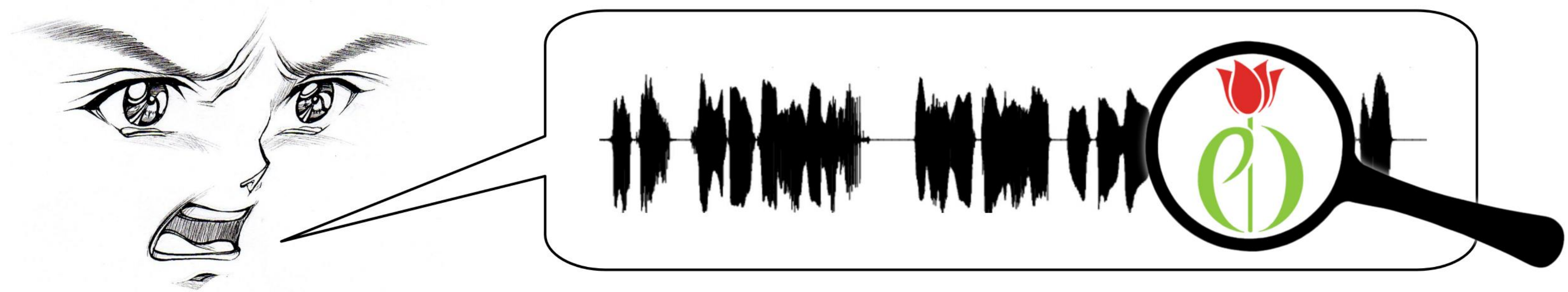
ktu kaunas university of technology 1922



Exploiting voice signal decomposition in expert system for Parkinson's disease detection

Aivaras Šimulis, Evaldas Vaičiukynas

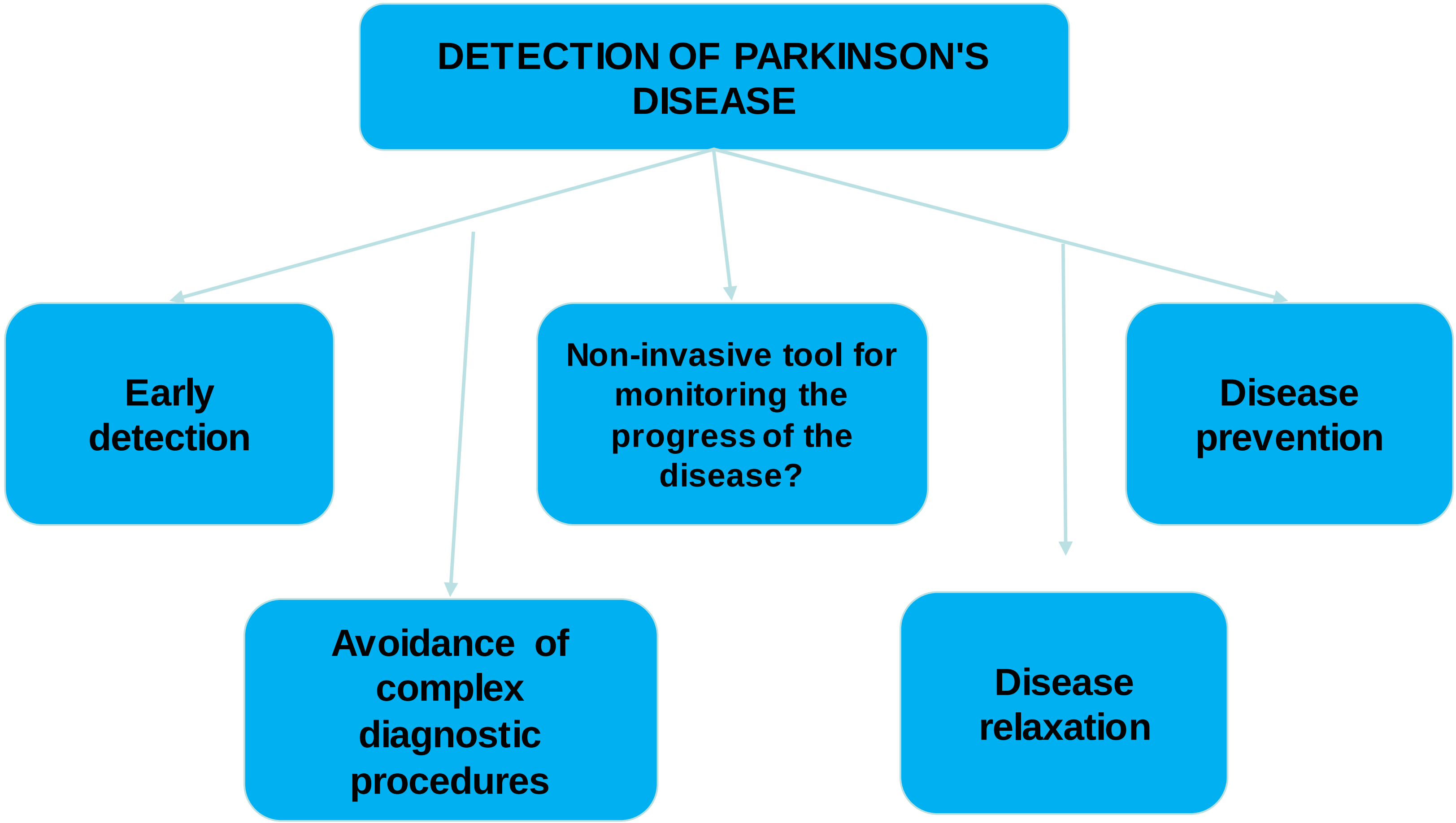
Department of Information Systems,
Kaunas University of Technology, Kaunas, Lithuania

IVUS 2017 conference (April 28, 2017) @ Kaunas, Lithuania

Problem, goal and keypoints of the research

- Problem
  - Parkinson's disease (PD) – the 2nd most common neurogenerative disease. Loss of dopaminergic neurons at the time of diagnosis can reach ~50% and neuroprotective strategies could be too late to slow it.
  - Speech anomalies in PD subjects can be observed ~5 years before clinical diagnosis, early objective diagnostic markers/tools are needed.
- Goal
  - Non-invasive tool for Parkinson's disease (PD) screening.
  - Medical decision support system for healthcare specialists.
    - Monitoring of PD progression, outcome of therapy or anti-PD drugs.
- Keypoints
  - Subject's data: audio signal (.wav) from acoustic cardioid microphone
    - ~3 sustained voice recordings – vowel "a" phonation at a comfortable pitch
  - Mode decomposition algorithms: empirical (EMD), variational (VMD).
  - Feature set: 12 Perceptual Linear Predictive Cepstral Coefficients
  - Random forest (RF) learner: base and meta (decision-level fusion).
  - RF variable importance from meta-learner for extra insights.
  - t-SNE on RF proximity matrix provide 2D mapping for recordings.

Problem and it‘s relevance



Research fields: voice signal analysis, pattern recognition, expert systems, computational intelligence.

Resarch object: voice recordings of healthy and Parkinson‘s disease individuals.

Related work and existing challenges

- Extensive overview – Orozco-Arroyave et al. 2016:
  - Orozco-Arroyave JR, Hönig F, Arias-Londono JD, Vargas-Bonilla JF, Daqrouq K, Skodda S, Rusz J, Nöth E (2016) **Automatic detection of Parkinson's disease in running speech spoken in three different languages**. *The Journal of the Acoustical Society of America* 139(1):481–500. doi:10.1121/1.4939739
  - recommends splitting speech into voiced / unvoiced parts

- Existing challenges in the related work:
  - Size of database is a major problem → unreliable estimates.
    - biggest database so far – 88 PD and 88 HC German subjects
  - Reported researches often lack proper* validation scheme.
    - *proper: leave-one-subject-out (or leave-one-individual-out)
  - Main emphasis on the extraction of various audio features.
  - Interspeech 2015 computational paralinguistics challenge !

Summary of the LUHS voice database

Recordings	Parkinson (PD)	Healthy (HC)	Total
Male	36 (107)	105 (312)	141 (419)
Female	39 (116)	203 (599)	242 (715)
Total:	**75 (223)**	**308 (911)**	383 (1134)

- Numbers denote amount of subjects (recordings):
  - PD – Parkinson’s disease patients
  - HC – healthy control subjects

- Recorded at: Department of Otorhinolarynology, Lithuanian University of Health Sciences (LUHS), Kaunas, Lithuania
- Microphone: AKG Perception 220 acoustic cardioid

Research methodology

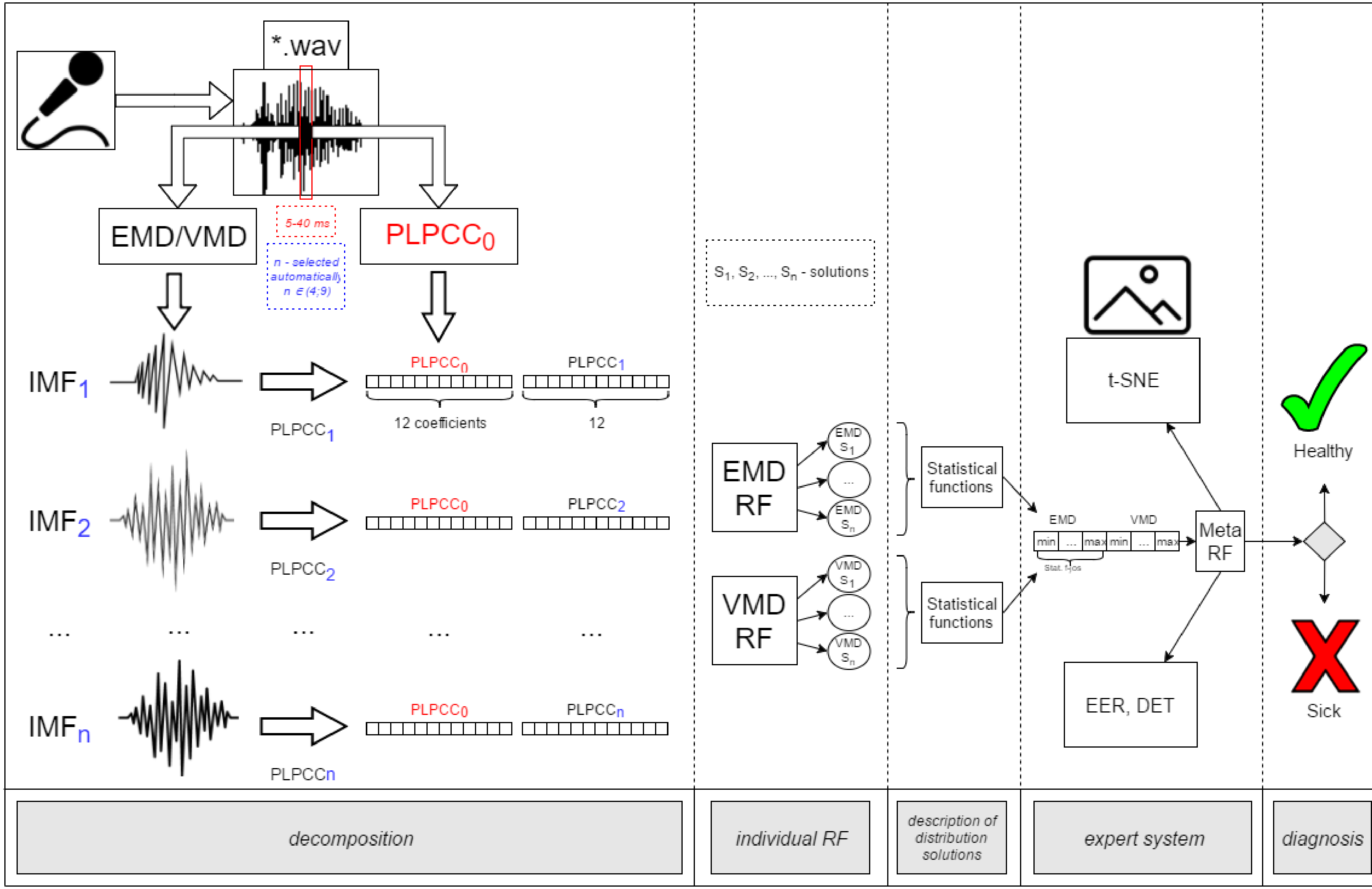
1. Statistical functions	2. Features	3. Detection	4. Validation	5. Prototype

• min / max
• mean
• median
• trimean
• std,
• IQR quartiles
• skewness
• kurtosis.
	• Decomposition:
	  • EMD → IMFs
	  • VMD → IMFs
	• Feature extraction
	  • PLPCC
		• Random forest
		• Ensemble of RFs
		  • base-RF
		  • meta-RF
			• RF out-of-bag
			• Equal error rate
			• Vizualization of proximity matrix from RF by t-SNE
			• Variable improtance
				• Service requirements
				• Prototype specification
				• Applicable technologies

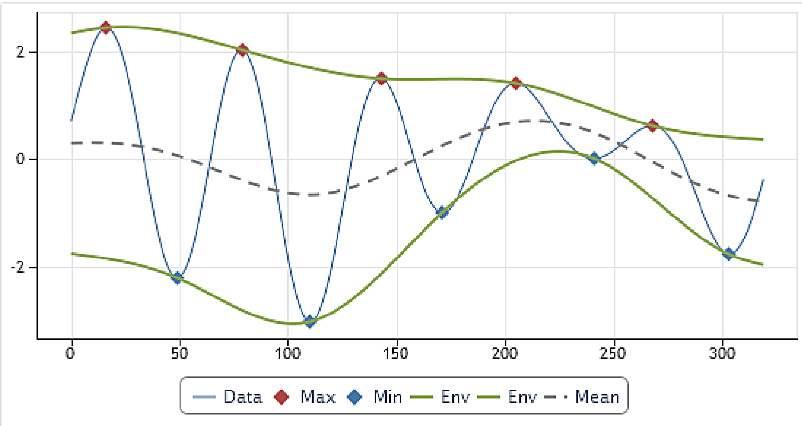
Explanation:

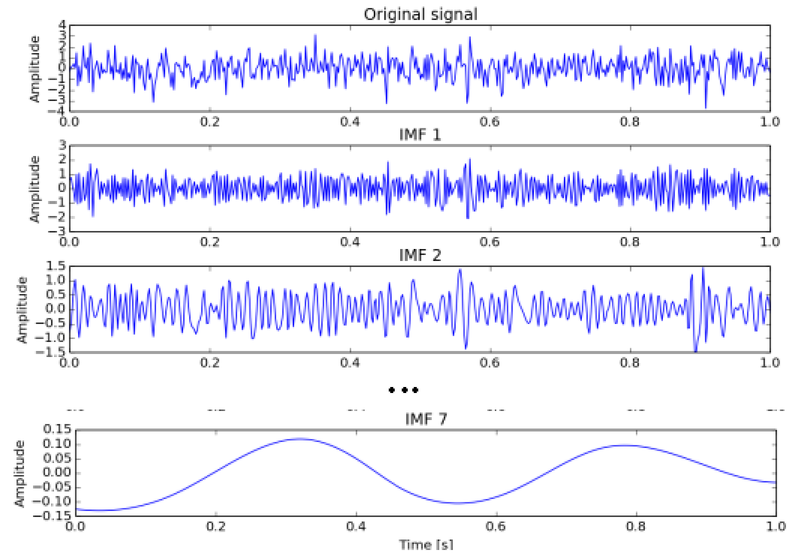
1. Statistical functions are used to compress:
   - (baseline) frame-based PLPCC features from all windows
   - (proposed) base-RF decisions from of all IMF components
2. PLPCC features are calculated for each IMF from decomposition
3. Random forest for Parkinson's disease detection
4. Evaluation of goodness-of-detection
5. Medical decision system prototype

Proposed solution: signal mode decomposition



Mode decomposition techniques considered

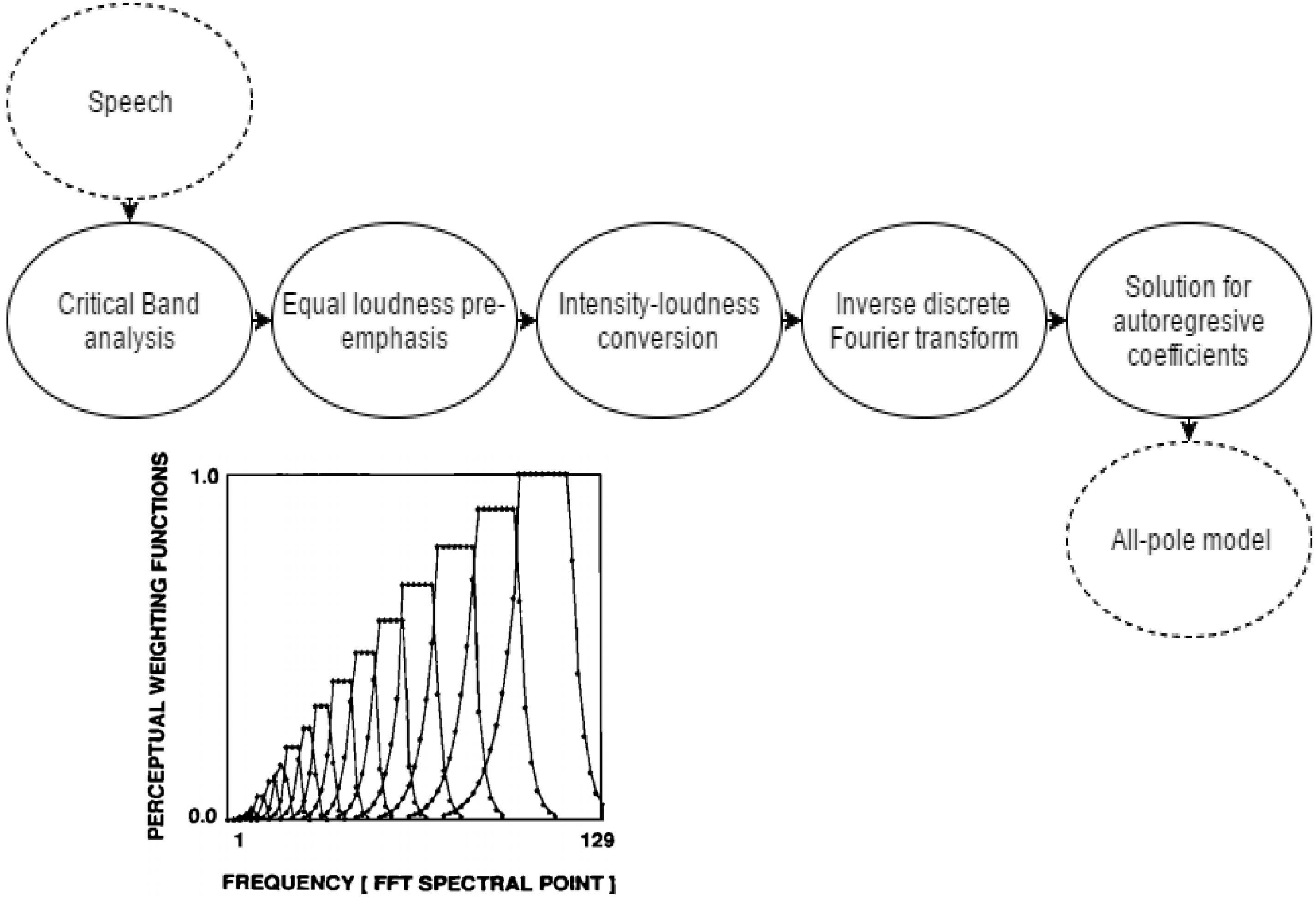




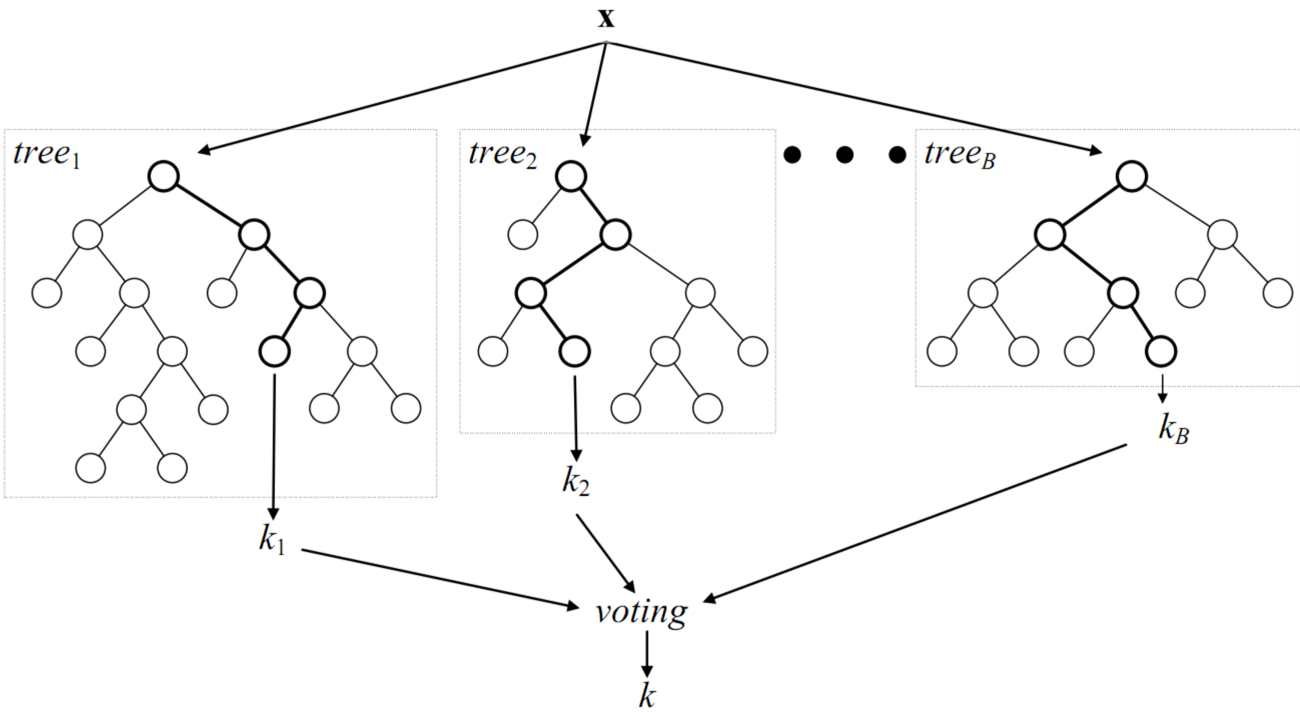
- empirical mode decomposition (EMD) – automatic number of IMFs
  - Torres ME, Colominas MA, Schlotthauer G (2011) **A complete Ensemble Empirical Mode decomposition with adaptive noise**. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4144-4147.
- variational model decomposition (VMD)
  - Dragomiretskiy K, Zosso D (2014) **Variational Mode Decomposition**. IEEE Transactions on Signal Processing, 62(3), 531–544.

Perceptual Linear Predictive Cepstral Coefficients

- compact “frequency signature” (Hermansky, 1990)
- alternative to Mel-Frequency Cepstral Coefficients



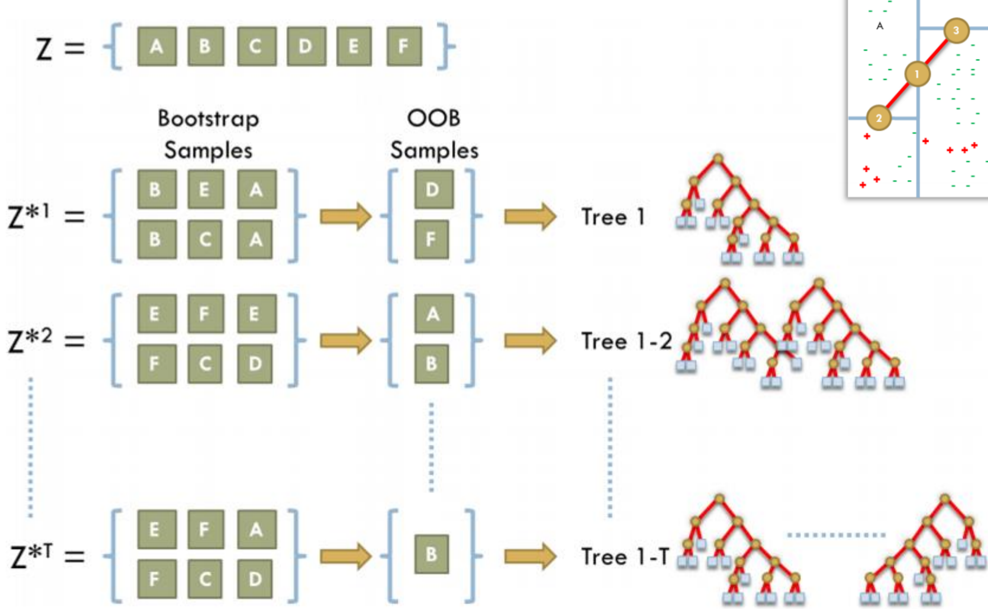
Detector – random forest (RF) (Breiman, 2001)

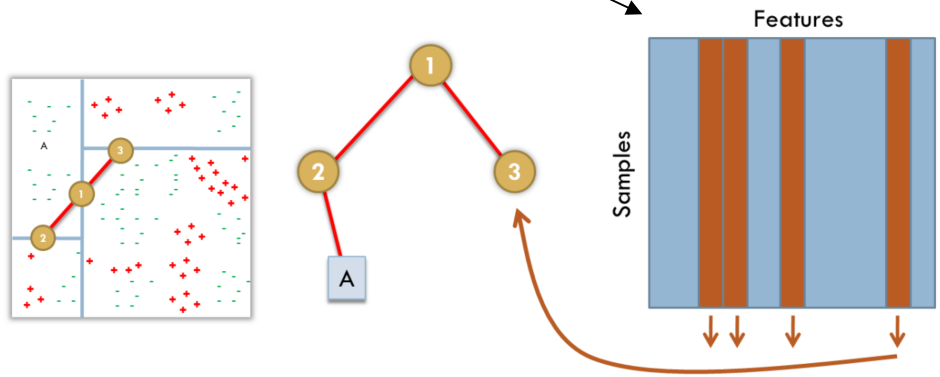


- RF is a committee of classification and regression trees (CART). Final decision *k* is derived by majority voting.
  - B (number of trees in the forest) = 5000
  - mtry (number of features for each node) = $\sqrt{p}$, $2\cdot\sqrt{p}$, $\frac{1}{2}\cdot p$
    - *p* – dimensionality of the object **x** (size of the feature vector)
- 2D map: proximity matrix → distance matrix → t-SNE

RF – ensemble of unpruned CART trees

- efficient and robust against over-fitting
  - features out-of-bag (OOB) validation - no k-fold CV needed
  - computes proximity matrix (~ similarity between recordings)
  - bootstrap aggregating + random subspace
    (instance bagging + feature bagging)





Decision-level fusion: base-RFs by meta-RF

- Base-learner RF is decomposition- and window-specific.
- Meta-learner RF combines decisions from base RFs.
- Input to the meta-learner is the difference between class posteriori probabilities obtained from the base-learner:

$$d(\{t_1, ..., t_L\}, \mathbf{x}) = \frac{\sum_{i=1}^{L} f(t_i, \mathbf{x}, c=2)}{L} - \frac{\sum_{i=1}^{L} f(t_i, \mathbf{x}, c=1)}{L}$$

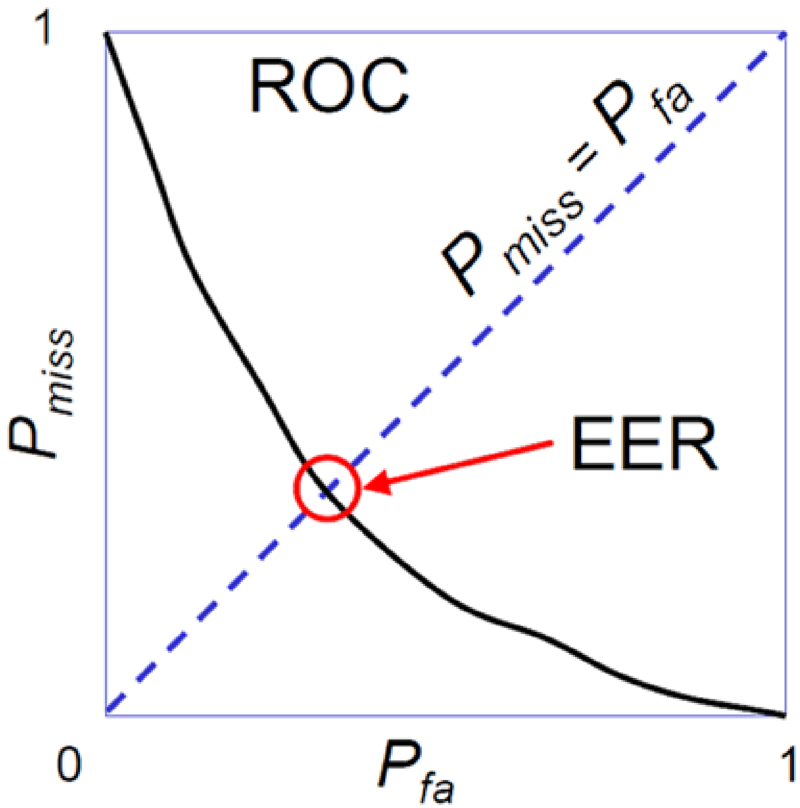
  - **x** – object being classified, L – number of trees $t_1,..t_L$ in the base RF for which object **x** is OOB, *c* is class label (1 – HC, 2 – PD)
  - f(t_i, **x**, *c*) stands for the *c*-th class frequency in the leaf node, into which **x** falls in the *i*-th tree t_i of the base RF:

$$f(t_i, \mathbf{x}, c) = \frac{n(t_i, \mathbf{x}, c)}{\sum_{j=1}^{C} n(t_i, \mathbf{x}, c_j)}$$

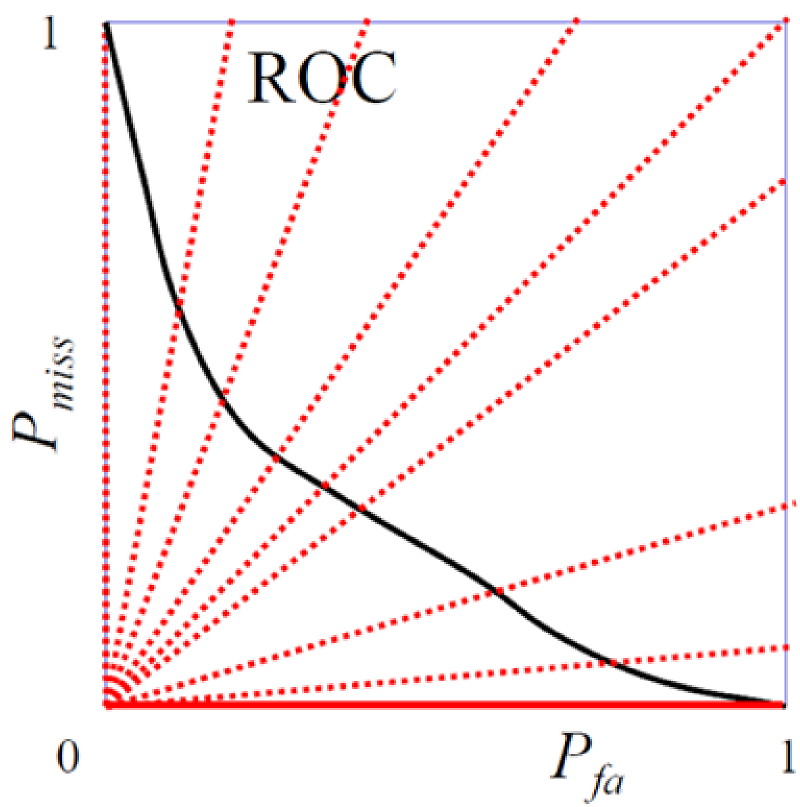
  - *C* – number of classes, n(t_i, **x**, *c*) – number of training data from class *c* falling into the same leaf node of t_i as **x**

Measuring the goodness of detection

- EER = [0 .. 50 .. 100%]
  - *equal error rate*
- ROC (DET) crosses diagonal
  - or sensitivity = specificity
  - or miss rate = false alarm



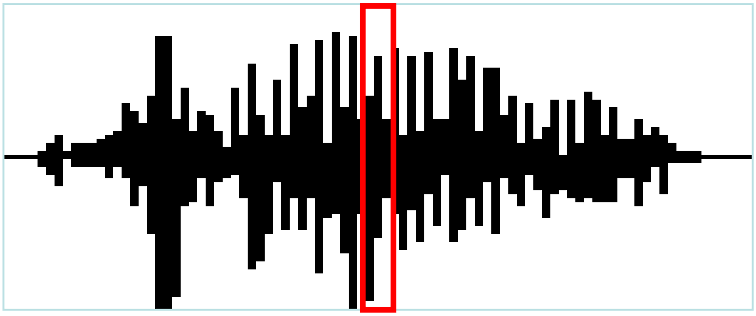
- C_{llr} = [0 .. $\log_2 N$ .. ∞]
  - *cost of log-likelihood-ratio*
  - *~ multi-class cross-entropy*
- integral of ROC curve & costs & class priors (universal measure)



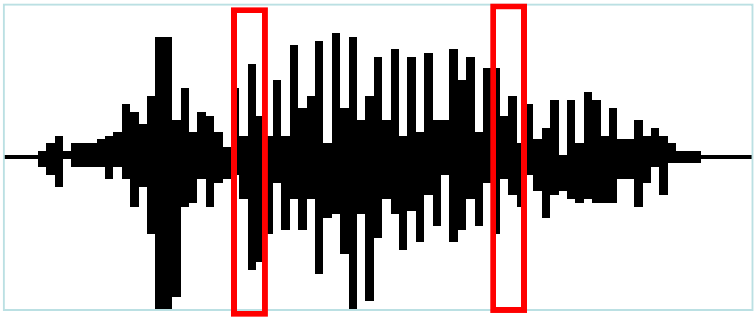
Experimental setup: windowing scheme

Window size (0.01 – 0.10 s)

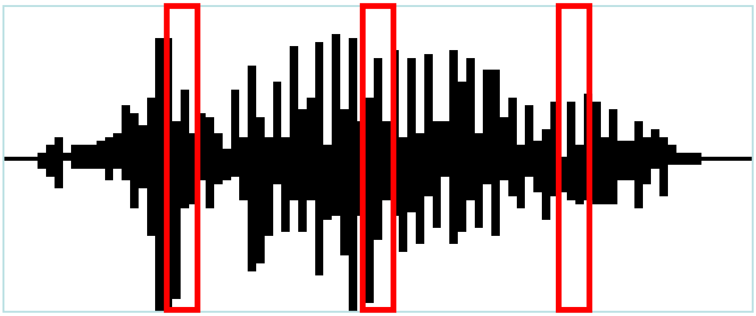
1 window



2 windows



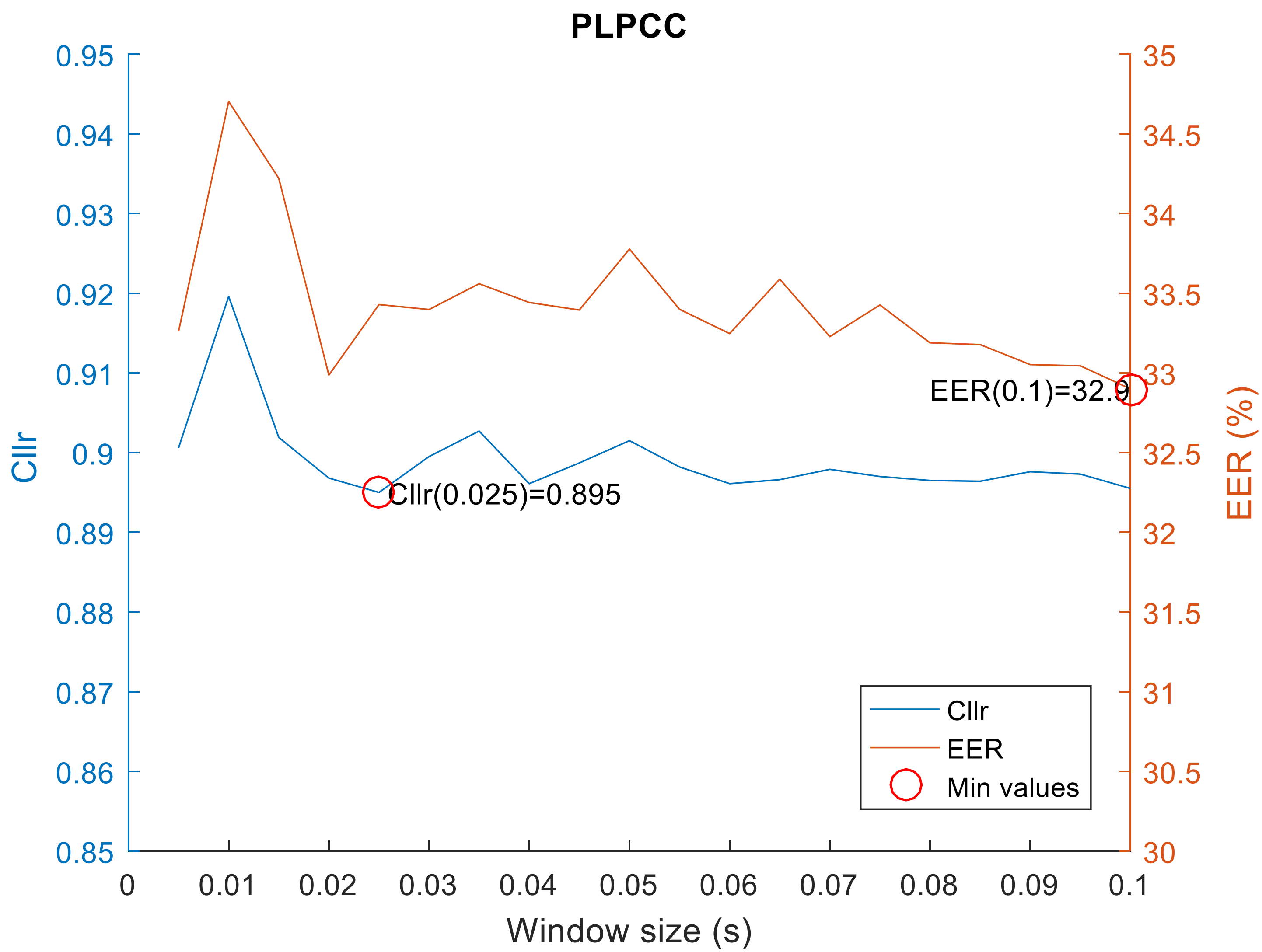
3 windows



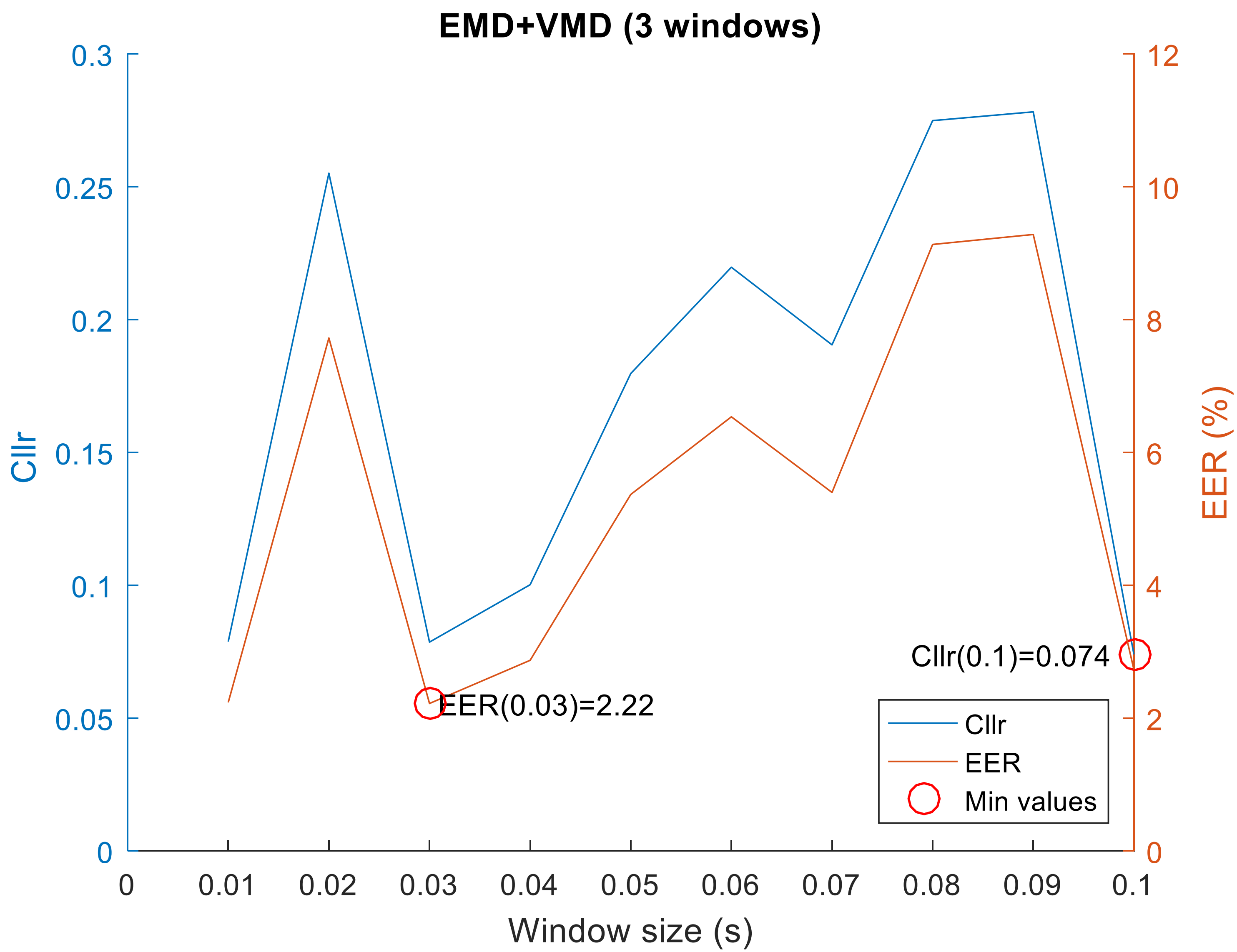
Experimental results: window size and quantity

Size (s)	Quantity	Type of decomposition used in decision-level fusion					
		EMD		*VMD*		*EMD+VMD*	
		Cllr	*EER*	*Cllr*	*EER*	*Cllr*	*EER*
0.01	1	0.718	26.23	0.237	6.65	0.215	6.10
	2	0.617	20.83	0.754	26.20	0.565	19.12
	3	0.115	3.03	0.245	6.71	0.079	2.24
0.02	1	0.448	13.96	0.682	23.03	0.375	12.42
	2	0.385	13.08	0.610	20.32	0.336	10.45
	3	0.321	9.70	0.499	16.66	0.255	7.72
0.03	1	0.456	13.92	***0.100***	***2.65***	0.085	***2.22***
	2	0.306	10.07	0.503	14.95	0.264	8.27
	3	0.342	11.32	0.115	3.33	0.079	***2.22***
0.04	1	0.419	11.77	0.586	19.08	0.359	11.51
	2	0.379	11.69	0.541	17.01	0.313	10.05
	3	***0.010***	***2.90***	0.451	13.70	0.100	2.87
0.05	1	0.355	10.71	0.583	19.02	0.303	8.53
	2	0.347	10.92	0.539	16.67	0.300	9.73
	3	0.216	5.80	0.449	13.36	0.180	5.37
0.06	1	0.316	10.08	0.555	17.16	0.266	7.88
	2	0.328	10.07	0.441	14.23	0.265	8.41
	3	0.244	7.39	0.419	13.24	0.220	6.54
0.07	1	0.405	11.84	0.561	17.15	0.349	10.61
	2	0.450	14.68	0.474	15.48	0.380	12.73
	3	0.236	6.07	0.356	12.15	0.190	5.40
0.08	1	0.379	12.09	0.529	16.85	0.309	9.93
	2	0.323	11.27	0.505	15.91	0.288	9.78
	3	0.297	9.87	0.448	14.11	0.275	9.13
0.09	1	0.443	13.34	0.591	19.59	0.368	11.24
	2	0.349	10.56	0.504	16.10	0.307	9.84
	3	0.317	10.11	0.458	13.90	0.278	9.28
0.10	1	0.363	10.78	0.495	16.21	0.252	7.66
	2	0.390	12.17	0.509	15.64	0.359	12.07
	3	0.254	7.21	0.109	3.86	***0.074***	2.73

Experimental results – baseline solution

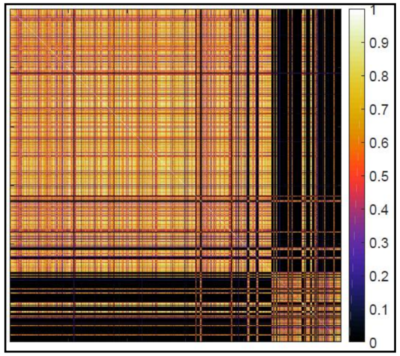


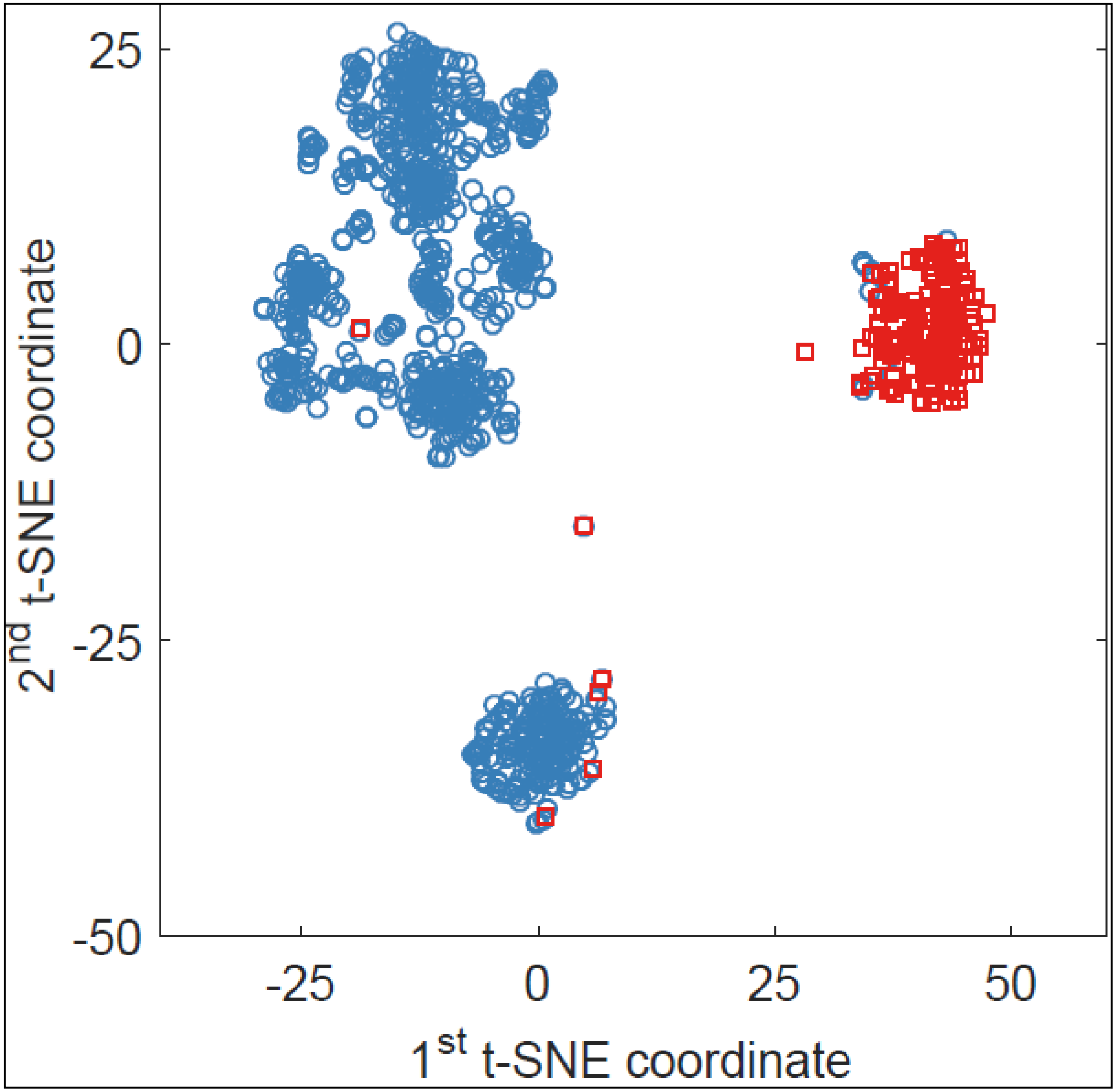
Experimental results – proposed solution



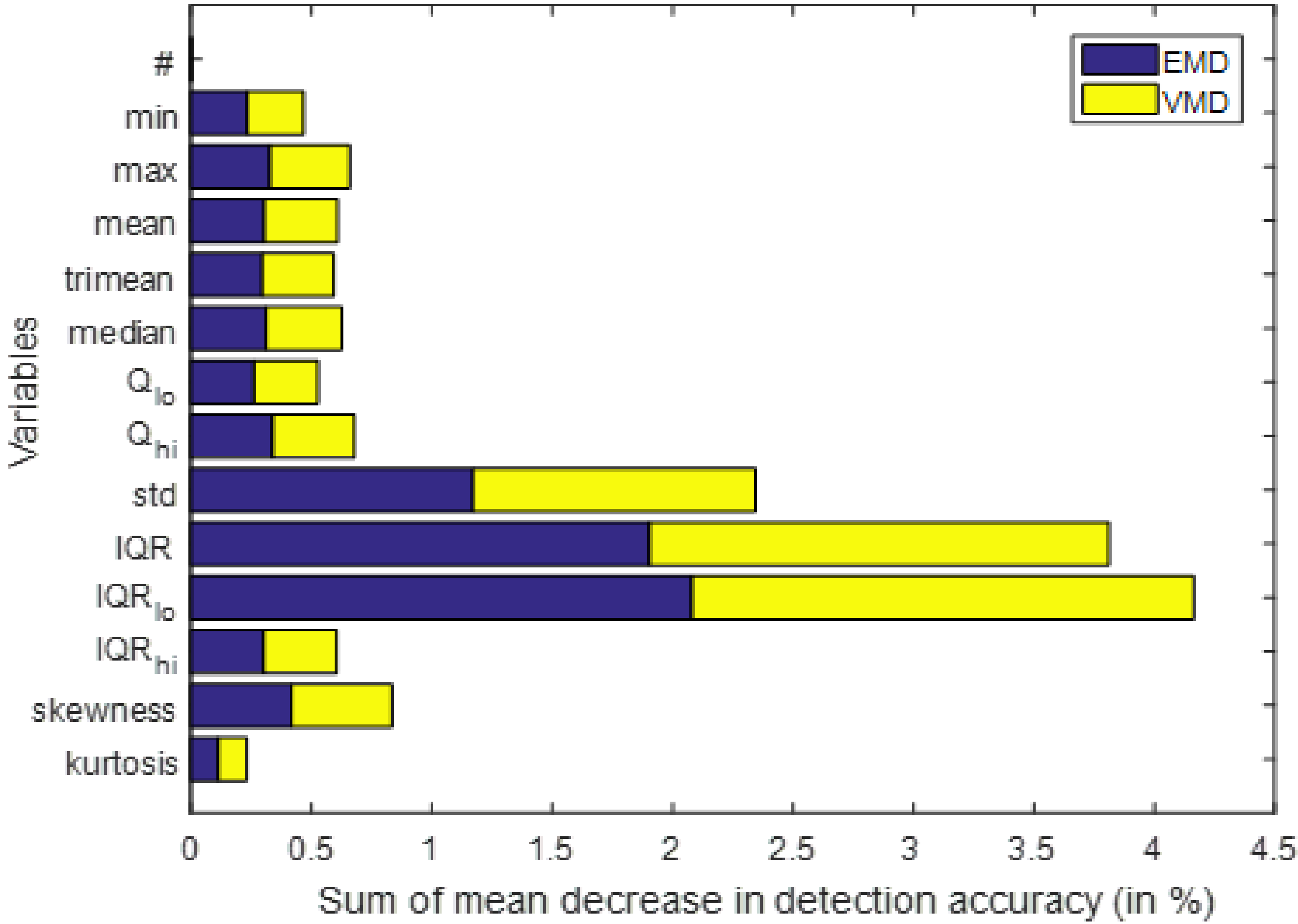
Medical decision support by visualization

- meta-RF proximity matrix → t-SNE → 2D map of .wavs



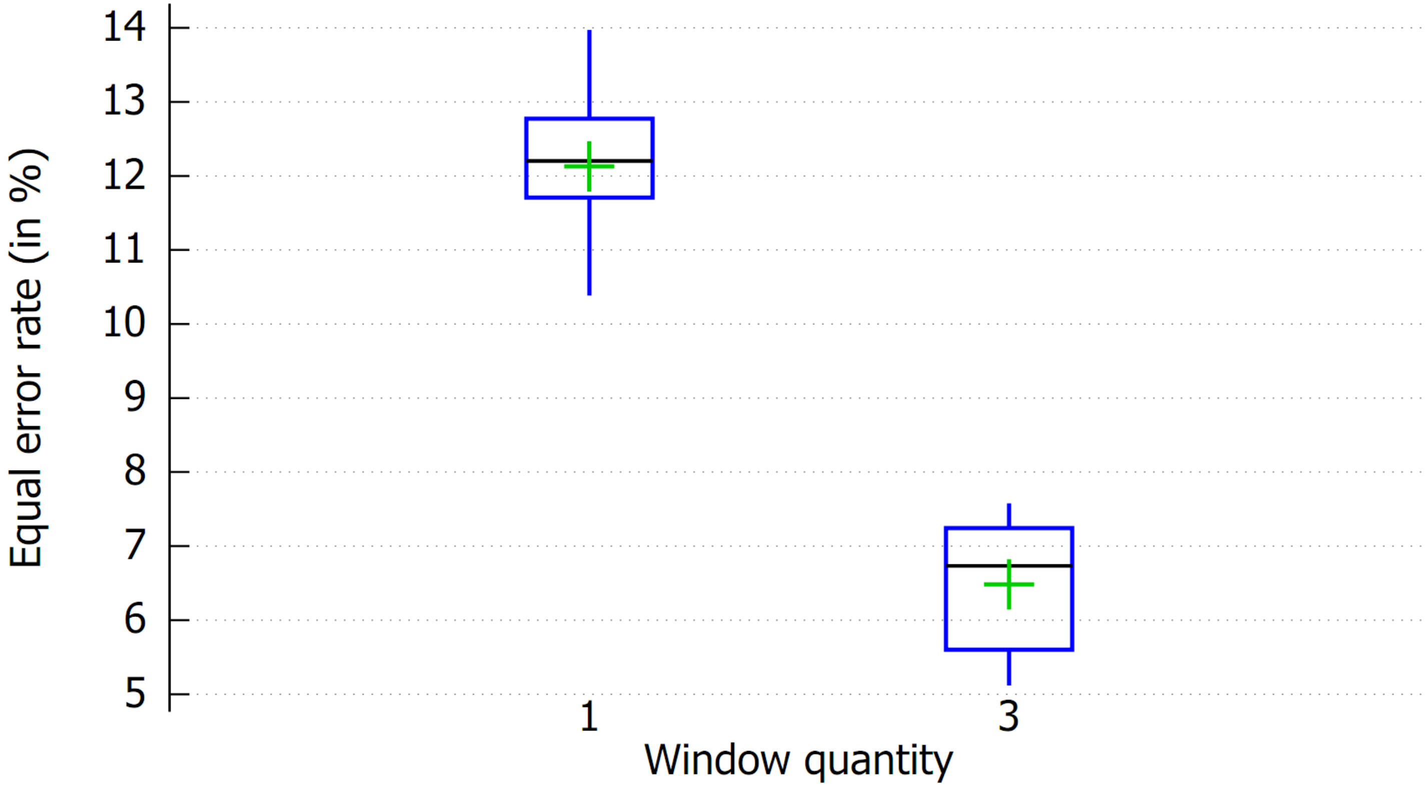


Variable importance from meta-RF



Sensitivity analysis – repeated detections

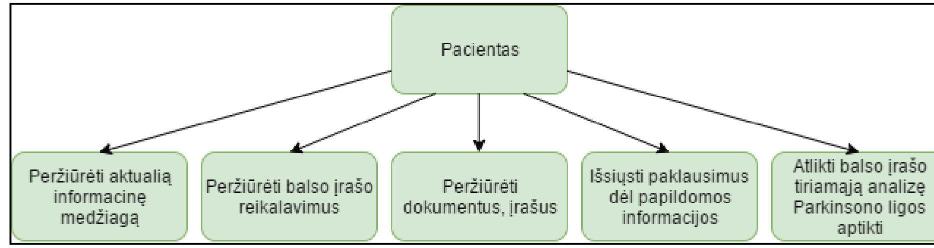
- (5 EMD+VMD runs) × (5 base-RF) × (5 meta-RF) = 125
- medians are significantly different by Wilcoxon rank-sum

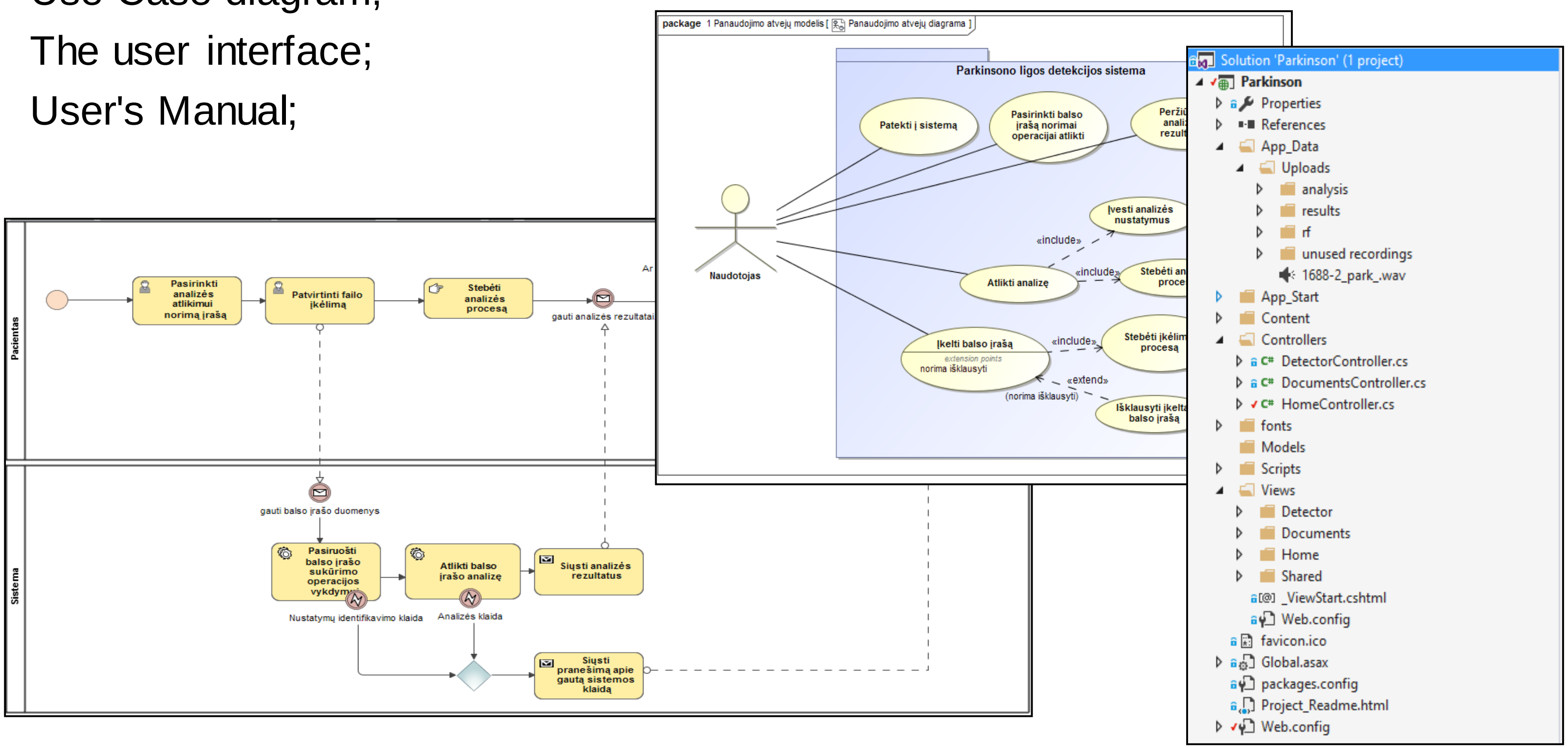


	mean	min	Q1	median	Q3	max
1	**12.129**	10.386	11.708	**12.203**	12.772	13.972
3	**6.4818**	5.1183	5.6006	**6.7329**	7.2433	7.5783

Designing expert system prototype

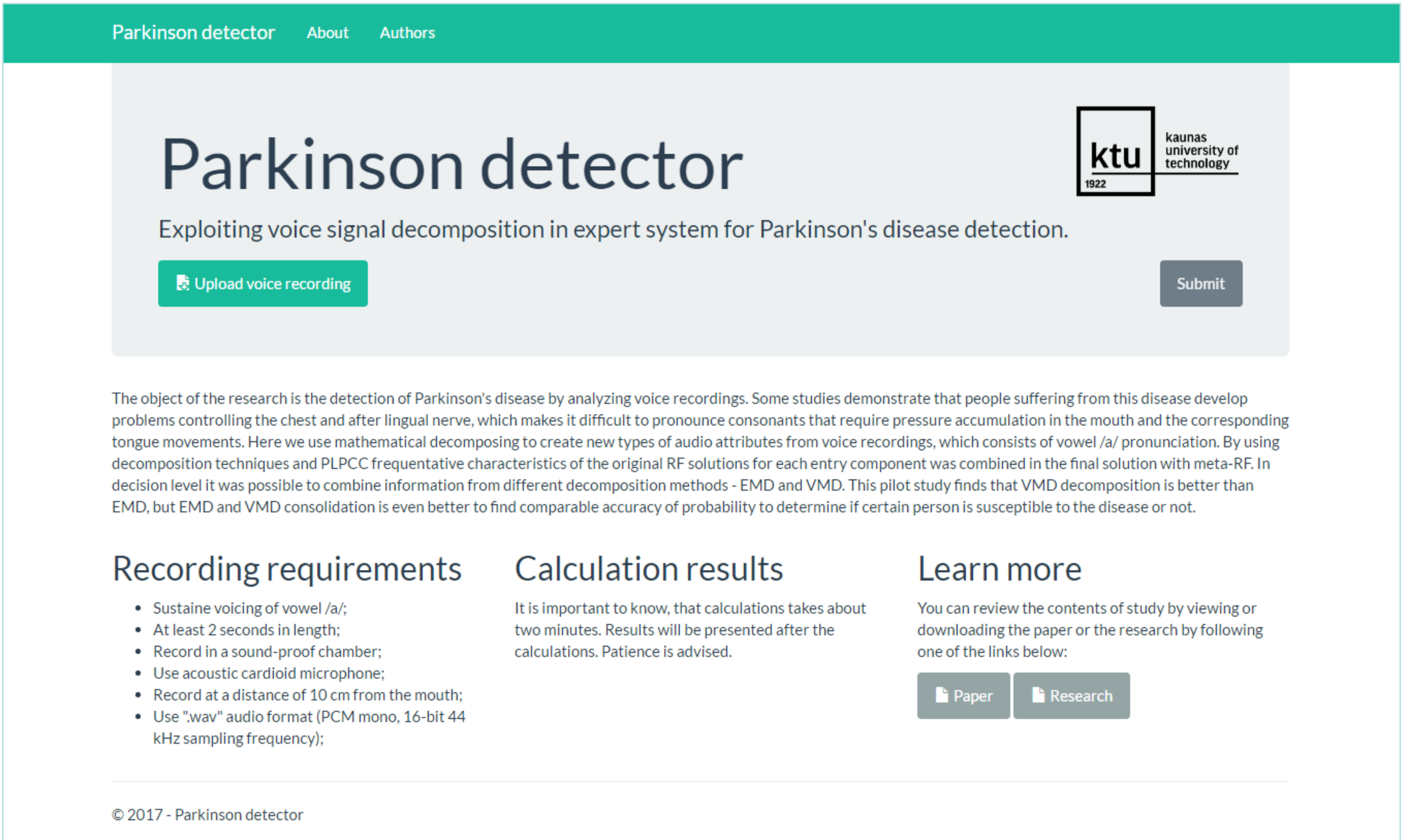
- Product definition, system limitations;
- Functional and non-functional requirements;
- Hierarchy chart;
- Technical specifications;
- The logical architecture;
- The business logic;
- Use Case diagram;
- The user interface;
- User's Manual;





Graphical user interface for expert system

- web-based solution to upload .wav and obtain diagnosis



Conclusions and future work

- Detection performance:
  - (baseline, without decomposition) best EER = ~33%
  - (proposed, 1 window of 30 ms) mean EER = ~12.1%
  - (proposed, 3 windows of 30 ms) mean EER = ~6.5%
- Fusion of 3 windows and EMD+VMD is recommended.
- Both decompositions are useful, VMD better than EMD.
- Variability-related statistics most important for meta-RF.

- Future work:
  - repeated mode decomposition due to numerical instability
    - base-RF would be trained on more IMF components
    - meta-RF would be enriched by more robust statistics
  - test with recordings done using smartphone microphone
  - test expert system with another voice database (mPower)