sample. This procedure was designed to assess inter-operator variability, in order to evaluate the consistency and reproducibility of the hybrid system when used by different. To quantify the agreement between the two transcriptions, the percentage of agreement was computed between the two operators' transcripts. The texts were first normalized (lowercasing, removal of punctuation and non-alphanumeric symbols) and tokenized in Python. The two token sequences were then aligned word by word using a gap-based alignment procedure, allowing insertions and deletions to be matched to empty positions. The percentage of agreement was defined as the number of aligned positions in which the two tokens were identical (excluding gaps) divided by the total number of aligned positions (including gaps).

7 Results

7.1 Participants characteristics

Sociodemographic and clinical characteristics of HC and PPA patients stratified according to the diagnosis are summarized in Table 7-1 and Table 7-2. HC and all groups of PPA were matched for age, sex and education as shown in Table 7-1.

Group		*HC*	*PPA*
Sex	*M*	43	39
	F	48	55
Age		67.47±5.45 (56.25–81.41)	68.88±8.46 (42.06–83.93)
Education		11.80±4.17 (5–22)	11.04±4.70 (3–22)
Disease duration		-	36.71±19.11 (3.02–126.59)

Table 7-1 Demographic characteristics of each group, reported as mean, standard deviation, minimum, and maximum values. Disease duration is expressed in months. Age and education were compared using the Kruskal–Wallis test, and sex differences were assessed with the chi-square test. Bonferroni correction was applied. No significant group differences were found (p > 0.05). HC = healthy controls; PPA = primary progressive aphasia.

When comparing each PPA variant with the HC, significant differences emerged in age for lvPPA group and in education for the nfvPPA. Regarding comparisons among the PPA variants, additional age significant differences were found between lvPPA and svPPA. Regarding education nfvPPA patients differed significantly from HC. Moreover, svPPA patients differed significantly from nfvPPA in both education and disease duration.

Group	*HC*	*nfvPPA*	*svPPA*	*lvPPA*
M	43	14	16	9
F	48	24	20	11
Age	67.47±5.45 (56.25–81.41)	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	11.80±4.17 (5–22)	9.11±4.83 * (3–22)	12.42±4.42 † (5–18)	12.25±3.77 (5–17)
Disease duration	-	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 7-2 Demographic characteristics for each group are reported as mean, standard deviation, minimum, and maximum values. Disease duration is expressed in months. Statistical analysis was performed using the chi-square test for sex differences and the Kruskal–Wallis test with subsequent multcompare *for age, education, and disease duration. Bonferroni correction was applied. Significant differences were found (p<0.05): * indicates a significant difference from the HC group; † indicates a significant difference from the nfvPPA group; ‡ indicates a significant difference from the svPPA group. HC=healthy controls, nfvPPA=non-fluent variant Primary Progressive Aphasia, svPPA=semantic variant Primary Progressive Aphasia, lvPPA=logoepnic variant Primary Progressive Aphasia*

Mean age, years of education and months of disease duration were calculated for each group, along with the standard deviation and range. Missing values were imputed using the median value of the corresponding group. Overall, women were more numerous than men in all conditions.

7.2 TBSS-GLM findings

In

Figure 7-1, voxels showing a significant difference in FA (p<0.05) for each contrast after TBSS analysis are displayed. The significance is color-coded accorded to the p-value, with red representing the highest significance.

- **nfvPPA vs others**
  Compared with HC, nfvPPA patients exhibited reduced FA in a wide range of fibers: inferior fronto-occipital fasciculus and superior longitudinal

fasciculus (temporal part) bilaterally, cingulate gyrus and uncinate fasciculus bilaterally, forceps minor, anterior thalamic radiation on the left and cingulum on the right. Relative to lvPPA, they showed decreased FA in the body of corpus callosum, forceps minor and left cingulate gyrus. Compared with svPPA, significant alterations were observed in the inferior fronto-occipital fasciculus bilaterally, in the forceps minor and the body of the corpus callosum, and in the left hemisphere in the superior longitudinal fasciculus (temporal part), inferior longitudinal fasciculus, anterior thalamic radiation and uncinate fasciculus.

- **svPPA vs others**
  Compared with HC, svPPA patients showed a pattern of decreased FA in the inferior longitudinal fasciculus and inferior fronto-occipital fasciculus bilaterally, in the body of the corpus callosum, anterior thalamic radiation and uncinate fasciculus bilaterally, forceps minor and in the left hemisphere in the cingulate gyrus, superior longitudinal fasciculus and its temporal part.
- **lvPPA vs others**
  Comparing lvPPA patients to HC, FA maps revealed a reduction affecting the inferior fronto-occipital fasciculus and superior longitudinal fasciculus (temporal part) bilaterally, the anterior thalamic radiation and uncinate fasciculus bilaterally and the forceps minor.

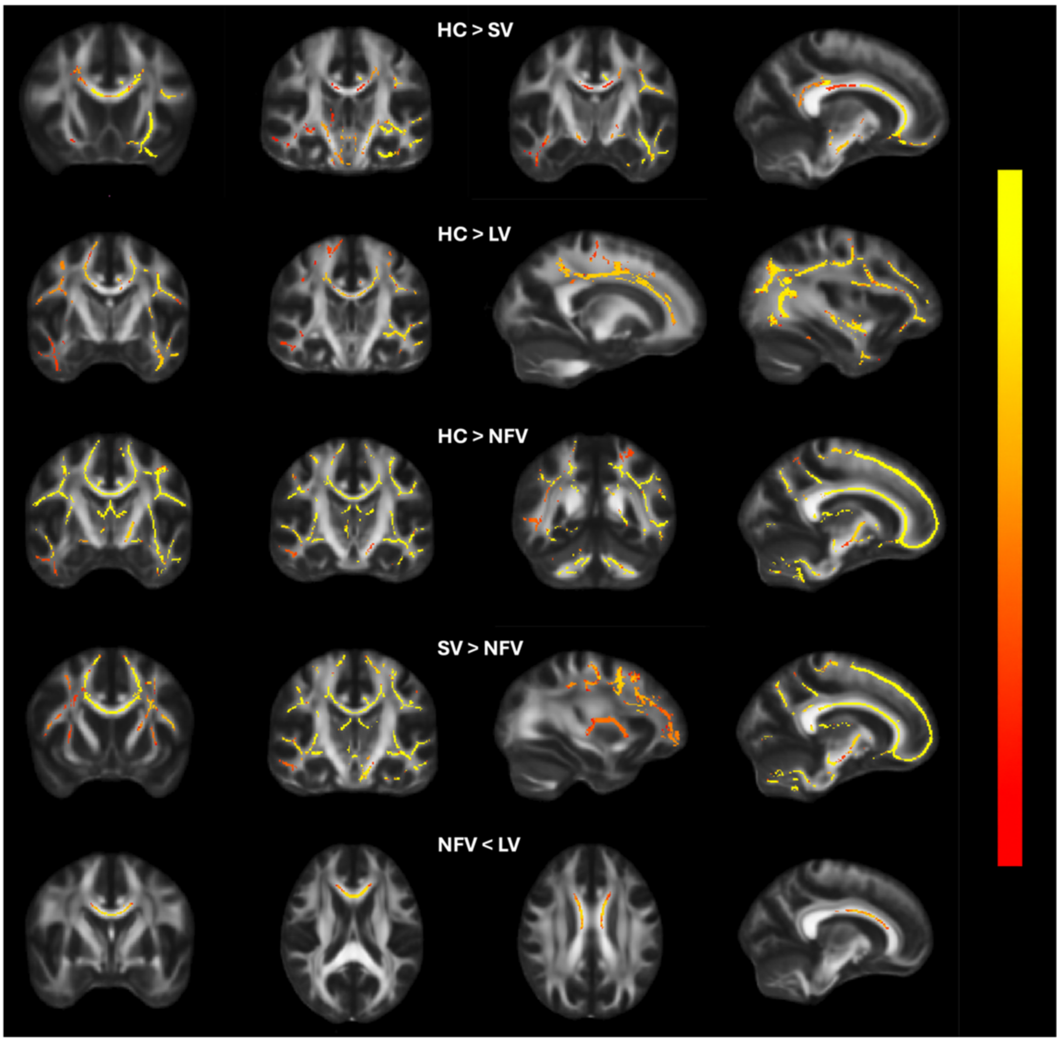


Figure 7-1 Significant FA differences between groups identified (p<0.05, TFCE-corrected). For all contrasts, coronal and sagittal views are shown; for the NFV < LV contrast, an axial view is also reported. HC = healthy controls; SV = semantic variant; LV = logopenic variant; NFV = non-fluent/agrammatic variant.

7.2.1 Feature extraction

From the significant TBSS contrasts, thirteen ROIs showing white-matter alterations were identified. These ROIs corresponded to the following anatomical regions:

- Left cingulate gyrus
- Right cingulate gyrus
- Corpus callosum
- Left frontal lobe
- Right frontal lobe

- Left insula
- Right insula
- Left occipital lobe
- Right occipital lobe
- Left parietal lobe
- Right parietal lobe
- Left temporal lobe
- Right temporal lobe

7.2.2 Feature selection

Cross-correlation analysis revealed no Pearson coefficients exceeding the 0.97 threshold; therefore, all thirteen features were retained for the subsequent analyses.

7.3 Audio preprocessing

To identify the most effective preprocessing pipeline, the WER values obtained on a sample of 20 HC from the seven approaches were compared.

The only method that showed statistically significant differences compared with all the others was the *noisereduce* approach, which performed significantly worse and was therefore excluded. [Figure 7-2] Among the remaining methods, none of which showed significant differences in terms of WER, the final choice was based on the lowest mean WER value, together with considerations of automation and processing speed. Pipelines requiring manual intervention (as those involving Audacity) were discarded.

Both the Method II and III pipelines were selected and used in all subsequent analyses.

METHOD	***Description***	***WER (%)***
I	Raw	14.4±13.0 (4.6–30.2)
II	16 kHz, mono	**11.8±6.5** (3.6–30.2)
III	Loudness normalization	**11.6±6.5** (4.5–19.0)
IV a	Audacity amplification	12.2±6.7 (4.5–31.0)
IV b	Audacity loudness normalization	12.0±6.7 (4.5–19.0)
V	amplification	11.9±6.9 (3.6–29.8)
VI	noisereduce	22.1±10.5* (8.2–50.0)

*Table 7-3 Mean WER values (± standard deviation), minimum and maximum for each method after outlier removal. Statistical analysis was performed using the Wilcoxon test with Bonferroni correction applied. * indicates a significant difference from all the other methods (p < 0.05). WER = word error rate.*

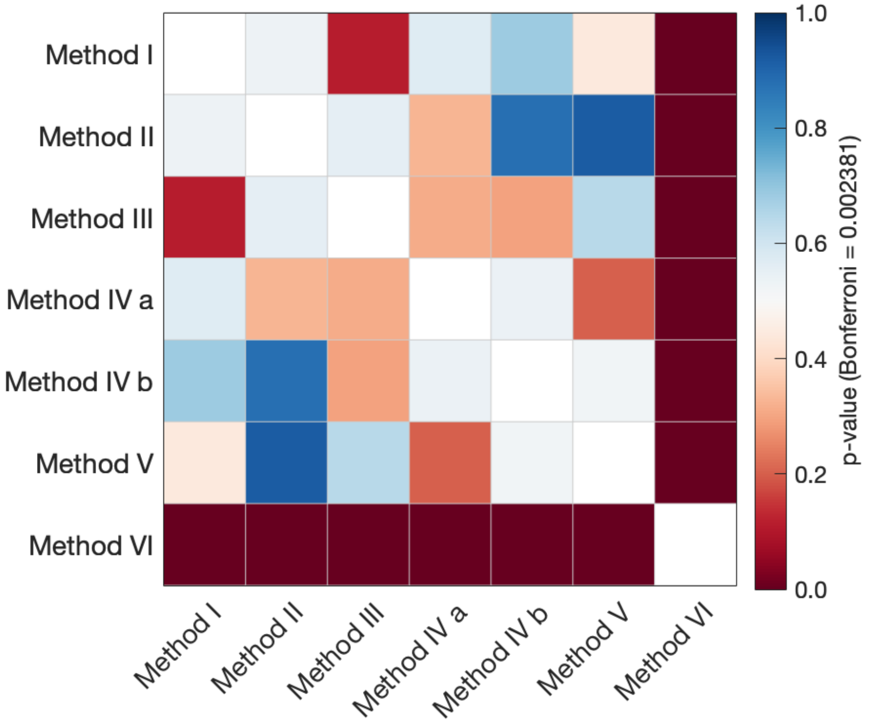


Figure 7-2 Colormap of significant pairwise comparisons between preprocessing methods. Bonferroni correction was applied. A significant difference was found between Method VI and all other methods (p < 0.05).

7.4 Automatic feature extraction

7.4.1 Utterances classifier

The performance of the best classifiers for utterance boundary identification is reported in Table 7-4. These correspond to the best-performing models obtained after hyperparameter tuning. Both kNN and RF achieved at least 80% balanced accuracy; however, the random forest classifier reached the highest value, with a balanced accuracy of 82.3% on the validation set.

The features selected by the model included: the duration of the pause following each word, the duration of the word itself, the standard deviation of pitch, and the mean pitch slope. These findings align with the guidelines of the *Manuale Analisi Eloquio*, which emphasize the importance of both temporal cues and pitch-related information in identifying utterance boundaries.

Classifier	***LR***	***RF***	***SVM***	***kNN***
Number of selected features	12	4	1	2
BA on validation set (%)	74.5	**82.3**	79.1	80.9

Table 7-4 Balanced accuracy averaged across cross-validation folds and the corresponding number of selected features after hyperparameter tuning for each classifier. LR = logistic regression; RF = random forest; SVM = support vector machines; kNN = k-nearest neighbors.

The final model was then evaluated on the test set in terms of balanced accuracy and two practical error metrics designed to assess the impact on the feature extraction pipeline: the absolute and relative error in the utterance count, and the absolute and relative error in the mean number of words per utterance.

Final model		RF
BA on test set (%)		72.5
Utterance count	***Absolute error***	15.8±15.0
	Relative error (%)	33.6±23.4
Mean word count per utterance	***Absolute error***	1.92±1.72
	Relative error (%)	51.6±50.4

Table 7-5 Final model performance on the test set, including balanced accuracy, the mean relative error in utterance count (± standard deviation), and the mean relative error in the mean number of words per utterance (± standard deviation). BA = balanced accuracy; RF = random forest.

7.4.2 High- and low-frequency verbs

The k-means algorithm applied to the list of Italian verbs sorted by frequency identified an optimal threshold separating high-frequency from low-frequency verbs. This procedure resulted in eleven verbs being classified as high-frequency: *andare*, *avere*, *dire*, *dovere*, *essere*, *fare*, *potere*, *sapere*, *stare*, *vedere* and *volere* [Figure 7-3]. These verbs were then used to compute the feature "percentage of high-frequency verbs".

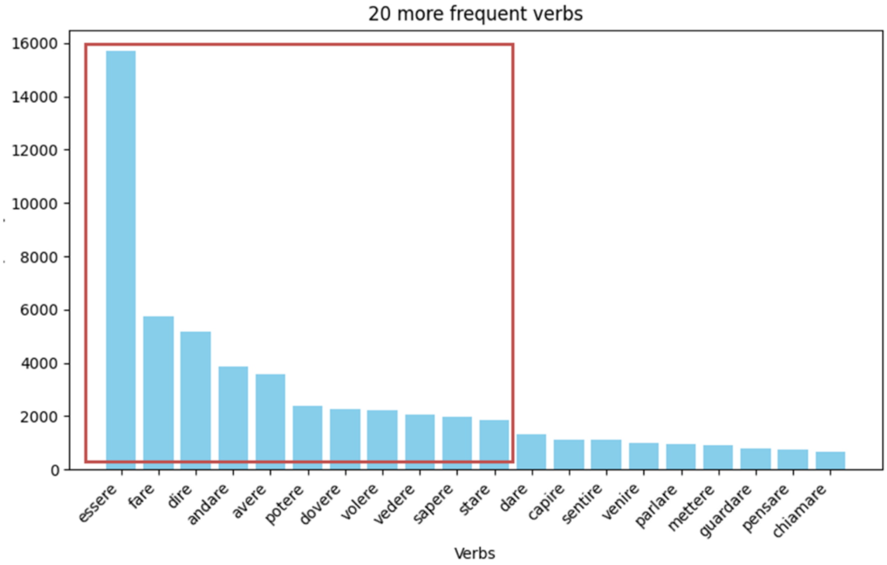


Figure 7-3 High-frequency verb cluster obtained from K-means. The figure shows the 20 most frequent verbs; the high-frequency cluster is highlighted in red.

7.5 Feature selection

Features that did not show at least one occurrence in the entire dataset were removed; these included *false starts* and *derived adverbs*.

Cross-correlation analysis identified two pairs of features exceeding the Pearson correlation coefficient threshold of 0.97, as reported in Table 7-6.

After evaluating these redundancies, logarithmic noun frequency and the number of closed-class words were removed, as noun frequency was already sufficiently representative, and the proportion between closed- and open-class words already accounted for the closed-class category.

Feature	*Feature*	*Pearson's coefficient (r)*
Noun frequency	Logarithmic noun frequency	0.988
Closed class words	Number of words	0.987

Table 7-6 Feature pairs showing Pearson's correlation coefficients above 0.97, with their corresponding r *values.*

7.6 Classification results

The performance of the multiclass classification models was evaluated using the mean balanced accuracy across the validation folds.

7.6.1 Classifiers based on TBSS-derived features

The performance of the classifiers. Trained using only imaging features is reported in Table 7-7. LR and SVM outperformed the other models, both achieving a balanced accuracy of approximately 75% across validation sets. In contrast, kNN showed clearly inferior performance, reaching only 58% of balanced accuracy.

Imaging classifier			
Classifier	*Hyperparameter*	*Number of selected features*	*BA on validation sets (%)*
LR	$\lambda = 0.005$	13	75.4±2.1
kNN	k=7	9	58.0±4.4
RF	K=250	8	62.1±3.2
SVM	C=1	11	**75.5±3.6**

Table 7-7 Results of the inner cross-validation for classifiers trained on imaging-derived features, expressed as mean balanced accuracy across validation sets (± standard deviation). BA = balanced accuracy; LR = logistic regression; kNN = k-nearest neighbors; RF = random forest; SVM = support vector machines.

7.6.2 Classifiers using linguistic features

The impact of different linguistic feature sets on classification performance, evaluated as balanced accuracy, was examined. Three different feature sets were compared:

- manually extracted features from the speech therapists' transcriptions [Table 7-8]
- automatically extracted features from transcription method II [Table 7-9]
- automatically extracted features from transcription method III [Table 7-10]

Linguistic classifier			
Classifier	***Hyperparameter***	***Number of selected features***	***BA on validation sets (%)***
LR	$\lambda = 0.01$	21	76.7±6.5
kNN	k=5	23	72.3±1.8
RF	K=50	15	77.1±2.9
SVM	C=0.1	13	**77.3±3.0**

Table 7-8 Results of classifiers trained on manually extracted linguistic features. Performance is expressed as mean balanced accuracy across validation sets (± standard deviation). BA = balanced accuracy; LR = logistic regression; kNN = k-nearest neighbors; RF = random forest; SVM = support vector machines.

Method II linguistic classifier			
Classifier	***Hyperparameter***	***Number of selected features***	***BA on validation sets (%)***
LR	$\lambda = 0.05$	15	78.0±3.4
kNN	k=3	26	74.3±2.6
RF	K=50	33	76.3±4.8
SVM	C=0.1	25	**81.4±7.6**

Table 7-9 Results of the inner cross-validation for linguistic classifiers trained on features extracted using transcription method II. Results are expressed as mean balanced accuracy across validation sets (± standard deviation). BA = balanced accuracy; LR = logistic regression; kNN = k-nearest neighbors; RF = random forest; SVM = support vector machines.

Method III linguistic classifier			
Classifier	***Hyperparameter***	***Number of selected features***	***BA on validation sets (%)***
LR	$\lambda = 0.05$	16	**78.0±4.4**
kNN	k=9	3	66.4±4.6
RF	K=50	30	75.4±2.7
SVM	C=1	13	74.6±4.2

Table 7-10 Results of the inner cross-validation for linguistic classifiers trained on features extracted using transcription method III. For each model, the table reports the tuned hyperparameters, the number of selected features, and the mean balanced accuracy across validation folds (± standard deviation). BA = balanced accuracy; LR = logistic regression; kNN = k-nearest neighbors; RF = random forest; SVM = support vector machines.

The best-performing classifiers were SVM and LR with L1 regularization. The best classifier using manually extracted features achieved a lower balanced accuracy than the best models based on automatically extracted features (both method II and III). Method II, in particular, stood out by surpassing the 80% balanced accuracy threshold.

7.6.3 Multimodal classifiers

Two feature sets were used as multimodal inputs:

- features combining linguistic method II and imaging-derived information
- features combining linguistic method III and imaging-derived information

Multimodal classification confirmed the superior performance of LR and SVM for this task, yielding a mean balanced accuracy of 81% for method II [Table 7-11] and 84% for method III [Table 7-12]. In both cases, the lowest mean balanced accuracy did not fall below 72%.

Method II linguistic and imaging classifier			
Classifier	***Hyperparameter***	***Number of selected features***	***BA on validation sets (%)***
LR	$\lambda = 0.0001$	29	**81.4±3.1**
kNN	k=7	41	72.5±3.5
RF	K=350	21	75.4±4.4
SVM	C=1	17	72.9±5.3

Table 7-11 Inner cross-validation results for classifiers trained on combined imaging and linguistic features obtained using transcription method II. For each model, the table reports the tuned hyperparameters, the number of selected features, and the mean balanced accuracy across validation folds (± standard deviation). BA = balanced accuracy; LR = logistic regression; kNN = k-nearest neighbors; RF = random forest; SVM = support vector machines.

Method III linguistic and imaging classifier			
Classifier	***Hyperparameter***	***Number of selected features***	***BA on validation sets (%)***
LR	$\lambda = 0.0001$	32	81.1±2.5
kNN	k=7	23	73.5±5.3
RF	K=250	46	75.1±1.3
SVM	C=1	12	**83.6±6.9**

Table 7-12 Inner cross-validation results for classifiers trained on combined imaging and linguistic features obtained using transcription method III. For each model, the table reports the tuned hyperparameters, the number of selected features, and the mean balanced accuracy across validation folds (± standard deviation). BA = balanced accuracy; LR = logistic regression; kNN = k-nearest neighbors; RF = random forest; SVM = support vector machines.

7.7 Results on the test set

The models identified as best-performing in the inner validation were then evaluated on the outer cross-validation test sets and assessed in terms of mean balanced accuracy ± standard deviation and mean macro F1-score ± standard deviation. All classifiers exceeded 80% mean balanced accuracy, except for the classifier based on manual features and the imaging-only classifier. Multimodal models showed the strongest performance, achieving values above 85% in both balanced accuracy and macro F1-score. [Table 7-13]

Comparison between the best classifiers				
Features type	***Classifier***	***Number of selected features***	***BA on test sets (%)***	***Macro F1 on test sets (%)***
Manual	SVM	13	76.5±10.4	77.2±10.4
linguistic method II	SVM	25	81.0±9.7	81.7±10.2
linguistic method III	LR	16	80.2±3.4	81.3±2.8
imaging	SVM	11	74.1±7.4	74.0±8.9
Multimodal method II	LR	29	**87.0±6.9**	**86.7±6.7**
Multimodal method III	SVM	12	85.1±11.5	85.1±11.2

Table 7-13 Best-performing classifier for each feature set, with outer cross-validation performance expressed as mean balanced accuracy and macro F1 on the test sets (± standard deviation). BA = balanced accuracy.

The best overall model was the LR multimodal classifier using the multimodal feature set from method II. It achieved a balanced accuracy of 87% and showed a lower standard deviation compared with the multimodal model based on method III. An aggregated and normalized confusion matrix across the five test sets is presented below. [Figure 7-4] The true class with the highest number of misclassifications was lvPPA. In contrast, 93% of nfvPPA cases and 95% of svPPA cases were correctly classified. Among the misclassified instances, nfvPPA and lvPPA errors were evenly split between the other two classes, whereas all misclassified svPPA cases were assigned to lvPPA.

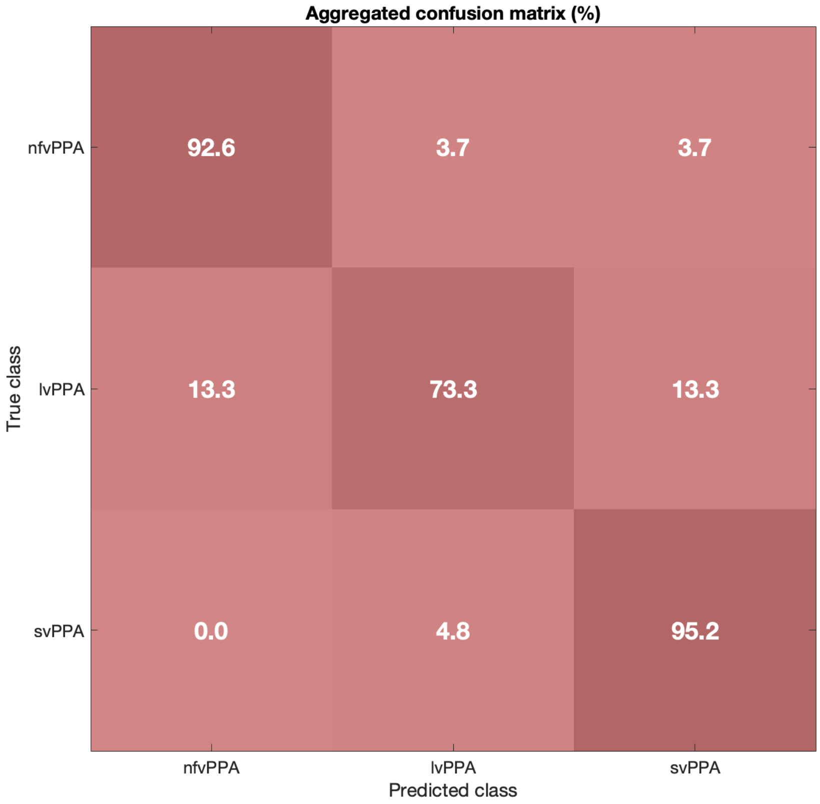


Figure 7-4 Confusion matrix of the best-performing model aggregated across test sets.

This LR model selected a total of 29 features, including 11 imaging-derived features (out of 13) and 18 linguistic features. Feature importance was quantified using the absolute value of the L1-regularized coefficients of the multinomial LR model, which provides a direct, model-intrinsic measure of each feature’s global relevance. Based on this criterion, the selected features were ranked by importance, as shown in Figure 7-5.

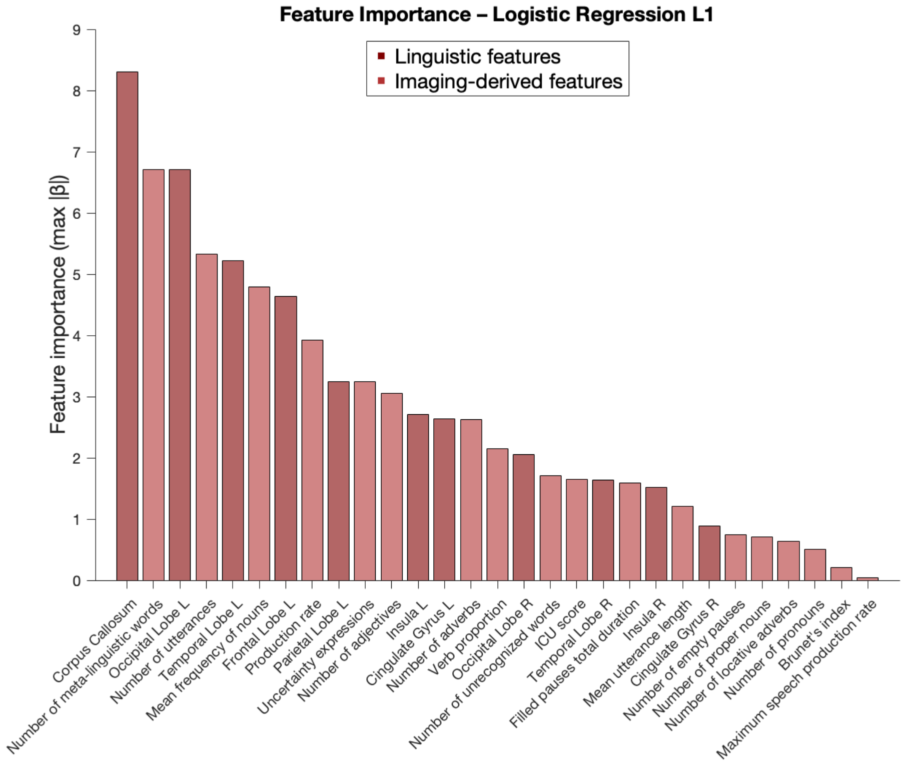


Figure 7-5 Selected features ordered by global importance based on L1-regularized coefficients.

7.7.1 Shapley Additive Explanations

To complement the global ranking obtained from L1 coefficients and facilitate interpretability of the best performing model, SHapley Additive exPlanations (SHAP) values were computed using the same trained model. The multiclass SHAP summary plot provides a class-specific view of feature impact, revealing feature–class associations that are not captured by coefficient-based importance alone. [Figure 7-6] In addition, separate violin plots were generated for each diagnostic group (nfvPPA, lvPPA, svPPA), illustrating the distribution of SHAP values across subjects and the directionality of each feature's effect. [Figure 7-7][Figure 7-8][Figure 7-9]

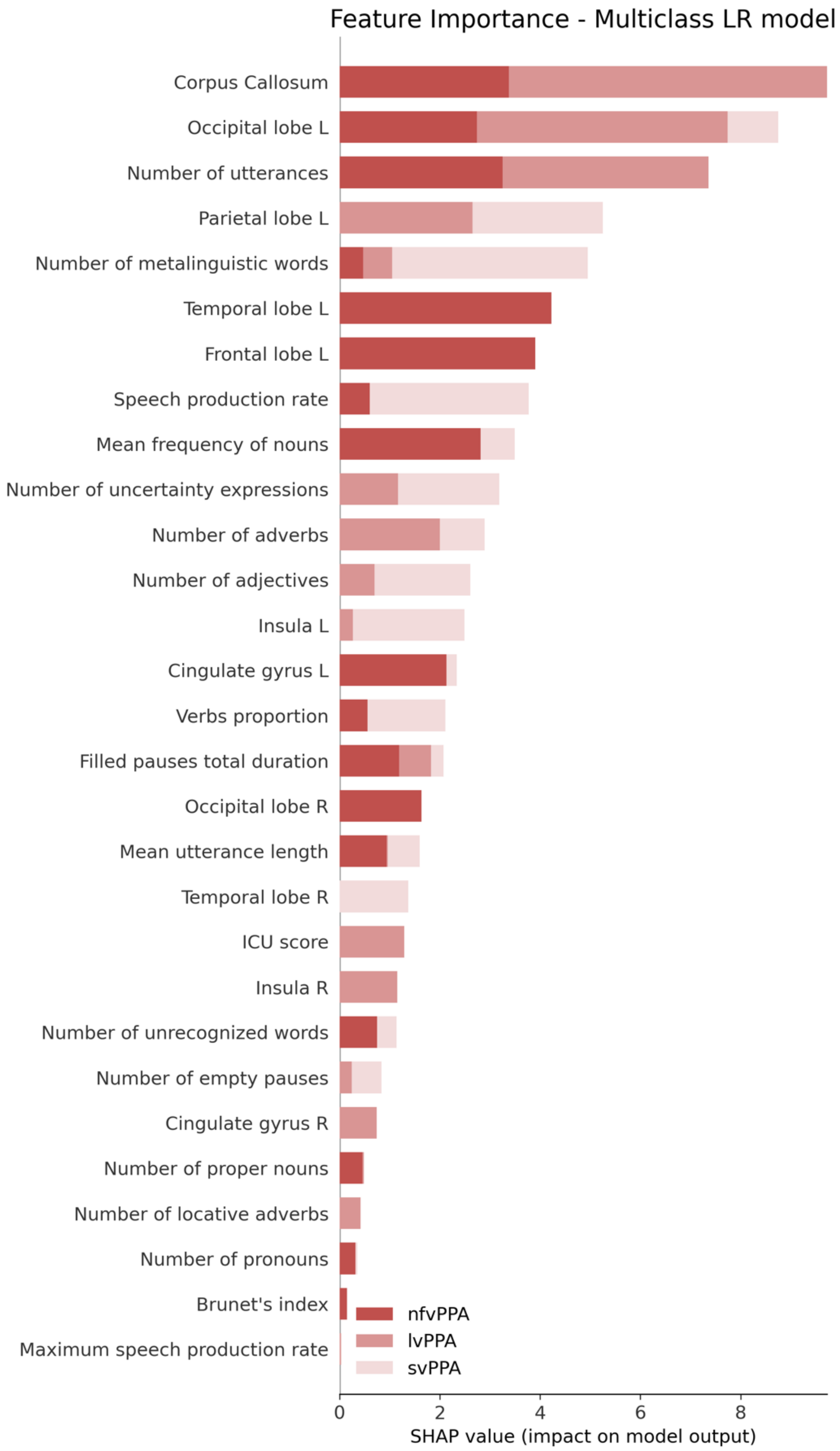


Figure 7-6 Multiclass SHAP feature-importance summary plot. LR = logistic regression; SHAP = Shapley Additive Explanations.

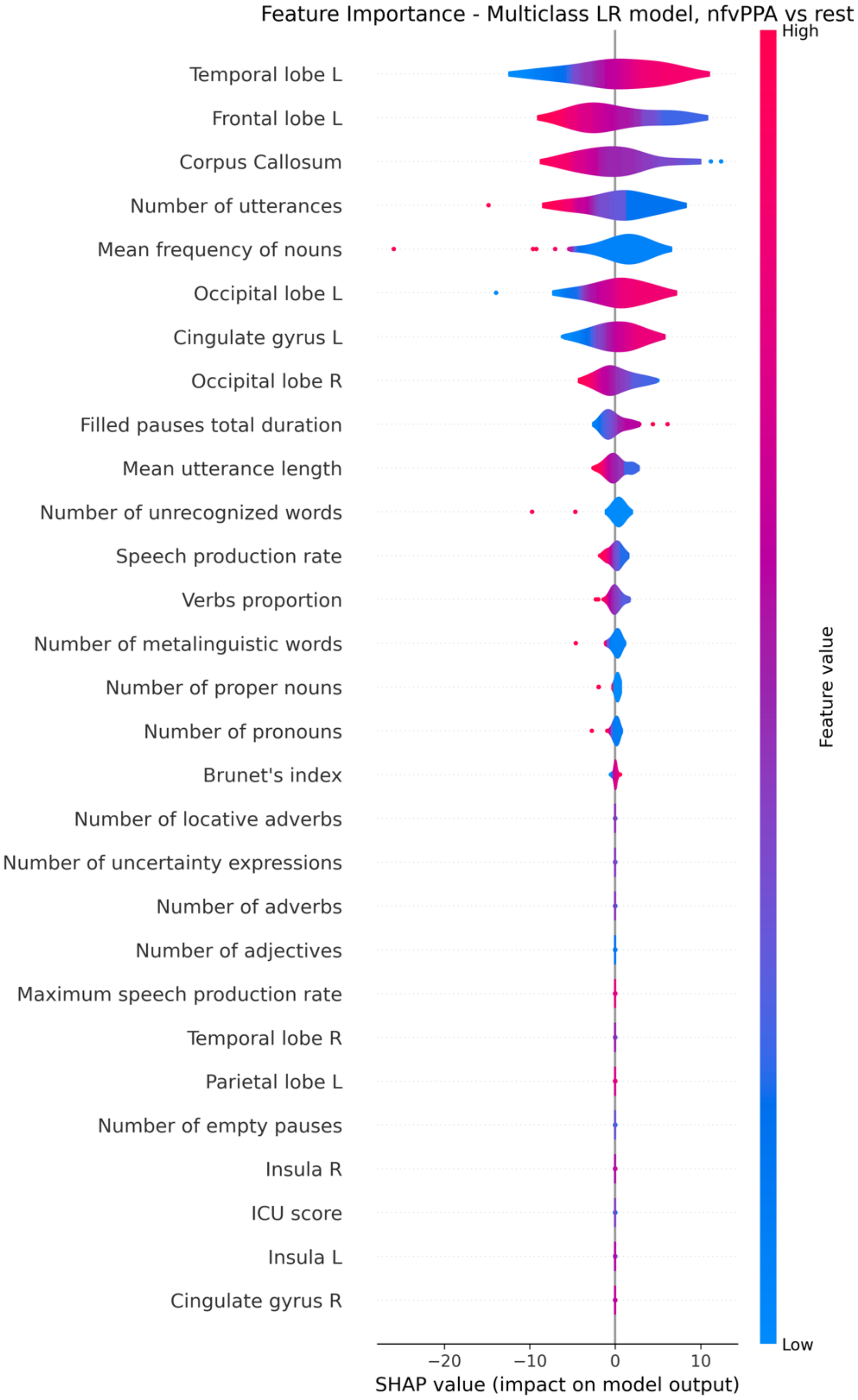


Figure 7-7 SHAP value distribution illustrating the impact of features on the prediction of nfvPPA versus the other classes. LR = logistic regression; SHAP = Shapley Additive Explanations; nfvPPA = non-fluent/agrammatic variant primary progressive aphasia.

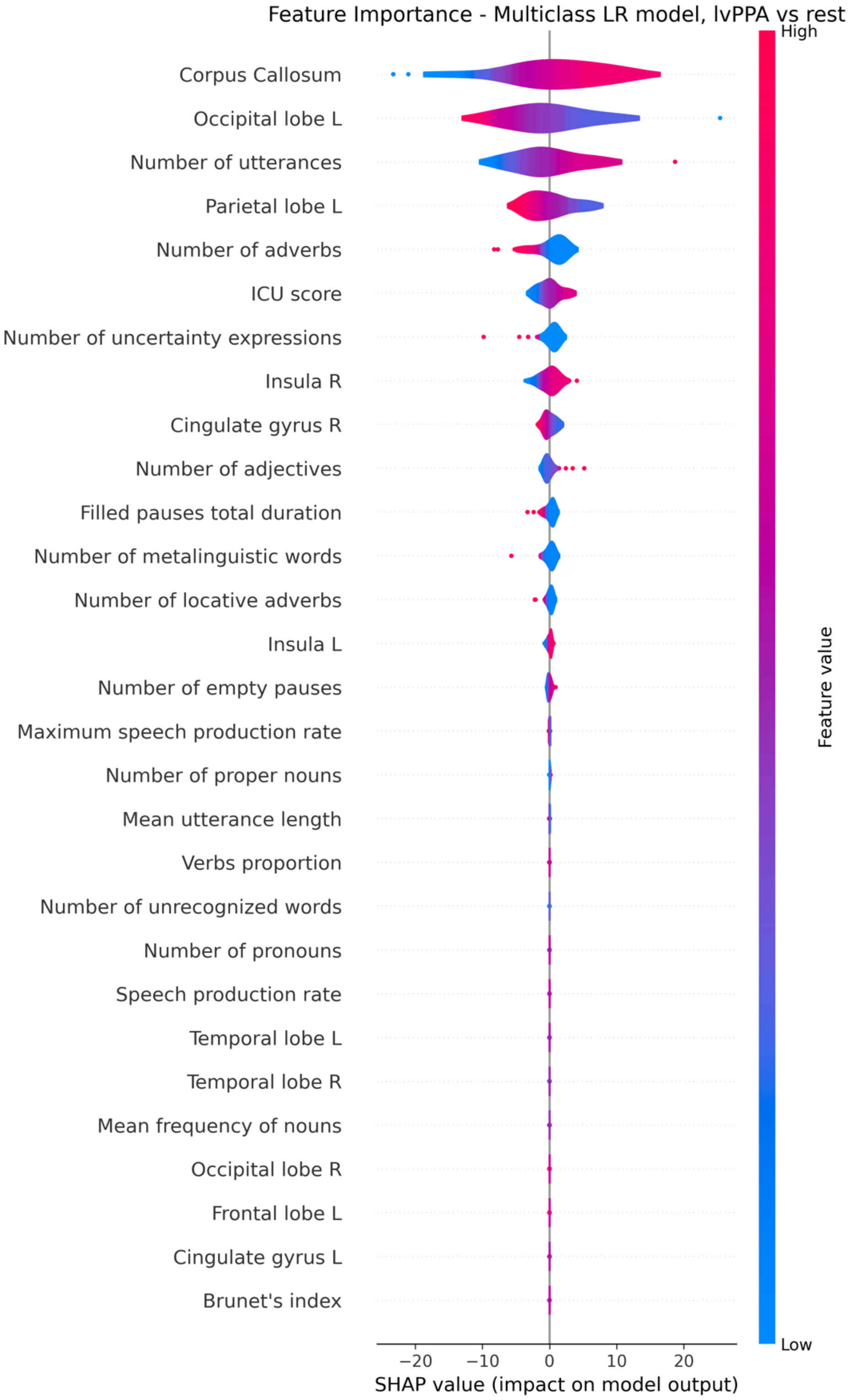


Figure 7-8 SHAP value distribution illustrating the impact of features on the prediction of lvPPA versus the other classes. LR = logistic regression; SHAP = Shapley Additive Explanations; lvPPA = logopenic variant primary progressive aphasia.

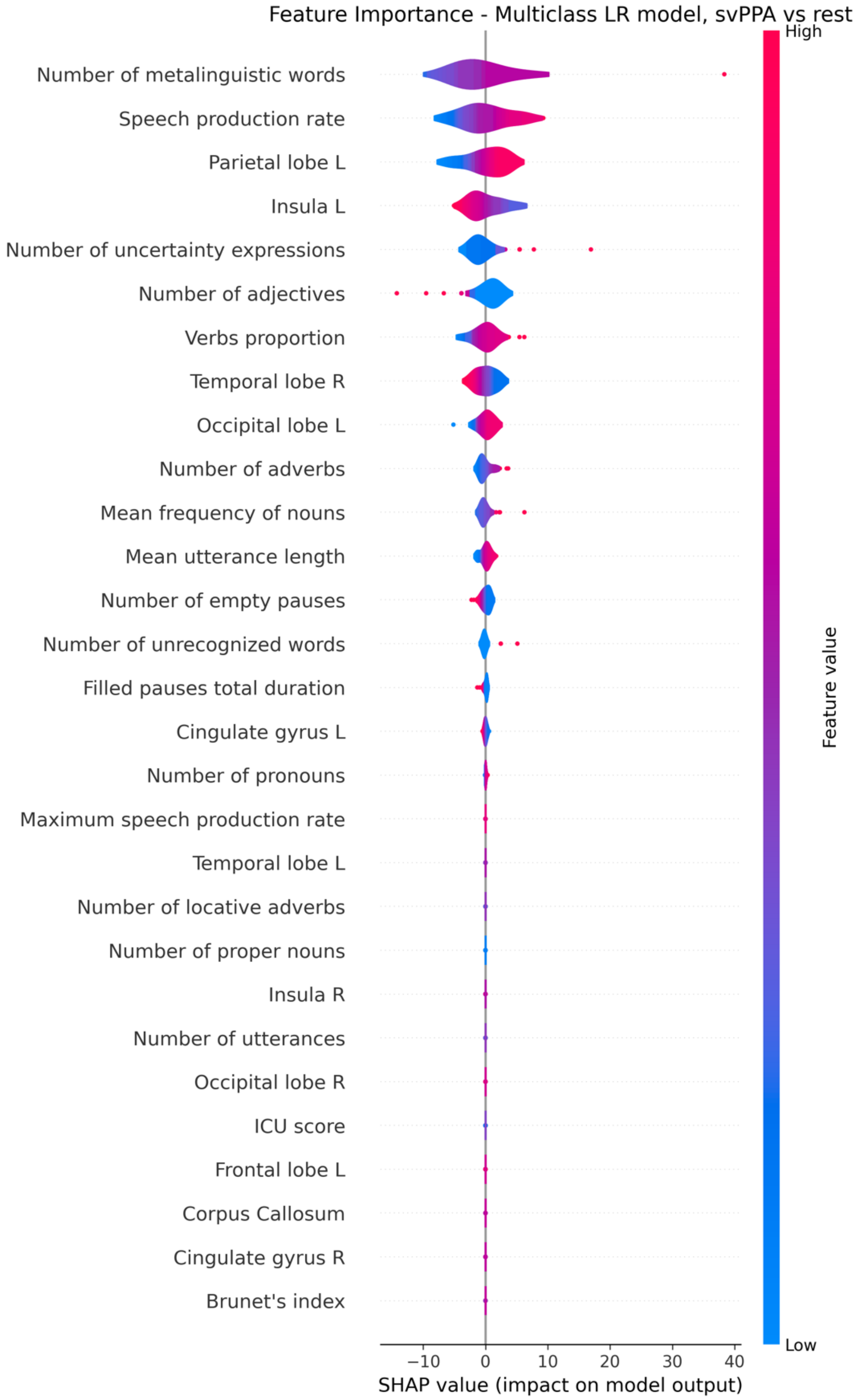


Figure 7-9 SHAP value distribution illustrating the impact of features on the prediction of svPPA versus the other classes. LR = logistic regression; SHAP = Shapley Additive Explanations; svPPA = semantic variant primary progressive aphasia.

7.8 Comparative analysis of manual vs automatic processing

7.8.1 Transcription and pauses

No statistically significant differences were observed between transcription Methods II and III within the same group. Between-group comparisons revealed a significant difference between HC and PPA for Method II. [Table 7-14]

	HC	*PPA*
Method II	11.8±6.5 (3.6–30.2)	19.4±9.5 * (7.4–37.4)
Method III	11.6±6.5 (4.5–30.6)	18.9±10.2 (6.9–42.9)

*Table 7-14 Mean WER and standard deviation for each group and transcription method. Statistical analysis was performed using the paired Wilcoxon test to compare transcription methods II and III within each group; between-group comparisons were conducted using the Wilcoxon rank-sum test in MATLAB. Bonferroni correction was applied. * indicates a significant difference from the HC group (p = 0.0494). HC = healthy controls; PPA = primary progressive aphasia.*

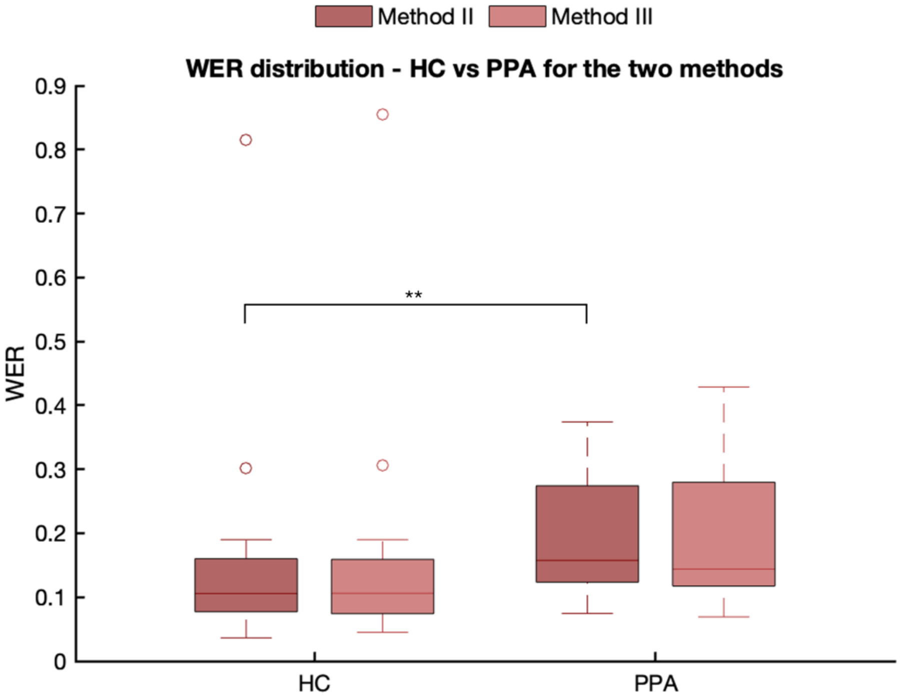


*Figure 7-10 WER distribution across groups for the two transcription methods. Statistically significant differences after Bonferroni correction are indicated by **.*

An additional analysis revealed no statistically significant differences among the PPA variants, nor between the two transcription methods within each variant. [Table 7-15]

	svPPA	*lvPPA*	*nfvPPA*
Method II	18.2±9.0 (10.4–34.4)	17.0±12.0 (7.4–37.4)	22.7±8.1 (14.0–33.8)
Method III	17.7±9.1 (10.2–28.3)	16.3±11.1 (6.9–32.7)	22.3±10.9 (13.0–42.9)

Table 7-15 WER of the three PPA variants, reported as mean ± standard deviation, minimum, and maximum values. Statistical analysis was performed using the Kruskal–Wallis test for between-variant comparisons and the paired Wilcoxon test to compare the two transcription methods. Bonferroni correction was applied. No significant differences were found. svPPA = semantic variant primary progressive aphasia; lvPPA = logopenic variant primary progressive aphasia; nfvPPA = non-fluent/agrammatic variant primary progressive aphasia.

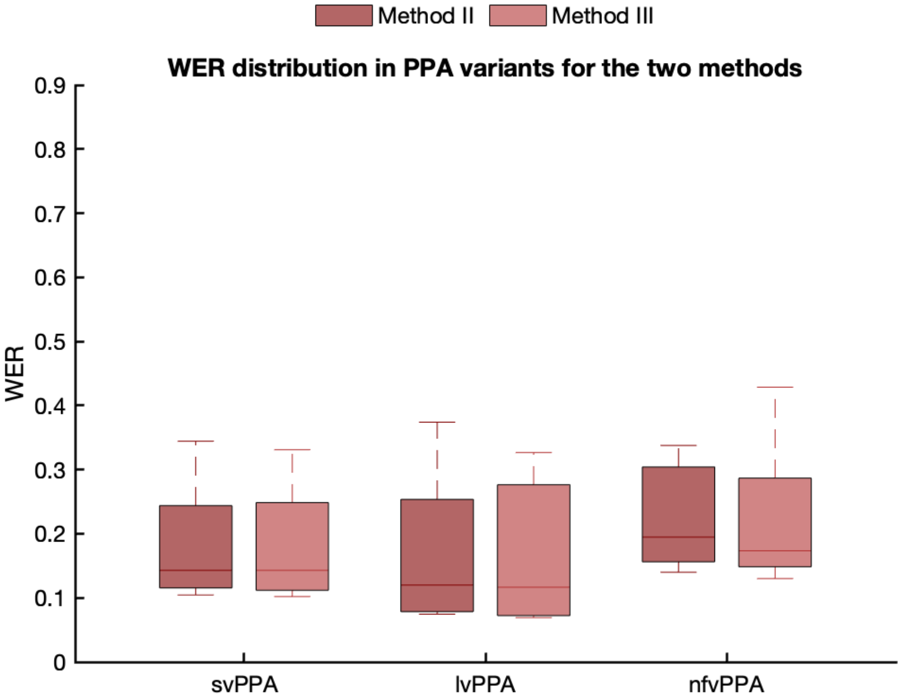


Figure 7-11 Boxplot of WER values across the three PPA variants. Statistical analysis was performed using the Kruskal–Wallis test for between-variant comparisons and the Wilcoxon test for between-method comparisons. No significant differences were found. svPPA = semantic variant primary progressive aphasia; lvPPA = logopenic variant primary progressive aphasia; nfvPPA = non-fluent/agrammatic variant primary progressive aphasia.

The pause-identification analysis further examined the performance of the automatic system relative to manual annotation. The results showed that the automatic system outperforms manual annotation in the detection of silent pauses but performs less effectively in identifying filled pauses. In particular, recall is lower than precision, indicating that missed detections of filled pauses (false negatives) occur more frequently than incorrect detections (false positives).

	Manual approach		***Automatic approach***	
	Precision (%)	*Recall (%)*	*Precision (%)*	*Recall (%)*
Empty-pauses	81.6±28.7	80.2±29.1	94.9±5.6	91.5±4.5
Filled-pauses	100	100	79.2±12.0	60.5±18.8

Table 7-16 Comparison between the automatic and manual approaches for the identification of empty and filled pauses, expressed in terms of precision and recall.

MAE between the count of words was 2.42±1.84 words per subject.

7.8.2 Linguistic feature agreement

The differences between manual and automatic linguistic feature extraction are illustrated in the Bland–Altman plots and in the table below. [Figure 7-12][Table 7-17]
Overall, the automatic system did not show substantial bias: mean relative error values were generally low, reaching a maximum of 26%. The largest discrepancies were observed for the number of verbs, the closed-class proportion, the verb proportion and the mean noun frequency.
The Bland-Altman plots showed a slightly positive bias for the number of verbs and mean noun frequency, indicating a mild overestimation by the automatic system. Conversely, a small negative bias emerged for the closed-class proportion and verb proportion, suggesting slight underestimation.

For the remaining features, the mean relative error was below 10%, indicating a good correspondence between automated and manual extraction. In these cases, the automatic system closely reproduced the manually derived values, with no appreciable systematic bias.

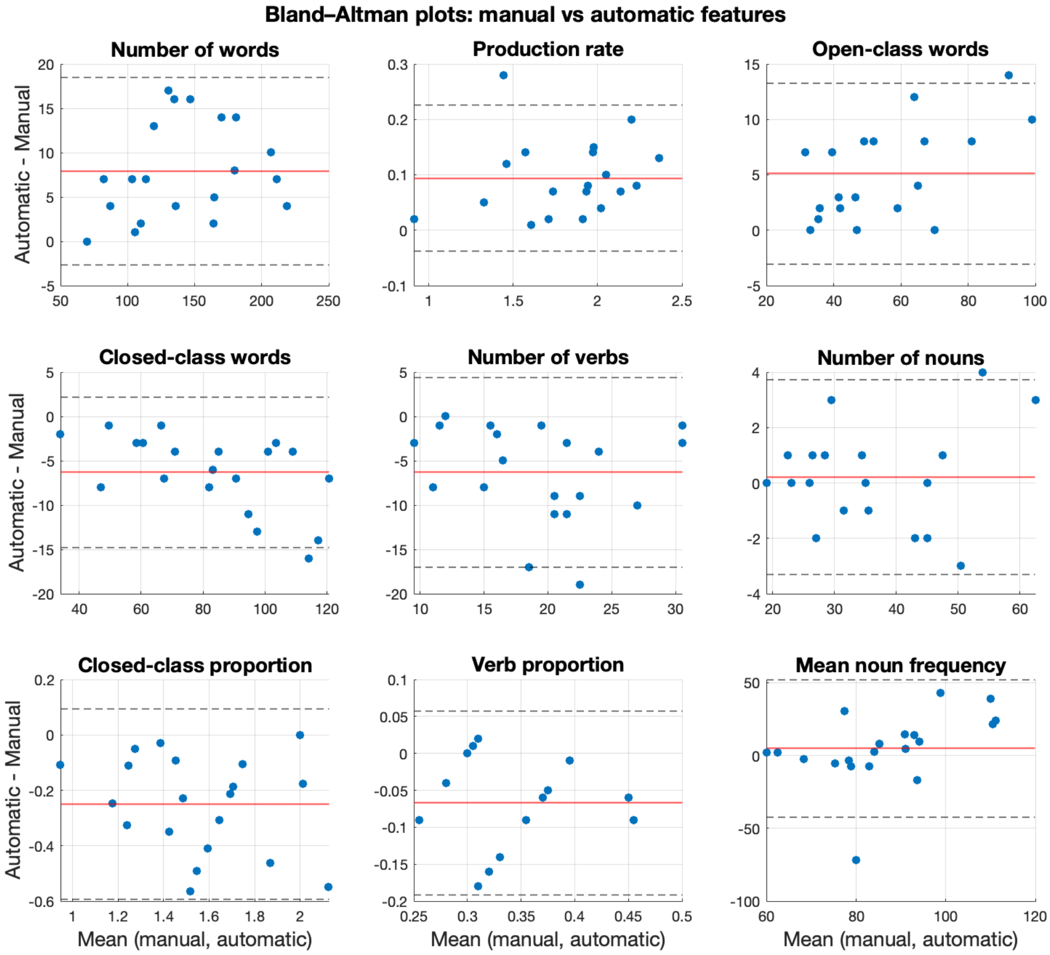


Figure 7-12 Bland–Altman plots comparing manual and automatic extraction for the nine linguistic features considered. Each subplot shows the mean of the two measurements (x-axis) and their difference (y-axis), along with the bias line and the limits of agreement.

Feature	*Mean relative error (%)*
Number of words	5.90
Speech production rate	5.45
Open-class words	9.73
Closed-class words	7.11
Number of verbs	26.37
Number of nouns	3.40
Closed class proportion	14.49
Verb proportion	18.18
Mean noun frequency	18.65

Table 7-17 Mean absolute relative error between automatically and manually extracted features..

7.8.3 WEB-APP

7.8.3.1 Efficiency evaluation

The speech therapists reported that the estimated time required for transcription with automatic feature extraction was substantially shorter than the time needed for the fully manual pipeline. They also found the availability of two transcription methods helpful, as each tended to capture different details that complemented one another.

7.8.3.2 Transcription analysis and Inter-rater variability

An analysis of the WER across ten subjects computed between the automatic transcription and the version corrected through the web application showed a significant difference between the two operators, while no significant differences were observed between the two methods when used by the same operator. Nevertheless, all WER values remained below 20%. [Figure 7-13]

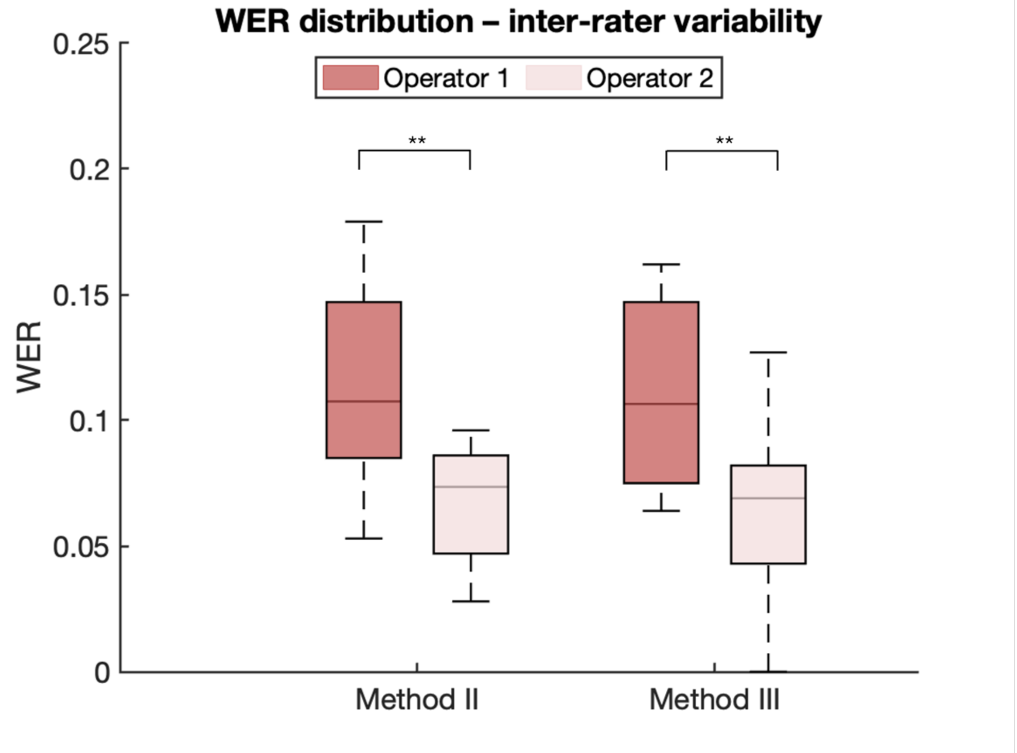


*Figure 7-13 Comparison of the two operators' performance across the two methods. Statistical analysis was performed using the Wilcoxon test. Statistically significant differences after Bonferroni correction are marked with "**". WER = word error rate.*

The percentage of agreement was above 80% for all subjects, with a mean value of 90.8±4.2 % for method II and 90.9±3.6 % for method III, as showed in Figure 7-14.

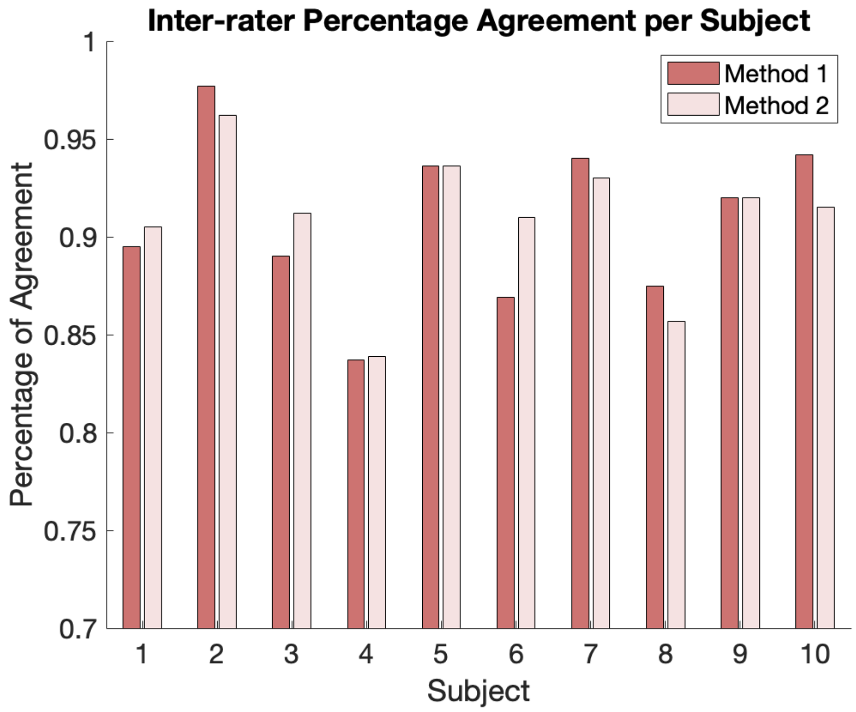


Figure 7-14 Inter-rater percentage of agreement per subject between the two operators for the two methods..

8 Discussion and conclusions

8.1 Discussion

Classification of PPA is clinically important, because accurate differentiation between variants has direct implications for prognosis, patient management and therapeutic planning. [227] Neuropsychological assessment and speech analysis, together with imaging confirmation, are usually used to distinguish the three variants.
However, manual transcription and manual feature extraction are extremely time-consuming, often requiring hours per subject, and strongly depend on the speech therapist's expertise. In contrast, machine-learning-based automation can significantly reduce processing time and increase reproducibility.

This study aimed to develop an automatic system capable of differentiating the three variants and to integrate the pipeline into a web application that combines automation with human expertise.

Diffusion tensor imaging and tract-based spatial statistics have been widely used to study white-matter tract alterations across variants, as each PPA subtype is characterized by damage in specific regions. [228] Recent studies have also focused on automating speech-based analysis, showing that ASR systems struggle to maintain fidelity when transcribing pathological speech, with typically results in higher WER compared with healthy controls. [229] Importantly, combining linguistic and imaging-derived features has been shown to enhance classification performance. [227]

In this work, the two preprocessing pipelines proved interchangeable, as no significant differences were detected. This result may be attributed to the heterogeneity of the audio sources (different recording devices, environments, and noise levels), which likely prevented any single method from adapting optimally to the full range of recordings.

The utterance classifier identified two pitch- and prosody-related features and two duration-related features as the most relevant, in alignment with the *Manuale Analisi Eloquio* guidelines for utterance-boundary identification.

Regarding TBSS analysis, significant group differences were observed. NfvPPA patients showed decreased FA predominantly in left-lateralized and anterior tracts. SvPPA patients, among other alterations, exhibited bilateral alterations in the inferior longitudinal fasciculus and the uncinate fasciculus, as well as damage to superior and inferior temporal white-matter pathways in the left hemisphere, which are likely involved in semantic and lexical processing. FA maps from lvPPA patients revealed bilateral alterations in the uncinate fasciculus and superior longitudinal fasciculus. These findings are consistent with previous research studies. [227][230]

Among the linguistic-only classifiers, those based on automatically extracted features outperformed the models using manual features, with SVM achieving a balanced accuracy of 81.4±7.6% across validation sets.
Imaging-only performance was lower, with a balanced accuracy of 75.5±3.6%, again with SVM yielding the best results. The multimodal classifier integrating both feature types further improved performance, highlighting Logistic Regression with L1 regularization as a suitable model for this task. Achieving 87.0±6.9% balanced accuracy across the test sets and a macro F1 score of 86.7±6.7%, the model demonstrated stable performance across all classes. These results can be considered robust, as they were obtained using nested cross-validation.

As shown by the aggregated confusion matrix, the model exhibited no substantial bias when classifying nfvPPA and lvPPA, as indicated by the equal proportion of false positives. In contrast, a difference emerged in the misclassification pattern of svPPA, with all errors occurring toward the lvPPA class.

A slight bias toward misclassifying lvPPA may be attributable to its lower representation within the dataset. Feature importance derived from the β coefficients indicated that left-hemisphere imaging features contributed more strongly than right-hemisphere ones, while both linguistic and imaging features were represented among the ten most influential predictors.

SHAP analysis further clarified the contribution of each feature to the model's ability to distinguish the three variants. The three most influential features overall (corpus callosum, left occipital lobe and number of utterances) had a strong impact on nfvPPA and lvPPA classifications but had minimal influence on svPPA.
For nfvPPA vs rest, imaging features dominated the explanation profiles, particularly those derived from the left temporal and frontal lobes (high temporal values and low frontal values). The number of utterances emerged as the most relevant linguistic predictor, with a low number of utterances and a low mean utterance length, consistent with previous studies [227].
For lvPPA vs rest, imaging features again appeared highly influential, specifically high corpus callosum values and low left occipital lobe values, followed by the number of utterances, with tended to be high in this group. [227] Interestingly, the ICU score showed a meaningful contribution to the model output: high ICU values tended to shift the classification toward the lvPPA variant.
For svPPA vs rest, linguistic features predominated. A high number of metalinguistic comments and an increased speech-production rate were the most

impactful predictors, followed by elevated values in the left parietal lobe and lower values in the insula. The association between frequent metalinguistic comments and svPPA is consistent with research findings, reporting that this variant tends to ask questions and produce spontaneous comments during discourse. [231]

The manual vs automatic comparison showed a mean WER of approximately 11-12% for HC and 18-19% for PPA, likely reflecting the fact that ASR systems are usually trained of healthy, fluent speech. No significant WER differences emerged across PPA variants, indicating consistent ASR performance across variant types. Regarding pause analysis, the automatic system performed well in detecting silent pauses, outperforming human annotation, thanks to the 100-ns timestamp resolution, which enabled precise pause-boundary identification (95% precision and 92% recall). Silero VAD detected filled pauses less reliably (79% precision and 60% recall), while the human operator was considered the ground truth. This lower performance is likely due to the fact that some vowel prolongations were interpreted by the ASR system as part of a word rather than as a pause. As a consequence, these segments were not given to Silero for classification, reducing the number of filled pauses available for correct identification.

The mean relative error between automatically and manually extracted features indicated good agreement, with a maximum relative error of 26%. Moreover, the automatic system showed consistency across subjects, as in the Bland-Altman plots no more than one fell outside the limits of agreement, confirming that at least 95% of the differences lay within the expected range.

The web application proved to be substantially faster than manual transcription, even when manual correction was included. Post-correction WER relative to the fully automatic transcription remained below 20%, suggesting that only limited adjustments were required. Inter-rater variability was notable, likely due to the two speech therapists having different levels of expertise, with the more experienced operator performing more corrections. Nevertheless, the percentage of agreement between the two operators exceeded 80% for all subjects, highlighting the tool's potential for standardization and efficiency.

8.2 Limitations

The main limitation of this study lies in the discrepancies between automatic and manual transcriptions, as these errors propagate to the subsequent feature extraction step. ASR systems tend to correct mispronounced words, omit disfluencies, and fail to report vowel lengthening. This removes clinically relevant information that could further improve classification performance.

This is mainly attributable to the fact that ASR systems are usually trained on healthy speech. Future work should explore fine-tuning existing ASR model weights or training a custom model, to better capture the characteristics of pathological speech.

To do so, a wide amount of data would be necessary. This links to a broader challenge related to rare neurodegenerative diseases: collecting large amounts of high-quality data is difficult. However, expanding the dataset would allow the model to learn more from the data and improve its ability to generalize.

8.3 Conclusions

Automating the speech-processing pipeline is a promising direction, with the potential to standardize analyses, reduce processing time and support diagnostic workflows, while complementing imaging-based classification. Future work will focus on enhancing ASR fidelity to pathological speech, enabling a more accurate representation of disfluencies and speech errors, and ultimately providing a more complete characterization of PPA speech patterns.

Bibliography

[1] D. C. Tippett, 'Classification of primary progressive aphasia: challenges and complexities', *F1000Research*, vol. 9, p. F1000 Faculty Rev-64, Jan. 2020, doi: 10.12688/f1000research.21184.1.

[2] M.-M. Mesulam, 'Primary progressive aphasia', *Ann. Neurol.*, vol. 49, no. 4, pp. 425–432, Apr. 2001, doi: 10.1002/ana.91.

[3] M.-M. Mesulam, 'Primary Progressive Aphasia: A 25-year Retrospective', *Alzheimer Dis. Assoc. Disord.*, vol. 21, no. 4, p. S8, Dec. 2007, doi: 10.1097/WAD.0b013e31815bf7e1.

[4] M. M. Mesulam and S. Weintraub, 'Spectrum of primary progressive aphasia', *Baillieres Clin. Neurol.*, vol. 1, no. 3, pp. 583–609, Nov. 1992.

[5] M.-M. Mesulam, 'Slowly progressive aphasia without generalized dementia', *Ann. Neurol.*, vol. 11, no. 6, pp. 592–598, 1982, doi: 10.1002/ana.410110607.

[6] A. PICK, 'Uber die Beziehungen der senilen Hirnatrophie zur Aphasie', *Prag Med Wchnschr*, vol. 17, pp. 165–167, 1892.

[7] P. Sérieux, 'Sur un cas de surdite verbale pure', *Rev Med*, vol. 13, pp. 733–750, 1893.

[8] M. L. Gorno-Tempini *et al.*, 'Classification of primary progressive aphasia and its variants', *Neurology*, vol. 76, no. 11, pp. 1006–1014, Mar. 2011, doi: 10.1212/WNL.0b013e31821103e6.

[9] J. R. Hodges, K. Patterson, S. Oxbury, and E. Funnell, 'Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy', *Brain*, vol. 115, no. 6, pp. 1783–1806, 1992.

[10] M. Grossman *et al.*, 'Progressive Nonfluent Aphasia: Language, Cognitive, and PET Measures Contrasted with Probable Alzheimer's Disease', *J. Cogn. Neurosci.*, vol. 8, no. 2, pp. 135–154, Mar. 1996, doi: 10.1162/jocn.1996.8.2.135.

[11] M. Grossman and S. Ash, 'Primary Progressive Aphasia: A Review', *Neurocase*, vol. 10, no. 1, pp. 3–18, Feb. 2004, doi: 10.1080/13554790490960440.

[12] 'Cognition and anatomy in three variants of primary progressive aphasia - Gorno-Tempini - 2004 - Annals of Neurology - Wiley Online Library'. Accessed: Oct. 10, 2025. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ana.10825

[13] M. Montembeault, S. M. Brambati, M. L. Gorno-Tempini, and R. Migliaccio, 'Clinical, Anatomical, and Pathological Features in the Three Variants of Primary Progressive Aphasia: A Review', *Front. Neurol.*, vol. 9, p. 692, Aug. 2018, doi: 10.3389/fneur.2018.00692.

[14] J. R. Hodges *et al.*, ‘Clinicopathological correlates in frontotemporal dementia’, *Ann. Neurol.*, vol. 56, no. 3, pp. 399–406, Sept. 2004, doi: 10.1002/ana.20203.

[15] D. C. Tippett and Z. Keser, ‘Chapter 6 - Clinical and neuroimaging characteristics of primary progressive aphasia’, in *Handbook of Clinical Neurology*, vol. 185, A. E. Hillis and J. Fridriksson, Eds, in Aphasia, vol. 185. , Elsevier, 2022, pp. 81–97. doi: 10.1016/B978-0-12-823384-9.00016-5.

[16] ‘Longitudinal Changes in Cognition, Behaviours, and Functional Abilities in the Three Main Variants of Primary Progressive Aphasia: A Literature Review’. Accessed: Oct. 30, 2025. [Online]. Available: https://www.mdpi.com/2076-3425/11/9/1209

[17] P. Bekkhus-Wetterberg, A. Brækhus, E. G. Müller, M. I. Norvik, I. E. Winsnes, and T. B. Wyller, ‘Primary progressive aphasia’, *Tidsskr. Den Nor. Legeforening*, Nov. 2022, doi: 10.4045/tidsskr.22.0100.

[18] A. Mouton *et al.*, ‘The course of primary progressive aphasia diagnosis: a cross-sectional study’, *Alzheimers Res. Ther.*, vol. 14, no. 1, p. 64, May 2022, doi: 10.1186/s13195-022-01007-6.

[19] T. Kiymaz, M. Z. Khan Suheb, F. Lui, and O. De Jesus, ‘Primary Progressive Aphasia’, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025. Accessed: Oct. 11, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK563145/

[20] M. Pengo *et al.*, ‘Sex influences clinical phenotype in frontotemporal dementia’, *Neurol. Sci.*, vol. 43, no. 9, pp. 5281–5287, Sept. 2022, doi: 10.1007/s10072-022-06185-7.

[21] E. G. Spinelli *et al.*, ‘Typical and atypical pathology in primary progressive aphasia variants’, *Ann. Neurol.*, vol. 81, no. 3, pp. 430–443, 2017, doi: 10.1002/ana.24885.

[22] H. Ulugut and Y. A. L. Pijnenburg, ‘Frontotemporal dementia: Past, present, and future’, *Alzheimers Dement.*, vol. 19, no. 11, pp. 5253–5263, 2023, doi: 10.1002/alz.13363.

[23] M. González-Sánchez *et al.*, ‘TARDBP mutation associated with semantic variant primary progressive aphasia, case report and review of the literature’, *Neurocase*, vol. 24, no. 5–6, pp. 301–305, 2018, doi: 10.1080/13554794.2019.1581225.

[24] M. Mesulam *et al.*, ‘Progranulin mutations in primary progressive aphasia: the PPA1 and PPA3 families’, *Arch. Neurol.*, vol. 64, no. 1, pp. 43–47, Jan. 2007, doi: 10.1001/archneur.64.1.43.

[25] K. Younes and B. L. Miller, ‘Frontotemporal Dementia: Neuropathology, Genetics, Neuroimaging, and Treatments’, *Psychiatr. Clin. North Am.*, vol. 43, no. 2, pp. 331–344, June 2020, doi: 10.1016/j.psc.2020.02.006.

[26] M. A. Santos-Santos *et al.*, ‘Rates of Amyloid Imaging Positivity in Patients With Primary Progressive Aphasia’, *JAMA Neurol.*, vol. 75, no. 3, pp. 342–352, Mar. 2018, doi: 10.1001/jamaneurol.2017.4309.

[27] M. Teichmann, R. Migliaccio, A. Kas, and B. Dubois, ‘Logopenic progressive aphasia beyond Alzheimer’s—an evolution towards dementia with Lewy bodies’, *J. Neurol. Neurosurg. Psychiatry*, vol. 84, no. 1, pp. 113–114, Jan. 2013, doi: 10.1136/jnnp-2012-302638.

[28] F. Conca, V. Esposito, G. Giusto, S. F. Cappa, and E. Catricalà, ‘Characterization of the logopenic variant of Primary Progressive Aphasia: A systematic review and meta-analysis’, *Ageing Res. Rev.*, vol. 82, p. 101760, Dec. 2022, doi: 10.1016/j.arr.2022.101760.

[29] M. Chu *et al.*, ‘Atrophy network mapping of clinical subtypes and main symptoms in frontotemporal dementia’, *Brain*, vol. 147, no. 9, pp. 3048–3058, 2024.

[30] C. R. Marshall *et al.*, ‘Primary progressive aphasia: a clinical approach’, *J. Neurol.*, vol. 265, no. 6, pp. 1474–1490, June 2018, doi: 10.1007/s00415-018-8762-6.

[31] M. P. Broca, ‘REMARQUES SUR LE SIÉGE DE LA FACULTÉ DU LANGAGE ARTICULÉ, SUIVIES D’UNE OBSERVATION D’APHÉMIE (PERTE DE LA PAROLE)’.

[32] C. Wernicke, *Der aphasische Symptomencomplex*. 1874.

[33] ‘Key Difference Between Broca’s and Wernicke’s Aphasia’. Accessed: Oct. 30, 2025. [Online]. Available: https://lonestarneurology.net/blog/brocas-vs-wernickes-aphasia/

[34] J. R. Binder, J. A. Frost, T. A. Hammeke, R. W. Cox, S. M. Rao, and T. Prieto, ‘Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging’, *J. Neurosci.*, vol. 17, no. 1, pp. 353–362, Jan. 1997, doi: 10.1523/JNEUROSCI.17-01-00353.1997.

[35] A. D. Friederici, ‘The Brain Basis of Language Processing: From Structure to Function’, *Physiol. Rev.*, vol. 91, no. 4, pp. 1357–1392, Oct. 2011, doi: 10.1152/physrev.00006.2011.

[36] P. M. Beeson, K. Rising, and J. Volk, ‘Writing treatment for severe aphasia: who benefits?’, *J. Speech Lang. Hear. Res. JSLHR*, vol. 46, no. 5, pp. 1038–1060, Oct. 2003, doi: 10.1044/1092-4388(2003/083).

[37] M. L. Henry, M. V. Meese, S. Truong, M. C. Babiak, B. L. Miller, and M. L. Gorno-Tempini, ‘Treatment for apraxia of speech in nonfluent variant primary progressive aphasia’, *Behav. Neurol.*, vol. 26, no. 1–2, pp. 77–88, 2013, doi: 10.3233/BEN-2012-120260.

[38] D. C. Tippett, A. E. Hillis, and K. Tsapkini, ‘Treatment of Primary Progressive Aphasia’, *Curr. Treat. Options Neurol.*, vol. 17, no. 8, p. 34, June 2015, doi: 10.1007/s11940-015-0362-5.

[39] R. K. Soliman *et al.*, ‘Effects of tDCS on Language Recovery in Post-Stroke Aphasia: A Pilot Study Investigating Clinical Parameters and White Matter Change with Diffusion Imaging’, *Brain Sci.*, vol. 11, no. 10, p. 1277, Oct. 2021, doi: 10.3390/brainsci11101277.

[40] D. A. Decker and K. M. Heilman, ‘Steroid treatment of primary progressive aphasia’, *Arch. Neurol.*, vol. 65, no. 11, pp. 1533–1535, Nov. 2008, doi: 10.1001/archneur.65.11.1533.

[41] W. R. Shankle *et al.*, ‘Omental therapy for primary progressive aphasia with tau negative histopathology: 3 year study’, *Neurol. Res.*, vol. 31, no. 7, pp. 766–769, Sept. 2009, doi: 10.1179/174313209X382511.

[42] M. L. Henry and S. M. Grasso, ‘Assessment of Individuals with Primary Progressive Aphasia’, *Semin. Speech Lang.*, vol. 39, pp. 231–241, June 2018, doi: 10.1055/s-0038-1660782.

[43] ‘BDAE the Boston Diagnostic Aphasia Examination - University of Texas at Dallas’. Accessed: Oct. 13, 2025. [Online]. Available: https://utdallas.primo.exlibrisgroup.com/discovery/fulldisplay/alma9927692213901421/01UT_DALLAS:UTDALMA

[44] A. Kertesz, P. McMonagle, M. Blair, W. Davidson, and D. G. Munoz, ‘The evolution and pathology of frontotemporal dementia’, *Brain J. Neurol.*, vol. 128, no. Pt 9, pp. 1996–2005, Sept. 2005, doi: 10.1093/brain/awh598.

[45] S. Savage, S. Hsieh, F. Leslie, D. Foxe, O. Piguet, and J. R. Hodges, ‘Distinguishing subtypes in primary progressive aphasia: application of the Sydney language battery’, *Dement. Geriatr. Cogn. Disord.*, vol. 35, no. 3–4, pp. 208–218, 2013, doi: 10.1159/000346389.

[46] J. R. Hodges, M. Martinos, A. M. Woollams, K. Patterson, and A.-L. R. Adlam, ‘Repeat and Point: differentiating semantic dementia from progressive non-fluent aphasia’, *Cortex J. Devoted Study Nerv. Syst. Behav.*, vol. 44, no. 9, pp. 1265–1270, Oct. 2008, doi: 10.1016/j.cortex.2007.08.018.

[47] D. Sapolsky, K. Domoto-Reilly, and B. C. Dickerson, ‘Use of the Progressive Aphasia Severity Scale (PASS) in monitoring speech and language status in PPA’, *Aphasiology*, vol. 28, no. 8–9, pp. 993–1003, Jan. 2014, doi: 10.1080/02687038.2014.931563.

[48] C. E. Leyton *et al.*, ‘Subtypes of progressive aphasia: application of the International Consensus Criteria and validation using β-amyloid imaging’, *Brain J. Neurol.*, vol. 134, no. Pt 10, pp. 3030–3043, Oct. 2011, doi: 10.1093/brain/awr216.

[49] A. M. Butts, M. M. Machulda, J. R. Duffy, E. A. Strand, J. L. Whitwell, and K. A. Josephs, ‘Neuropsychological Profiles Differ among the Three Variants of Primary Progressive Aphasia’, *J. Int. Neuropsychol. Soc. JINS*, vol. 21, no. 6, pp. 429–435, July 2015, doi: 10.1017/S1355617715000399.

[50] S. Weintraub *et al.*, ‘Verbal and Nonverbal Memory in Primary Progressive Aphasia: The Three Words-Three Shapes Test’, *Behav. Neurol.*, vol. 26, no. 1–2, pp. 67–76, 2013, doi: 10.3233/BEN-2012-110239.

[51] ‘Rey, A. (1964) L’examen clinique en psychologie (The Clinical Psychological Examination). Presse Universitaires de France, Paris. - References - Scientific Research Publishing’. Accessed: Oct. 13, 2025. [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=1656248

[52] D. Wechsler, *Wechsler memory scale*. in Wechsler memory scale. San Antonio, TX, US: Psychological Corporation, 1945.

[53] 'Primary Progressive Aphasia: Losing the Ability to Communicate', Frontiers for Young Minds. Accessed: Oct. 24, 2025. [Online]. Available: https://kids.frontiersin.org/articles/10.3389/frym.2023.1054532

[54] C. K. Thompson and J. E. Mack, 'Grammatical impairments in PPA', *Aphasiology*, vol. 28, no. 8–9, pp. 1018–1037, Sept. 2014, doi: 10.1080/02687038.2014.912744.

[55] M. Montembeault, S. M. Brambati, M. L. Gorno-Tempini, and R. Migliaccio, 'Clinical, Anatomical, and Pathological Features in the Three Variants of Primary Progressive Aphasia: A Review', *Front. Neurol.*, vol. 9, Aug. 2018, doi: 10.3389/fneur.2018.00692.

[56] R.-J. M. van Geuns *et al.*, 'Basic principles of magnetic resonance imaging', *Prog. Cardiovasc. Dis.*, vol. 42, no. 2, pp. 149–156, Sept. 1999, doi: 10.1016/S0033-0620(99)70014-9.

[57] V. P. B. Grover, J. M. Tognarelli, M. M. E. Crossey, I. J. Cox, S. D. Taylor-Robinson, and M. J. W. McPhail, 'Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians', *J. Clin. Exp. Hepatol.*, vol. 5, no. 3, pp. 246–255, Sept. 2015, doi: 10.1016/j.jceh.2015.08.001.

[58] V. Mlynárik, 'Introduction to nuclear magnetic resonance', *Anal. Biochem.*, vol. 529, pp. 4–9, July 2017, doi: 10.1016/j.ab.2016.05.006.

[59] 'Basic Concepts of MR Imaging, Diffusion MR Imaging, and Diffusion Tensor Imaging - Magnetic Resonance Imaging Clinics'. Accessed: Oct. 22, 2025. [Online]. Available: https://www.mri.theclinics.com/article/S1064-9689(10)00074-7/fulltext

[60] L. D. Hall, 'Nuclear Magnetic Resonance', in *Advances in Carbohydrate Chemistry*, vol. 19, M. L. Wolfrom, Ed., Academic Press, 1964, pp. 51–93. doi: 10.1016/S0096-5332(08)60279-9.

[61] S. E. Forshult, *Magnetic Resonance Imaging – MRI – An Overview*. Fakulteten för teknik- och naturvetenskap, 2007. Accessed: Oct. 22, 2025. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-1203

[62] 'Net Magnetisation, rf pulses and flip angle. a) At equilibrium, the net...', ResearchGate. Accessed: Oct. 24, 2025. [Online]. Available: https://www.researchgate.net/figure/Net-Magnetisation-rf-pulses-and-flip-angle-a-At-equilibrium-the-net-magnetisation-Mo_fig2_49645994

[63] 'T1 relaxation process. Diagram showing the process of T1 relaxation...', ResearchGate. Accessed: Oct. 24, 2025. [Online]. Available: https://www.researchgate.net/figure/T1-relaxation-process-Diagram-showing-the-process-of-T1-relaxation-after-a-90-rf-pulse_fig3_49645994

[64] 'Transverse (T2 and T2*) relaxation processes. A diagram showing the...', ResearchGate. Accessed: Oct. 24, 2025. [Online]. Available: https://www.researchgate.net/figure/Transverse-T2-and-T2-relaxation-processes-A-diagram-showing-the-process-of-transverse_fig4_49645994

[65] 'Relaxation Times » IT'IS Foundation'. Accessed: Oct. 13, 2025. [Online]. Available: https://itis.swiss/virtual-population/tissue-properties/database/relaxation-times/

[66] 'Generating a spin echo. The presence of magnetic field inhomogeneities...', ResearchGate. Accessed: Oct. 24, 2025. [Online]. Available: https://www.researchgate.net/figure/Generating-a-spin-echo-The-presence-of-magnetic-field-inhomogeneities-causes-additional_fig6_49645994

[67] K. Möllenhoff, A.-M. Oros-Peusquens, and N. J. Shah, 'Introduction to the Basics of Magnetic Resonance Imaging', in *Molecular Imaging in the Clinical Neurosciences*, G. Gründer, Ed., Totowa, NJ: Humana Press, 2012, pp. 75–98. doi: 10.1007/7657_2012_56.

[68] 'Generating a gradient echo. This diagram show how the reversal of a...', ResearchGate. Accessed: Oct. 24, 2025. [Online]. Available: https://www.researchgate.net/figure/Generating-a-gradient-echo-This-diagram-show-how-the-reversal-of-a-magnetic-field_fig5_49645994

[69] J. P. Mugler III and J. R. Brookeman, 'Rapid three-dimensional T1-weighted MR imaging with the MP-RAGE sequence', *J. Magn. Reson. Imaging*, vol. 1, no. 5, pp. 561–567, 1991, doi: 10.1002/jmri.1880010509.

[70] J. P. Marques, T. Kober, G. Krueger, W. van der Zwaag, P.-F. Van de Moortele, and R. Gruetter, 'MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field', *NeuroImage*, vol. 49, no. 2, pp. 1271–1281, Jan. 2010, doi: 10.1016/j.neuroimage.2009.10.002.

[71] T. Yokoo *et al.*, 'A quantitative approach to sequence and image weighting', *J. Comput. Assist. Tomogr.*, vol. 34, no. 3, pp. 317–331, 2010.

[72] A. Zimny, L. Zińska, J. Bladowska, M. Neska-Matuszewska, and M. Sąsiadek, 'Intracranial lesions with high signal intensity on T1-weighted MR images – review of pathologies', *Pol. J. Radiol.*, vol. 78, no. 4, pp. 36–46, 2013, doi: 10.12659/PJR.889663.

[73] K.-A. Mardal, M. E. Rognes, T. B. Thompson, and L. M. Valnes, *Mathematical Modeling of the Human Brain: From Magnetic Resonance Images to Finite Element Simulation*. Springer Nature, 2022. doi: 10.1007/978-3-030-95136-8.

[74] G. Helms, 'Segmentation of human brain using structural MRI', *Magn. Reson. Mater. Phys. Biol. Med.*, vol. 29, no. 2, pp. 111–124, Apr. 2016, doi: 10.1007/s10334-015-0518-z.

[75] F. H. Alhazmi, O. M. Abdulaal, A. A. Qurashi, K. M. Aloufi, and V. Sluming, 'The effect of the MR pulse sequence on the regional corpus callosum morphometry', *Insights Imaging*, vol. 11, no. 1, p. 17, Feb. 2020, doi: 10.1186/s13244-019-0821-8.

[76] M. Symms, H. R. Jäger, K. Schmierer, and T. A. Yousry, 'A review of structural magnetic resonance neuroimaging', *J. Neurol. Neurosurg. Psychiatry*, vol. 75, no. 9, pp. 1235–1244, Sept. 2004, doi: 10.1136/jnnp.2003.032714.

[77] ‘MRI Basics’. Accessed: Oct. 24, 2025. [Online]. Available: https://case.edu/med/neurology/NR/MRI%20Basics.htm

[78] G.-H. Jahng, S. Park, C.-W. Ryu, and Z.-H. Cho, ‘Magnetic Resonance Imaging: Historical Overview, Technical Developments, and Clinical Applications’, *Prog. Med. Phys.*, vol. 31, no. 3, pp. 35–53, Sept. 2020, doi: 10.14316/pmp.2020.31.3.35.

[79] ‘Einstein, A. (1956) Investigations on the Theory of the Brownian Movement. Courier Corporation. - References - Scientific Research Publishing’. Accessed: Oct. 14, 2025. [Online]. Available: https://www.scirp.org/reference/referencespapers?referenceid=2402183

[80] D. L. Bihan and E. Breton, ‘Method to measure the molecular diffusion and/or perfusion parameters of live tissue’, US4809701A, Mar. 07, 1989 Accessed: Oct. 14, 2025. [Online]. Available: https://patents.google.com/patent/US4809701A/en

[81] P. J. Basser, J. Mattiello, and D. Lebihan, ‘Estimation of the Effective Self-Diffusion *Tensor* from the NMR Spin Echo’, *J. Magn. Reson. B*, vol. 103, no. 3, pp. 247–254, Mar. 1994, doi: 10.1006/jmrb.1994.1037.

[82] ‘MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. | Radiology’. Accessed: Oct. 14, 2025. [Online]. Available: https://pubs.rsna.org/doi/abs/10.1148/radiology.161.2.3763909

[83] Y. Assaf, H. Johansen-Berg, and M. Thiebaut de Schotten, ‘The role of diffusion MRI in neuroscience’, *NMR Biomed.*, vol. 32, no. 4, p. e3762, 2019, doi: 10.1002/nbm.3762.

[84] ‘Diffusion Models and Metrics’, Practicum. Accessed: Oct. 24, 2025. [Online]. Available: http://practicum.labsolver.org/metrics.html

[85] A. Asokan *et al.*, ‘Deep into diffusion tensor imaging.’, ECR 2019 EPOS. Accessed: Oct. 14, 2025. [Online]. Available: https://epos.myesr.org/poster/esr/ecr2019/C-2698

[86] P. Kochunov *et al.*, ‘Fractional anisotropy of water diffusion in cerebral white matter across the lifespan’, *Neurobiol. Aging*, vol. 33, no. 1, pp. 9–20, Jan. 2012, doi: 10.1016/j.neurobiolaging.2010.01.014.

[87] ‘DTI’, Questions and Answers in MRI. Accessed: Oct. 24, 2025. [Online]. Available: http://mriquestions.com/dti-tensor-imaging.html

[88] ‘Figure 4. (A) Diffusion-weighted image (DWI) of the brain; (B)...’, ResearchGate. Accessed: Oct. 24, 2025. [Online]. Available: https://www.researchgate.net/figure/A-Diffusion-weighted-image-DWI-of-the-brain-B-diffusion-tensor-image-DTI-of-the_fig4_327877744

[89] C. D. Schaper, ‘Analytic Model of fMRI BOLD Signals for Separable Metrics of Neural and Metabolic Activity’, Mar. 09, 2019, *bioRxiv*. doi: 10.1101/573006.

[90] A. Routier *et al.*, ‘Structural, Microstructural, and Metabolic Alterations in Primary Progressive Aphasia Variants’, *Front. Neurol.*, vol. 9, Sept. 2018, doi: 10.3389/fneur.2018.00766.

[91] F. Agosta, S. Galantucci, and M. Filippi, ‘Advanced magnetic resonance imaging of neurodegenerative diseases’, *Neurol. Sci.*, vol. 38, no. 1, pp. 41–51, Jan. 2017, doi: 10.1007/s10072-016-2764-x.

[92] F. Agosta *et al.*, ‘Differentiation between Subtypes of Primary Progressive Aphasia by Using Cortical Thickness and Diffusion-Tensor MR Imaging Measures’, *Radiology*, Feb. 2015, doi: 10.1148/radiol.15141869.

[93] S. M. Smith *et al.*, ‘Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data’, *NeuroImage*, vol. 31, no. 4, pp. 1487–1505, July 2006, doi: 10.1016/j.neuroimage.2006.02.024.

[94] M. Bach *et al.*, ‘Methodological considerations on tract-based spatial statistics (TBSS)’, *NeuroImage*, vol. 100, pp. 358–369, Oct. 2014, doi: 10.1016/j.neuroimage.2014.06.021.

[95] ‘Comparison of skeletons and mis-registrations observed in TBSS...’, ResearchGate. Accessed: Oct. 15, 2025. [Online]. Available: https://www.researchgate.net/figure/Comparison-of-skeletons-and-mis-registrations-observed-in-TBSS-analysis-a-Demonstrates_fig1_313039526

[96] Y. Zhang and F. Zhan, ‘Diffusion tensor imaging (DTI) Analysis Based on Tract-based spatial statistics (TBSS) and Classification Using Multi-Metric in Alzheimer’s Disease’, *J. Integr. Neurosci.*, vol. 22, no. 4, p. 101, July 2023, doi: 10.31083/j.jin2204101.

[97] G. C. Schwindt *et al.*, ‘Whole-brain white matter disruption in semantic and nonfluent variants of primary progressive aphasia’, *Hum. Brain Mapp.*, vol. 34, no. 4, pp. 973–984, 2013, doi: 10.1002/hbm.21484.

[98] A. Abraham, ‘Nature and scope of AI techniques’, *Handb. Meas. Syst. Des.*, 2005, Accessed: Oct. 15, 2025. [Online]. Available: http://isda03.softcomputing.net/ci_chapter.pdf

[99] W. S. McCulloch and W. Pitts, ‘A logical calculus of the ideas immanent in nervous activity’, *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943, doi: 10.1007/BF02478259.

[100] ‘Biological neuron versus McCulloch and Pitts’ artificial neuron model’, ResearchGate. Accessed: Nov. 01, 2025. [Online]. Available: https://www.researchgate.net/figure/Biological-neuron-versus-McCulloch-and-Pitts-artificial-neuron-model_fig2_359233566

[101] A. M. Turing, ‘Computing machinery and intelligence (1950)’, *The Essen*, 2004, Accessed: Oct. 15, 2025. [Online]. Available: https://www.edwardfrenkel.com/turing-intelligence.pdf

[102] S. Cave *et al.*, ‘The Meanings of AI: A Cross-Cultural Comparison’, in *Imagining AI*, 1st edn, S. Cave and K. Dihal, Eds, Oxford University PressOxford, 2023, pp. 16–36. doi: 10.1093/oso/9780192865366.003.0002.

[103] C. Zhang and Y. Lu, ‘Study on artificial intelligence: The state of the art and future prospects’, *J. Ind. Inf. Integr.*, vol. 23, p. 100224, Sept. 2021, doi: 10.1016/j.jii.2021.100224.

[104] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, ‘A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August

31, 1955', *AI Mag.*, vol. 27, no. 4, pp. 12–12, Dec. 2006, doi: 10.1609/aimag.v27i4.1904.

[105] F. Rosenblatt, 'The perceptron: A probabilistic model for information storage and organization in the brain', *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958, doi: 10.1037/h0042519.

[106] M. Kuipers and R. Prasad, 'Journey of Artificial Intelligence', *Wirel. Pers. Commun.*, vol. 123, Apr. 2022, doi: 10.1007/s11277-021-09288-0.

[107] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. in Prentice Hall series in artificial intelligence. Upper Saddle River: Prentice Hall, 1995.

[108] L. Rubinger, A. Gazendam, S. Ekhtiari, and M. Bhandari, 'Machine learning and artificial intelligence in research and healthcare', *Injury*, vol. 54, pp. S69–S73, May 2023, doi: 10.1016/j.injury.2022.01.046.

[109] Werbos, 'Backpropagation: past and future', in *IEEE 1988 International Conference on Neural Networks*, July 1988, pp. 343–353 vol.1. doi: 10.1109/ICNN.1988.23866.

[110] A. Toosi, A. G. Bottino, B. Saboury, E. Siegel, and A. Rahmim, 'A Brief History of AI: How to Prevent Another Winter (A Critical Review)', *PET Clin.*, vol. 16, no. 4, pp. 449–469, Oct. 2021, doi: 10.1016/j.cpet.2021.07.001.

[111] L. Rubinger, A. Gazendam, S. Ekhtiari, and M. Bhandari, 'Machine learning and artificial intelligence in research and healthcare', *Injury*, vol. 54, pp. S69–S73, May 2023, doi: 10.1016/j.injury.2022.01.046.

[112] 'Brief History of Artificial Intelligence - Neuroimaging Clinics'. Accessed: Oct. 20, 2025. [Online]. Available: https://www.neuroimaging.theclinics.com/article/S1052-5149(20)30054-X/abstract

[113] 'What Is Machine Learning? | SpringerLink'. Accessed: Nov. 01, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-18305-3_1

[114] M. Kubat, *An Introduction to Machine Learning*. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-63913-0.

[115] J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenburg, T. Verbelen, and J. S. Rellermeyer, 'A Survey on Distributed Machine Learning', *ACM Comput Surv*, vol. 53, no. 2, p. 30:1-30:33, Mar. 2020, doi: 10.1145/3377454.

[116] E. F. Morales and H. J. Escalante, 'Chapter 6 - A brief introduction to supervised, unsupervised, and reinforcement learning', in *Biosignal Processing and Classification Using Computational Learning and Intelligence*, A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, and O. Mendoza-Montoya, Eds, Academic Press, 2022, pp. 111–129. doi: 10.1016/B978-0-12-820125-1.00017-8.

[117] X. (Jerry) Zhu, 'Semi-Supervised Learning Literature Survey', University of Wisconsin-Madison Department of Computer Sciences, Technical Report, 2005. Accessed: Nov. 01, 2025. [Online]. Available: https://minds.wisconsin.edu/handle/1793/60444

[118] admin, ‘Supervised vs Unsupervised Learning: A Beginner’s Guide👨‍💻’. Accessed: Nov. 01, 2025. [Online]. Available: https://techiefeather.com/supervised-vs-unsupervised-learning/

[119] ‘Reinforcement Learning - GeeksforGeeks’. Accessed: Nov. 01, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/what-is-reinforcement-learning/

[120] E. Pellegrini *et al.*, ‘Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review’, *Alzheimers Dement. Amst. Neth.*, vol. 10, pp. 519–535, 2018, doi: 10.1016/j.dadm.2018.07.004.

[121] J. P. A. van Soest, A. L. A. J. Dekker, E. Roelofs, and G. Nalbantov, ‘Application of Machine Learning for Multicenter Learning’, in *Machine Learning in Radiation Oncology: Theory and Applications*, I. El Naqa, R. Li, and M. J. Murphy, Eds, Cham: Springer International Publishing, 2015, pp. 71–97. doi: 10.1007/978-3-319-18305-3_6.

[122] S. Raschka, ‘Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning’, Nov. 11, 2020, *arXiv*: arXiv:1811.12808. doi: 10.48550/arXiv.1811.12808.

[123] I. O. Muraina, ‘IDEAL DATASET SPLITTING RATIOS IN MACHINE LEARNING ALGORITHMS: GENERAL CONCERNS FOR DATA SCIENTISTS AND DATA ANALYSTS’.

[124] M. Huang, ‘Survey of System Design for Distributed ML & FL’, NetH-Lab. Accessed: Nov. 01, 2025. [Online]. Available: https://neth-lab.netlify.app/publication/21-12-31-survey-of-system-design-for-distributed-ml-and-fl/

[125] X. Ying, ‘An Overview of Overfitting and its Solutions’, *J. Phys. Conf. Ser.*, vol. 1168, no. 2, p. 022022, Feb. 2019, doi: 10.1088/1742-6596/1168/2/022022.

[126] E. Ohiri, ‘What is underfitting and overfitting in machine learning?’, CUDO Compute. Accessed: Nov. 01, 2025. [Online]. Available: https://www.cudocompute.com/blog/overfitting-and-underfitting-in-machine-learning-causes-indicators-and-how

[127] A. O. Salau and S. Jain, ‘Feature Extraction: A Survey of the Types, Techniques, Applications’, in *2019 International Conference on Signal Processing and Communication (ICSC)*, Mar. 2019, pp. 158–164. doi: 10.1109/ICSC45622.2019.8938371.

[128] ‘Feature Extraction in Machine Learning: A Complete Guide’. Accessed: Nov. 01, 2025. [Online]. Available: https://www.datacamp.com/tutorial/feature-extraction-machine-learning?dc_referrer=https%3A%2F%2Fwww.google.com%2F

[129] M. Suresh, ‘Feature Engineering — Overview’, Medium. Accessed: Nov. 01, 2025. [Online]. Available: https://monicasuresh.medium.com/feature-engineering-overview-43c77e2f86e3

[130] I. Guyon and A. Elisseeff, ‘An Introduction to Variable and Feature Selection’, *J. Mach. Learn. Res.*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[131] 'Introduction to Feature Selection - MATLAB & Simulink'. Accessed: Nov. 01, 2025. [Online]. Available: https://www.mathworks.com/help/stats/feature-selection.html

[132] A. Kaur, K. Guleria, and N. K. Trivedi, 'Feature selection in machine learning: Methods and comparison', in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, 2021, pp. 789–795. Accessed: Oct. 09, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9404623/

[133] A. Jović, K. Brkić, and N. Bogunović, 'A review of feature selection methods with applications', in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2015, pp. 1200–1205. doi: 10.1109/MIPRO.2015.7160458.

[134] L. Xie, Z. Li, Y. Zhou, Y. He, and J. Zhu, 'Computational Diagnostic Techniques for Electrocardiogram Signal Analysis', *Sensors*, Nov. 2020.

[135] K. Razzaq and M. Shah, 'Machine Learning and Deep Learning Paradigms: From Techniques to Practical Applications and Research Frontiers', *Computers*, vol. 14, no. 3, p. 93, Mar. 2025, doi: 10.3390/computers14030093.

[136] '🔍 Linear vs Logistic Regression – Explained with Key Differences | UpdateGadh'. Accessed: Nov. 02, 2025. [Online]. Available: https://updategadh.com/machine-learning-tutorial/linear-vs-logistic-regression/

[137] J. K. Harris, 'Primer on binary logistic regression', *Fam. Med. Community Health*, vol. 9, no. Suppl 1, p. e001290, Dec. 2021, doi: 10.1136/fmch-2021-001290.

[138] A. El-Koka, K.-H. Cha, and D.-K. Kang, 'Regularization parameter tuning optimization approach in logistic regression', in *2013 15th International Conference on Advanced Communications Technology (ICACT)*, Jan. 2013, pp. 13–18. Accessed: Oct. 10, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6488130

[139] R. Tibshirani, 'Regression Shrinkage and Selection Via the Lasso', *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[140] V. Jain, A. Phophalia, and J. S. Bhatt, 'Investigation of a Joint Splitting Criteria for Decision Tree Classifier Use of Information Gain and Gini Index', in *TENCON 2018 - 2018 IEEE Region 10 Conference*, Oct. 2018, pp. 2187–2192. doi: 10.1109/TENCON.2018.8650485.

[141] B. de Ville, 'Decision trees', *WIREs Comput. Stat.*, vol. 5, no. 6, pp. 448–455, 2013, doi: 10.1002/wics.1278.

[142] A. Singh, 'Decision Tree in Machine Learning', Applied AI Blog. Accessed: Oct. 09, 2025. [Online]. Available: https://www.appliedaicourse.com/blog/decision-tree-in-machine-learning/

[143] L. Breiman, 'Random Forests', *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[144] B. T. Padovese and L. R. Padovese, ‘A machine learning approach to the recognition of Brazilian Atlantic Forest Parrot species’, *bioRxiv*, pp. 2019–12, 2019.

[145] C. Cortes and V. Vapnik, ‘Support-vector networks’, *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sept. 1995, doi: 10.1007/BF00994018.

[146] D. Bzdok, M. Krzywinski, and N. Altman, ‘Machine learning: supervised methods’, *Nat. Methods*, vol. 15, no. 1, p. 5, 2018.

[147] B. Kallfelz Sirmacek, N. Botteghi, and S. Sanchez, ‘SEQUENTIAL IMAGE PROCESSING METHODS FOR IMPROVING SEMANTIC VIDEO SEGMENTATION ALGORITHMS A PREPRINT’, Oct. 2019.

[148] R. Salomon, ‘Evolutionary algorithms and gradient search: similarities and differences’, *IEEE Trans. Evol. Comput.*, vol. 2, no. 2, pp. 45–55, July 1998, doi: 10.1109/4235.728207.

[149] ‘Evolutionary Algorithms’, DeepAI. Accessed: Nov. 02, 2025. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/evolutionary-algorithms

[150] ‘Stochastic Gradient Descent: Understanding the Basics’. Accessed: Nov. 02, 2025. [Online]. Available: https://botpenguin.com/glossary/stochastic-gradient-descent

[151] ‘ANGUITA, Davide, et al. The’K’in K-fold Cross Validation.... - Google Scholar’. Accessed: Oct. 09, 2025. [Online]. Available: https://scholar.google.com/scholar?hl=it&as_sdt=0%2C5&q=ANGUITA%2C+Davide%2C+et+al.+The%27K%27in+K-fold+Cross+Validation.+In%3A+Esann.+2012.+p.+441-446.&btnG=

[152] ‘The Necessity of Leave One Subject Out (LOSO) Cross Validation for EEG Disease Diagnosis | SpringerLink’. Accessed: Nov. 02, 2025. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-86993-9_50

[153] T. Hastie, R. Tibshirani, and J. Friedman, ‘The elements of statistical learning’. Springer series in statistics New-York, 2009. Accessed: Oct. 09, 2025. [Online]. Available: https://www.academia.edu/download/31156736/10.1.1.158.8831.pdf

[154] T. Kaneva, B. Evstatiev, I. Valova, N. Valov, and K. Gabrovska-Evstatieva, ‘Comparing Different Evaluation Metrics with the Grid Search Method for Classification of Highly Imbalanced Data’, in *2024 8th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, 2024, pp. 1–5. Accessed: Oct. 09, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10757282/

[155] S. Swaminathan and B. R. Tantri, ‘Confusion Matrix-Based Performance Evaluation Metrics’, *Afr. J. Biomed. Res.*, vol. 27, pp. 4023–4031, Nov. 2024, doi: 10.53555/AJBR.v27i4S.4345.

[156] ‘How to explain the ROC AUC score and ROC curve?’ Accessed: Nov. 02, 2025. [Online]. Available: https://www.evidentlyai.com/classification-metrics/explain-roc-curve

[157] 'Compare Deep Learning Models Using ROC Curves - MATLAB & Simulink'. Accessed: Nov. 02, 2025. [Online]. Available: https://it.mathworks.com/help/deeplearning/ug/compare-deep-learning-models-using-ROC-curves.html

[158] M. L. Giger, 'Machine Learning in Medical Imaging', *J. Am. Coll. Radiol.*, vol. 15, no. 3, Part B, pp. 512–520, Mar. 2018, doi: 10.1016/j.jacr.2017.12.028.

[159] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, 'Linguistic Features Identify Alzheimer's Disease in Narrative Speech', *J. Alzheimer's Dis.*, vol. 49, no. 2, pp. 407–422, Jan. 2016, doi: 10.3233/JAD-150520.

[160] S. Ash, P. Moore, L. Vesely, and M. Grossman, 'The decline of narrative discourse in Alzheimer's disease', *Brain Lang.*, vol. 103, no. 1–2, pp. 181–182, 2007.

[161] A. Parikh, L. ten Bosch, H. van den Heuvel, and C. Tejedor-García, 'Comparing Modular and End-To-End Approaches in ASR for Well-Resourced and Low-Resourced Languages', in *Proceedings Of The 6th International Conference On Natural Language And Speech Processing (ICNLSP 2023)*, 2023, pp. 266–273. Accessed: Oct. 09, 2025. [Online]. Available: https://aclanthology.org/2023.icnlsp-1.28.pdf

[162] Z. Brahmi, M. Mahyoob, M. Al-Sarem, J. Algaraady, K. Bousselmi, and A. Alblwi, 'Exploring the Role of Machine Learning in Diagnosing and Treating Speech Disorders: A Systematic Literature Review', *Psychol. Res. Behav. Manag.*, vol. 17, pp. 2205–2232, Dec. 2024, doi: 10.2147/PRBM.S460283.

[163] F. van Leeuwen, 'A 101 guide to the FDA regulatory process for AI radiology software'. Accessed: Nov. 02, 2025. [Online]. Available: https://www.quantib.com/blog/a-101-guide-to-the-fda-regulatory-process-for-ai-radiology-software

[164] M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, 'Automatic speech recognition: a survey', *Multimed. Tools Appl.*, vol. 80, no. 6, pp. 9411–9457, Mar. 2021, doi: 10.1007/s11042-020-10073-7.

[165] H. Liu, B. MacWhinney, D. Fromm, and A. Lanzi, 'Automation of Language Sample Analysis', *J. Speech Lang. Hear. Res.*, vol. 66, no. 7, pp. 2421–2433, July 2023, doi: 10.1044/2023_JSLHR-22-00642.

[166] G. Hemakumar and P. Punitha, 'Speech recognition technology: a survey on Indian languages', *Int. J. Inf. Sci. Intell. Syst.*, vol. 2, no. 4, pp. 1–38, 2013.

[167] K. H. Davis, R. Biddulph, and S. Balashek, 'Automatic Recognition of Spoken Digits', *J. Acoust. Soc. Am.*, vol. 24, no. 6, pp. 637–642, Nov. 1952, doi: 10.1121/1.1906946.

[168] J. W. Forgie and C. D. Forgie, 'Results Obtained from a Vowel Recognition Computer Program', *J. Acoust. Soc. Am.*, vol. 31, no. 11, pp. 1480–1489, Nov. 1959, doi: 10.1121/1.1907653.

[169] B. Yegnanarayana and R. N. J. Veldhuis, 'Extraction of vocal-tract system characteristics from speech signals', *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 313–327, July 1998, doi: 10.1109/89.701359.

[170] X. Tang, ‘Hybrid Hidden Markov Model and Artificial Neural Network for Automatic Speech Recognition’, in *2009 Pacific-Asia Conference on Circuits, Communications and Systems*, May 2009, pp. 682–685. doi: 10.1109/PACCS.2009.138.

[171] A. C. Morris, V. Maier, and P. Green, ‘From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition’, in *Interspeech 2004*, ISCA, Oct. 2004, pp. 2765–2768. doi: 10.21437/Interspeech.2004-668.

[172] S. S. Mahmoud, R. F. Pallaud, A. Kumar, S. Faisal, Y. Wang, and Q. Fang, ‘A comparative investigation of automatic speech recognition platforms for aphasia assessment batteries’, *Sensors*, vol. 23, no. 2, p. 857, 2023.

[173] ‘Soniox. Italian Speech Recognition Benchmark 2023 - Cerca con Google’. Accessed: Oct. 09, 2025. [Online]. Available: https://www.google.com/search?q=Soniox.+Italian+Speech+Recognition+Benchmark+2023&rlz=1C5CHFA_enIT916IT917&oq=Soniox.+Italian+Speech+Recognition+Benchmark+2023&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIGCAEQRRg80gEHNTM0ajBqN6gCALACAA&sourceid=chrome&ie=UTF-8

[174] M. Labied, A. Belangour, M. Banane, and A. Erraissi, ‘An overview of Automatic Speech Recognition Preprocessing Techniques’, in *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, Mar. 2022, pp. 804–809. doi: 10.1109/DASA54658.2022.9765043.

[175] R. M. Patil and C. M. Patil, ‘Unveiling the State-of-the-Art: A Comprehensive Survey on Voice Activity Detection Techniques’, in *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*, July 2024, pp. 1–5. doi: 10.1109/APCIT62007.2024.10673721.

[176] Z. Lin, Z. Chen, B. Zeng, L. Chen, and J. Cai, ‘Performance Optimization in the Cascade of VAD and ASR Systems: A Study on Evaluation and Alignment Strategies’, in *2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec. 2024, pp. 1–6. doi: 10.1109/APSIPAASC63619.2025.10848640.

[177] C. A. *et al.*, ‘Automatic Silence Detection Employing Artificial Intelligence for Clinical Context Analyses’, in *2024 3rd International Congress of Biomedical Engineering and Bioengineering (CIIBBI)*, Nov. 2024, pp. 1–6. doi: 10.1109/CIIBBI63846.2024.10785143.

[178] J. Li *et al.*, ‘A Comparative Study of Acoustic and Linguistic Features Classification for Alzheimer’s Disease Detection’, in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, June 2021, pp. 6423–6427. doi: 10.1109/ICASSP39728.2021.9414147.

[179] S. Cho *et al.*, ‘Automatic classification of AD pathology in FTD phenotypes using natural speech’, *Alzheimers Dement.*, vol. 20, no. 5, pp. 3416–3428, 2024, doi: 10.1002/alz.13748.

[180] E. D. Liddy, ‘Natural Language Processing’.

[181] J. Hirschberg and C. D. Manning, ‘Advances in natural language processing’, *Science*, vol. 349, no. 6245, pp. 261–266, July 2015, doi: 10.1126/science.aaa8685.

[182] S. Kadre, S. Kadre, and S. Dey, ‘Syntactic and Semantic Techniques in NLP’, in *Mastering Text Analytics : A Hands-on Guide to NLP Using Python*, S. Kadre, S. Kadre, and S. Dey, Eds, Berkeley, CA: Apress, 2025, pp. 211–274. doi: 10.1007/979-8-8688-1582-9_6.

[183] G. Gosztolya, V. Vincze, L. Tóth, M. Pákáski, J. Kálmán, and I. Hoffmann, ‘Identifying Mild Cognitive Impairment and mild Alzheimer’s disease based on spontaneous speech using ASR and linguistic features’, *Comput. Speech Lang.*, vol. 53, pp. 181–197, Jan. 2019, doi: 10.1016/j.csl.2018.07.007.

[184] A. Slegers *et al.*, ‘Connected speech markers of amyloid burden in primary progressive aphasia’, *Cortex*, vol. 145, pp. 160–168, 2021.

[185] A. Khodabakhsh, F. Yesil, E. Guner, and C. Demiroglu, ‘Evaluation of linguistic and prosodic features for detection of Alzheimer’s disease in Turkish conversational speech’, *EURASIP J. Audio Speech Music Process.*, vol. 2015, no. 1, p. 9, Mar. 2015, doi: 10.1186/s13636-015-0052-y.

[186] N. Rezaii *et al.*, ‘Artificial intelligence classifies primary progressive aphasia from connected speech’, *Brain*, vol. 147, no. 9, pp. 3070–3082, Sept. 2024, doi: 10.1093/brain/awae196.

[187] A. M. Jensen, H. J. Chenery, and D. A. Copland, ‘A comparison of picture description abilities in individuals with vascular subcortical lesions and Huntington’s Disease’, *J. Commun. Disord.*, vol. 39, no. 1, pp. 62–77, Jan. 2006, doi: 10.1016/j.jcomdis.2005.07.001.

[188] S. Galantucci *et al.*, ‘White matter damage in primary progressive aphasias: a diffusion tensor tractography study’, *Brain*, vol. 134, no. 10, pp. 3011–3029, Oct. 2011, doi: 10.1093/brain/awr099.

[189] F. Quek *et al.*, ‘Multimodal human discourse: gesture and speech’, *ACM Trans Comput-Hum Interact*, vol. 9, no. 3, pp. 171–193, Sept. 2002, doi: 10.1145/568513.568514.

[190] M. Turk, ‘Multimodal interaction: A review’, *Pattern Recognit. Lett.*, vol. 36, pp. 189–195, Jan. 2014, doi: 10.1016/j.patrec.2013.07.003.

[191] F. Cutugno, V. A. Leano, R. Rinaldi, and G. Mignini, ‘Multimodal framework for mobile interaction’, in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, in AVI ’12. New York, NY, USA: Association for Computing Machinery, May 2012, pp. 197–203. doi: 10.1145/2254556.2254592.

[192] I. Guyon and A. Elisseeff, ‘An Introduction to Variable and Feature Selection’.

[193] L. M. Pereira, A. Salazar, and L. Vergara, ‘On Comparing Early and Late Fusion Methods’, in *Advances in Computational Intelligence*, I. Rojas, G. Joya, and A. Catala, Eds, Cham: Springer Nature Switzerland, 2023, pp. 365–378. doi: 10.1007/978-3-031-43085-5_29.

[194] F. Eyben and B. Schuller, 'openSMILE:): the Munich open-source large-scale multimedia feature extractor', *SIGMultimedia Rec*, vol. 6, no. 4, pp. 4–13, Jan. 2015, doi: 10.1145/2729095.2729097.

[195] F. Agosta *et al.*, 'Added Value of Multimodal Structural MRI to the Clinical Diagnosis of Primary Progressive Aphasia Variants (P3.182)', *Neurology*, vol. 90, no. 15_supplement, p. P3.182, Apr. 2018, doi: 10.1212/WNL.90.15_supplement.P3.182.

[196] G. Kim, J. G. Choi, M. Ku, H. Cho, and S. Lim, 'A Multimodal Deep Learning-Based Fault Detection Model for a Plastic Injection Molding Process', *IEEE Access*, vol. 9, pp. 132455–132467, 2021, doi: 10.1109/ACCESS.2021.3115665.

[197] M. F. Folstein, S. E. Folstein, and P. R. McHugh, '"Mini-mental state"', *J. Psychiatr. Res.*, vol. 12, no. 3, pp. 189–198, Nov. 1975, doi: 10.1016/0022-3956(75)90026-6.

[198] I. Appollonio *et al.*, 'The Frontal Assessment Battery (FAB): normative values in an Italian population sample', *Neurol. Sci.*, vol. 26, no. 2, pp. 108–116, June 2005, doi: 10.1007/s10072-005-0443-4.

[199] A. Orsini, D. Grossi, E. Capitani, M. Laiacona, C. Papagno, and G. Vallar, 'Verbal and spatial immediate memory span: Normative data from 1355 adults and 1112 children', *Ital. J. Neurol. Sci.*, vol. 8, no. 6, pp. 537–548, Dec. 1987, doi: 10.1007/BF02333660.

[200] A. Rey, *L'examen clinique en psychologie. [The clinical examination in psychology.]*. in L'examen clinique en psychologie. Oxford, England: Presses Universitaries De France, 1958, p. 222.

[201] P. Caffarra, G. Vezzadini, F. Dieci, F. Zonato, and A. Venneri, 'Rey-Osterrieth complex figure: normative values in an Italian population sample', *Neurol. Sci.*, vol. 22, no. 6, pp. 443–447, Mar. 2002, doi: 10.1007/s100720200003.

[202] H. SPINNLER, 'Standardizzazione e taratura italiana di test neuropsicologici', *Ital J Neurol Sci*, vol. 6, no. 0, pp. 21–120, 1987.

[203] M. Monaco, A. Costa, C. Caltagirone, and G. A. Carlesimo, 'Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population', *Neurol. Sci.*, vol. 34, no. 5, pp. 749–754, May 2013, doi: 10.1007/s10072-012-1130-x.

[204] A. R. Giovagnoli, M. Del Pesce, S. Mascheroni, M. Simoncelli, M. Laiacona, and E. Capitani, 'Trail making test: normative values from 287 normal adult controls', *Ital. J. Neurol. Sci.*, vol. 17, no. 4, pp. 305–309, Aug. 1996, doi: 10.1007/BF01997792.

[205] A. Basso, E. Capitani, and M. Laiacona, 'Raven's coloured progressive matrices: normative values on 305 adult normal controls', *Funct. Neurol.*, vol. 2, no. 2, pp. 189–194, Apr. 1987.

[206] P. J. Manos, 'Ten-point clock test sensitivity for Alzheimer's Disease in patients with MMSE scores greater than 23', *Int. J. Geriatr. Psychiatry*, vol.

14, no. 6, pp. 454–458, 1999, doi: 10.1002/(SICI)1099-1166(199906)14:6%3C454::AID-GPS951%3E3.0.CO;2-N.

[207] ‘Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function | JAMA | JAMA Network’. Accessed: Nov. 14, 2025. [Online]. Available: https://jamanetwork.com/journals/jama/article-abstract/666768

[208] M. P. Lawton and E. M. Brody, ‘Assessment of older people: Self-maintaining and instrumental activities of daily living’, *The Gerontologist*, vol. 9, no. 3, Pt 1, pp. 179–186, 1969, doi: 10.1093/geront/9.3_Part_1.179.

[209] ‘The Neuropsychiatric Inventory | Neurology’. Accessed: Nov. 14, 2025. [Online]. Available: https://www.neurology.org/doi/abs/10.1212/wnl.44.12.2308?casa_token=YUDCbyADyz4AAAAA:ySTdavN5RrY849izVGWbLPehDYqwRN6G_pOOMo5jvDgTJrYbWHguPosU96o-RnHifGTIvfAbgv8lzfU

[210] A. Alberici *et al.*, ‘The Frontal Behavioural Inventory (Italian version) differentiates frontotemporal lobar degeneration variants from Alzheimer’s disease’, *Neurol. Sci.*, vol. 28, no. 2, pp. 80–86, Apr. 2007, doi: 10.1007/s10072-007-0791-3.

[211] B. Borroni *et al.*, ‘The FTLD-modified Clinical Dementia Rating scale is a reliable tool for defining disease severity in Frontotemporal Lobar Degeneration: evidence from a brain SPECT study’, *Eur. J. Neurol.*, vol. 17, no. 5, pp. 703–707, 2010, doi: 10.1111/j.1468-1331.2009.02911.x.

[212] C. P. Hughes, L. Berg, W. Danziger, L. A. Coben, and R. L. Martin, ‘A New Clinical Scale for the Staging of Dementia’, *Br. J. Psychiatry*, vol. 140, no. 6, pp. 566–572, June 1982, doi: 10.1192/bjp.140.6.566.

[213] A. De Renzi and L. A. Vignolo, ‘Token test: A sensitive test to detect receptive disturbances in aphasics’, *Brain J. Neurol.*, vol. 85, pp. 665–678, 1962, doi: 10.1093/brain/85.4.665.

[214] E. Catricalà, P. A. Della Rosa, V. Ginex, Z. Mussetti, V. Plebani, and S. F. Cappa, ‘An Italian battery for the assessment of semantic memory disorders’, *Neurol. Sci.*, vol. 34, no. 6, pp. 985–993, June 2013, doi: 10.1007/s10072-012-1181-z.

[215] N. Gamboz, E. Coluccia, A. Iavarone, and M. A. Brandimonte, ‘Normative data for the Pyramids and Palm Trees Test in the elderly Italian population’, *Neurol. Sci.*, vol. 30, no. 6, pp. 453–458, Dec. 2009, doi: 10.1007/s10072-009-0130-y.

[216] ‘(PDF) New Normative Data for the Italian Version of the Aachen Aphasia Test [A.A.T.]’, *ResearchGate*, Aug. 2025, Accessed: Nov. 15, 2025. [Online]. Available: https://www.researchgate.net/publication/279555790_New_Normative_Data_for_the_Italian_Version_of_the_Aachen_Aphasia_Test_AAT

[217] G. Novelli, C. Papagno, E. Capitani, and M. Laiacona, ‘Tre test clinici di memoria verbale a lungo termine: Taratura su soggetti normali. / Three clinical

tests for the assessment of verbal long-term memory function: Norms from 320 normal subjects.', *Arch. Psicol. Neurol. Psichiatr.*, pp. 278–296, Jan. 1970.

[218] 'REAPER | Audio Production Without Limits'. Accessed: Nov. 08, 2025. [Online]. Available: https://www.reaper.fm/index.php

[219] E. Zwyssig, M. Lincoln, and S. Renals, 'A digital microphone array for distant speech recognition', in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2010, pp. 5106–5109. doi: 10.1109/ICASSP.2010.5495040.

[220] A. Czyżewski *et al.*, 'A Comprehensive Polish Medical Speech Dataset for Enhancing Automatic Medical Dictation', *Sci. Data*, vol. 12, p. 1436, Aug. 2025, doi: 10.1038/s41597-025-05776-1.

[221] V. C M, S. Pal, N. Mantri, and G. K. Agrawal, 'Effect of Loudspeaker Emitted Speech on ASR performance', presented at the Proc. Interspeech 2025, 2025, pp. 3170–3173. doi: 10.21437/Interspeech.2025-2470.

[222] 'Audio Loudness Normalization With FFmpeg', notes.txt. Accessed: Nov. 07, 2025. [Online]. Available: http://peterforgacs.github.io/2018/05/20/Audio-normalization-with-ffmpeg/index.html

[223] 'Noise Reduction - Audacity Manual'. Accessed: Nov. 07, 2025. [Online]. Available: https://manual.audacityteam.org/man/noise_reduction.html

[224] *noisereduce: Noise reduction using Spectral Gating in Python*. Python. Accessed: Nov. 07, 2025. [OS Independent]. Available: https://github.com/timsainb/noisereduce

[225] PatrickFarley, 'Speech to text overview - Speech service - Azure AI services'. Accessed: Nov. 07, 2025. [Online]. Available: https://learn.microsoft.com/en-us/azure/ai-services/speech-service/speech-to-text

[226] N. Ö. Doğan, 'Bland-Altman analysis: A paradigm to understand correlation and agreement', *Turk. J. Emerg. Med.*, vol. 18, no. 4, pp. 139–141, Dec. 2018, doi: 10.1016/j.tjem.2018.09.001.

[227] E. Canu *et al.*, 'Connected Speech Alterations and Progression in Patients With Primary Progressive Aphasia Variants', *Neurology*, vol. 104, no. 9, p. e213524, May 2025, doi: 10.1212/WNL.0000000000213524.

[228] C. J. Mahoney *et al.*, 'White matter tract signatures of the progressive aphasias', *Neurobiol. Aging*, vol. 34, no. 6, pp. 1687–1699, June 2013, doi: 10.1016/j.neurobiolaging.2012.12.002.

[229] K. A. Tetzloff *et al.*, 'Automatic Speech Recognition in Primary Progressive Apraxia of Speech', *J. Speech Lang. Hear. Res.*, vol. 67, no. 9, pp. 2964–2976, Sept. 2024, doi: 10.1044/2024_JSLHR-24-00049.

[230] F. Agosta *et al.*, 'Language networks in semantic dementia', *Brain*, vol. 133, no. 1, pp. 286–299, Jan. 2010, doi: 10.1093/brain/awp233.

[231] M. J. Strong, *Amyotrophic Lateral Sclerosis and the Frontotemporal Dementias*. OUP Oxford, 2012.

Dedications

Giunta alla conclusione di questo lavoro, desidero esprimere la mia gratitudine a tutte le persone che hanno contribuito, in modi diversi, a rendere possibile questo percorso.

Ringrazio il Professor Molinari per la sua disponibilità, professionalità e attenzione durante tutte le fasi della tesi.

Rivolgo un sentito ringraziamento al Chiar.mo Professor Filippi per avermi dato l'opportunità di prendere parte a questo progetto e di entrare in un contesto scientifico così stimolante.

Desidero inoltre ringraziare la Professoressa Agosta, la cui competenza e dedizione rappresentano un esempio costante e fonte di ispirazione.

Un ringraziamento speciale va a Silvia, che è stata una guida preziosa e un punto di riferimento sempre presente. Le sono profondamente grata per le opportunità che mi ha offerto, per il tempo dedicatomi e per aver accompagnato la mia crescita sia scientifica sia personale.

Grazie di cuore a Simonetta, per la sua disponibilità, i consigli e il sostegno con cui mi ha seguita lungo tutto il percorso.

Il mio grazie va anche a Laura e Anna, per la loro gentilezza e per la condivisione di idee e momenti che hanno reso il lavoro più leggero e sereno.

Un pensiero affettuoso è per tutti i ragazzi del CAB: grazie per la compagnia, per i sorrisi, per le pause pranzo che spezzavano le giornate più intense e per la vostra capacità di rendere ogni momento più leggero.

Infine, grazie con tutto il cuore ad Alice e Francesco: condividere questa esperienza con voi è stato un dono prezioso, che ha reso questo percorso più ricco, più bello e infinitamente più significativo.