BIOSIGNALS 2026
19TH INTERNATIONAL CONFERENCE ON BIO-INSPIRED SYSTEMS AND SIGNAL PROCESSING
2 - 4 MARCH, 2026
MARBELLA, SPAIN
BIOSTEC 2026

Neuromorphic Speech Lab
NeuSpe
©CTB-UPM

Neuro-Acoustic Intelligence: A Unified Phonation-Encephalography Framework for Multi-Disorder Monitoring via Explainable AI

Pedro Gómez Vilda[1,2]

[1]NeuSpeLab, CTB, UPM, 28220 Pozuelo de Alarcón, Madrid, Spain

[2]E.T.S.I. Informática, Universidad Rey Juan Carlos, 28933 Móstoles, Madrid, Spain

E-mail: pedro.gomezv@upm.es

Motivations

Phonation alterations, such as **musculus vocalis dystonia**, vocal fold imbalance, and tremor in voice (altered neuromotor feedback) are some indicators of **neurodegeneration**.

Precise **knowledge of the neural processes** regulating the respiratory and phonation systems is essential to explain the effects of **PD or ASD** on speech in terms of neurological circuit modeling.

Statistical machine learning and neuromechanistic models for biomedical applications can provide an **elevated level of insight** to characterize the underlying mechanisms in **neuromotor and cognitive-related disorders**.

Objectives

To **describe phonation based on a neuromechanical model** in terms of amplitude distributions of the neuromechanical activity of the vocal folds during the emission of sustained vowels.

To propose the **concept of the phonation encephalographic activity** (phEGA) estimated from the laryngeal neuromotor drive which modulates vocal-fold tension, mapped onto the frequency bands used in EEG.

To test **phEG bands** of the vocal fold body stress (VFBS) **as potential predictors** of **altered neuromotor and/or cognitive behavior** from specific disorders, such as PD or ASD in contrast with normotypical phonation.

Why phonation?

Phonation: High-Complexity Neuromotor Trait

A tightly coordinated behavior **that exposes subtle impairments** in **motor planning and execution**.

Direct Window Into Cortical and Subcortical Dynamics

phEG bands **capture oscillatory patterns** linked to timing, inhibition, and sensorimotor integration.

Sensitive to Early Neurological Decline

Neural coupling might reveal **preclinical changes** in disorders like Parkinson's, ALS, and dystonia.

Integrates Motor, Cognitive, and Emotional Systems

Phonation simultaneously reflects **executive control, cognitive load, and affective modulation**.

Non-Invasive, Scalable, and Easy to Elicit

Ideal for **screening, longitudinal monitoring, and telemedicine applications**.

Provides Neural Signatures Beyond Acoustics Alone

phEG–voice coupling **uncovers mechanisms that acoustic features cannot explain**.

A Multidimensional Biomarker for Clinical and Research Use

Combines **acoustic, neuromotor, and neurocognitive information in a single, unified measure**.

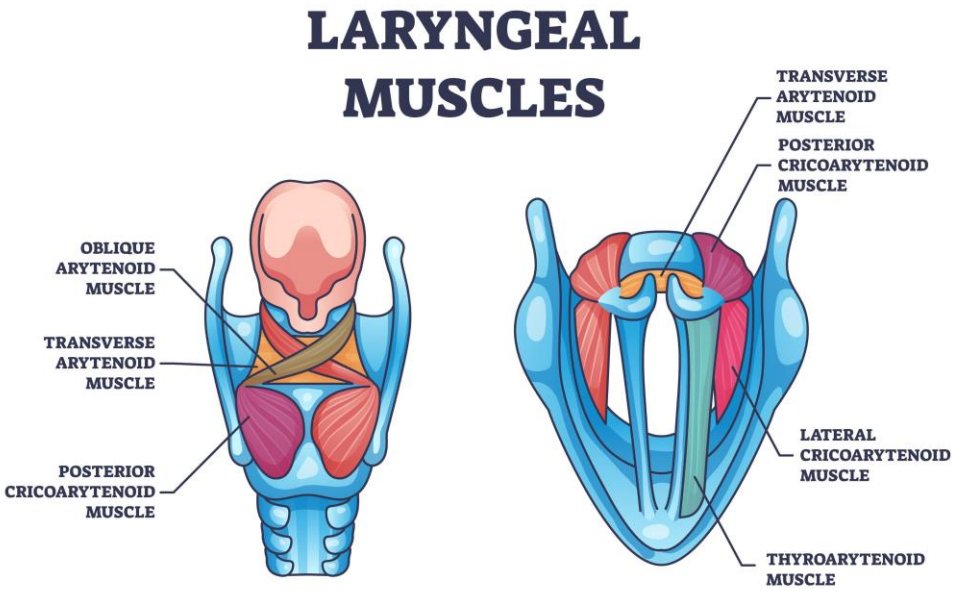
Larynx Muscles

The study utilizes the agonist-antagonist relationship of laryngeal muscles as predictive features:

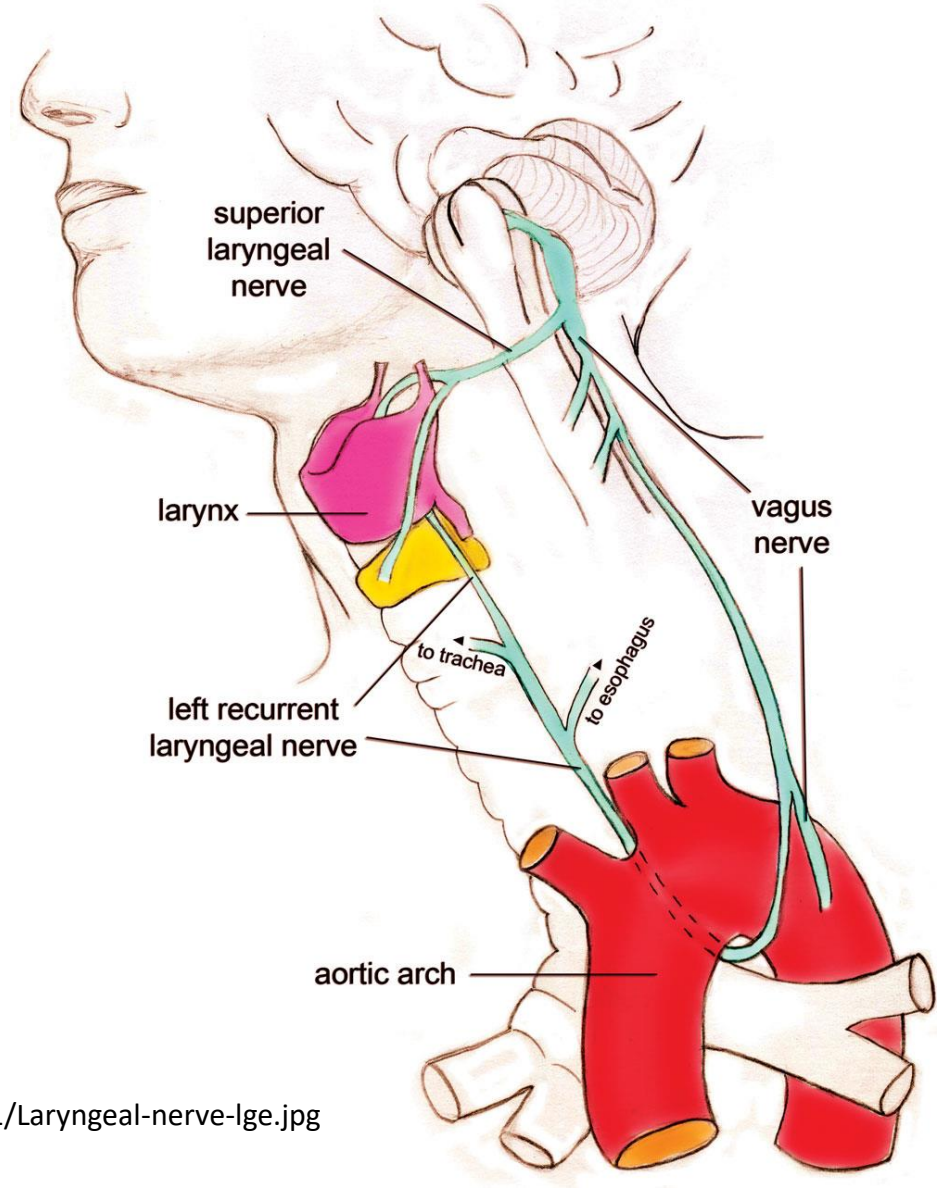
Cricothyroid (CT): Acts as the vocal fold *stretcher* Contraction increases tension and length.

Thyroarytenoid (TA): Acts as the vocal fold *relaxer* and thickener. Shortens and slackens the folds.

These dynamics are estimated via **Dynamic Neuromotor Alteration (DNMA)** correlates.



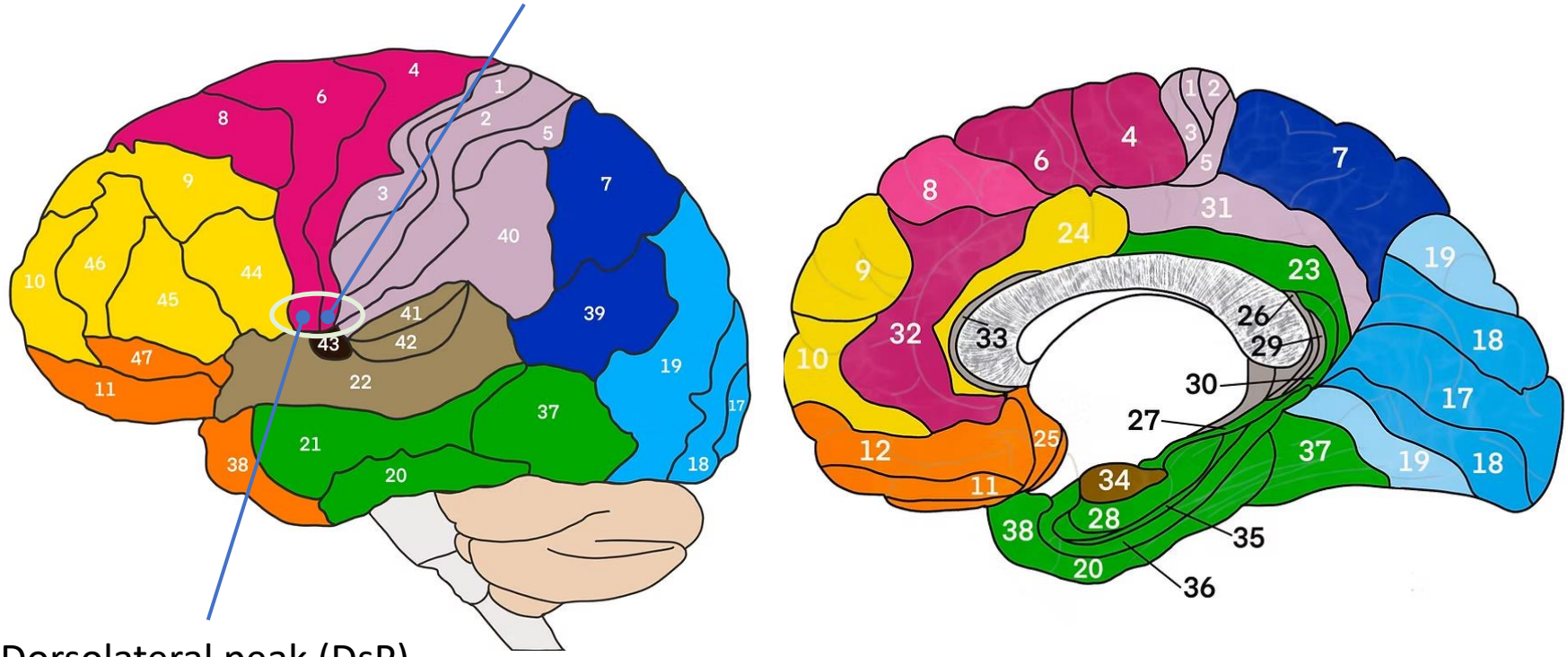
Laryngeal nerves



https://dl0.creation.com/articles/p097/c09721/Laryngeal-nerve-lge.jpg

Brodmann areas

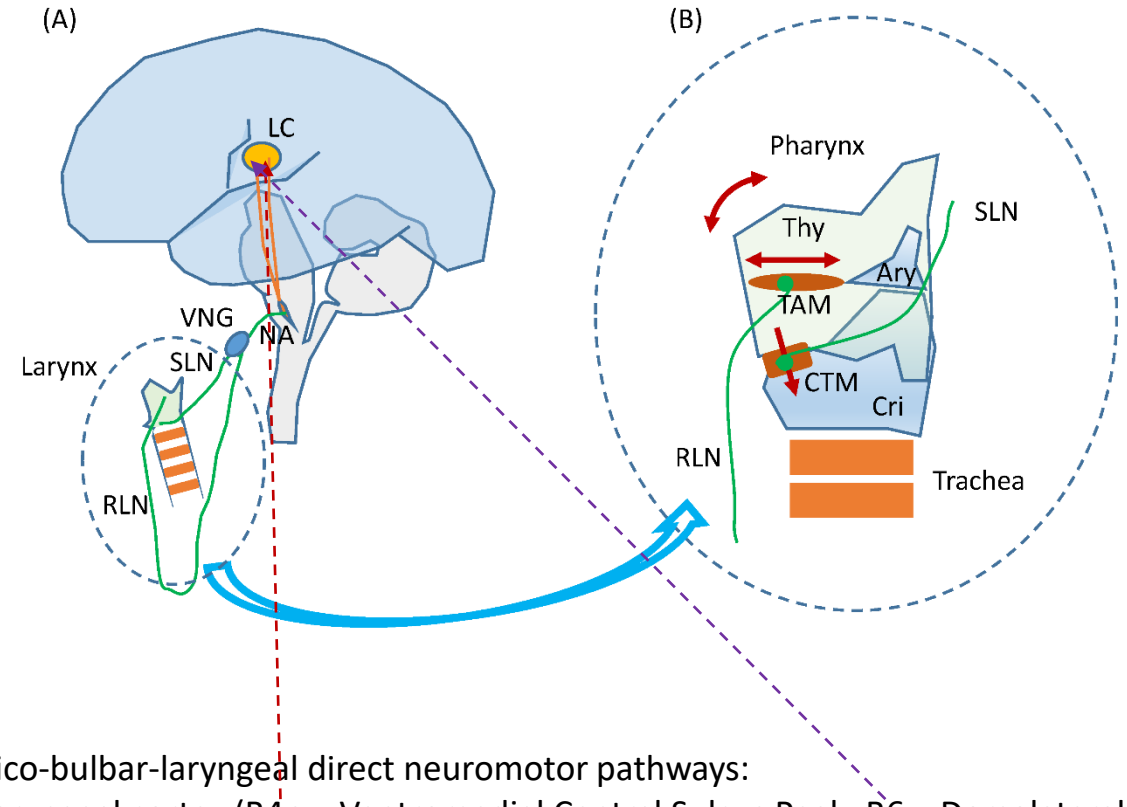
Ventromedial central sulcus peak (VmCSP)



Dorsolateral peak (DsP)

https://www.biosourcesoftware.com/post/comprehensive-brodmann-area-guide

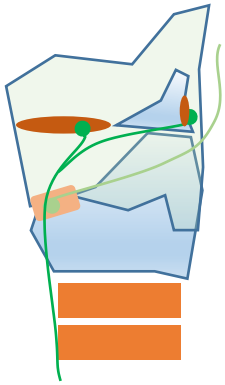
Laryngeal agonist-antagonist control



Thy: thyroid cartilage.
Ary: arytenoid cartilages.
Cri: cricoid cartilage.
TAM: thyroarytenoid muscle.
CTM: cricothyroid muscle.

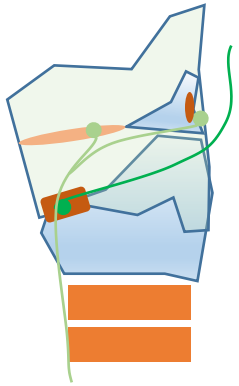
Cortico-bulbar-laryngeal direct neuromotor pathways:
LC: laryngeal cortex (B4p – Ventromedial Central Sulcus Peak; B6 – Dorsolateral Peak).
NA: nucleus ambiguous in medulla oblongata.
VNG: extracranial vagus nerve ganglia.
SLN: superior laryngeal nerve.
RLN: retrolaryngeal nerve.
The agonist-antagonist role that CTM and TAM will play in the cricothyroid joint is represented by the curved double-arrow line.

Cryco-Thyro-Aritenoid Joint



CTA: VF relaxer

RLN Active, SLN: Inactive



CTA: VF tensor

RLN: Inactive, SLN: Active

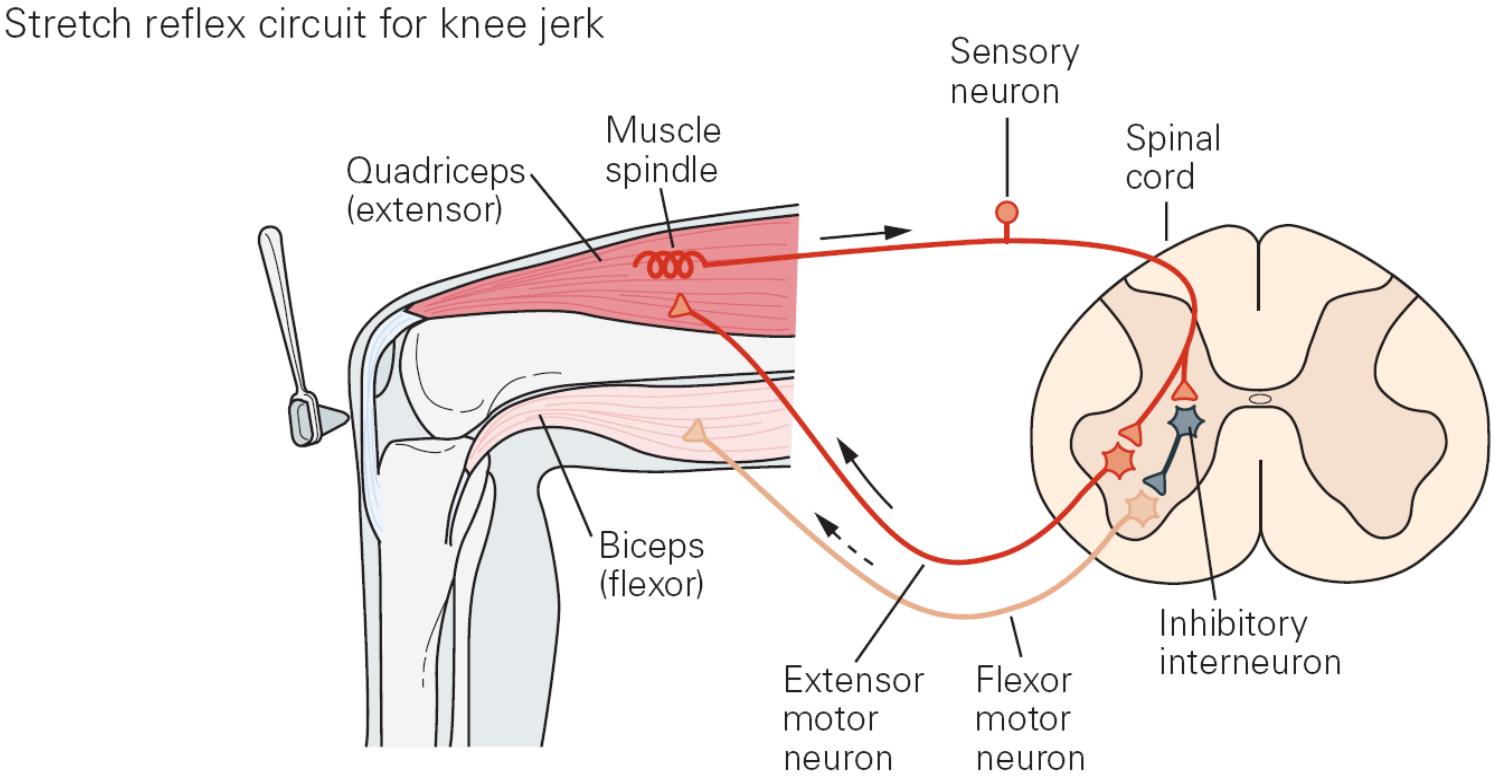
Agonist-Antagonist Neuromotor Activity

Muscles contract under neuromotor activity

To produce controllable movement muscles must act in pairs

Excitation and inhibition motor neurons intervene in the process

(E. Kandel, Ed. "Principles of Neural Science")



Working hypothesis

The differentiated neuromotor activity of **agonist** (ct: cricothyroid ta: thyroarytenoid) from **antagonist** (ta: thyroarytenoid) can be estimated by the **unbiased vocal fold body stiffess** (VFBS) transforming neural discharges into muscle contraction:

Direct Neuromechanical Pathway:

NMA → LPA→ VFBS → GS→ RP

Inverse Neuromechanical Estimation:

RP → GS → VFBS→ phEGB

NMA: neuromotor activity;

LPA: Laryngeal pathway activity;

VFBS: Vocal Fold Body Stiffness;

GS: Glottal Source;

RP: Recorded Phonation;

phEG: phonation Encephalographic Bands;

Cortico-limbical oscillatory loops in phEG

A. Motor Loop Disorders: Parkinson's Disease (PD) & ALS

- **The Mechanism:** Pathological synchronization in the basal ganglia-cortical loops.
- **phEG Application:** PD is characterized by *hypophonia* and reduced vocal pitch range. The uSI can detect the "locked" motor rhythms before clinical dysarthria becomes obvious. In ALS, phEG can monitor the early degradation of the bulbar motor neurons.

B. Cognitive & Connectivity Disorders: Alzheimer's (AD) & Dementia

- **The Mechanism:** Cortical atrophy and breakdown of long-range synchronization.
- **phEG Application:** AD affects the timing and "prosodic fluidness" of speech. The uSI could serve as a marker for the loss of temporal coordination between cognitive intent (frontal) and motor execution (laryngeal).

C. Affective & Limbic Disorders: Depression & Anxiety

- **The Mechanism:** Chronic cortical-limbic disconnect; "Limbic Overdrive."
- **phEG Application:** Depression often results in a "flat" or "monotone" vocal signature. Your theory suggests this is a failure of the limbic system to provide the necessary "desynchronizing" signals to the motor cortex. The uSI would show a lack of spectral volatility.

D. Social Cognition & Affective Processing: Schizophrenia

- **The Mechanism:** Atypical integration of sensory-motor and social-emotional information.
- **phEG Application:** Similar to ASD, Schizophrenia involves deficits in "vocal affect" (prosody). phEG markers can distinguish between the social-motor failure of ASD vs. the disorganized neural firing of Schizophrenia.

Fundamentals of Binary Detection: Completion and Balancing

حساب الجبر و المقابلة‘

Al-Kitāb al-mukhtaṣar fī hisāb al-ŷabr wa'l-muqābala

The Compendious Book on **Calculation by Completion and Balancing**

Abu Abdallah Muḥammad Ibn Mūsā **Al-Khwarizmī** (Abu Ŷāffar) 780-850 aD

Born in Khiva (χiβa) of **Khorezm** (χorazm), Uzbekistan, in the northwest of the country, near the Amu Darya River.

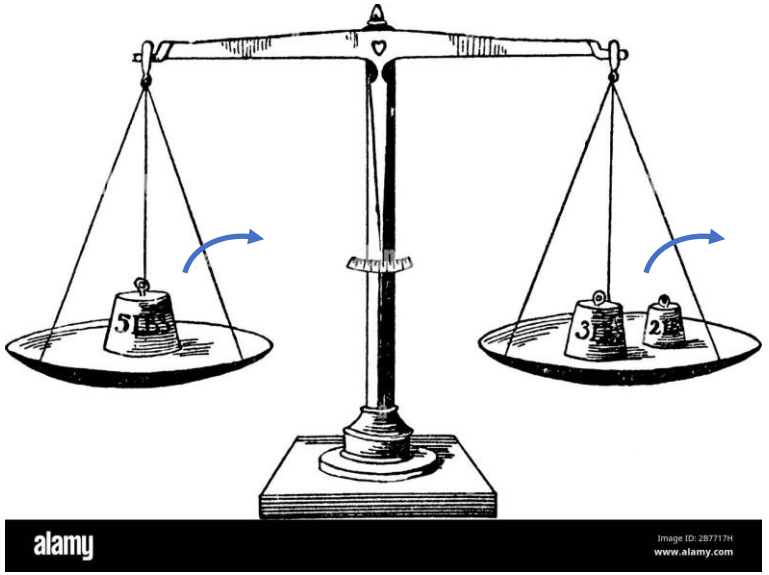
He proposed **algebraic concepts** and Hindu-Arab **numeral system** to Western knowledge.

It was translated into latin by **Gherardo Cremonensis** (Cremona, Italy, 1114 – Toledo, Spain, 1187)

The balance scales: The mother of Algebra

A classical two-plate balance scale is made of:

- Beam
- Fulcrum
- Pillar
- **Two pans**
- **Pointer**
- Base
- Standard weights
- Calibration screws



$$a + b = c; \text{ to solve for } a:$$

$$a + b - b = c - b \rightarrow a = c - b$$

An algorithm is a recipe…

Algorithm is a latinization of the demonym in Abu Abdallah Muḥammad Ibn Mūsā's name: **Al-Khwarizmī** (alχωarizmí), the person from **Khorazm**.

In Spanish the Word 'guarismo' stands for 'digit, or number figure', 'algarismo', in ancient Spanish, and in Catalan 'guarisme' is used.

Western science owes an enduring debt to Hindus, Persians, Arabs, and to scholars like Gherardo of Cremona, among others.

Essentially, an algorithm is a step-by-step recipe for cooking numbers, following the ancient rules of Al-Jabr…

The menu: PD stratification

The study involved 48 participants stratified into gender-balanced cohorts, comparing older Parkinson's and healthy control groups (mean ages 63–69) against a younger normotypical reference population (mean ages 32–35):

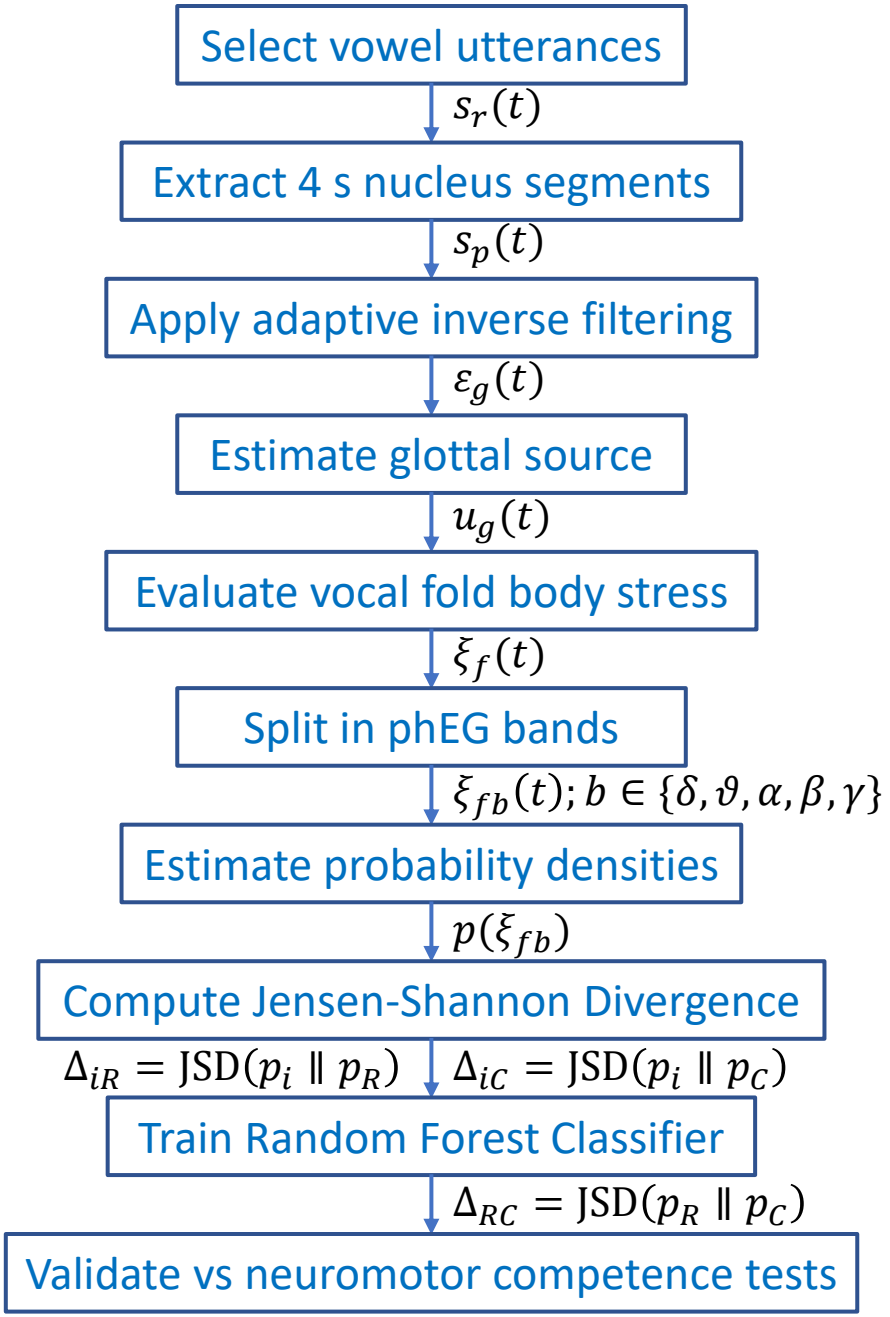
- **Dual-Site Data Sourcing**: Integration of the **PARCZ** database (Brno, Czech Republic) for pathological/control cohorts and the **HUGMM** database (Madrid, Spain) for stringent normotypical references.
- **Cohort Stratification**: Comprehensive analysis of sustained **[a:]** vowels from Parkinson's Disease (**PD**) patients, age-matched **Healthy Controls**, and a middle-aged **Reference** group.
- **Standardized Recording Protocol**: High-fidelity acquisition in sound-isolated environments (**< 30 dB SPL**) using cardioid microphones at a fixed **20 cm** distance to ensure signal consistency.
- **Signal Refinement**: Processing of **4-second** stable vowel segments, downsampled to **16 kHz** to focus on clinically relevant frequency components while maintaining 16-bit resolution.
- **Neuromechanical Inversion**: Implementation of **24th-order adaptive inverse filtering** to decouple the oro-naso-pharyngeal tract (ONPT) and isolate the **glottal correlate (VFBS)**.

Demographical description

Age distribution of participants			
Males	MHC	MPD	MRS
No. Part.	8	8	8
mean age	64.88	65.13	32.13
std. age	9.51	8.74	8.94
Females	FHC	FPD	FRS
No. Part.	8	8	8
mean age	62.75	69.25	35.38
std. age	2.87	7.11	12.84

Labels: FPD as P1xxx, MPD as P2xxx, FHC as C1xxx, MHC as C2xxx, MRS as N10xx and FRS as N11xx.

Processing pipe-line



The First Ingredient: phEG bands

The VFBS is split onto the following bands by a bank of band-pass filters:

$$\xi_{\delta}(t) = \mathcal{F}_{\Omega_{\delta}}\{\xi_b(t)\}/\bar{\xi}_b \,;\, 0 < \Omega_{\delta} \leq 4\ Hz$$

$$\xi_{\vartheta}(t) = \mathcal{F}_{\Omega_{\vartheta}}\{\xi_b(t)\}/\bar{\xi}_b \,;\, 4 < \Omega_{\vartheta} \leq 8\ Hz$$

$$\xi_{\alpha}(t) = \mathcal{F}_{\Omega_{\alpha}}\{\xi_b(t)\}/\bar{\xi}_b \,;\, 8 < \Omega_{\alpha} \leq 16\ Hz$$

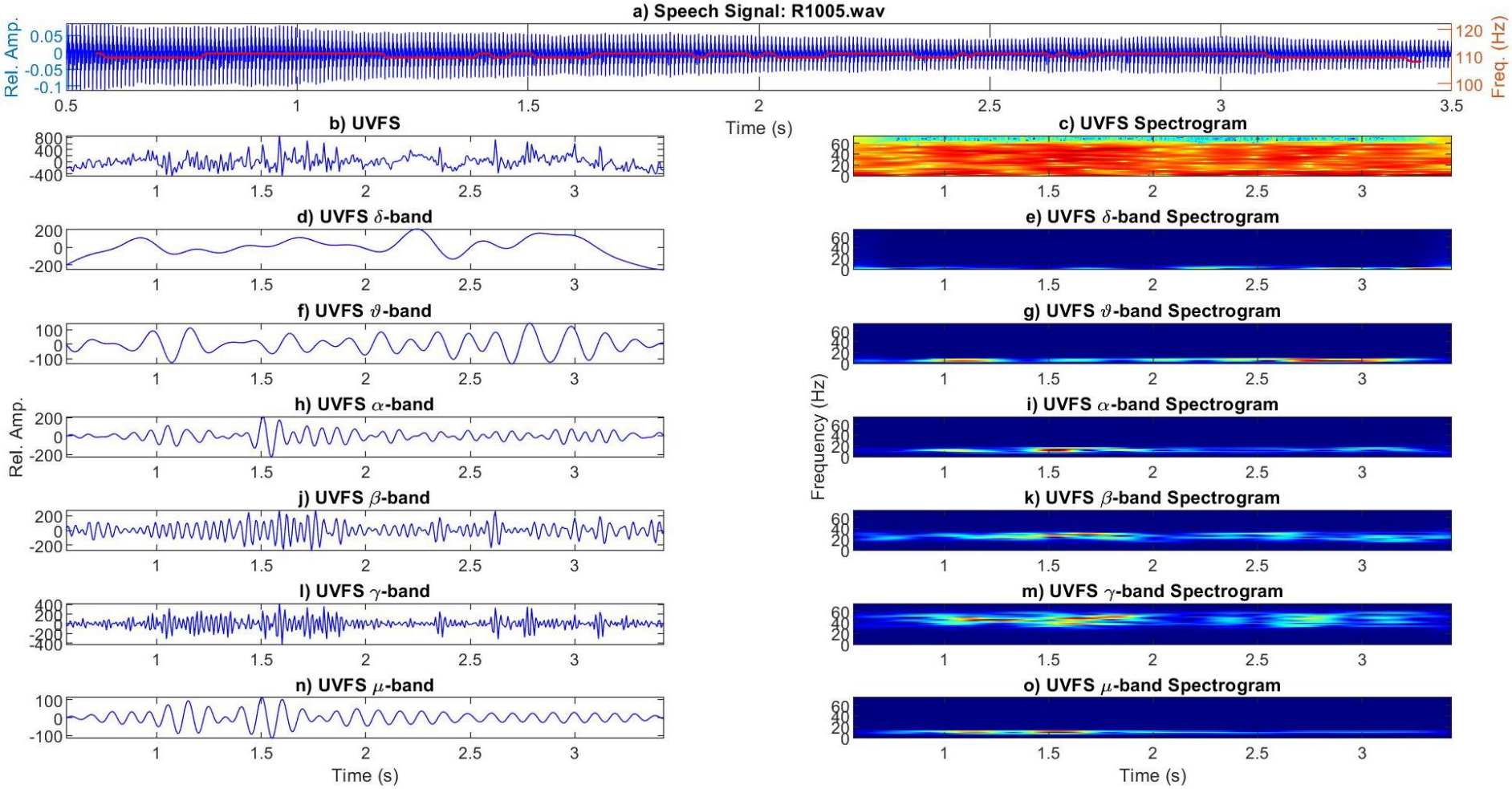
$$\xi_{\beta}(t) = \mathcal{F}_{\Omega_{\beta}}\{\xi_b(t)\}/\bar{\xi}_b \,;\, 16 < \Omega_{\beta} \leq 32\ Hz$$

$$\xi_{\gamma}(t) = \mathcal{F}_{\Omega_{\gamma}}\{\xi_b(t)\}/\bar{\xi}_b \,;\, 32 < \Omega_{\gamma} \leq 64\ Hz$$

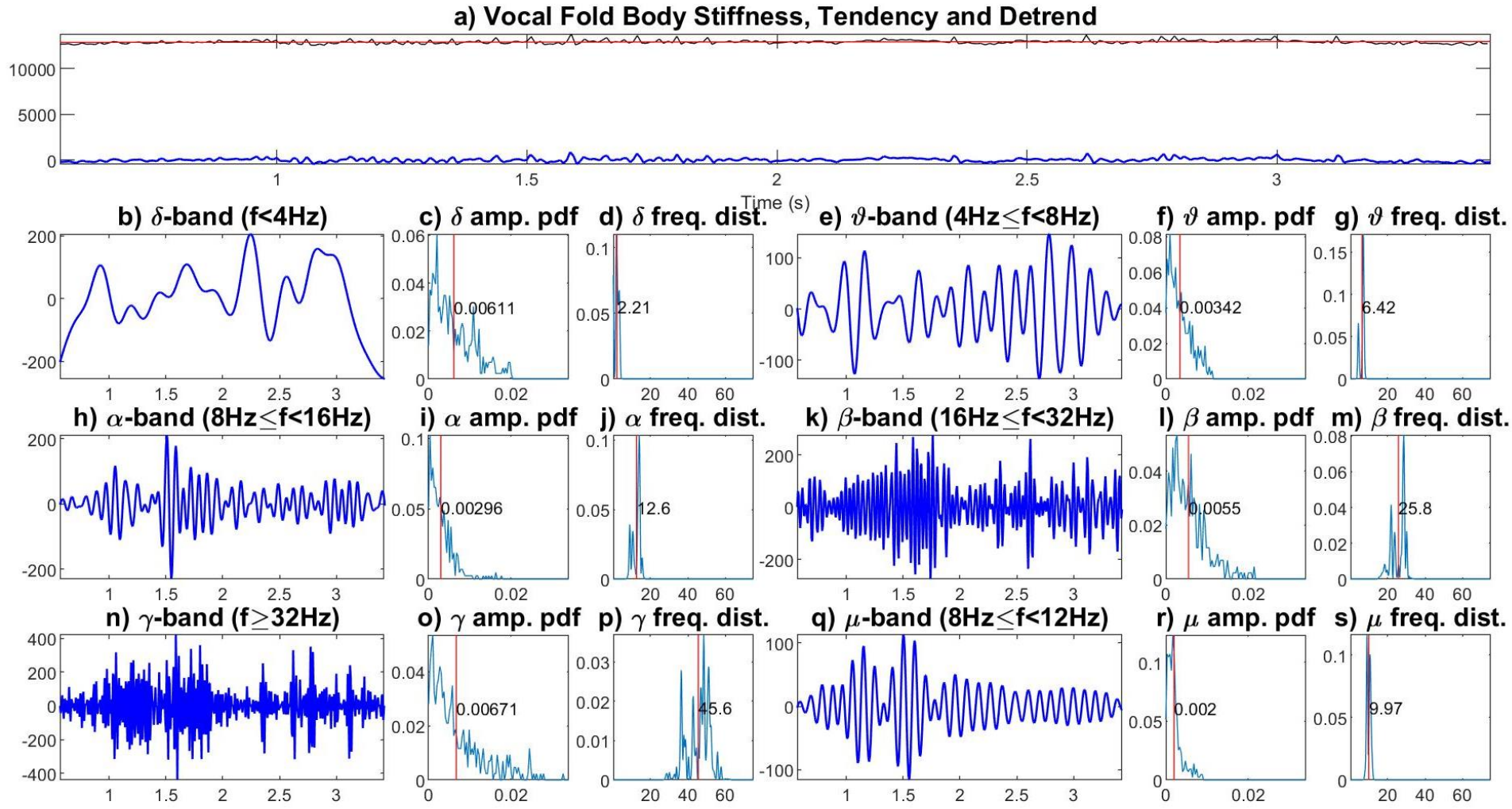
$$\xi_{\alpha l}(t) = \xi_{\mu}(t) = \mathcal{F}_{\Omega_{\mu}}\{\xi_b(t)\}/\bar{\xi}_b \,;\, 8 < \Omega_{\mu} \leq 12\ Hz$$

$$\xi_{\alpha h}(t) = \mathcal{F}_{\Omega_{\mu}}\{\xi_b(t)\}/\bar{\xi}_b \,;\, 12 < \Omega_{\mu} \leq 16\ Hz$$

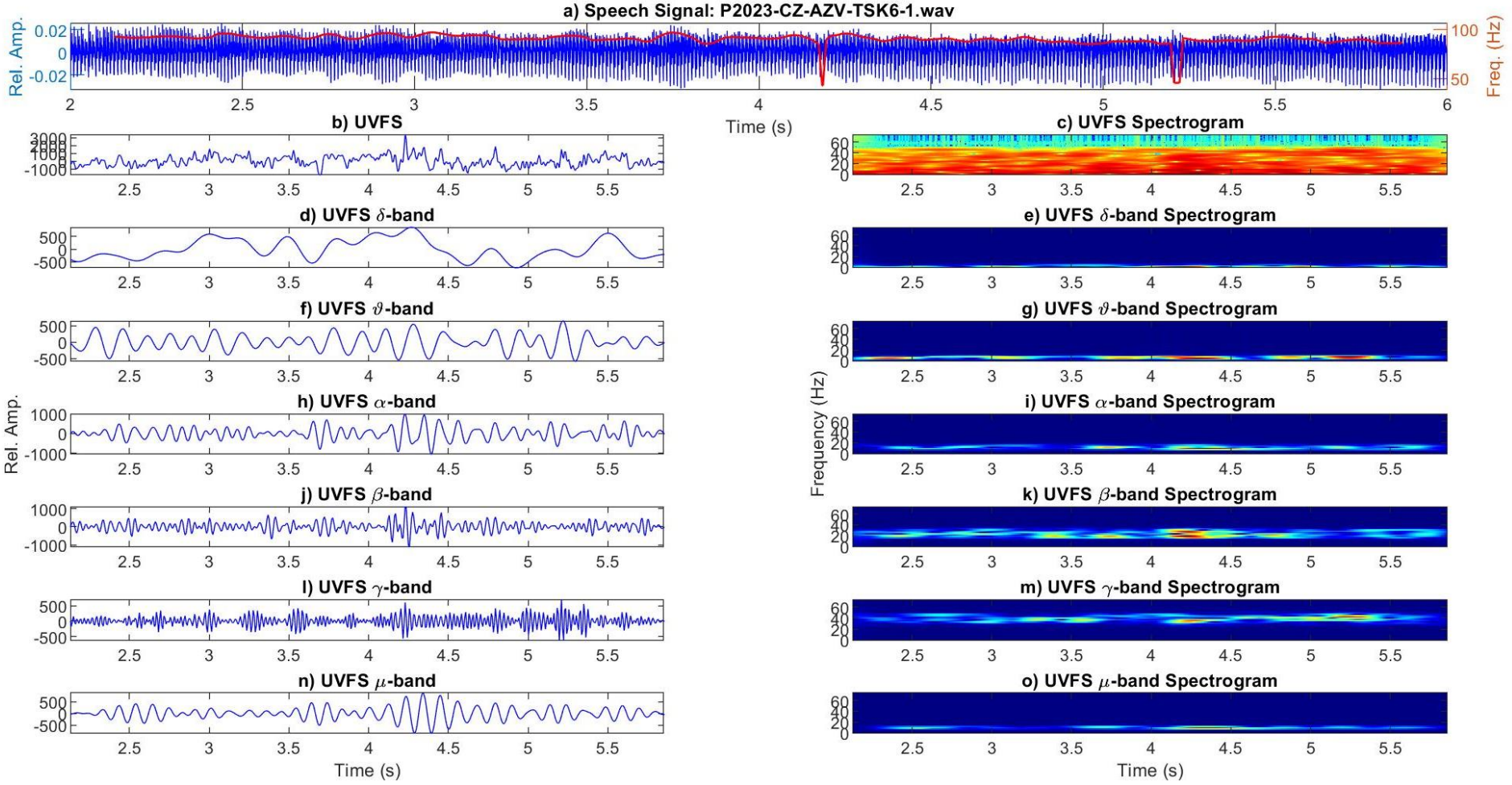
Male NP phEG Spectrograms (e.g.)



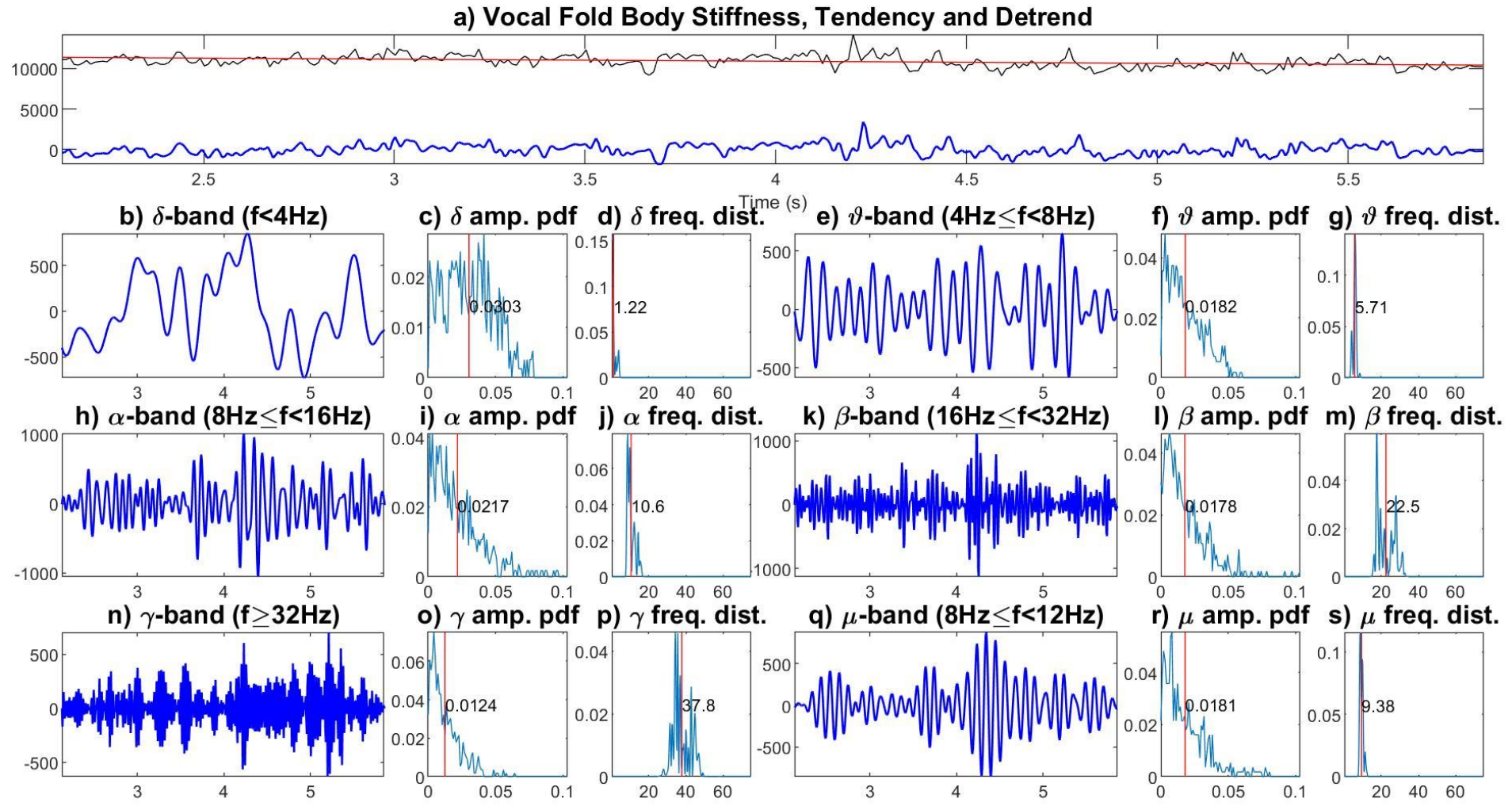
Male NP phEG Signals (e.g.)



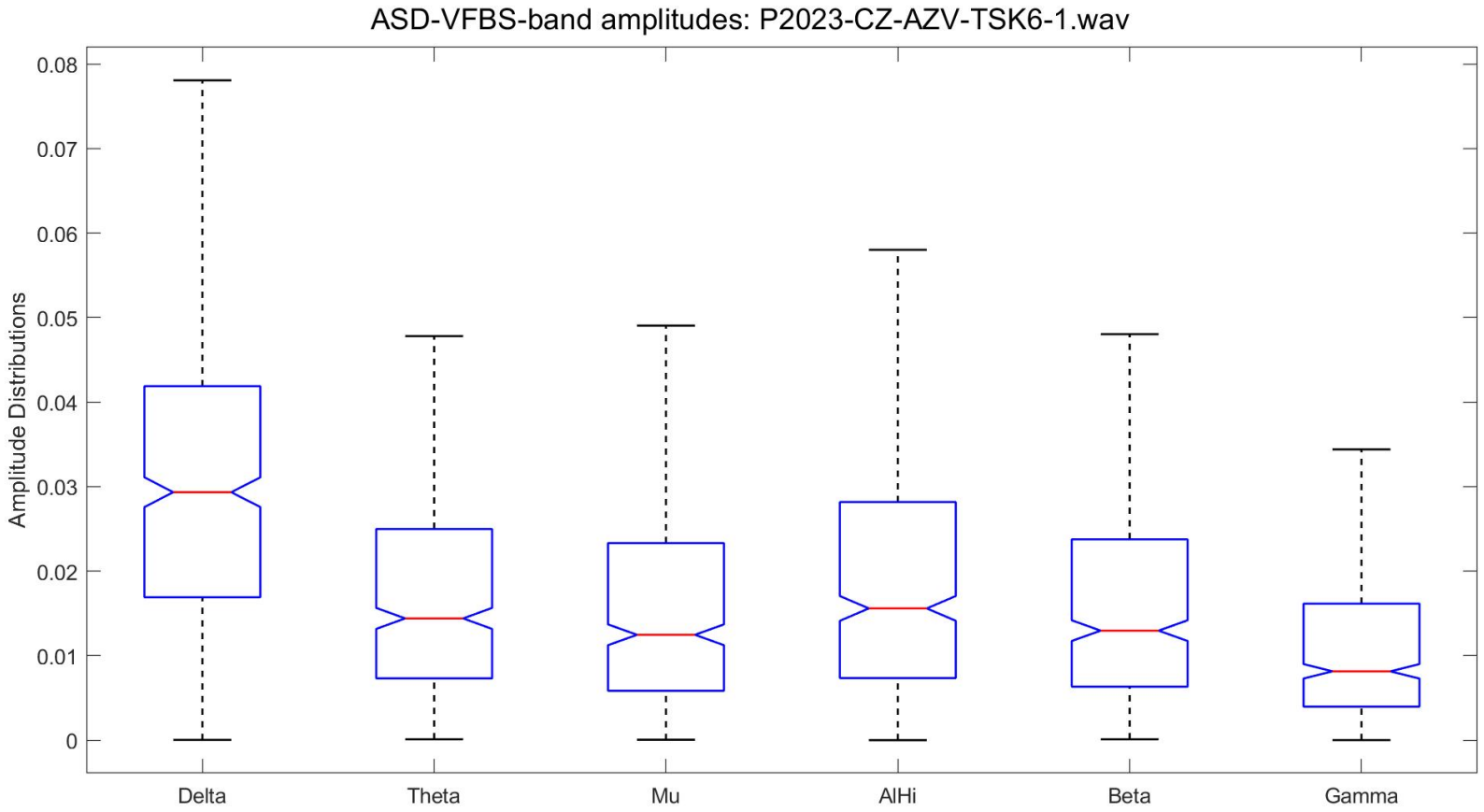
Male PD phEG Spectrograms (e.g.)



Male PD phEG Signals (e.g.)



Male PD Band Distributions (e.g.)



The α complex

Alpha-Band Activity (8–12 Hz): The Sensory Gate - The α rhythm is primarily interpreted through the Inhibition-Timing Hypothesis. It acts as a mechanism for "gating" information by actively inhibiting task-irrelevant cortical areas.

Mechanism: High α power represents a functional "pulsed inhibition" that reduces the firing rate of local neurons. By increasing α power in regions not needed for a task (e.g., visual cortex during a speech task), the brain increases the "signal-to-noise" ratio in relevant areas.

Phonation Context: α modulation **controls cortico-limbic communication** and internal "noise," allowing the motor cortex to focus exclusively on laryngeal control.

Reference: Klimesch, W. (2012). α-band oscillations, working memory, and control: The inhibition-timing hypothesis. State of the Art, 36(2), 322–337.2.

The β complex

Beta-Band Activity (13–30 Hz): The Status Quo - As established in PD research, β oscillations serve as the "Status Quo" signal, maintaining the current sensorimotor or cognitive state.

Mechanism: β activity is dominant during tonic (steady) muscle contractions. It signals that the current motor set is sufficient and should not be changed. It acts as a top-down inhibitory signal from the motor cortex and basal ganglia that prevents the "leakage" of new, unplanned movements.

Phonation Context: It provides the **"locked-in" tension required for a sustained [a:]** vowel, but in Parkinson's, this becomes a pathological "trap" that prevents motor flexibility.

Reference: Engel, A. K., & Fries, P. (2010). Beta-band oscillations—signalling the status quo? Current Opinion in Neurobiology, 20(2), 156–165.3.

The γ complex

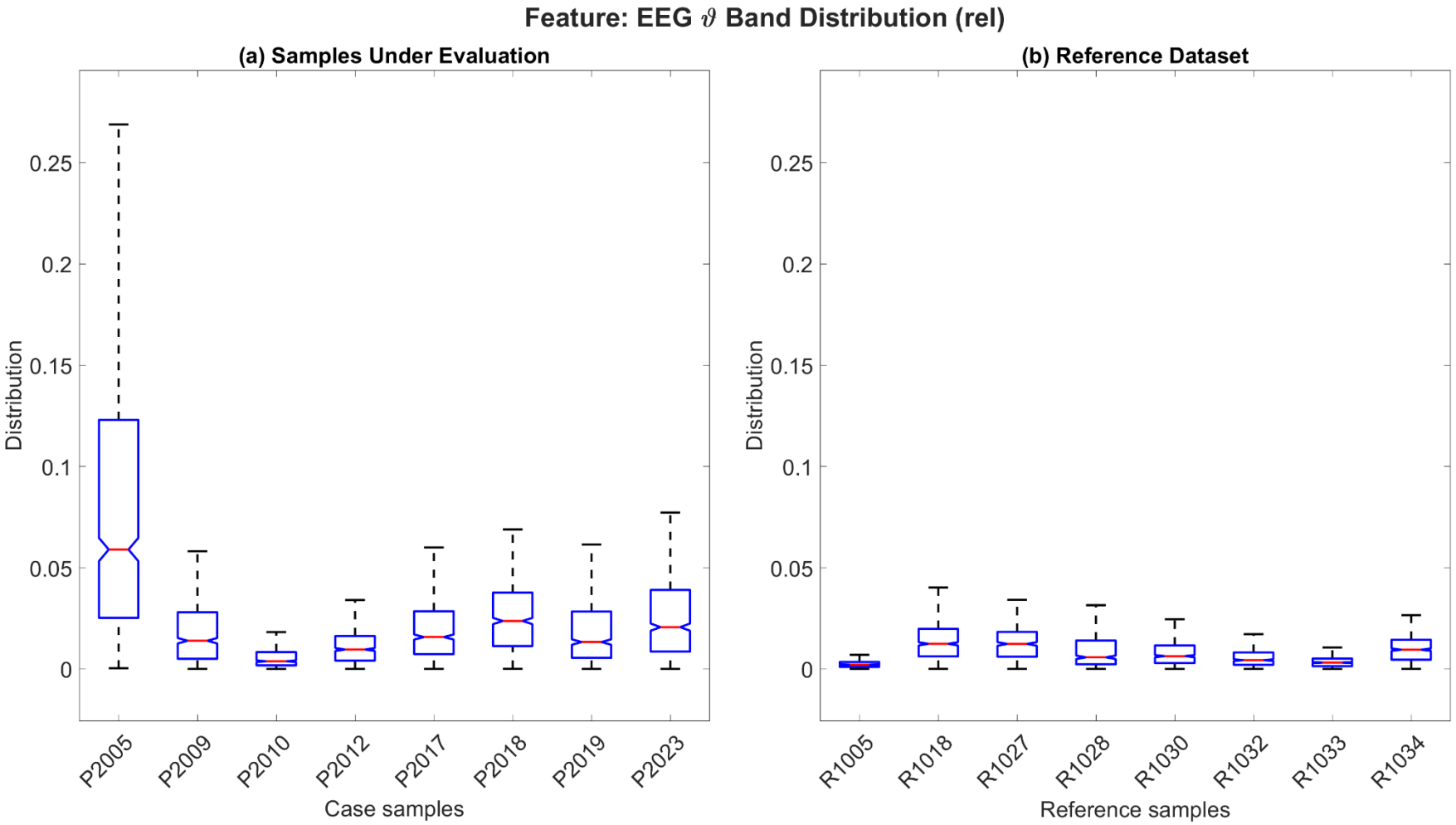
Gamma-Band Activity (>30 Hz): Local ComputationThe γ rhythm is the engine of local information processing and "binding.“

Mechanism: γ oscillations are generated by the fast-spiking interneuron-pyramidal cell (ING/PING) loops. They represent the active firing of neural assemblies as they process specific sensory or motor data. γ is often "nested" within slower ϑ or α waves (Phase-Amplitude Coupling), where the slower wave acts as the carrier and the γ wave carries the specific data packet.

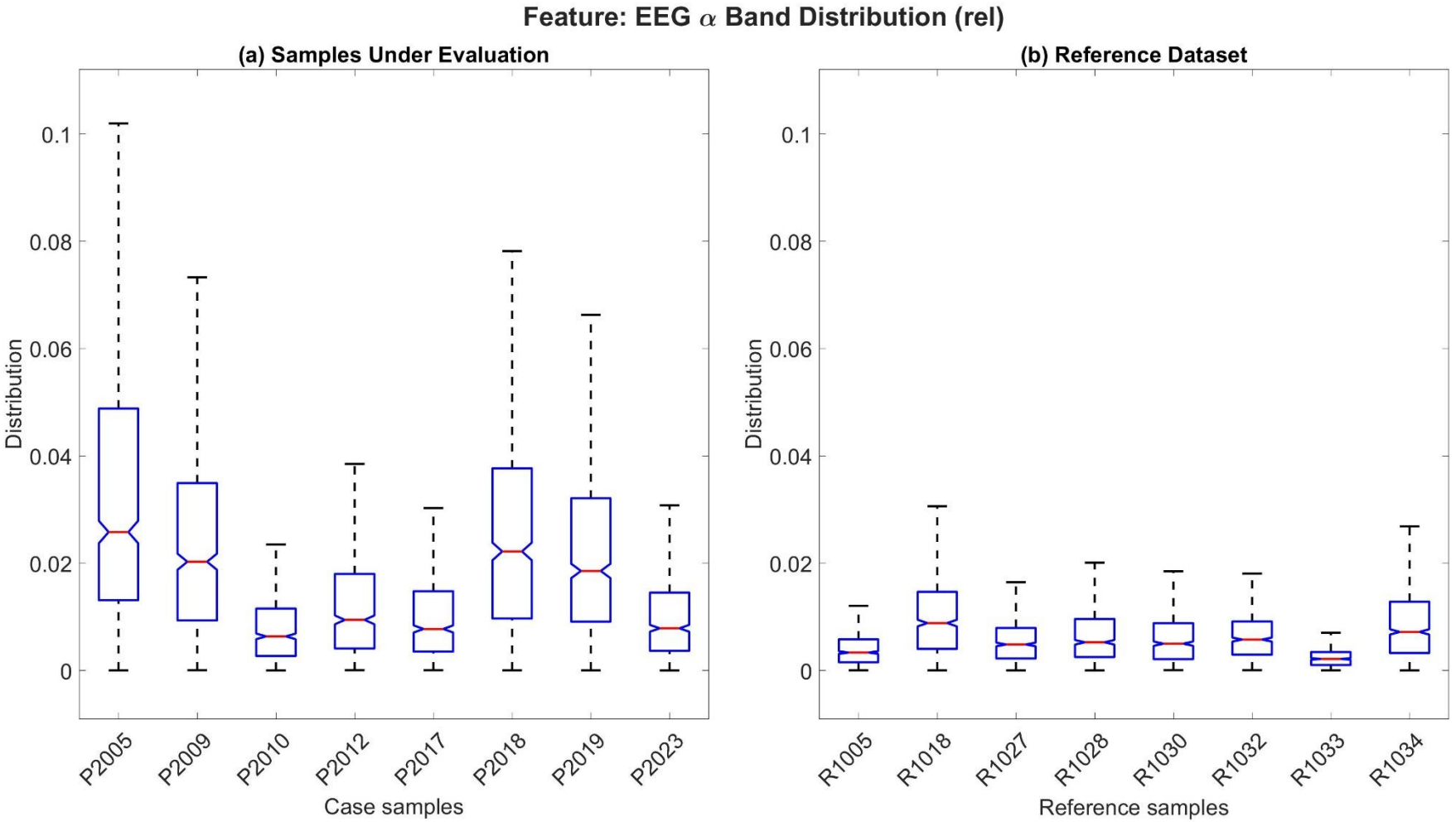
Phonation Context: γ power would reflect the high-speed, localized processing of **fine-grained acoustic feedback and the precise timing of vocal fold micro-adjustments**.

Reference: Buzsáki, G., & Wang, X. J. (2012). Mechanisms of gamma oscillations. Annual Review of Neuroscience, 35, 203–225.

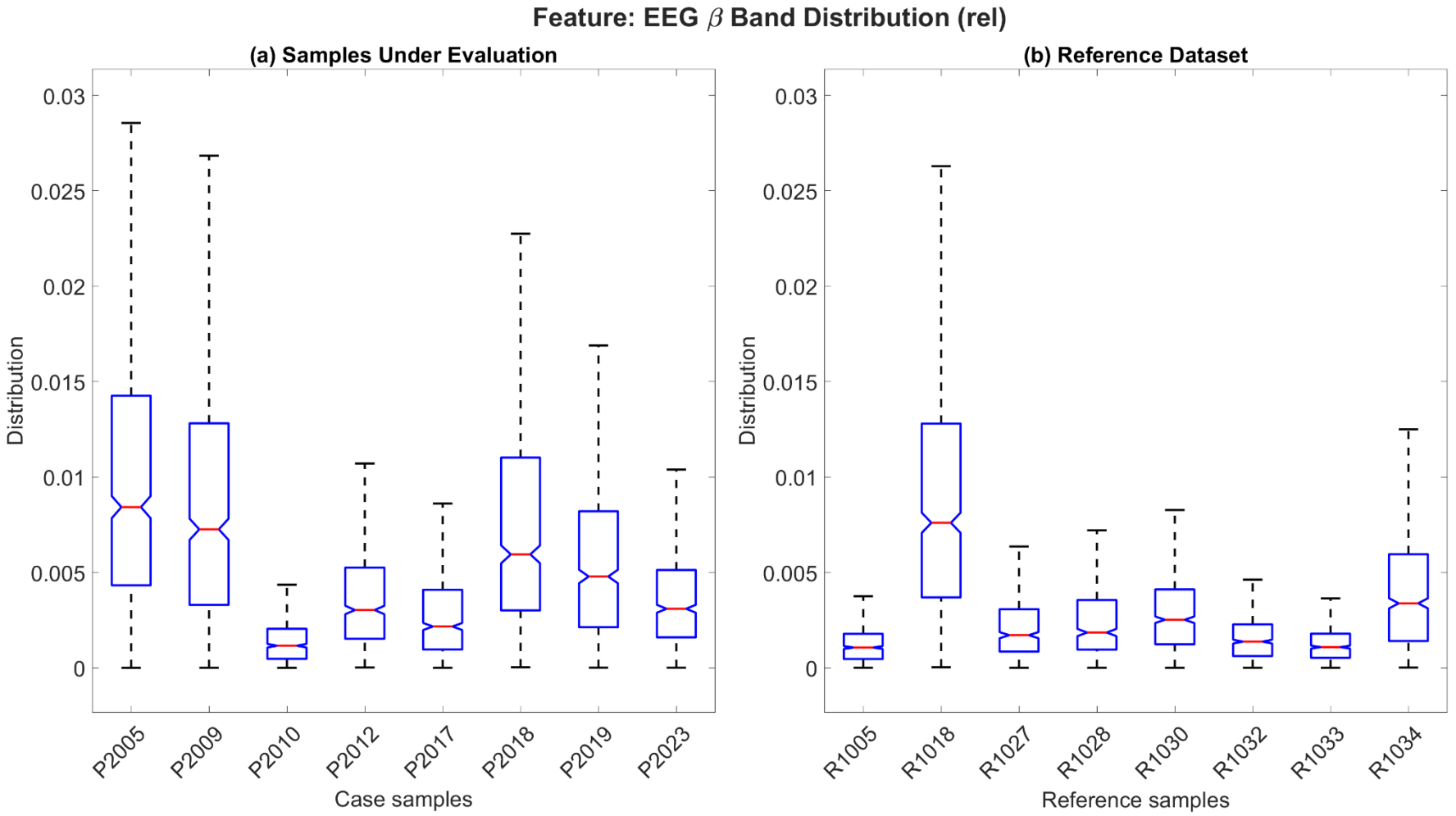
Results: Male Feature Distributions



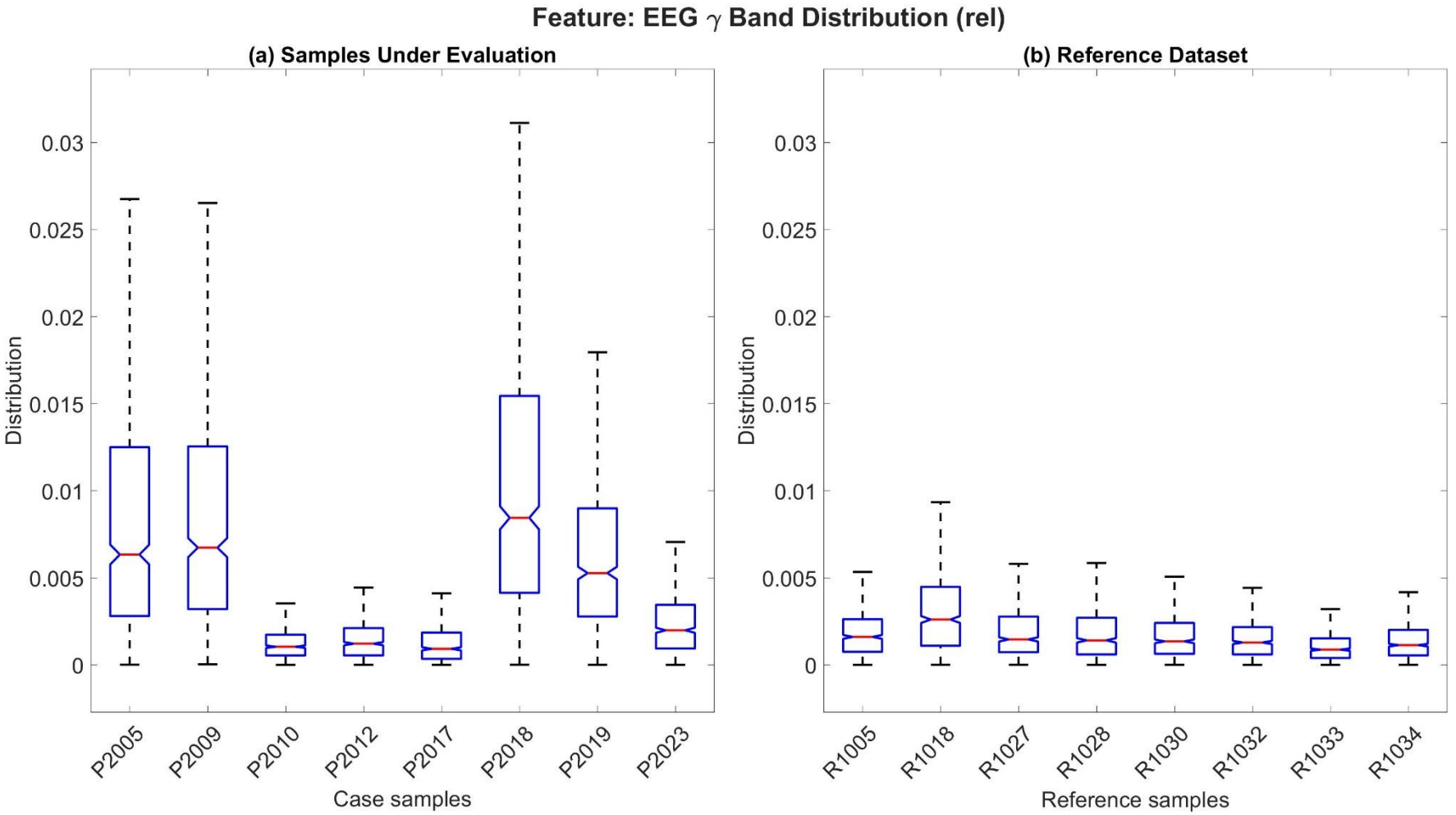
Results: Male Feature Distributions



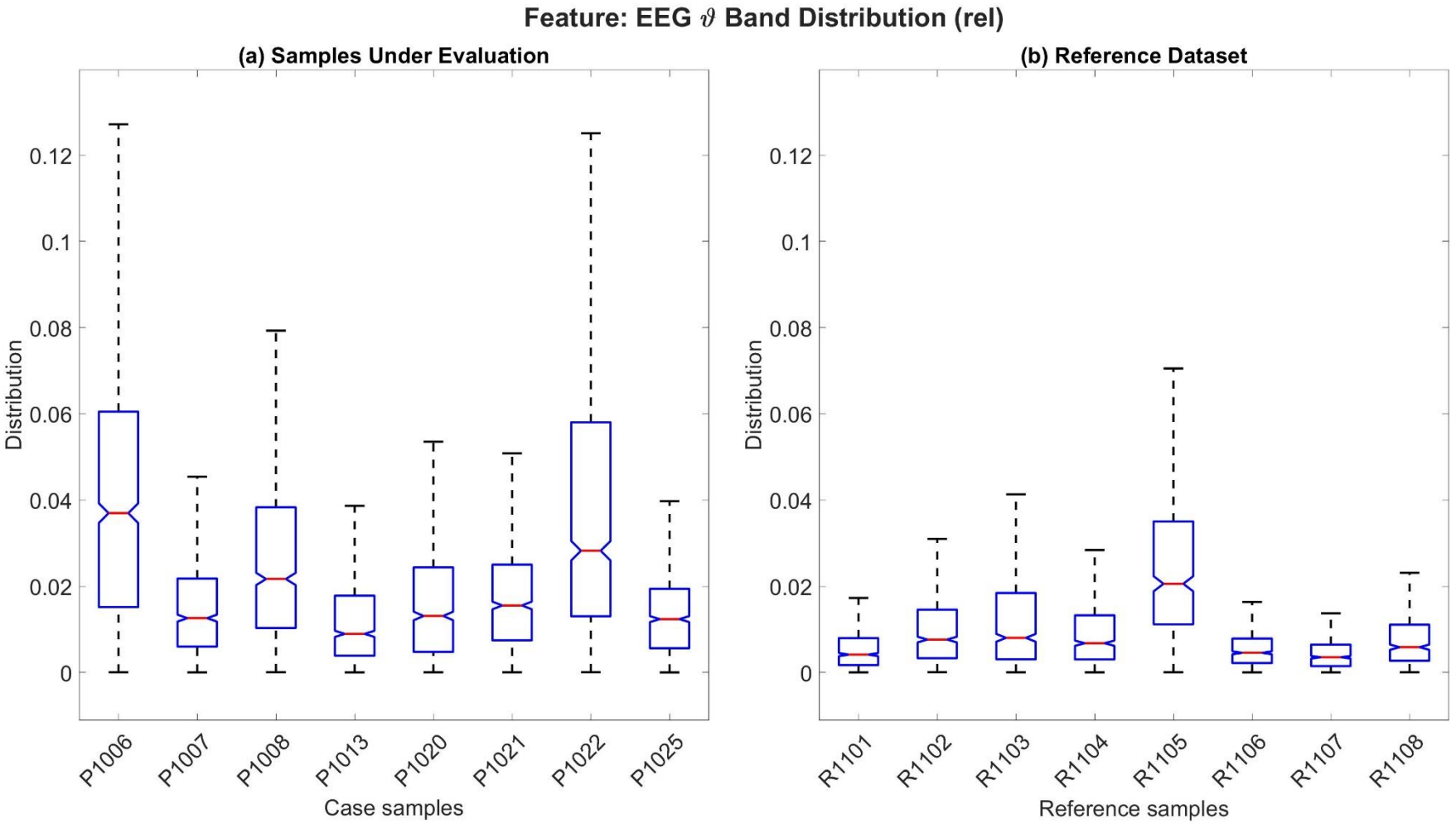
Results: Male Feature Distributions



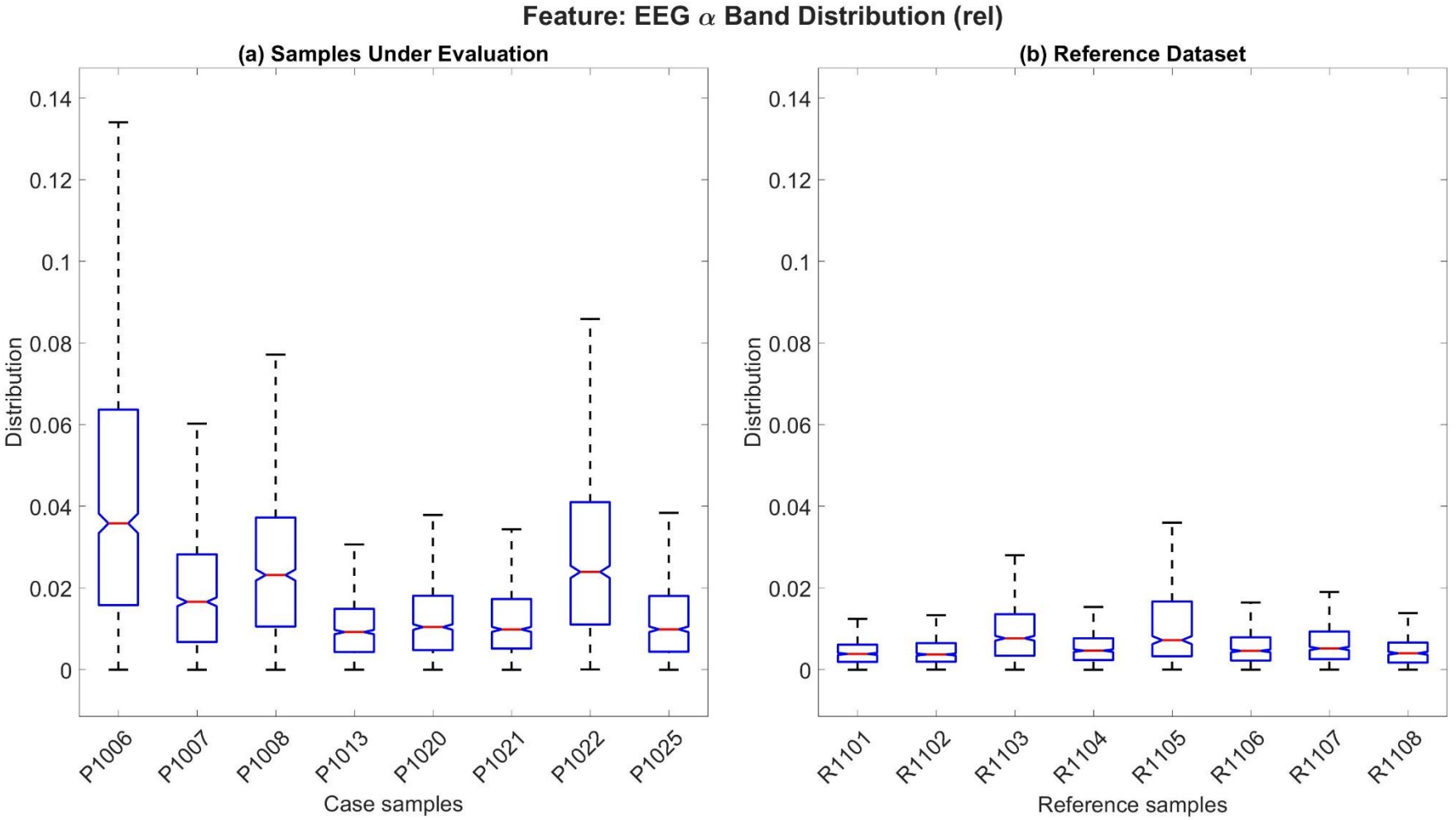
Results: Male Feature Distributions



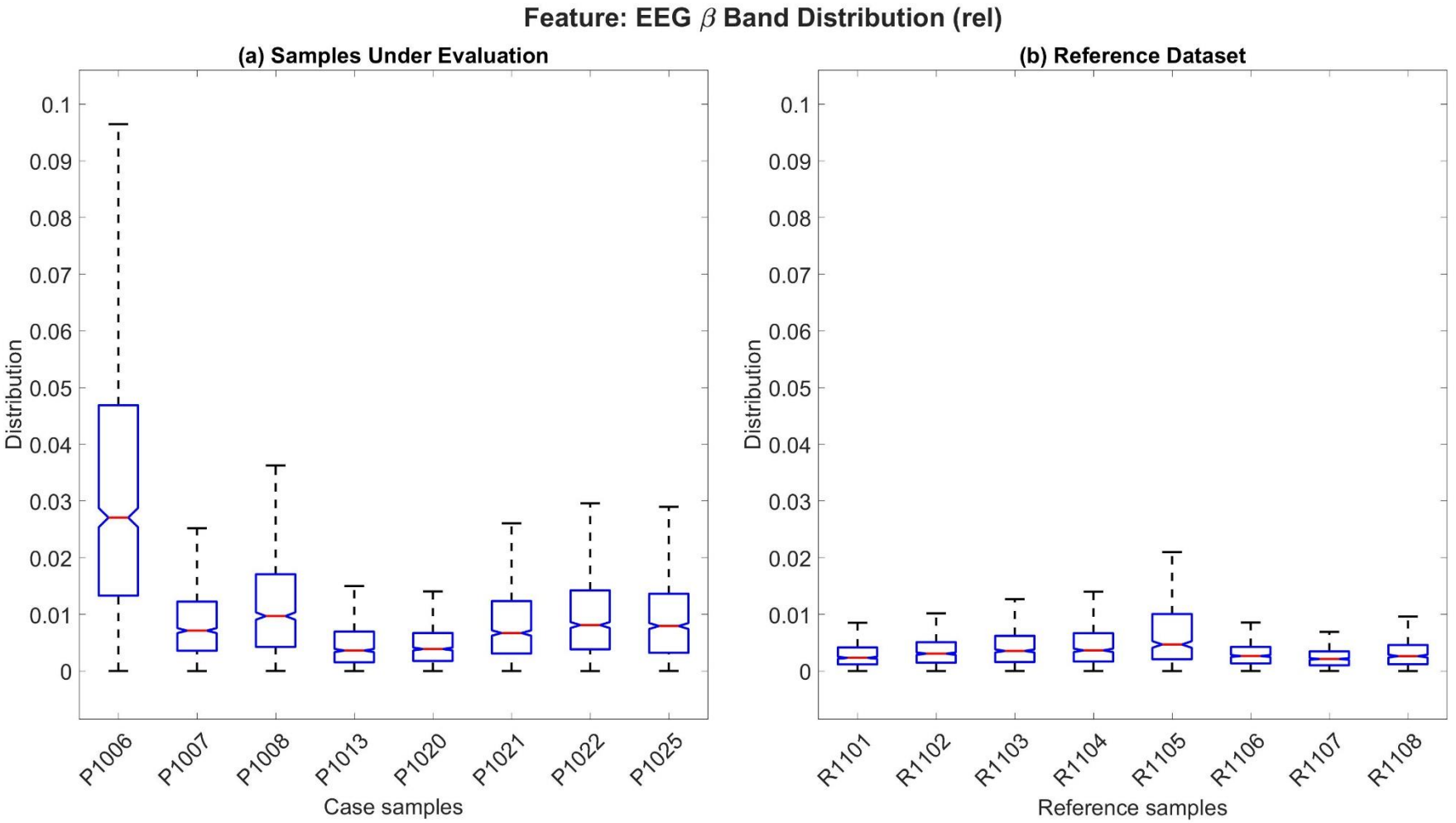
Results: Female Feature Distributions



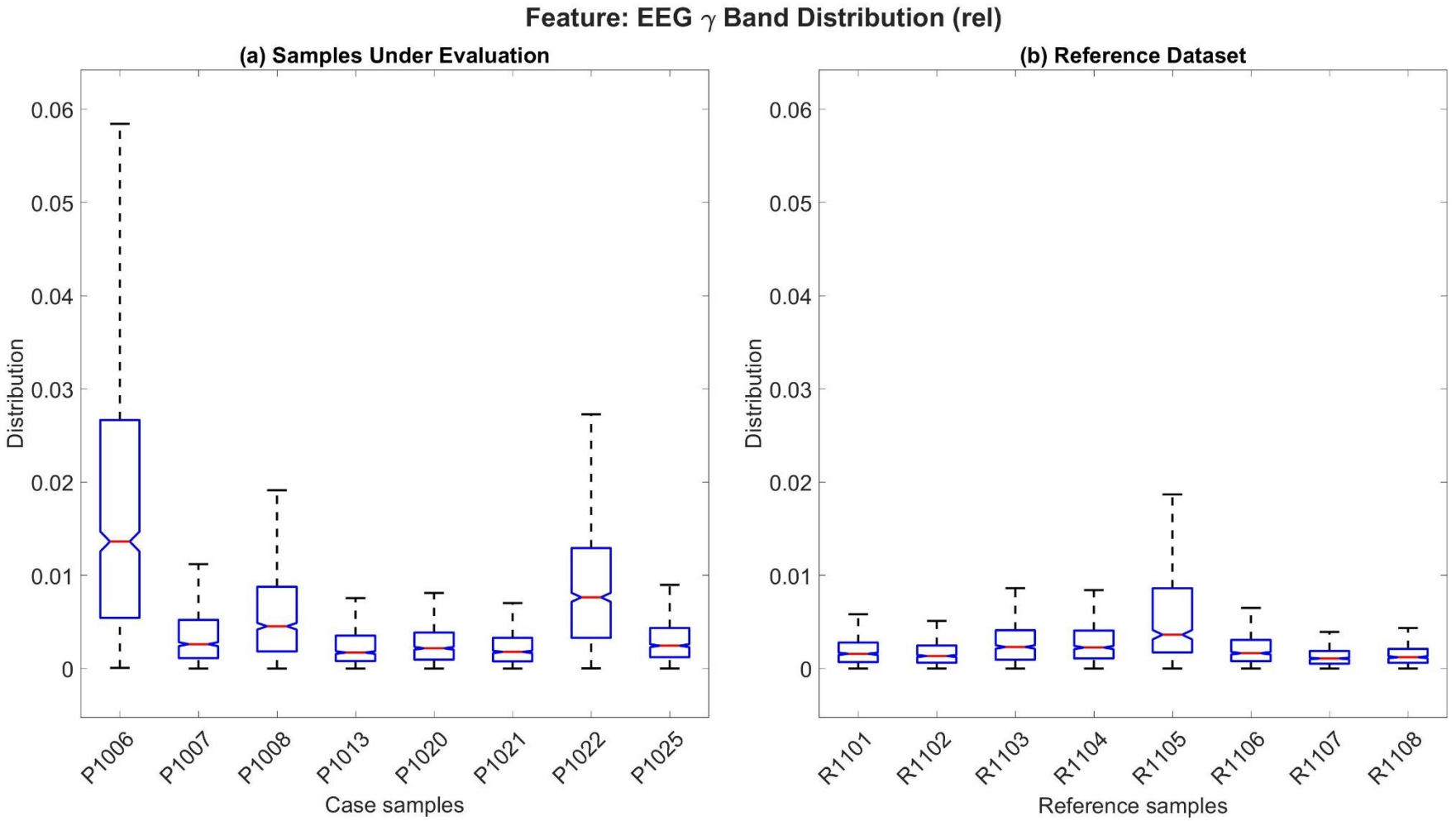
Results: Female Feature Distributions



Results: Female Feature Distributions



Results: Female Feature Distributions



The second ingredient: Jensen-Shannon Metrics

$$\Delta_{JS}(p(\xi)|q(\xi)) = \frac{\Delta_{KL}(p(\xi)|g(\xi)) + \Delta_{KL}(q(\xi)|g(\xi))}{2};$$

$$g(\xi) = \frac{p(\xi) + q(\xi)}{2}$$

$$\Delta_{KL}(p(\xi)|q(\xi)) = \int_0^{\infty} p(\xi) abs\left\{log\frac{p(\xi)}{q(\xi)}\right\} d\xi$$

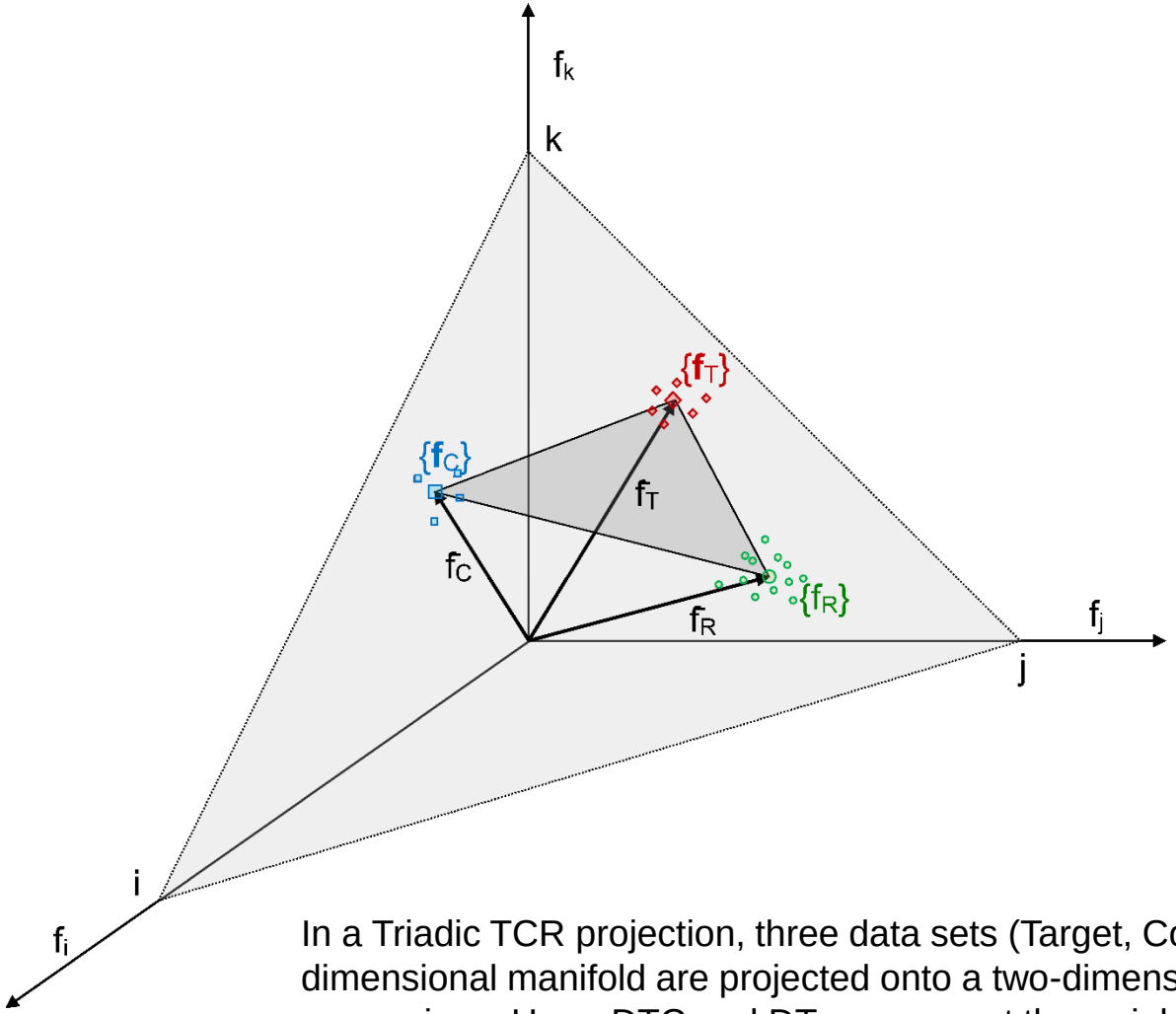
$$\boldsymbol{\Delta_{iR}} = \mathbf{JSD}(\boldsymbol{p_i} \parallel \boldsymbol{p_R}) = \frac{1}{2} D_{\mathrm{KL}}(p_i \parallel M_{iR}) + \frac{1}{2} D_{\mathrm{KL}}(p_R \parallel M_{iR});$$

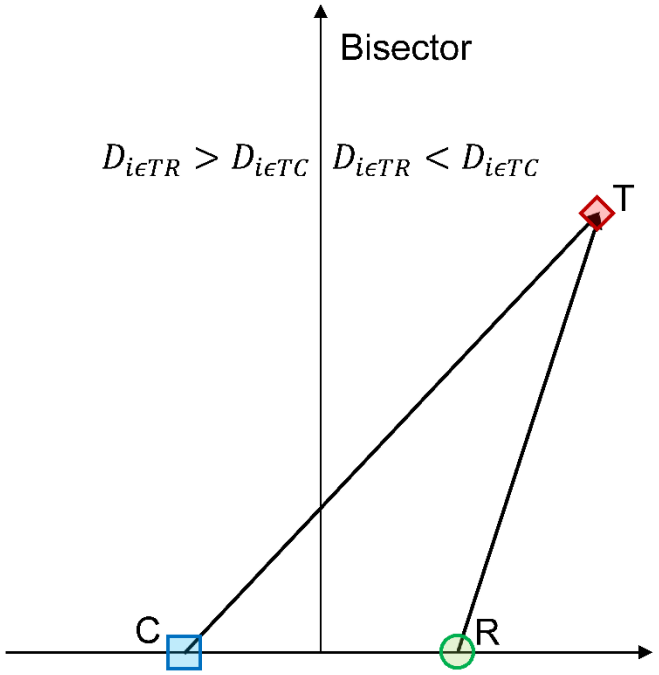
$$M_{iR}(x) = \frac{1}{2}\left(p_i(x) + p_R(x)\right);$$

$$\boldsymbol{\Delta_{iC}} = \mathbf{JSD}(\boldsymbol{p_i} \parallel \boldsymbol{p_C}) = \frac{1}{2} D_{\mathrm{KL}}(p_i \parallel M_{iC}) + \frac{1}{2} D_{\mathrm{KL}}(p_C \parallel M_{iC});$$

$$M_{iC}(x) = \frac{1}{2}\left(p_i(x) + p_C(x)\right);$$

Bisector Clustering by Triadic Projection

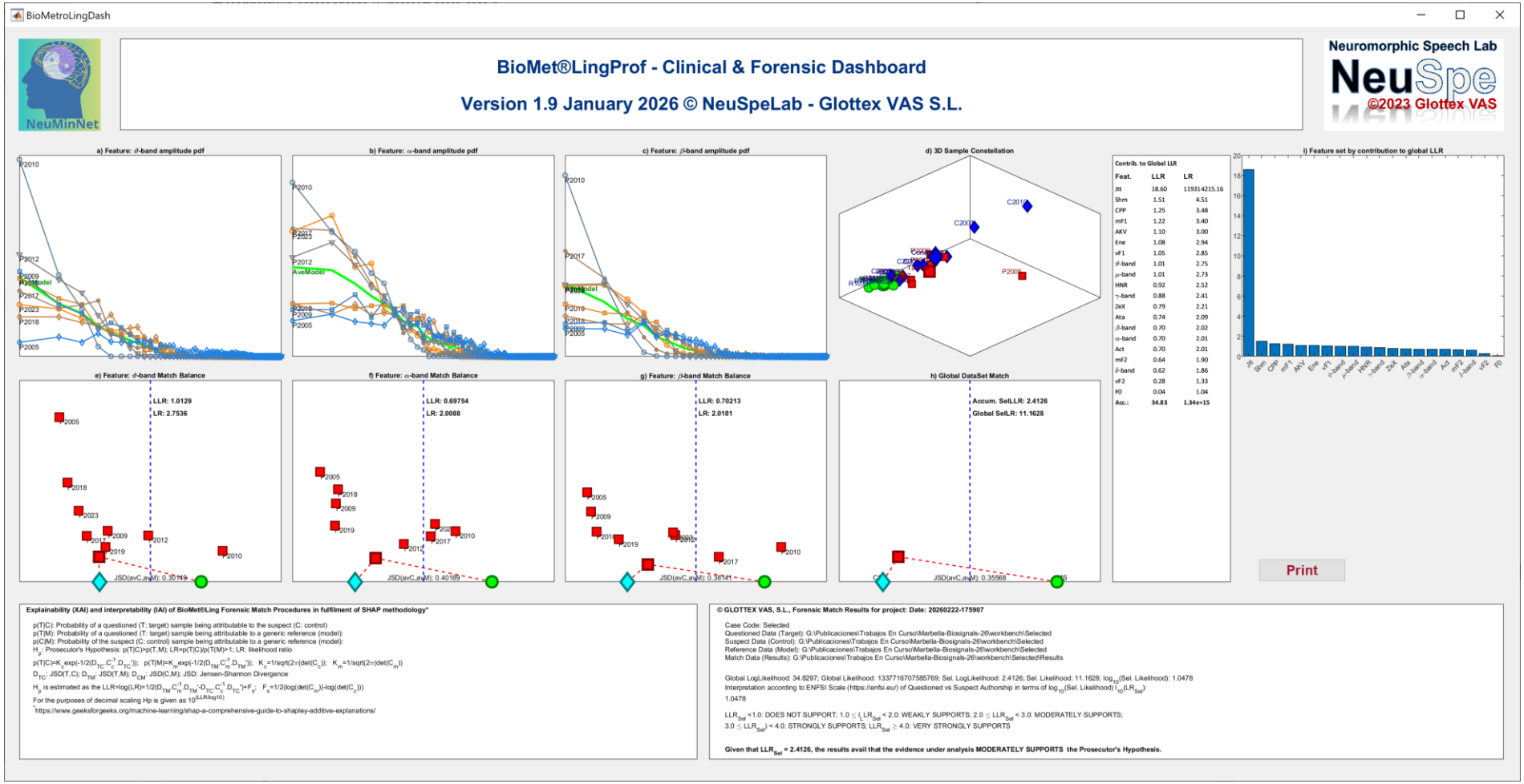




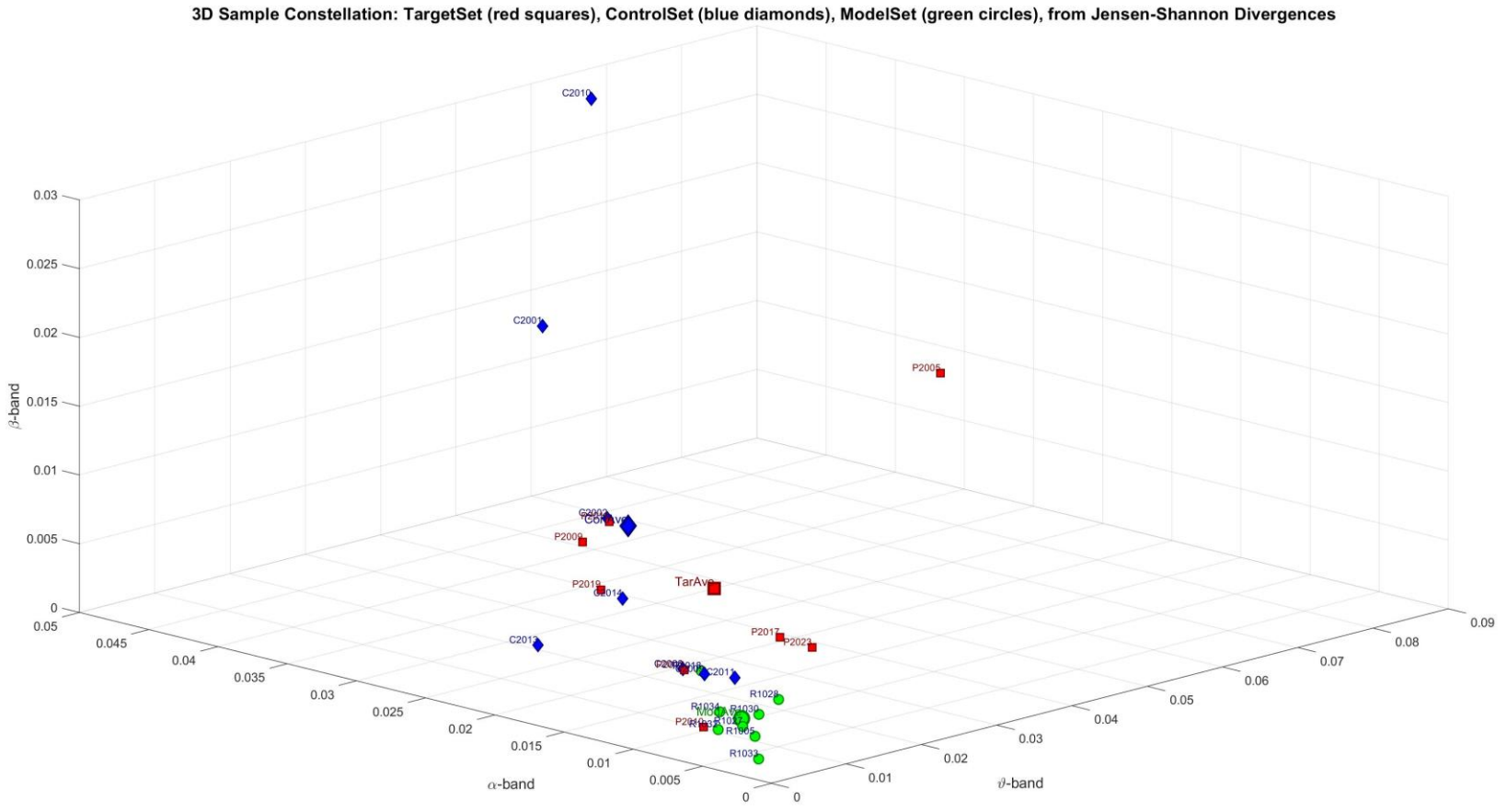
In a Triadic TCR projection, three data sets (Target, Control, and Reference) defined in an n-dimensional manifold are projected onto a two-dimensional plane using a balance-scale comparison. Here, DTC and DTm represent the weights placed on the scale pans:

$$either\ D_{i\epsilon TS|RS} < D_{i\epsilon TS|CS}; or\ D_{i\epsilon TS|RS} > D_{i\epsilon TS|CS}$$

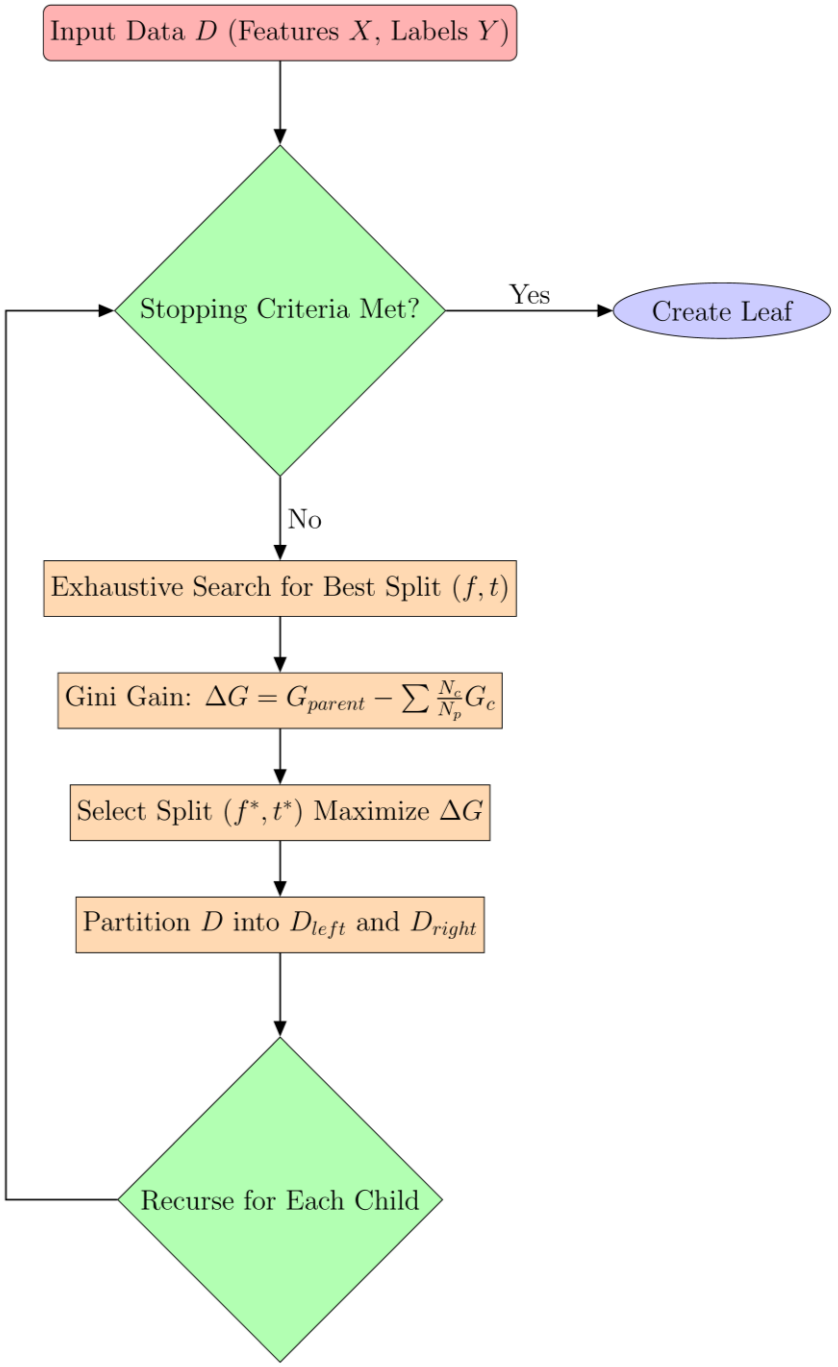
The Holy Grial of Clinical XAI: The Dashboard



Graphical Information: The Missing Link in Trustworthy Clinical AI

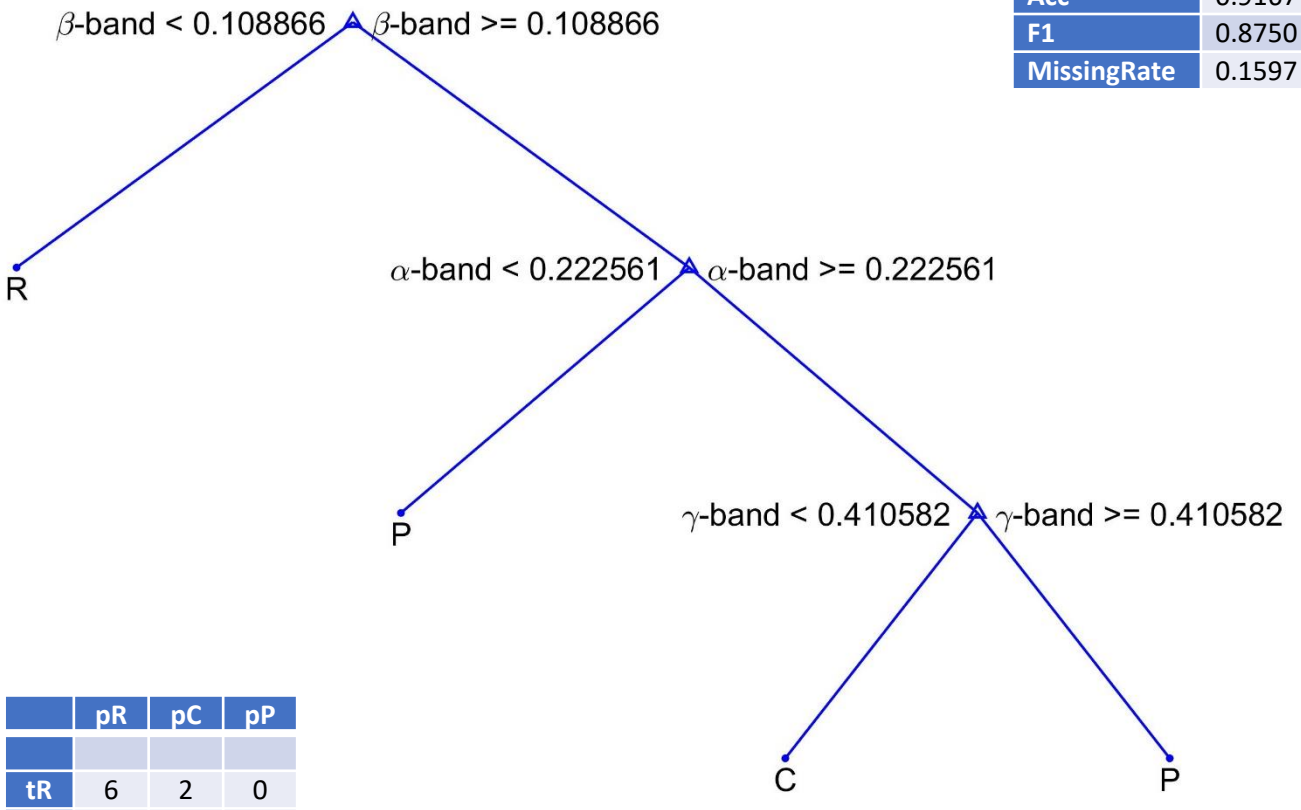


The third ingredient: Decision Trees



DTC (δ, ϑ, α, β, γ): males

Sample	Pred. Class	Global JSD
R1034	R	0.38286
R1030	R	0.4007
R1027	R	0.47614
R1032	R	0.4818
R1028	R	0.51784
P2010	P	0.53924
R1005	R	0.60855
P2023	P	0.71066
P2012	C	0.72022
C2009	C	0.76245
C2008	C	0.79614
P2017	P	0.81142
C2013	C	0.81513
R1033	C	0.81588
C2011	C	0.8299
R1018	C	0.91726
C2014	C	0.9944
C2002	C	1.0346
P2018	P	1.1928
P2019	P	1.2669
P2009	P	1.284
C2010	P	1.3084
C2001	C	1.4187
P2005	P	1.4602

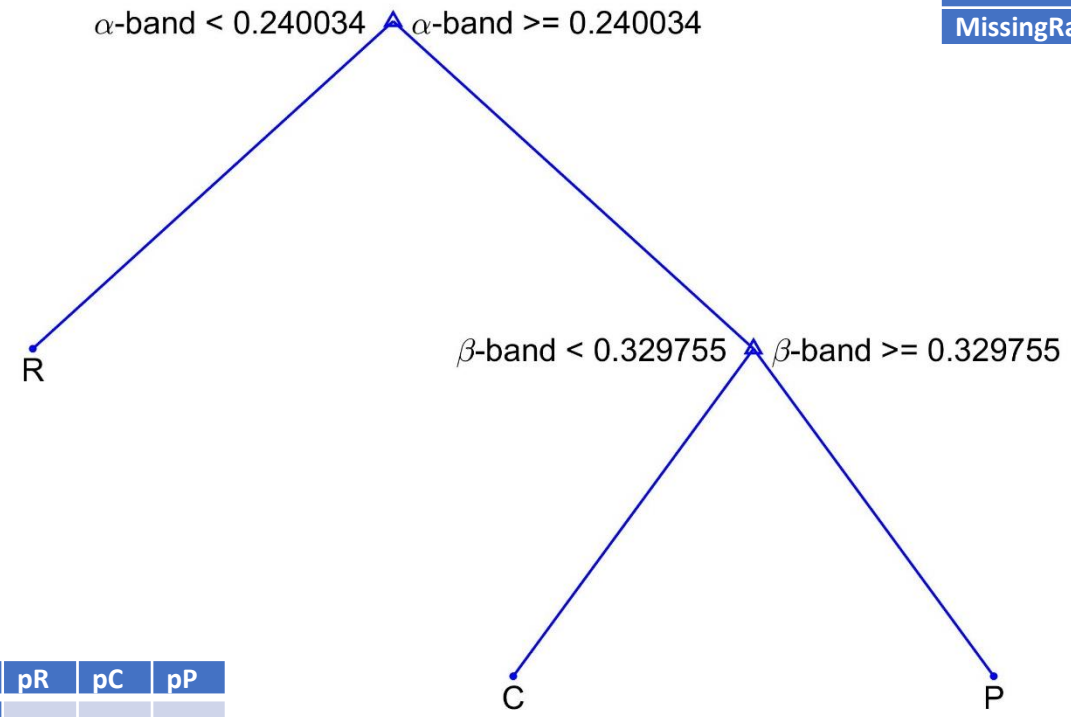


Sen	0.8750
Spe	0.9375
Acc	0.9167
F1	0.8750
MissingRate	0.1597

	pR	pC	pP
tR	6	2	0
tC	0	7	1
tP	0	1	7

DTC (δ, ϑ, α, β, γ): females

Sample	Pred. Class	Global JSD
R1106	R	0.26479
C1005	R	0.27534
R1101	R	0.3211
R1104	R	0.33511
R1103	R	0.3378
C1012	R	0.34736
P1013	C	0.35211
R1108	R	0.35427
R1102	R	0.35779
R1107	R	0.42915
C1007	C	0.46454
P1020	C	0.47889
C1006	R	0.48837
C1018	C	0.52185
P1021	P	0.59771
P1025	P	0.60236
P1007	P	0.67157
R1105	R	0.68746
C1003	C	0.81282
C1004	C	0.83537
C1017	P	0.91549
P1008	P	0.91635
P1022	P	1.0326
P1006	P	1.1491



	Sen	0.7500
	Spe	0.9375
	Acc	0.8750
	F1	0.8000
	MissingRate	0.2431

	pR	pC	pP
tR	8	0	0
tC	3	4	1
tP	0	2	6

Another menu: ASD longitudinal tracking

In ASD the relationship between the theta, mu (low alpha) and high alpha, known as the u_{shape} Synchronization Index (uSI) plays a relevant role:

$$uSI = \xi_{\mu} - (\xi_{\vartheta} + \xi_{\alpha h})/2 = (3\xi_{\mu} - \xi_{\vartheta} + 2\xi_{\alpha})/2$$

The long term evolution of uSI is thought to play a very relevant role in understanding and predictive sudden behavior changes in patients.

The difficulties in the study lay on the quite **different characteristics of behavioral patterns** between different individuals, which **hamper transversal studies**.

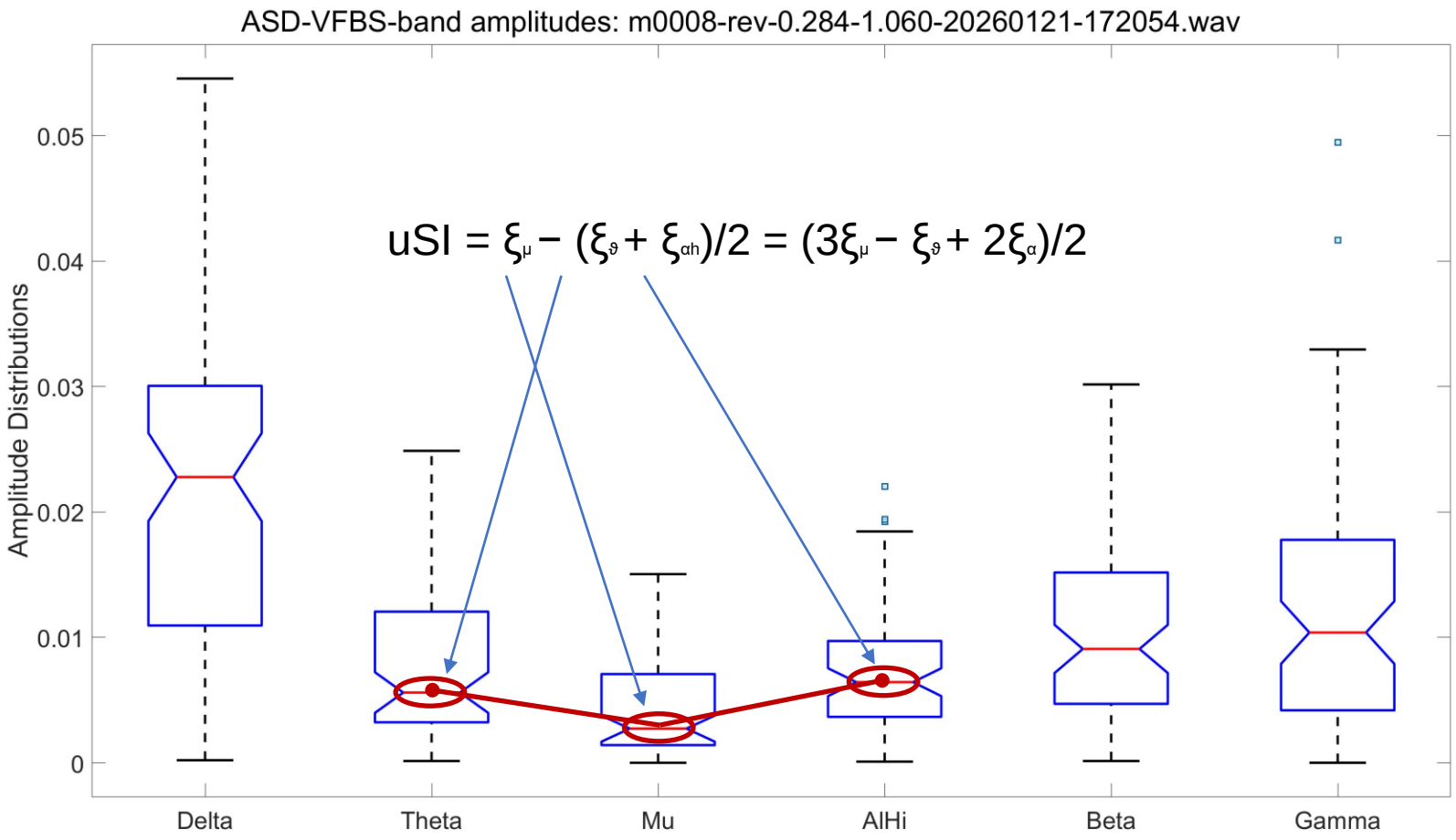
Longitudinal studies might benefit from Data Augmentation by Bootstrapping following the integration of Long Short Term Models and Decision Trees or Random Forests

Two study cases

Characteristics of participants. F: female; M: Male; CARS: childhood autism rating scale (non autistic condition <30, moderate 30-36, severe: >36); DEX: Dysexecutive Questionnaire (optimum condition <10, normal condition 10-18, moderate dysexecutive 19-28, severe cognitive alteration >28)

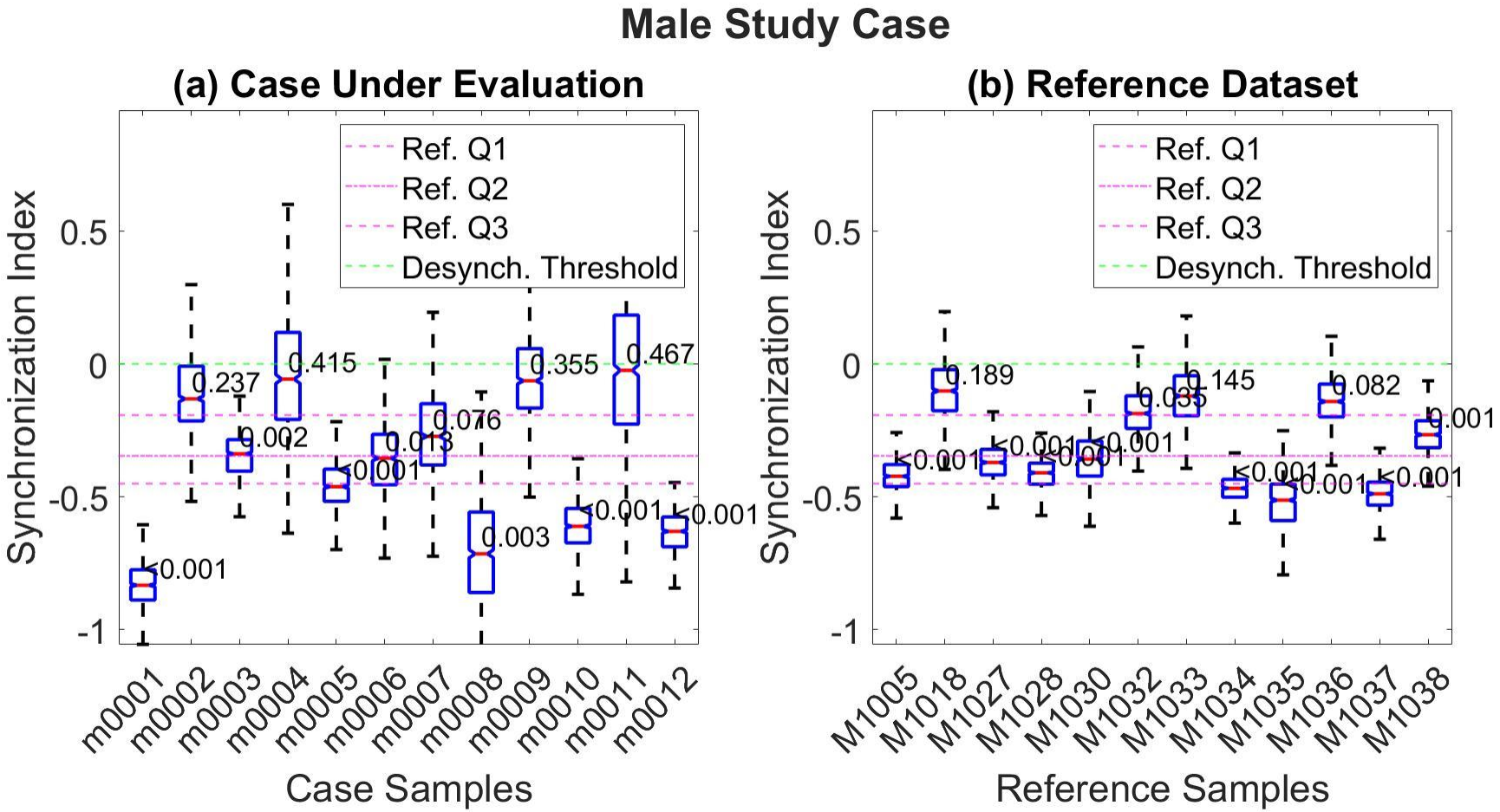
Gender	F	M
Age	45	40
Diagnosis	ASD	ASD
CARS	30	30
DEX	45	44
Comorbidities	Intellectual Disability	Severe Intellectual Disability

uSI estimation from phEG bands

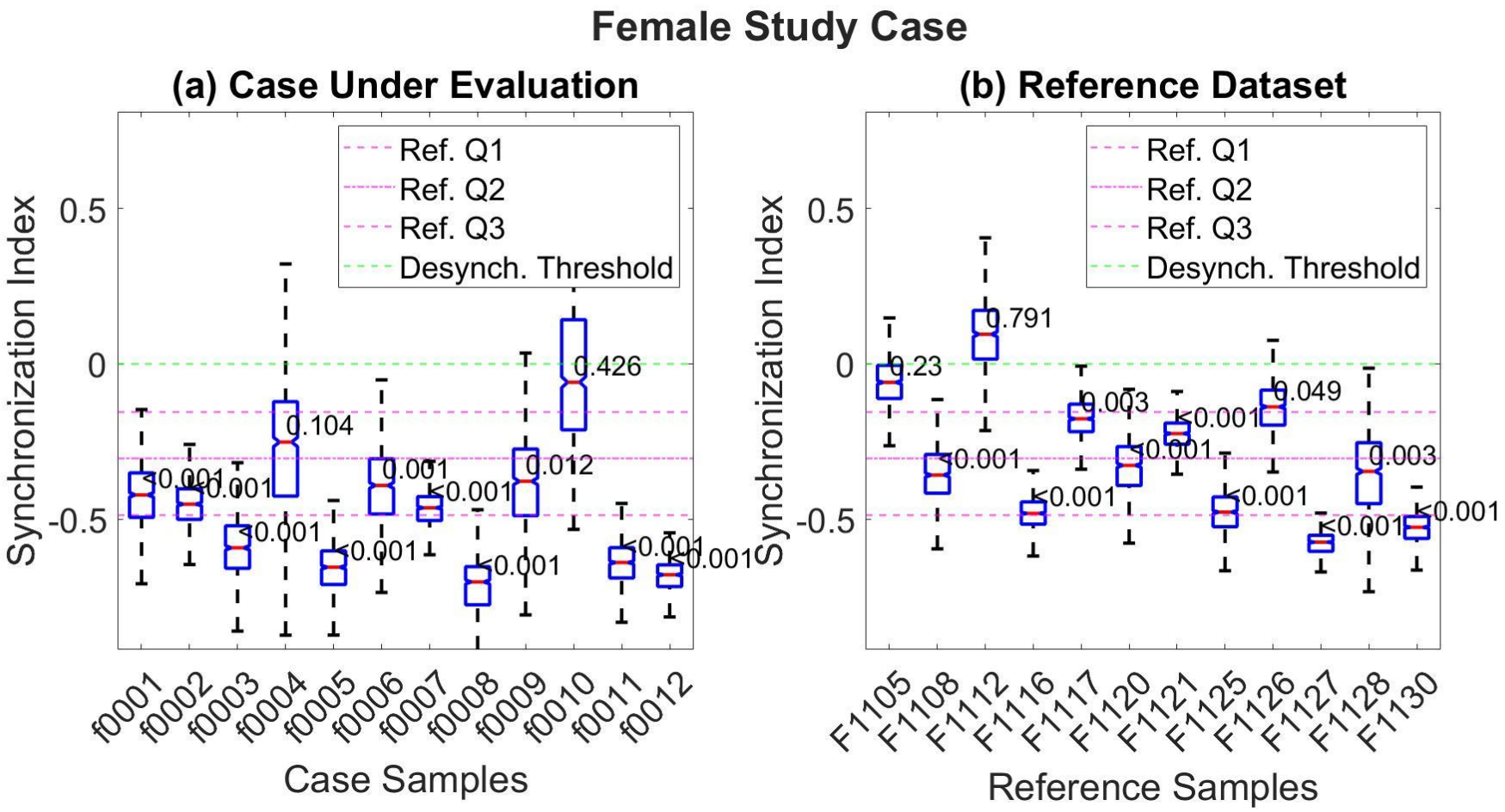


Negative uSI: high desynchronization in the mu band (neurotypical response)
Near zero uSI: Potential ASD marker (absence of mirror system response))

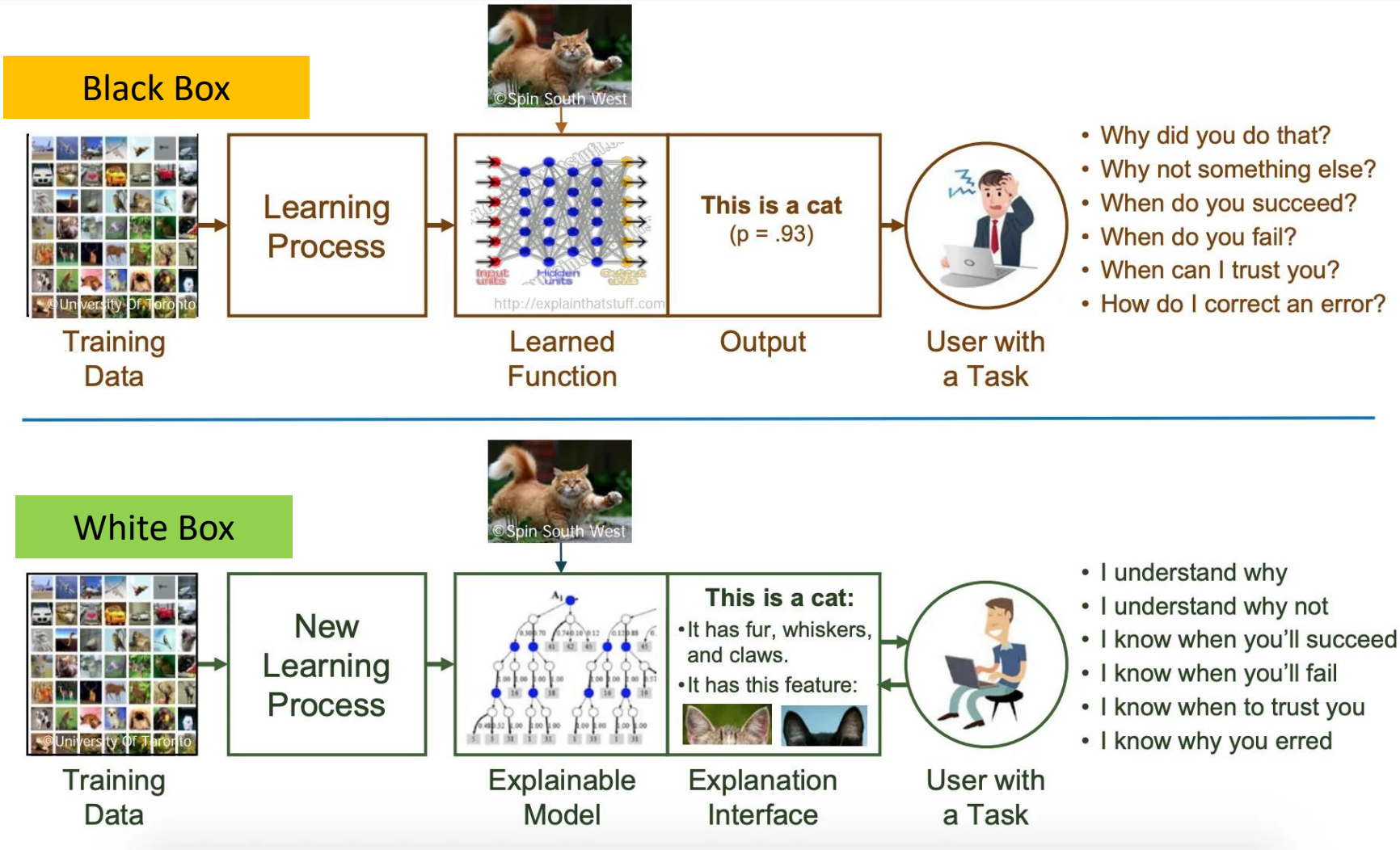
uSI ASD male case



uSI ASD female case

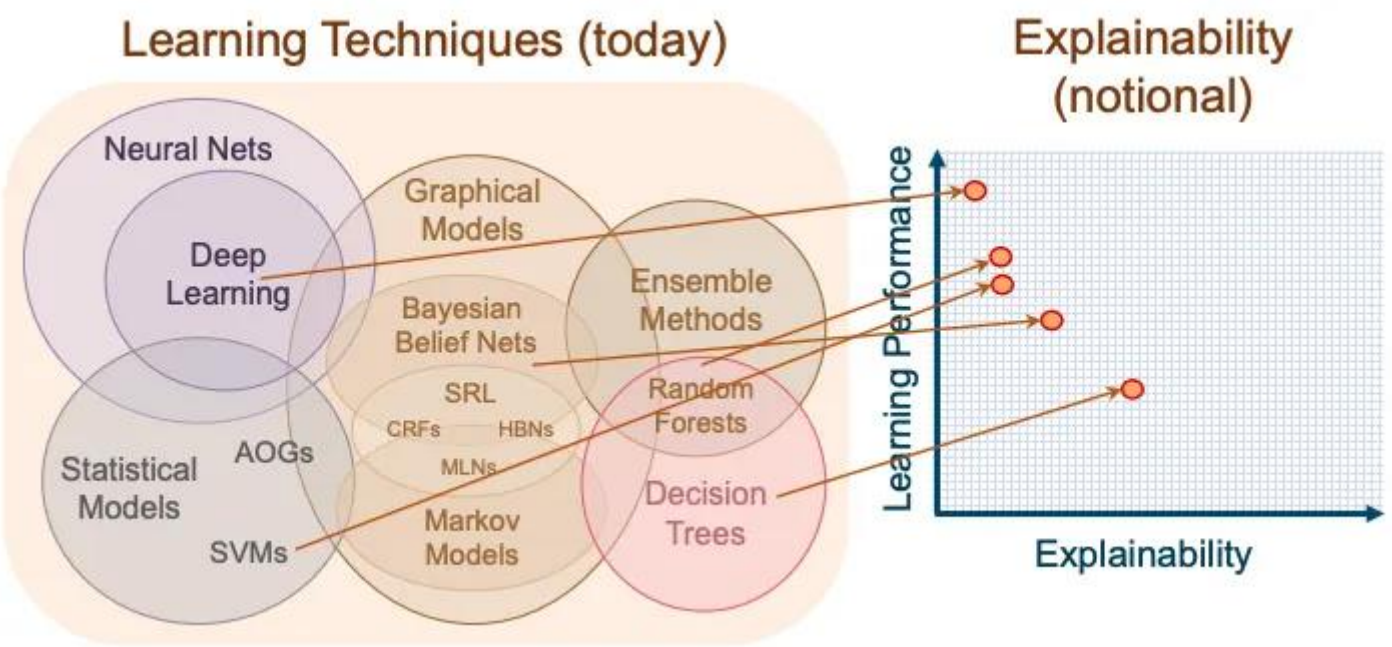


Overcoming BBvsWB AI? Explainable (XAI)



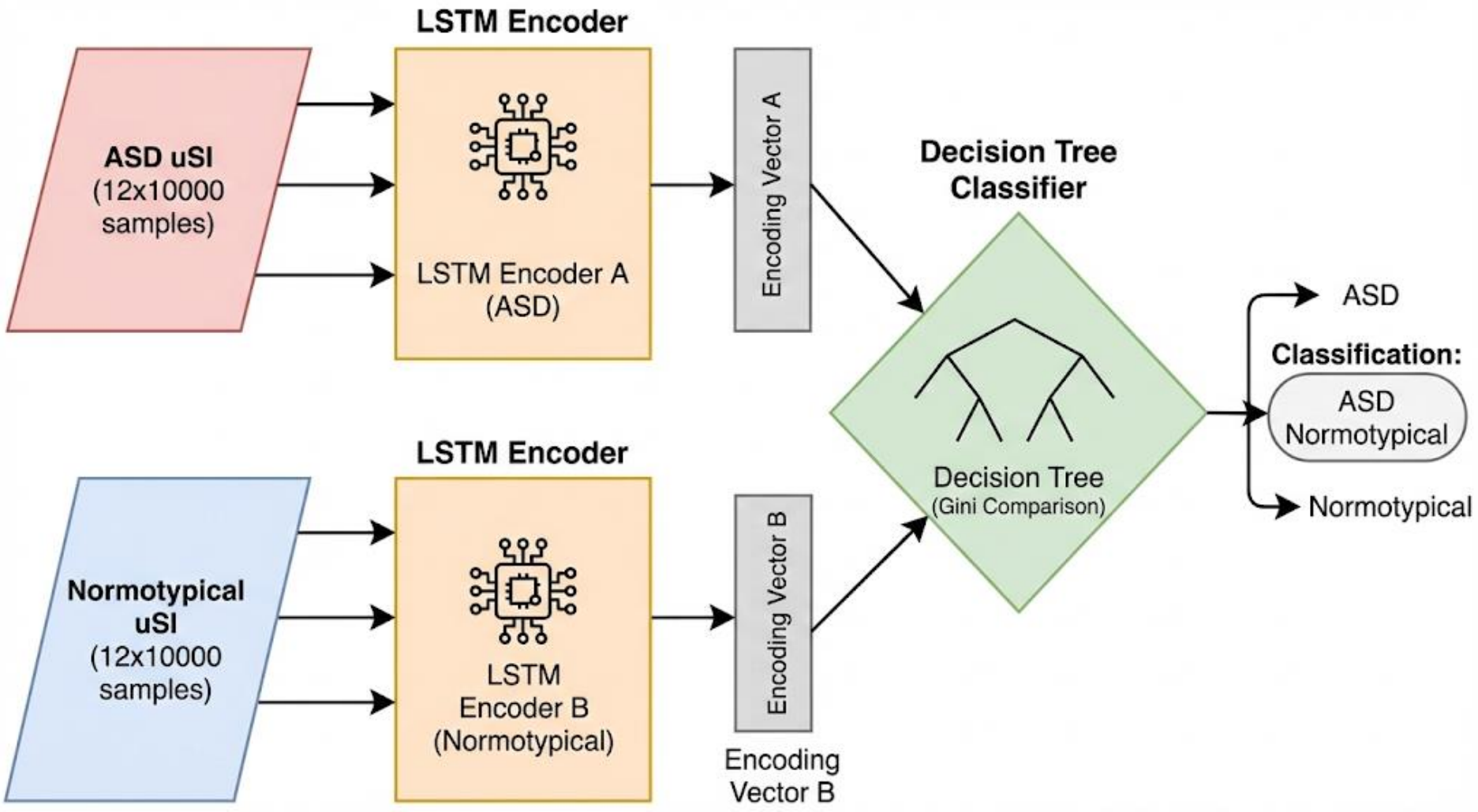
https://lawtomated.com/explainable-ai-all-you-need-to-know-the-what-how-why-of-explainable-ai/

More on Explainability

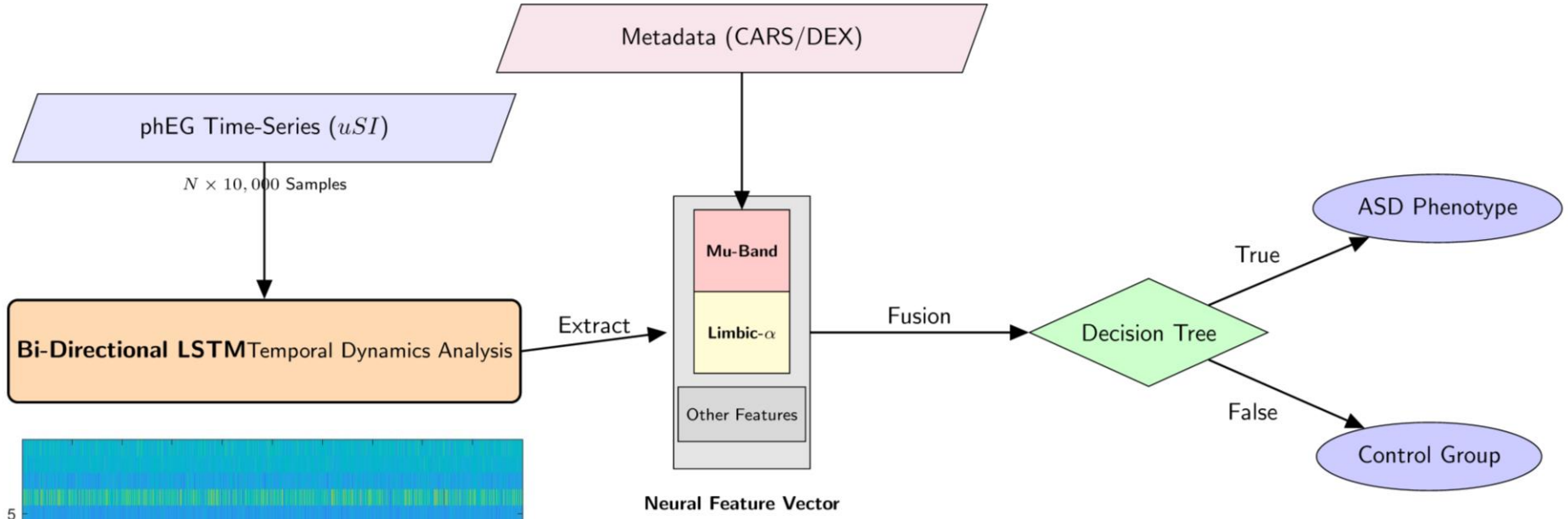


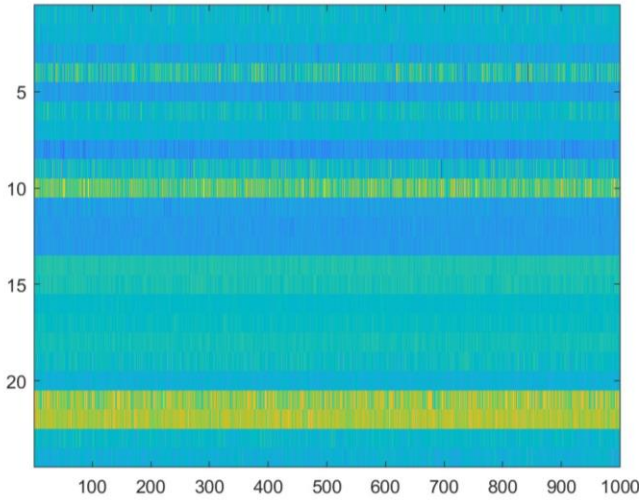
https://lawtomated.com/explainable-ai-all-you-need-to-know-the-what-how-why-of-explainable-ai/

Integrating black & white box AIs



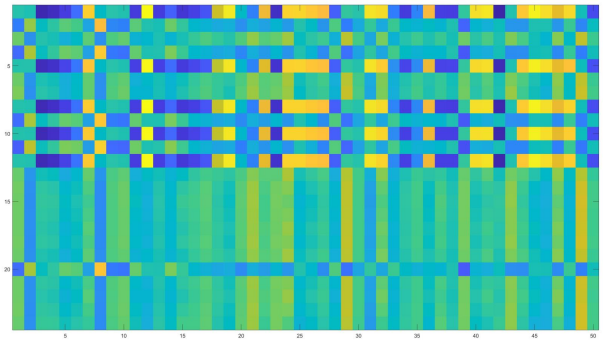
Physiology-Informed Hybrid Model Data Reduction



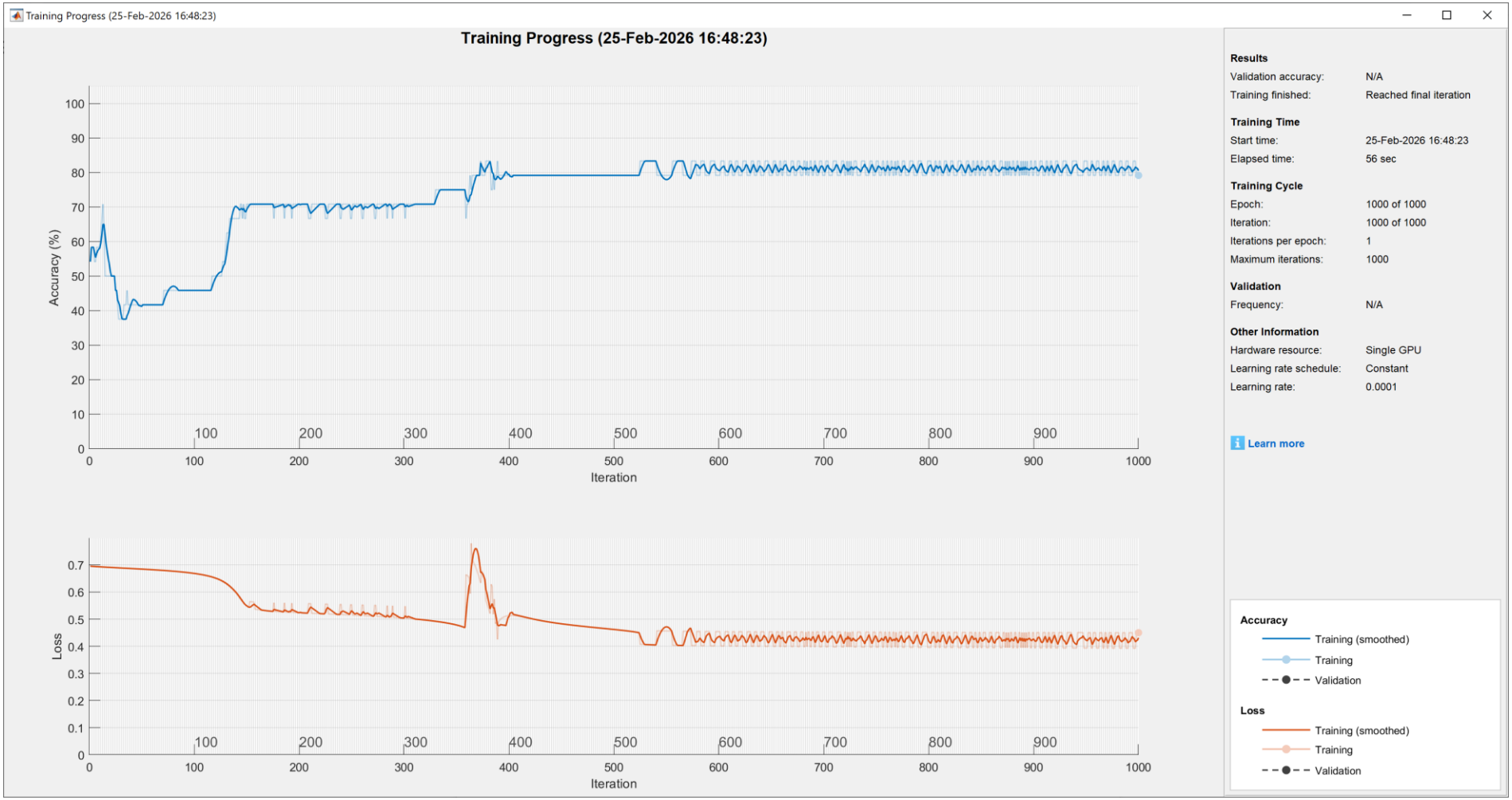


← Bootstrapped uSI

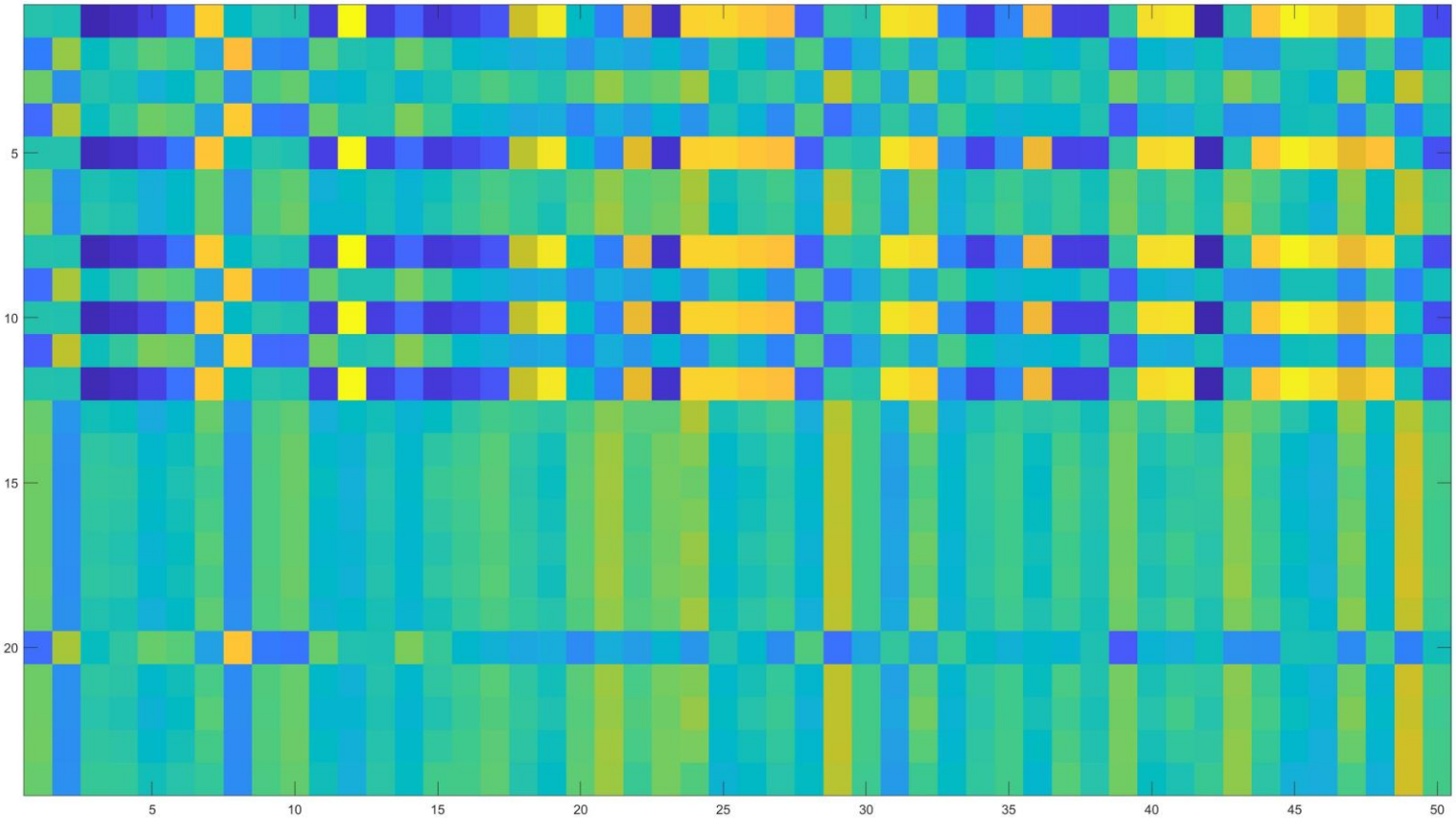
Encoded uSI →



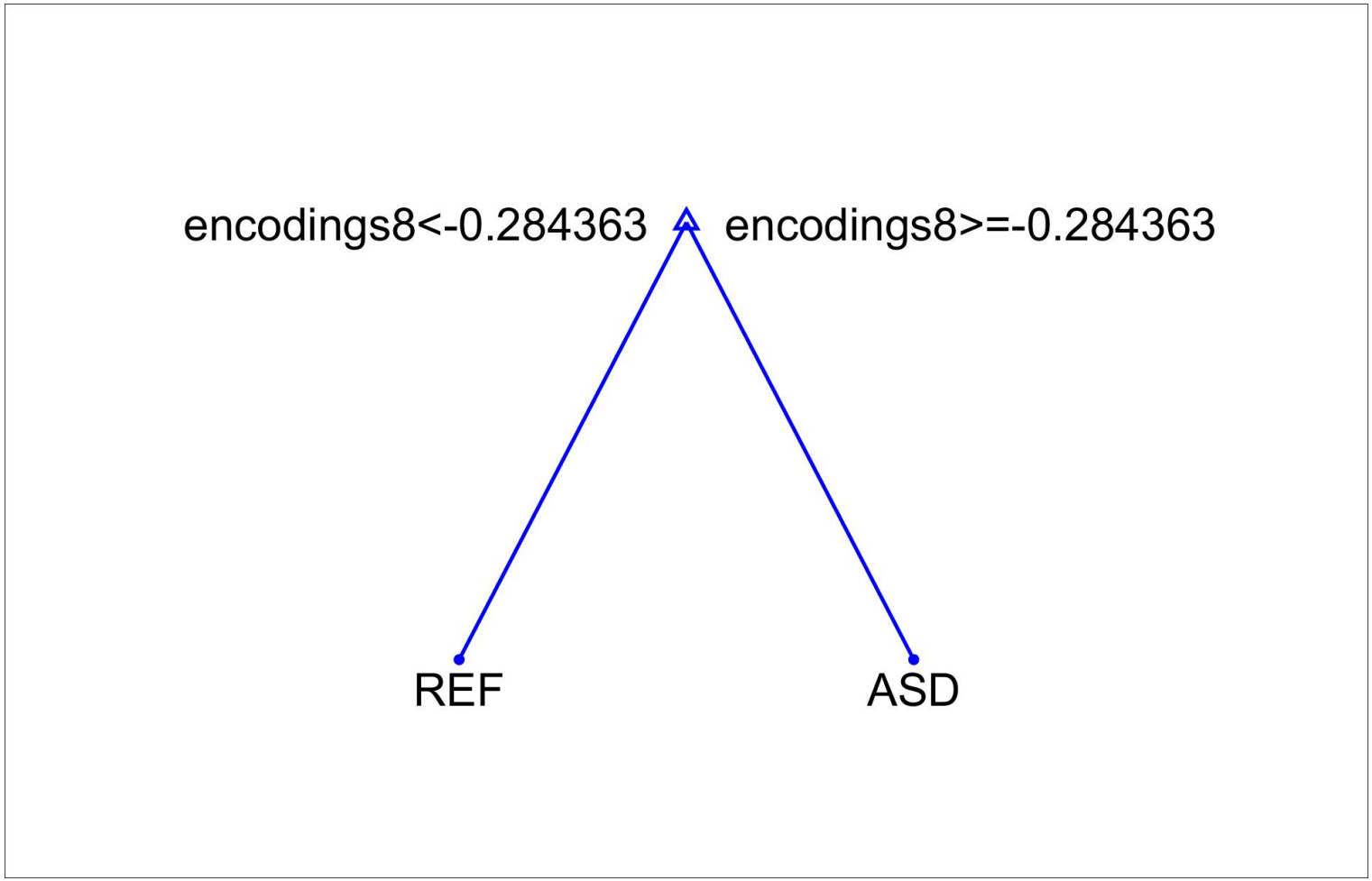
LSTM training profile



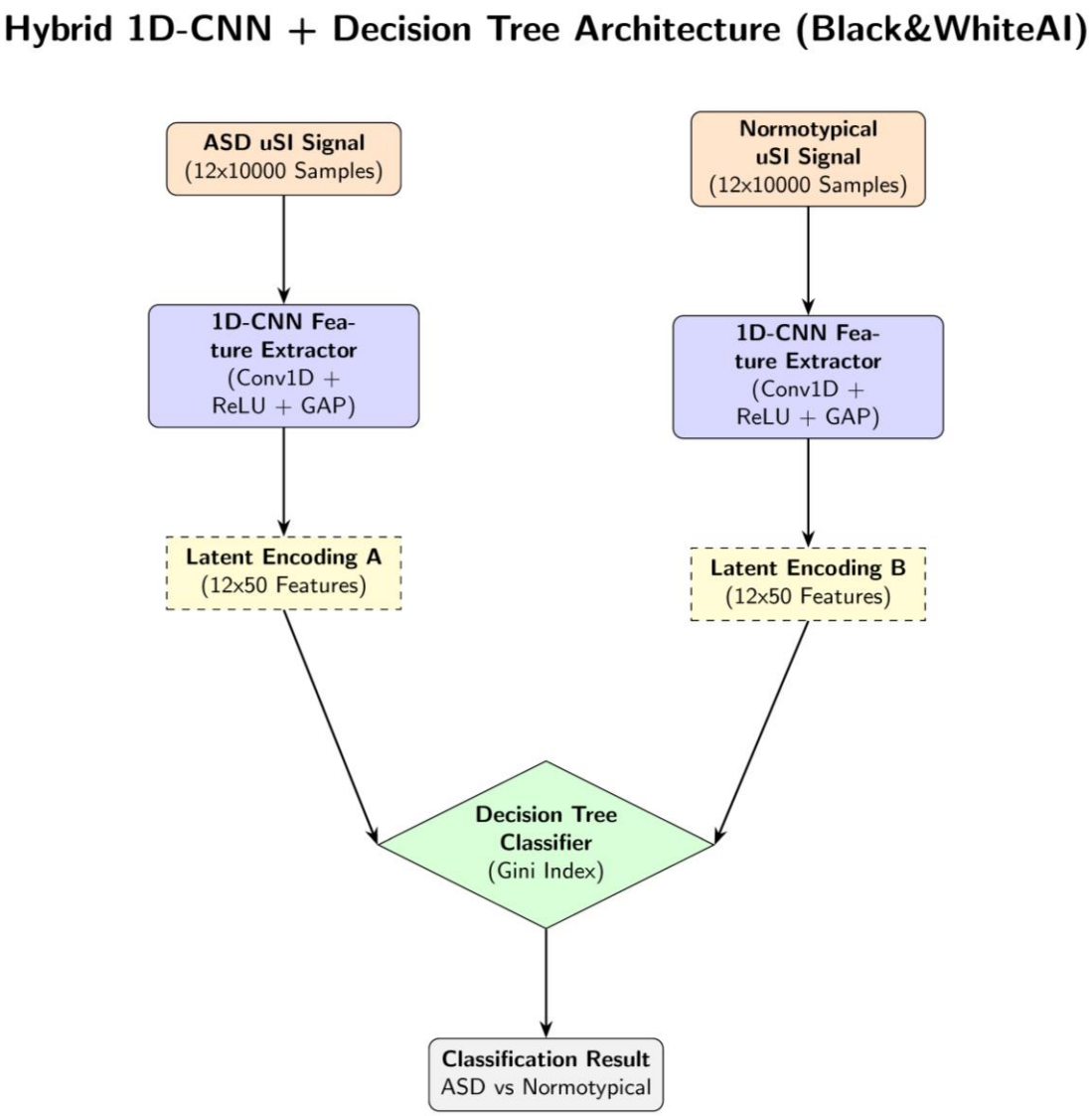
LTMS encodings



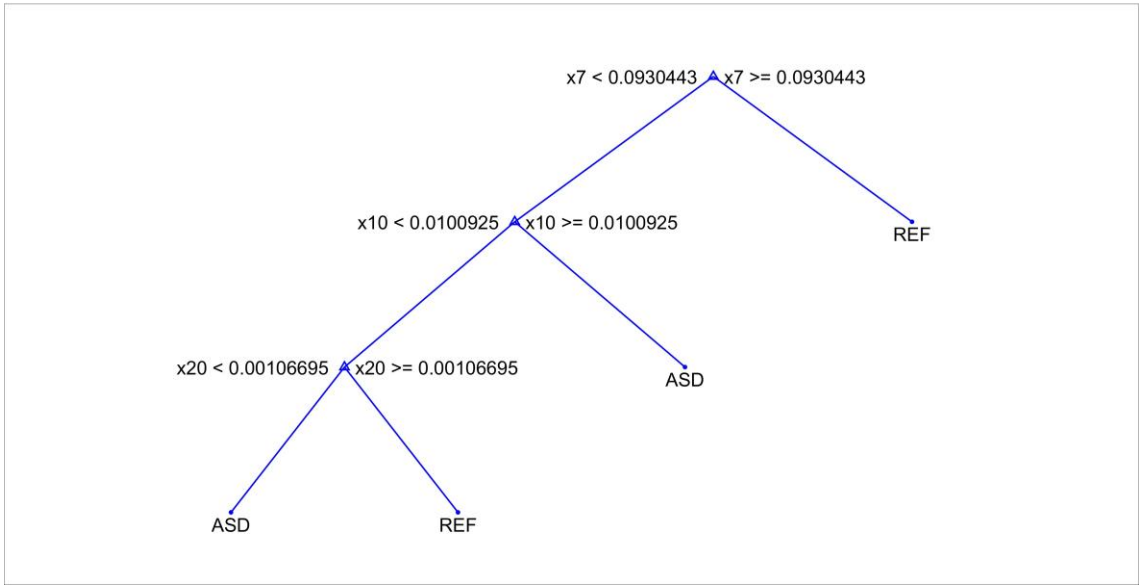
LSTM joint DT: female case



CNN-DT architecture



Female case CNN with LOOCV



LOOCV is particularly valuable when:

Datasets are very small, making traditional train/test splits unstable.

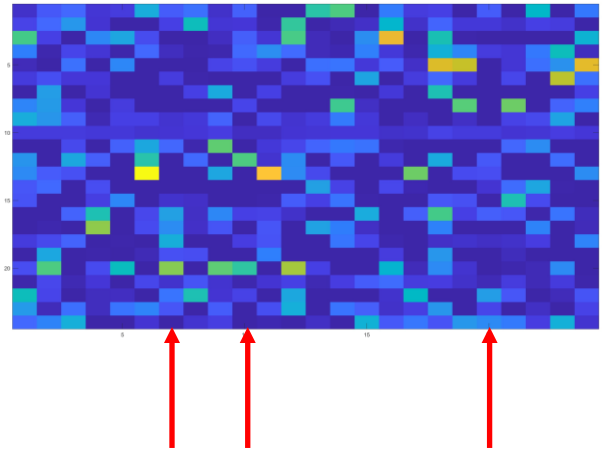
Every sample is precious, such as in clinical, biomedical, or rare-condition datasets.

Model variance is high, and you want a stable estimate of performance.

Needing to detect influential samples, since each sample becomes a test case.

One pending battle: CNN-DT + SHAP

CNN Encodings



- SHAP: SHapley Additive exPlanations, a unified framework for interpreting machine-learning models by assigning individual, case-specific contributions to each feature.
- Shapley values quantify how much each “player” contributes to the final outcome.
- SHAP answers a simple but powerful question: “For this specific prediction, how much did each feature push the model up or down relative to the baseline?”
- It decomposes a prediction into additive components: prediction=baseline+$\sum_i \phi_i$ where each ϕ_i is the Shapley value of feature i.

From Diseases to Trajectories

Why Individual Time-Courses Must Prevail in Multivariate Syndromes as ASD?

Osler's* claim that "**there are no diseases, only sick patients**" resonates strongly in multivariate syndromes like ASD, where transversal snapshots flatten the individuality of each case.

Only **longitudinal characterization** honors the evolving, person-specific trajectories that define illness.

The clinical reality Osler remarked: **illness lives in patients, not in categories**.

Individual Trajectories as the True Unit of Explanation: From Osler to SHAP in Multivariate Syndromes.

*Regius Professor of Medicine at **Oxford University William Osler (1849–1919)** was a Canadian physician, often referred to as the **"Father of Modern Medicine"**

Formal definition of Semantics

In the B&WAI framework, the "Semantics" of any architecture is the process of assigning clinical meaning (M) to raw numerical data (S) via Latent Encoding (*F*).

S (Syntax): The set of all well-formed formulas or strings allowed by the grammar

M (Model/Universe): A non-empty set representing the domain of discourse (the "world" we are describing).

F (Interpretation Function): A function that maps every constant in S to an element in M, otherwise, every n-ary predicate to a set of n-tuples in M.

Therefore, the following definitions apply:

An architecture (*F*) is considered “White Box” iff it is bijective, and “Black Box” otherwise

Semantics vs Performance

In the context of Black&WhiteAI, this contraposition carries significant weight for architectural evaluation:

White Box (*F* is bijective): There is a **one-to-one correspondence between the signal features and the model's internal state**. The process is fully reversible and transparent; given an encoding, we can uniquely identify the input features that created it.

Black Box (*F* is not bijective): The mapping is likely many-to-one (non-injective). **Multiple distinct signal patterns might collapse into the same latent encoding**, causing a "loss of information" or "opacity" where **the specific causal path from X to Y cannot be uniquely inverted (an ill-posed problem)**.

Bridging the gap between Black-Box complexity and White-Box interpretability is not merely an advantage, but a prerequisite for clinical AI.

The transparency required for clinical adoption of AI is sought by utilizing the **uSI as a White-Box feature** to explain **Black-Box Deep Learning classifications**,.

For those still skeptical: Nested Learning

Uniform and Reusable Structure

Neuroplasticity is the brain's ability to reorganize itself via mechanisms like forming new synapses, strengthening/weakening existing ones, rerouting signals through alternate pathways, etc. Such ability requires uniform and reusable structure across brain.

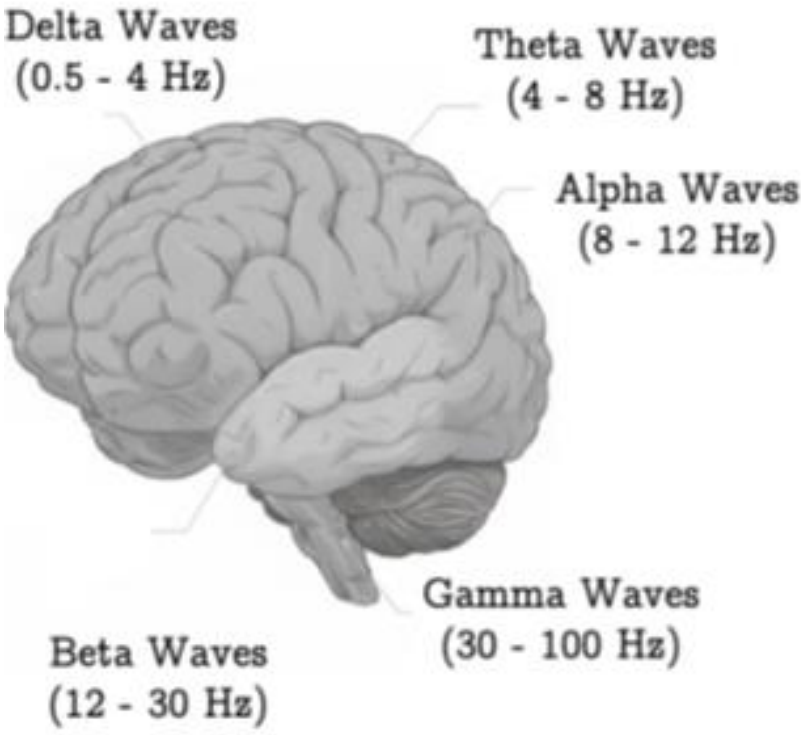
In Nested Learning (NL), architectures are decomposed into a set of neurons (i.e., linear or locally deep MLPs), each of which with its own context flow and objective. This design provides a uniform and reusable structure for learning.

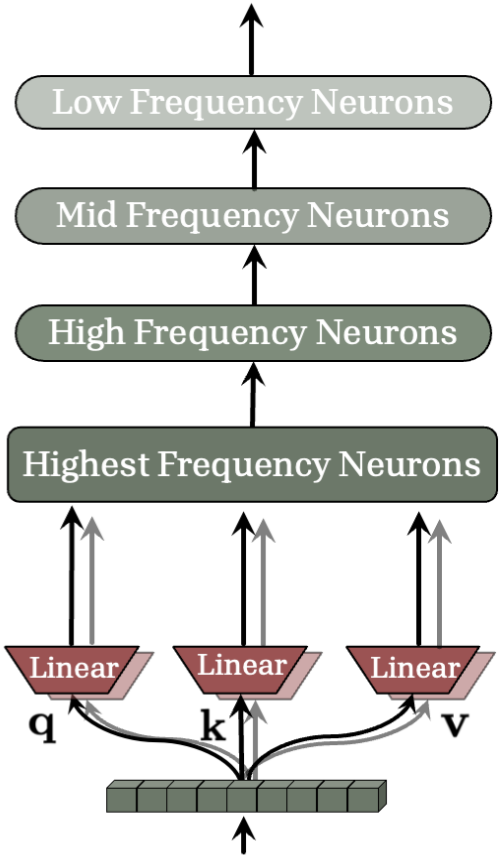
Multi Time Scale Update

Brain oscillations (or brain waves) critical for the brain to coordinate its activity. Notably, the brain does not rely on a single centralized clock to synchronize every neuron: the earlier layers update their activity quickly in high-frequency cycles, whereas later layers integrate information over longer, slower cycles.

In NL, parameters in each *"level"* are updated with their own specific frequency and does not rely on a single centralized clock. The HOPE's design allows the earlier layers update their activity quickly in high-frequency cycles, whereas later layers integrate information over longer, slower cycles.

Attention in/to NL





Behrouz, A., Razaviyayn, M., Zhong, P., & Mirrokni, V. (2025). Nested learning: The illusion of deep learning architectures. arXiv. https://doi.org/10.48550/arXiv.2512.24

Conclusions (I)

Findings: phEG-derived ϑ, α, β, and γ distributions help stratifying disordered, control, and normative cohorts using interpretable CART models and JSD criteria.

Explainable AI (XAI) via CART: Adoption of Classification and Regression Trees (CART) to prioritize architectural transparency, allowing clinicians to audit the "decision logic" behind each diagnosis.

Insights: Distinct amplitude and synchronization profiles serve as sensitive markers for idiosyncratic neuromotor signatures, revealing significant participant-level divergences in motor-cognitive function.

Biomarkers: The research validates β-band motor maintenance and α-band cortico-limbical connection as critical biomarkers, where their coupling with ϑ-band reflects the patient's overall neuromotor integrity.

Methodology: Utilizing explainable AI to decode laryngeal biomechanics via probability density functions offers a robust, non-invasive window into pathological oscillatory loops usually hidden from view.

Conclusions (II)

XAI strengthens the conclusions by emphasizing that clinical AI systems must preserve a clear and inspectable chain between the features they analyze and the decisions they produce.

In heterogeneous and multiform disorders, where symptom expression varies widely across individuals, XAI transparency becomes essential for ensuring that classifications are grounded in clinically meaningful signals rather than artifacts or hidden shortcuts.

By prioritizing architectural clarity and accountability, XAI enables clinicians to trace, audit, and validate the internal decision logic of the model, supporting safer deployment and more trustworthy interpretations.

When this feature–decision link is obscured within opaque data-handling pipelines, the model's outputs may remain technically accurate but lose clinical legitimacy, limiting their usefulness in real-world diagnostic and therapeutic contexts.

There is still a hope for model patronized petty AI and shallow learning

Conclusions (III)

The extreme heterogeneity of complex syndromes seen in disorders such as ASD fragments the phenotype space into numerous low-density regions, making it nearly impossible to assemble the large, homogeneous datasets required for high-capacity deep learning models to generalize reliably.

This **structural sparsity** is not a temporary limitation of recruitment but an intrinsic property of the disorder itself, which produces a combinatorial explosion of phenotypic variants that outpaces any realistic data-collection effort.

Paradoxically, this very constraint preserves a meaningful role for model-patronized "petty AI" and shallow learning methods: **approaches that operate with explicit features, controlled inductive biases, and interpretable decision boundaries remain better suited to sparse, high-variability clinical landscapes**.

In these contexts, models that reason over curated representations rather than massive data volumes can offer greater stability, transparency, and clinical trustworthiness—reminding us that there is still justified hope for carefully designed shallow learning in domains where deep learning's appetite for data cannot be met.

Future work

A unified phEG methodology is to be defined on non-invasive mapping of laryngeal neuromotor drive to established EEG frequency bands, providing a **"direct window" into cortical and subcortical dynamics**.

To establish a foundational framework for developing targeted therapeutic strategies and advancing the clinical **comprehension of neurodegenerative motor and cognitive dysregulation**.

Validation of phEG-band correlates against **speech-EEG synchronous databases**, such as from the work of Verwoert et al. 2022*

More powerful associations between **shallow-learning** methods and **explainable interfaces** are being sought and tested.

*M. Verwoert, et al., Dataset of speech production in intracranial electroencephalography, Sci. Data 9 (2022) 434

Acknowledgments

This study was funded by project no. LX22NPO5107 (MEYS), from the European Union – Next Generation EU, by project no. CZ.02.01.01/00/23_025/0008726 (A Lifetime with Language: The Nature and Ontogeny of Linguistic Communication), co-funded by the European Union, and by Grant PID2023-152984OB-I00 (CaRaHAVoz) funded by the State Research Agency of Spain (Agencia Estatal de Investigación—AEI), within the Network Cyted NT4SM (#225RT0169).

Thanks for your attention

pedro.gomezv@upm.es