PHD COMPUTER SCIENCE

Robust Learning Methods for Imprecise Data and Cautious Inference

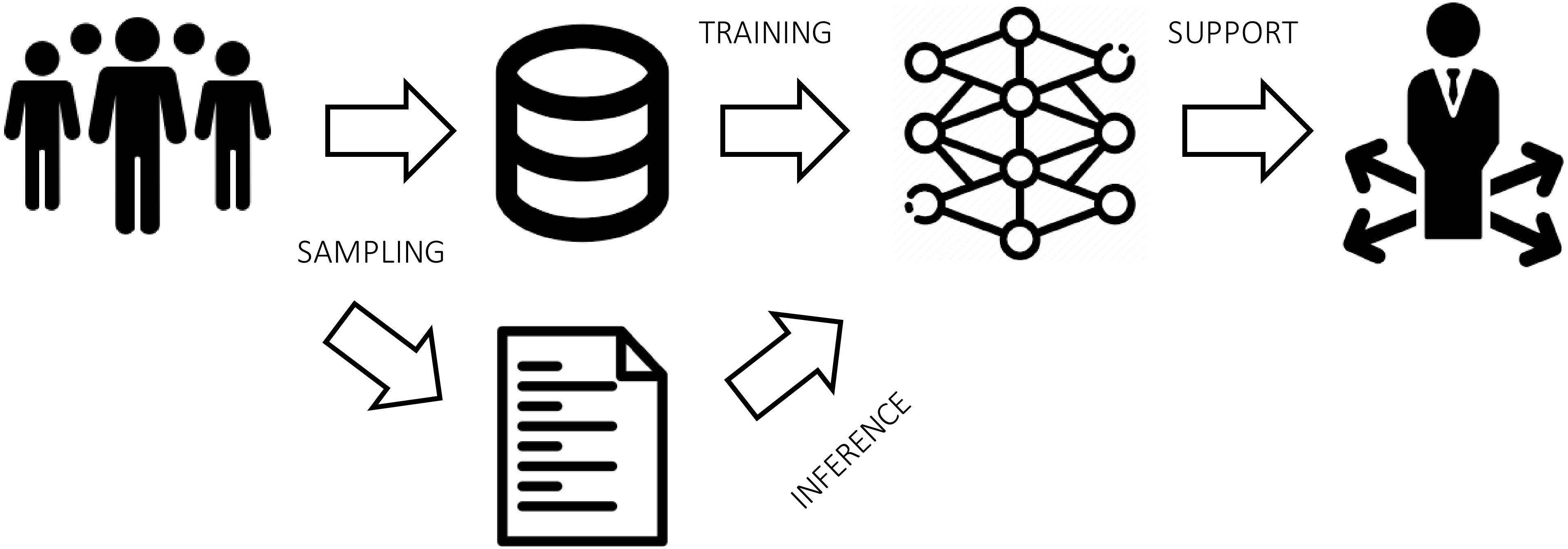
Andrea Campagner

PhD Candidate in Computer Science (XXXV cycle)

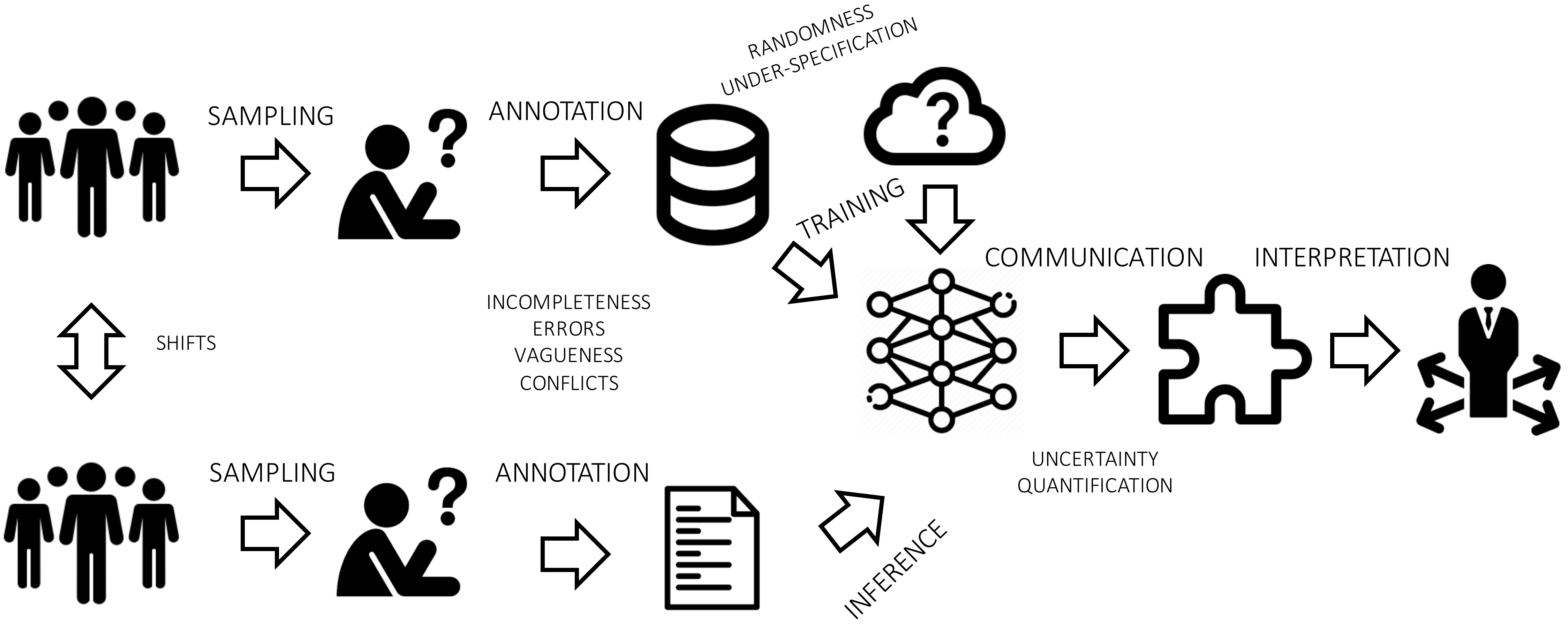
Supervisors: Davide Ciucci, Federico Cabitza

Tutor: Gianluigi Ciocca

Motivation: The Machine Learning Process (Ideal)



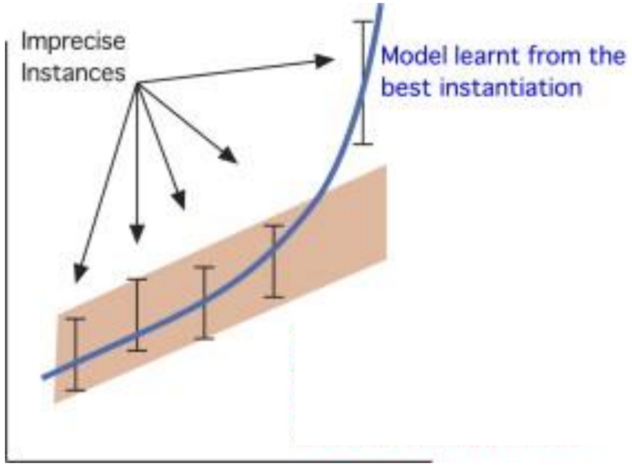
Motivation: The Machine Learning Process (Real)

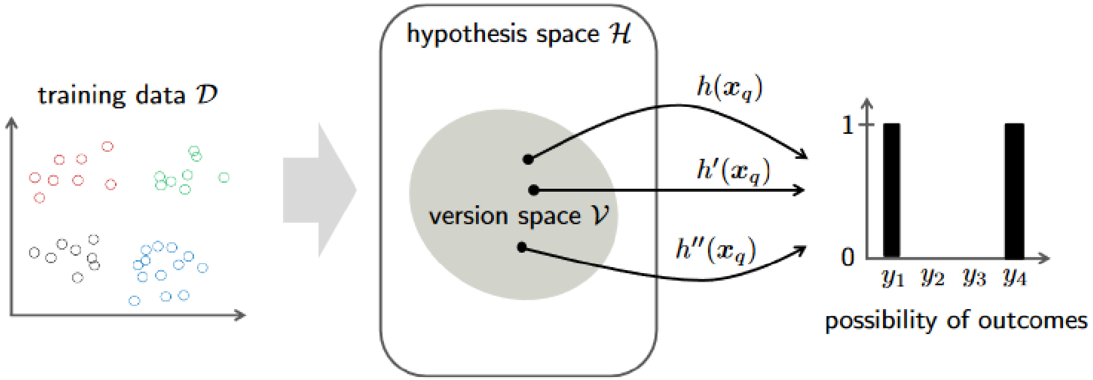


Basile, V. (2020). It's the End of the Gold Standard as We Know It. In *International Conference of the Italian Association for Artificial Intelligence*

Cabitza, F., Ciucci, D., & Rasoini, R. (2019). A giant with feet of clay: On the validity of the data that feed machine learning in medicine. *Organizing for the digital world*

Hildebrandt, M. (2020). The Issue of Bias. The Framing Powers of ML. *Machine Learning and Society: Impact, Trust, Transparency*





Imprecision in Machine Learning

Research Aim

Develop theory and robust algorithms **for** Imprecision in ML

- Imprecision in the Input: **Learning from fuzzy labels**
- Imprecision in the Output: **Cautious Inference**

Validation & Evaluation on Real-World Benchmarks

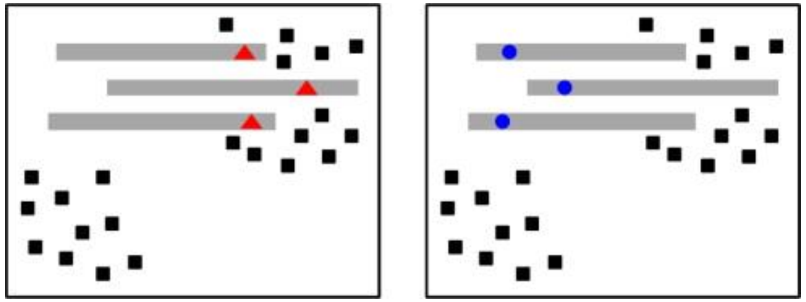
- Focus on **medical tasks:** *COVID, Biological Variability, Tumor Cells, Kyphosis, Knee MRI*
- Collaborations: **IRCCS Istituto Ortopedico Galeazzi, IRCCS Ospedale San Raffaele**

Learning from Imprecise Data

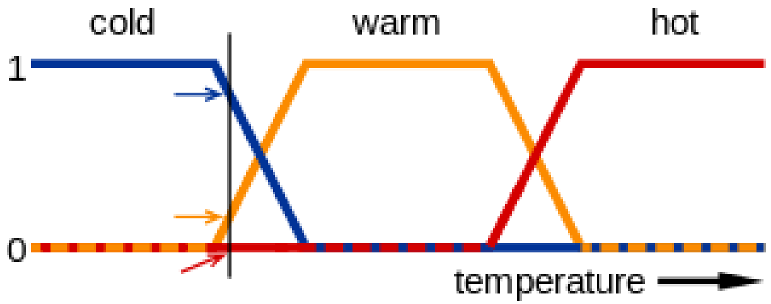
GENERALITY

Height	Weight	Country	Place	Number of days	Some column
12.0	35.0	India	Bengaluru	1.0	NaN
NaN	36.0	US	New York	2.0	NaN
13.0	32.0	UK	London	NaN	NaN
15.0	NaN	France	Paris	4.0	NaN
16.0	39.0	US	California	5.0	12.0
NaN	NaN	NaN	Mumbai	NaN	NaN
NaN	NaN	NaN	NaN	6.0	NaN

MISSING DATA
SEMI-SUPERVISED LEARNING

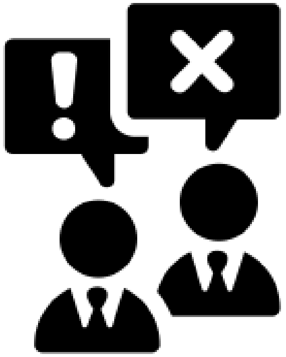


SET-VALUED DATA
SUPERSET LEARNING



FUZZY DATA
LEARNING FROM FUZZY LABELS

Learning from Fuzzy Labels: Why is it Important?

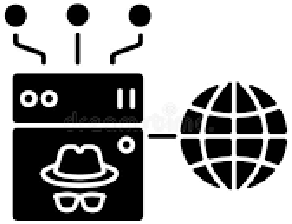


Can arise in many natural ways...

- Expert elicitation
- Multi-source Information Fusion
- Subjective Information

...and settings:

- Medicine
- Natural Language Processing



Cabitza, F., Campagner, A., Basile, V. (2023). Toward a Perspectivist Turn in Ground Truthing for Predictive Computing. AAAI 2023 (Accepted)
Campagner, A., Famiglini, L., Carobene, A., Cabitza, F. (2023). Everything is Varied: The Surprising Impact of Individual Variation on ML Reliability in Medicine. IEEE Transactions on Neural Networks and Learning Systems (Under Review)
Campagner, A., Ciucci, D., Svensson, C. M., et al. (2021). Ground truthing from multi-rater labeling with three-way decision and possibility theory. Information Sciences, 545, 771-790
Lienen, J., Hüllermeier, E. (2021). From label smoothing to label relaxation. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*
Cabitza, F., Ciucci, D., & Rasoini, R. (2019). A giant with feet of clay: On the validity of the data that feed machine learning in medicine. *Organizing for the digital world*
Svensson, CM., Hübler, R., Figge, MT. (2015). Automated classification of circulating tumor cells and the impact of interobserver variability on classifier training and performance. *Journal of Immunology Research*

Learning from Fuzzy Labels: Why is it Important?

Can arise in many natural ways...

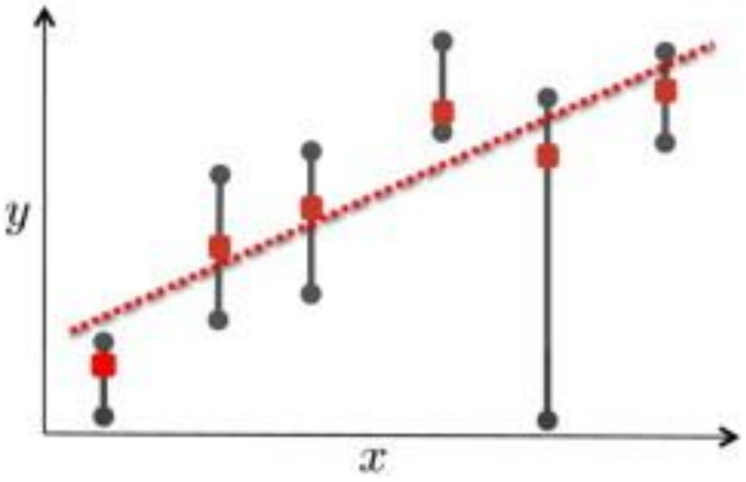
Standard ML tools cannot be applied

- Pre-processing
- Generative assumptions

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	$\{0 : 0.5, 1 : 1.0\}$
x_3	0	1	1	0	0
x_4	0	1	1	1	$\{0 : 1.0, 1 : 0.5\}$
x_5	0	1	0	1	1
x_6	0	1	0	0	$\{0 : 0.5, 1 : 1.0\}$

de Campos, CP., Antonucci, A. (2015). Imprecision in Machine Learning and AI. *IEEE Intelligent Informatics Bulletin*
Couso, I., Borgelt, C., Hullermeier, E., Kruse, R. (2019). Fuzzy sets in data analysis: From statistical foundations to machine learning. *IEEE Computational Intelligence Magazine*

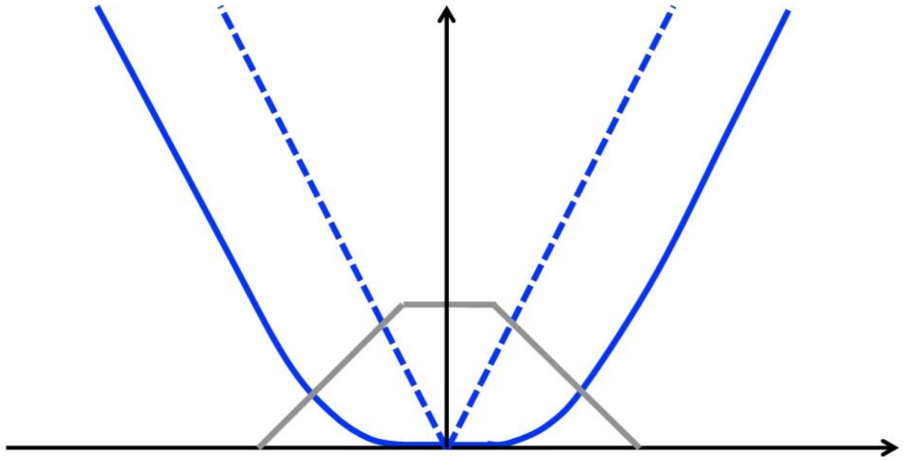
Learning from Fuzzy Labels: State-of-the-Art



Most popular approaches:

Generalized Risk Minimization

Instance-based Learning



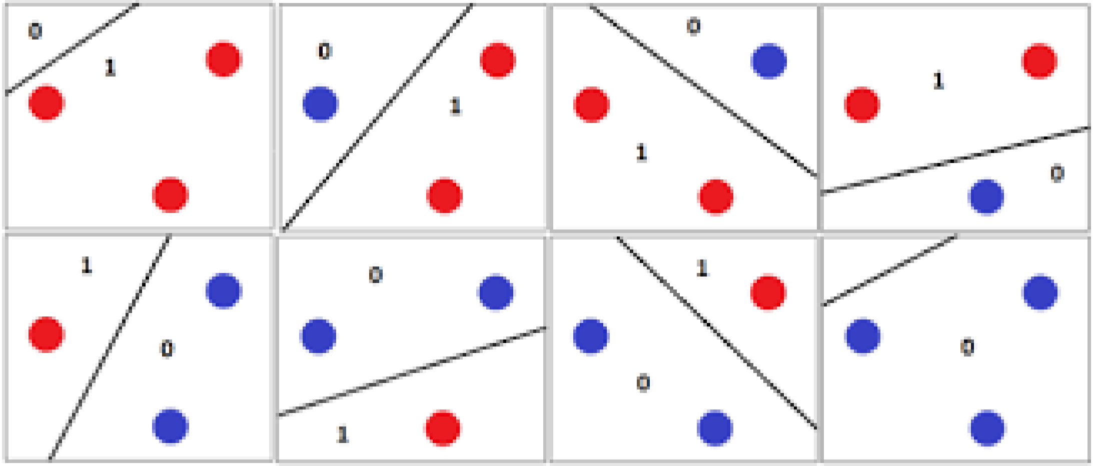
Lots of work (both theoretical and empirical) on semi-supervised and superset learning... not so much in learning from fuzzy labels!

Cabannes, V., Rudi, A., Bach, F. (2020). Structured prediction with partial labelling through the infimum loss. *International Conference on Machine Learning*
Hüllermeier, E. (2014). Learning from imprecise and fuzzy observations: Data disambiguation through generalized loss minimization. *International Journal of Approximate Reasoning*
Hüllermeier, E., Cheng, W. (2015). Superset learning based on generalized loss minimization. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*
Hüllermeier, E., Destercke, S., Couso, I. (2019). Learning from imprecise data: adjustments of optimistic and pessimistic variants. *International Conference on Scalable Uncertainty Management*
Quost, B., Denoeux, T. (2016). Clustering and classification of fuzzy data using the fuzzy EM algorithm. *Fuzzy Sets and Systems*

Learning from Fuzzy Labels: Research Gaps (RQ1)

Limited **theoretical understanding**

- **Learning theory:** when is it possible? With which resources (samples/approximation error)?
- **Computational complexity:** when is it possible to *do it efficiently?*

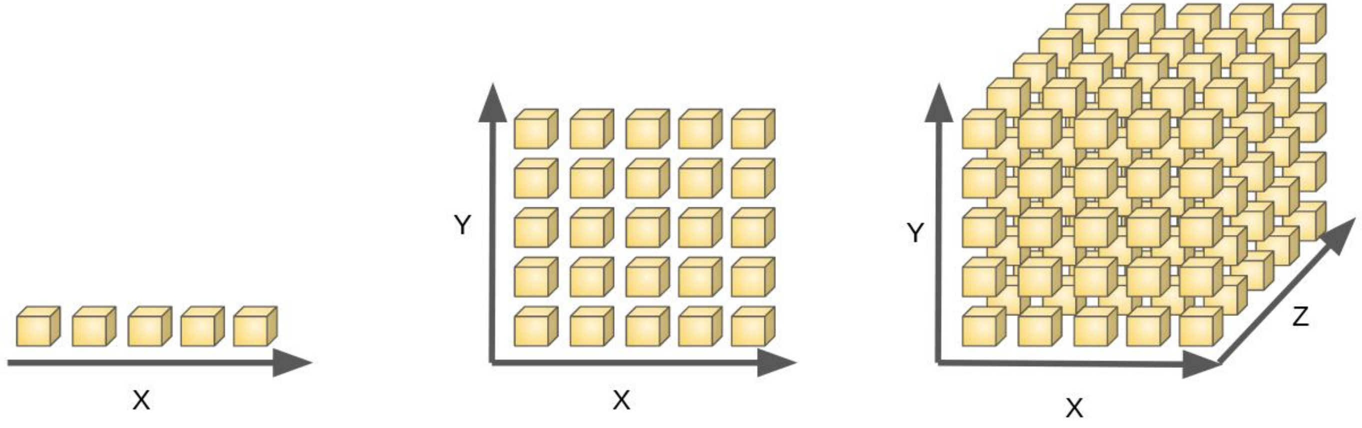


Cabannnes, V., Rudi, A., Bach, F. (2020). Structured prediction with partial labelling through the infimum loss. *International Conference on Machine Learning*.
Liu, L., Dietterich, T. (2014). Learnability of the superset label learning problem. *International Conference on Machine Learning*
Ma, G., Liu, F., Zhang, G., Lu, J. (2021). Learning from Imprecise Observations: An Estimation Error Bound based on Fuzzy Random Variables. *2021 IEEE International Conference on Fuzzy Systems*

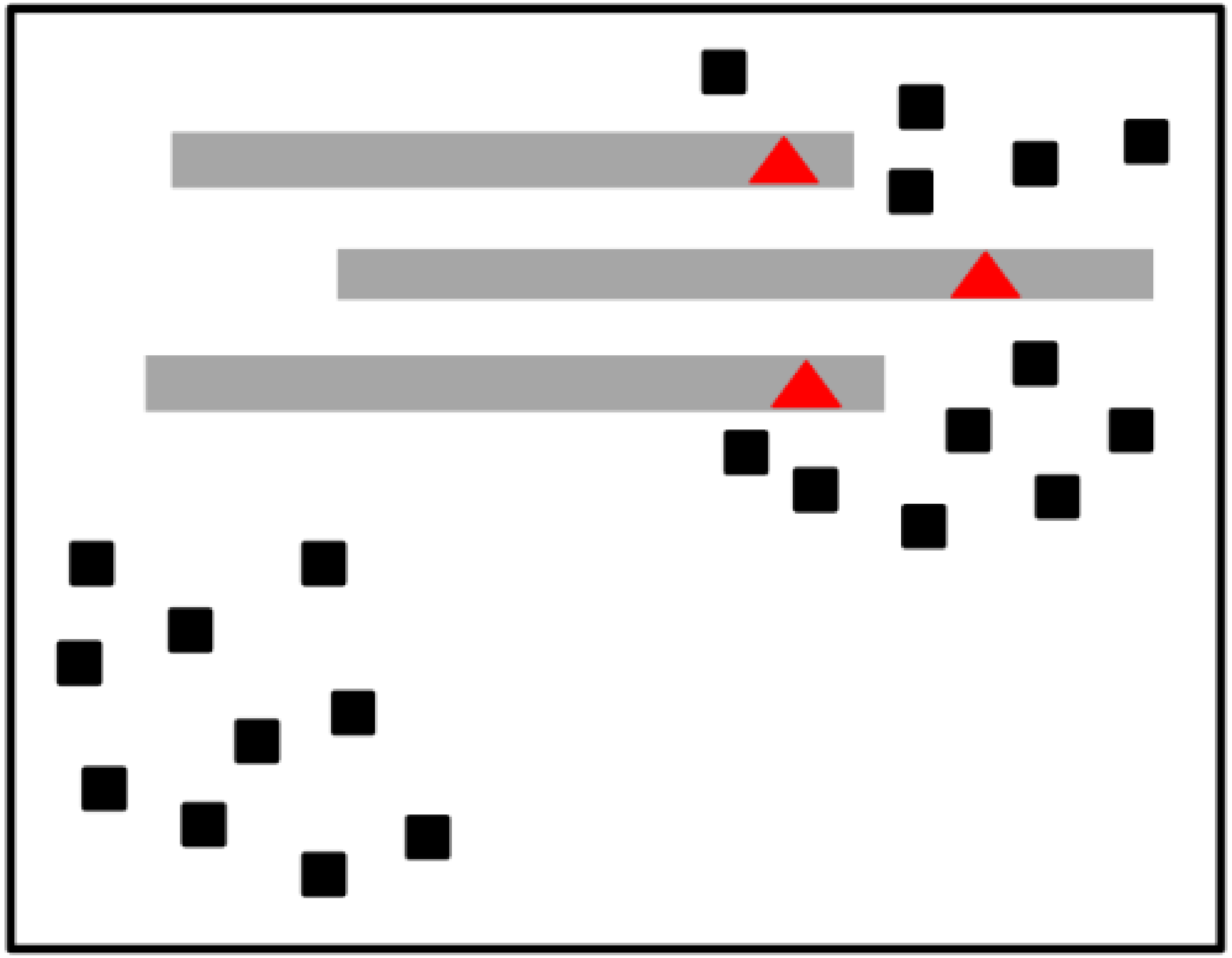
Learning from Fuzzy Labels: Research Gaps (RQ2)

Limited theoretical understanding

Lack of general methods to control **curse of dimensionality**



Billard, L., & Le-Rademacher, J. (2012). Principal component analysis for interval data. *Wiley Interdisciplinary Reviews: Computational Statistics*
Denoeux, T., & Masson, M. H. (2004). Principal component analysis of fuzzy data using autoassociative neural networks. *IEEE Transactions on Fuzzy Systems*
Douzal-Chouakria, A., Billard, L., & Diday, E. (2011). Principal component analysis for interval-valued observations. *Statistical Analysis and Data Mining: The ASA Data Science Journal*.
Li, M. L., Di Mauro, F., Candan, K. S., & Sapino, M. L. (2019). Matrix factorization with interval-valued data. *IEEE Transactions on Knowledge and Data Engineering*.
Wu, J. H., Zhang, M. L. (2019). Disambiguation enabled linear discriminant analysis for partial label dimensionality reduction. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*



For **GRM**, a PAC-learning style bound: *polynomial number of instances is sufficient for achieving desired accuracy...*

$$m_* = O\left(\frac{1}{(\epsilon\theta_{\mathcal{D}})^2}\left(d \cdot ln\left(\frac{d \cdot |Y|}{(\epsilon\theta_{\mathcal{D}})^2}\right) + ln\left(\frac{1}{\delta}\right)\right)\right)$$

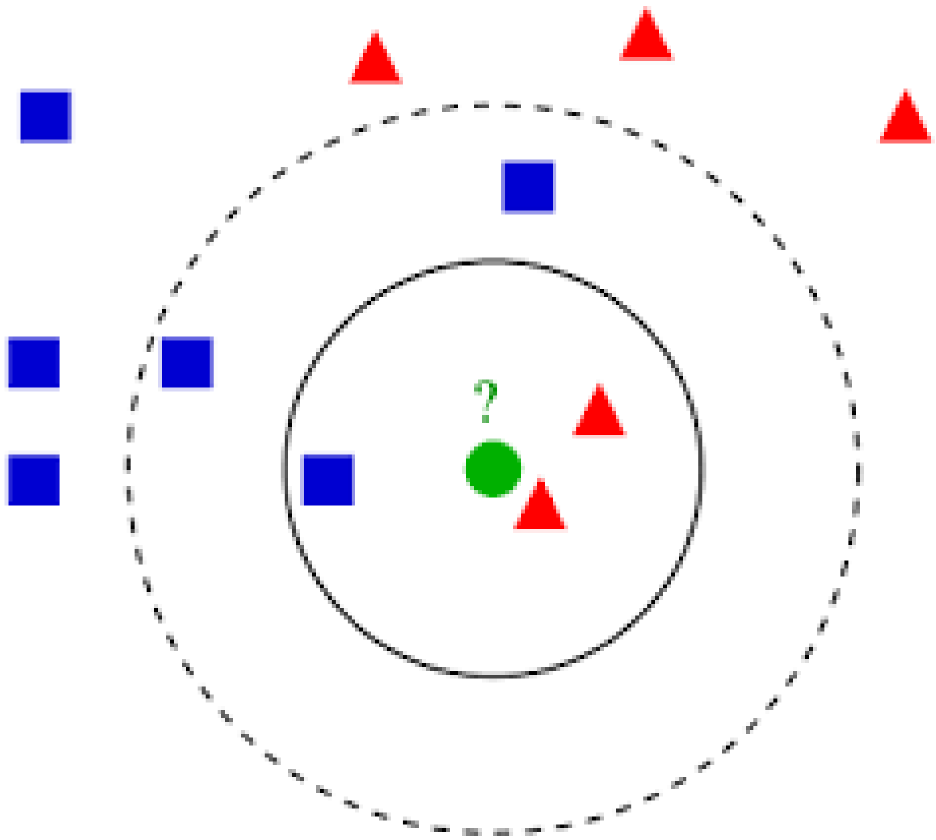
Conditional on the fuzzy labels being not too imprecise (there is information to be extracted!)

$$\theta_{\mathcal{D}} = log_2\left(\frac{2}{1+max\{\phi, 1-k\}}\right)$$

Disregarding the distribution parameters, *guarantees match those for supervised learning...*

However, learning with GRM is not computationally efficient (NP-hard) even for linear models...

Campagner, A. (2021). Learnability in "Learning from Fuzzy Labels". In 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6).



By constrast, instance- based learning **is time-efficient**...

But **not sample-efficient**: need *exponentially-many instances* to reach desired accuracy -> **curse of dimensionality**

$$
\begin{aligned}
\mathbf{E}[L_{\mathcal{D}}(F1-NN(S))] \quad & \leq (\alpha + 2k - k\alpha)|Y|L_{\mathbf{D}}^{2-Bayes} \\
& +(1 + k\alpha - k) \\
& +(1+\alpha+k\alpha)4c\sqrt{d}m^{\frac{-1}{d+1}}
\end{aligned}
$$

Campagner, A. (2021). Learnability in "Learning from Fuzzy Labels". In 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6).

Algorithm The RRL algorithm.

procedure RRL(S: dataset, n: ensemble size, $\mathcal{H}$: base function class)
    $Ensemble \leftarrow \emptyset$
    for all iterations $i = 1$ to n **do**
        Draw a boostrap sample S' from S
        $Tr_i \leftarrow \emptyset$
        for all $(x, \pi) \in S'$ **do**
            Sample $\alpha \sim Uniform[0, 1]$
            Add (x, y') to Tr_i, where $y' \sim Uniform[\pi^\alpha]$
        end for
        Add base model $h_i \in \mathcal{H}$ trained on Tr_i to $Ensemble$
    end for
    return $Ensemble$
end procedure

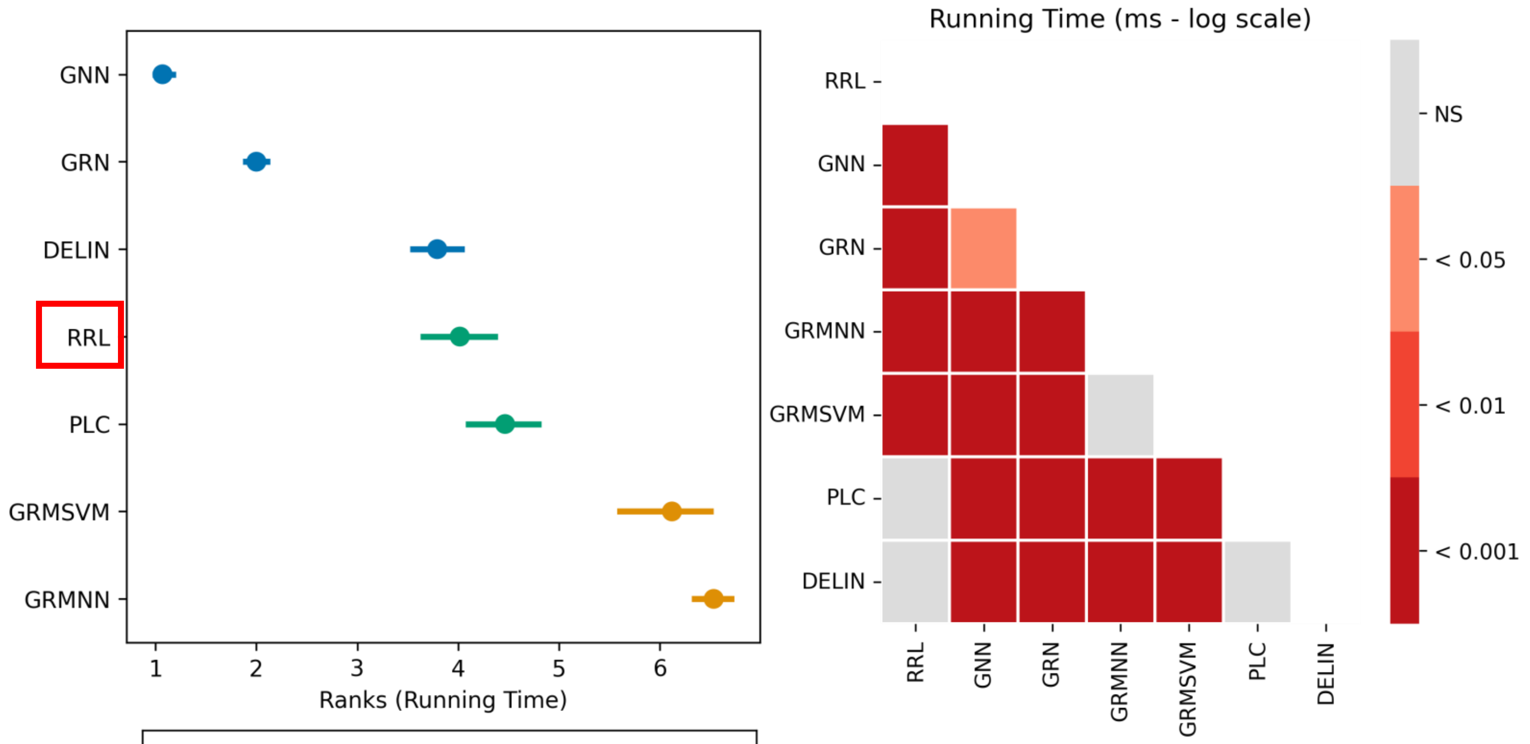
New Method: **RRL** (Random Resampling-based Learning)

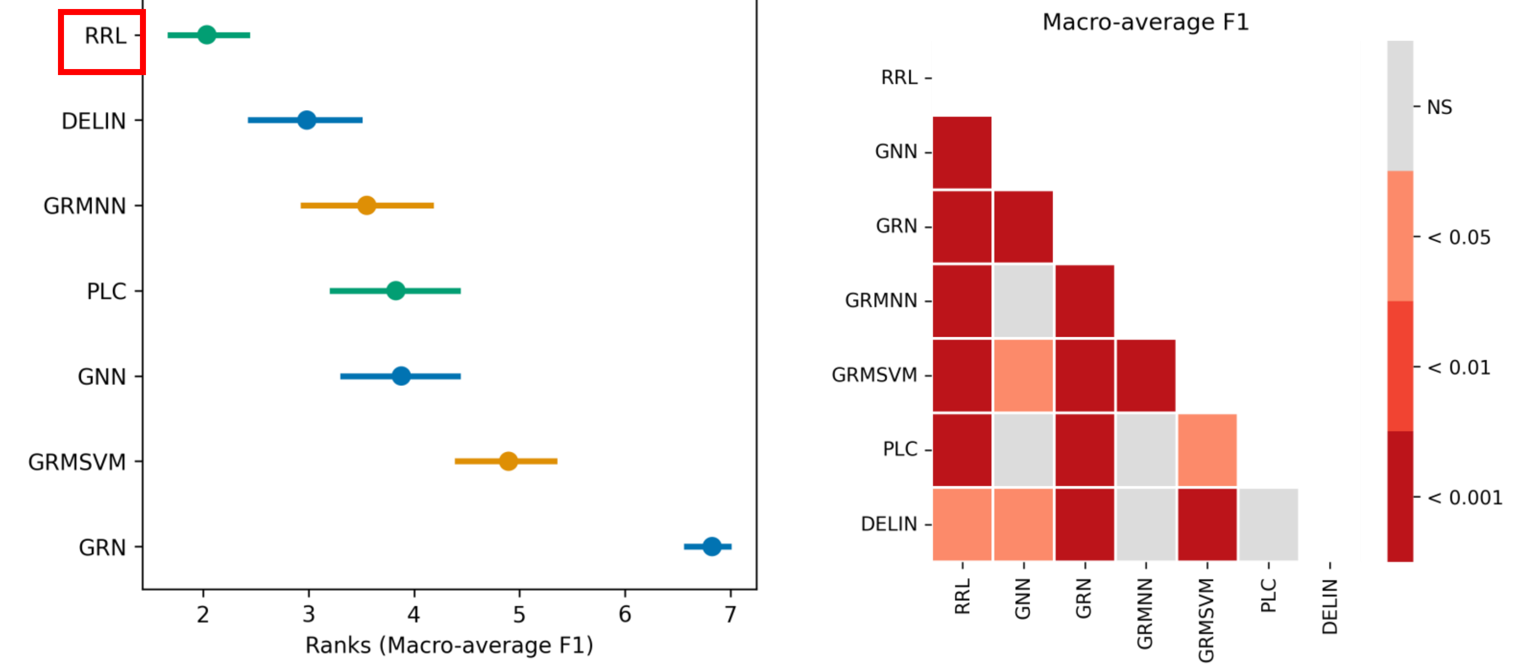
A *pseudo-label learning method* for learning from fuzzy labels
- Not require any ad-hoc algorithm implementation
- Grounds on imprecise probability theory for sampling
- Embarassingly parallel ensemble strategy

Strong theoretical guarantees (derive from sampling scheme)
- **Consistency** (convergence to Bayes classifier)
- **Finite-sample guarantees** (almost) matching those of GRM

$$\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right)\frac{|Y|L}{C}\sqrt{log\frac{2}{\delta}} + \sqrt{\frac{r \cdot ln(\frac{r|Y|^2}{\theta_{\mathcal{D}}^2}) + ln\frac{1}{\delta}}{m\theta_{\mathcal{D}}}} + \sqrt{\frac{K_n + ln\frac{m}{\delta}}{2(m-1)}}$$

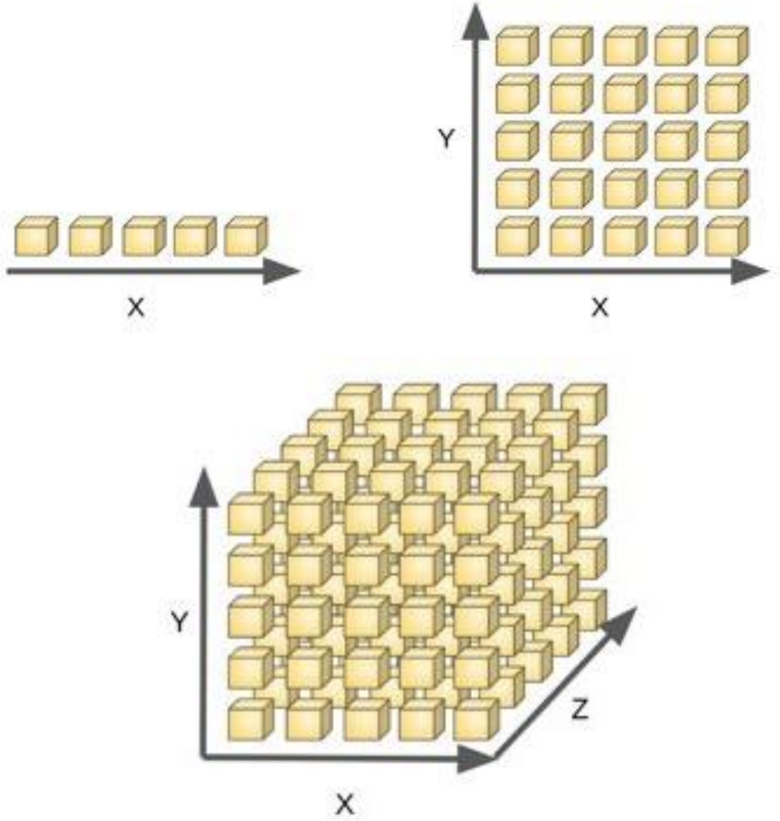
Campagner, A. (2023). **Learning from Fuzzy Labels: Theoretical Issues and Algorithmic Solutions**. International Journal of Approximate Reasoning (Under Review)





	Classes	Features	Instances
UCI datasets			
avila	10	10	20768
banknote	2	4	1372
cancerwisconsin	2	9	683
car	4	16	864
credit	2	61	1000
crowd	6	28	10845
diabetes	2	8	768
digits	10	62	5620
frog-family	4	22	7195
frog-genus	8	22	7195
frog-species	10	22	7195
hcv	4	12	582
htru	2	8	17898
ionosfera	2	33	351
iranian	2	45	7032
iris	3	4	150
mice	8	78	972
mushroom	6	99	5644
myocardial	2	111	1700
obesity	7	31	2111
occupancy	2	5	20560
pen	10	16	10992
robot	4	24	5456
sensorless	11	48	20000
shill	2	9	6321
sonar	2	60	208
vowel	11	9	990
wifi	4	7	2000
wine	3	13	178
Imprecise Datasets			
ctc	2	2500	617
covid	2	69	1624
mri	2	100	427
invasiveness	3	186	72
spine	7	14	120

Learning from Fuzzy Labels: Limitations



Lack of methods to control the curse of dimensionality. Impact on:

- Sample complexity and generalization...
- … but also computational complexity (e.g., norms in high-dimensional spaces)

Dimensionality reduction can reduce model complexity… however, existing SOTA method (**DELIN**) require *strong assumptions on the data* (linearity, normality, etc)!

Propose a non-parametric approach based on an extension of *Rough Set theory* (general approach to feature selection via **reducts**)

Wu, J. H., Zhang, M. L. (2019). Disambiguation enabled linear discriminant analysis for partial label dimensionality reduction. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*
Bao, W. X., Hang, J. Y., & Zhang, M. L. (2021, August). Partial label dimensionality reduction via confidence-based dependence maximization. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 46-54).
Zhang, M. L., Wu, J. H., & Bao, W. X. (2022). Disambiguation Enabled Linear Discriminant Analysis for Partial Label Dimensionality Reduction. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *16*(4), 1-18.

Possibilistic Decision Tables

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	$\{0 : 0.5, 1 : 1.0\}$
x_3	0	1	1	0	0
x_4	0	1	1	1	$\{0 : 1.0, 1 : 0.5\}$
x_5	0	1	0	1	1
x_6	0	1	0	0	$\{0 : 0.5, 1 : 1.0\}$

- The decision attribute is a possibility distribution
- Let $d(x)_y$ be the possibility of label *y* assigned to object *x*
  - We assume that the **real label** *t* has $d(x)_t > 0$
  - If $d(x)_a > d(x)_b$ then *a* is **more plausible** than *b*
  - All labels in the **strong 0-cut** are assumed possible

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	$\{0 : 0.5, 1 : 1.0\}$
x_3	0	1	1	0	0
x_4	0	1	1	1	$\{0 : 1.0, 1 : 0.5\}$
x_5	0	1	0	1	1
x_6	0	1	0	0	$\{0 : 0.5, 1 : 1.0\}$

$\Downarrow$

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	1
x_3	0	1	1	0	0
x_4	0	1	1	1	0
x_5	0	1	0	1	1
x_6	0	1	0	0	1

PDTs: Instantiations

- An instantiation is an assignment of labels which is compatible with the imprecise information
  - It is a "classic" Decision Table!
- We can define an order on instantiations:

  $I_1 <_C I_2$ *iff* $min_x\, d(x)_{y1} < min_x\, d(x)_{y2}$

- $<_C$ defines a *possibility distribution* p_C *over instantiations*

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	$\{0 : 0.5, 1 : 1.0\}$
x_3	0	1	1	0	0
x_4	0	1	1	1	$\{0 : 1.0, 1 : 0.5\}$
x_5	0	1	0	1	1
x_6	0	1	0	0	$\{0 : 0.5, 1 : 1.0\}$

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	$\{0, 1\}$
x_3	0	1	1	0	0
x_4	0	1	1	1	$\{0, 1\}$
x_5	0	1	0	1	1
x_6	0	1	0	0	$\{0, 1\}$

A Detour: Superset Decision Tables

- A specific case of PDT: sets of possible values
- **Superset Reduct**: a set of attributes *R* which is *a reduct for some instantiation*
- **Minimum Description Length (MDL) Reduct**: a *size-minimal Superset Reduct*
- Each α-cut of a PDT gives an SDT S^{α}
  - Take all labels with possibility degree greater than α

Campagner, A., Ciucci, D., & Hüllermeier, E. (2021). Rough set-based feature selection for weakly labeled data. *International Journal of Approximate Reasoning*.

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	$\{0:0.5, 1:1.0\}$
x_3	0	1	1	0	0
x_4	0	1	1	1	$\{0:1.0, 1:0.5\}$
x_5	0	1	0	1	1
x_6	0	1	0	0	$\{0:0.5, 1:1.0\}$

0.5-Possibilistic Reduct

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	$\{0, 1\}$
x_3	0	1	1	0	0
x_4	0	1	1	1	$\{0, 1\}$
x_5	0	1	0	1	1
x_6	0	1	0	0	$\{0, 1\}$

α-Possibilistic Reducts

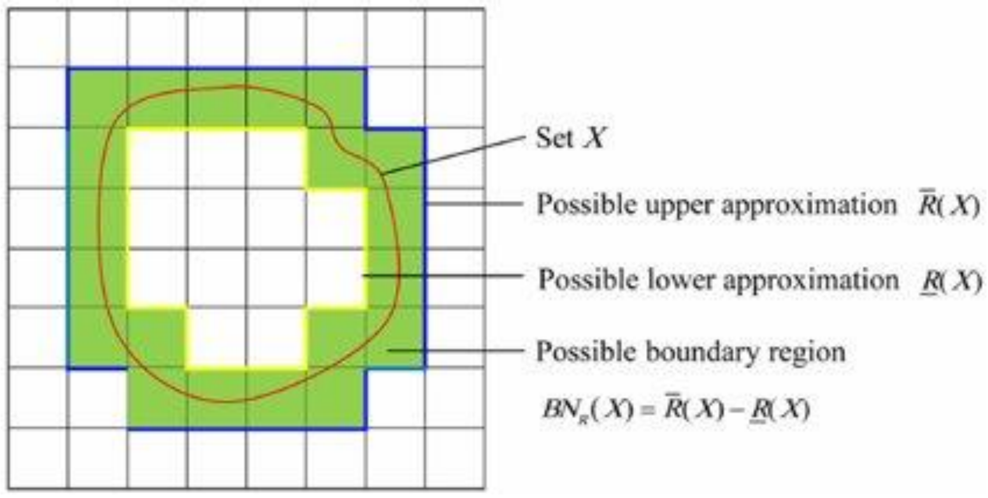
- An *α-Possibilistic Reduct* is a superset reduct for *SDT* S^{α}
- Similarly, we can define *α-MDL Reducts*
- This approach doesn't take into account the orders on the instantiations... but can be used as a starting point

	w	x	v	z	d
x_1	0	0	0	0	0
x_2	0	0	0	1	$\{0 : 0.5, 1 : 1.0\}$
x_3	0	1	1	0	0
x_4	0	1	1	1	$\{0 : 1.0, 1 : 0.5\}$
x_5	0	1	0	1	1
x_6	0	1	0	0	$\{0 : 0.5, 1 : 1.0\}$

Other Definitions of Reducts

- *R* is a *C-Reduct* if it is a reduct for an instantiation I_1 and $\nexists$ R' reduct for I_2 s.t. both $|R'| < |R|$ and $I_1 <_C I_2$
- R is a *λ-Reduct* if $sup_{I \in I(R)} (1-\lambda) p_C(I) - \lambda |R| / |Att|$ is maximal among all possible sets of attributes

Feature Selection: Some Theoretical Results



Theorem: *The problem of finding reducts in learning from fuzzy labels is polynomially reducible to computing standard reducts*

Corollary: *Finding reducts in learning from fuzzy labels is NP-hard: In particular, NP^{NP}-Complete*

In Supervised and Superset learning, greedy local search...

… but doing the same in learning from fuzzy labels ***is not easy***

Campagner, A., Ciucci, D., Hullermeier, E. (2021). Rough set-based feature selection for weakly labeled data. International Journal of Approximate Reasoning, 136, 150-167
Campagner, A., Ciucci, D. (2021). Feature selection and disambiguation in learning from fuzzy labels using rough sets. Lecture Notes in Computer Science, vol 12872

Genetic Rough Set Feature Selection

$$Fitness_C(\langle I, F\rangle) = \langle r, p\rangle,$$
$$Fitness_\lambda(\langle I, F\rangle) = (1-\lambda)p - \lambda\frac{r}{|Att|},$$
$$Fitness_D(\langle I, F\rangle) = \langle r, s\rangle,$$

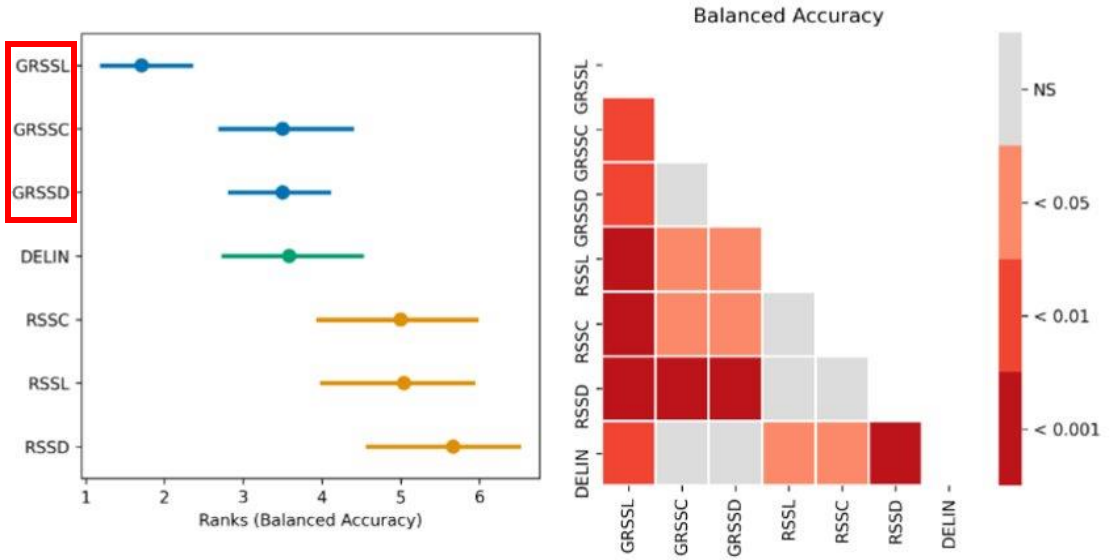
From local to **global approximation**: we encode our definitions of reducts in terms of loss/fitness functions

Optimize these loss functions using *genetic algorithms*

Complexity O(tnm), where t number of generations

Weak Convergence Guarantee: as t grows, the probability that a reduct is in population goes to 1!

Campagner, A., Ciucci, D. (2022). Rough-set Based Genetic Algorithms for Weakly Supervised Feature Selection. Communications in Computer and Information Science, vol 1602



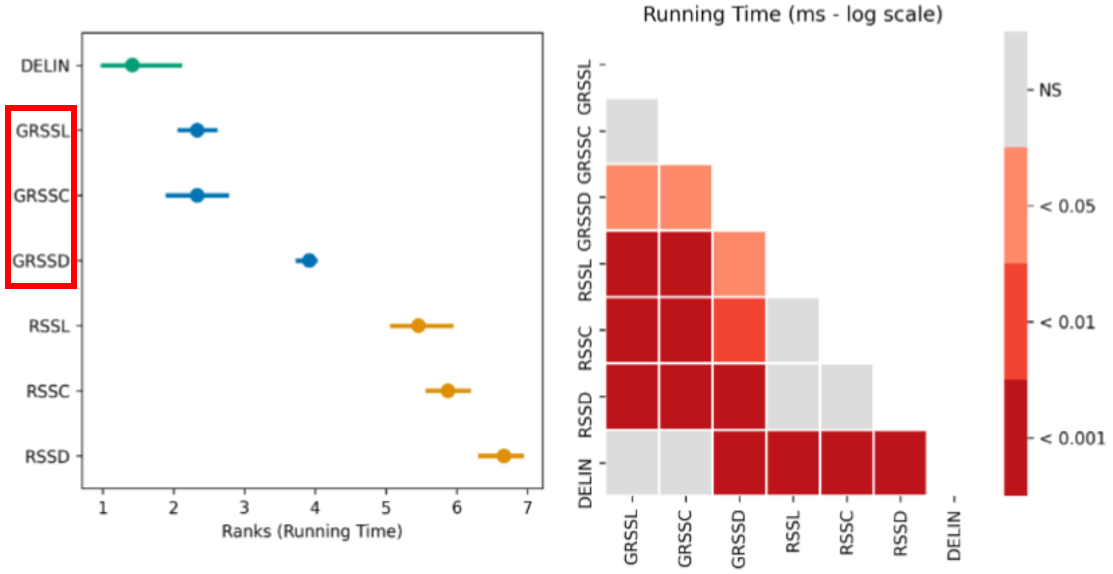


Table 1: List of datasets

Dataset	Instances	Features	Classes	Dataset	Instances	Features	Classes
Avila	20768	10	10	Myocardial	1700	111	2
Car	864	16	4	Pen	10992	16	10
Crowd	10845	28	6	Sensorless	20000	48	11
Frog Family	7195	22	4	Taiwan	6819	94	2
HCV	582	12	4	Wifi	2000	7	4
Iranian	7032	45	2	CTC	617	2500	2
Mushroom	5644	99	6	Kyphosis	120	14	7

Campagner, A., Ciucci, D. (2022). Rough-set Based Genetic Algorithms for Weakly Supervised Feature Selection. Communications in Computer and Information Science, vol 1602

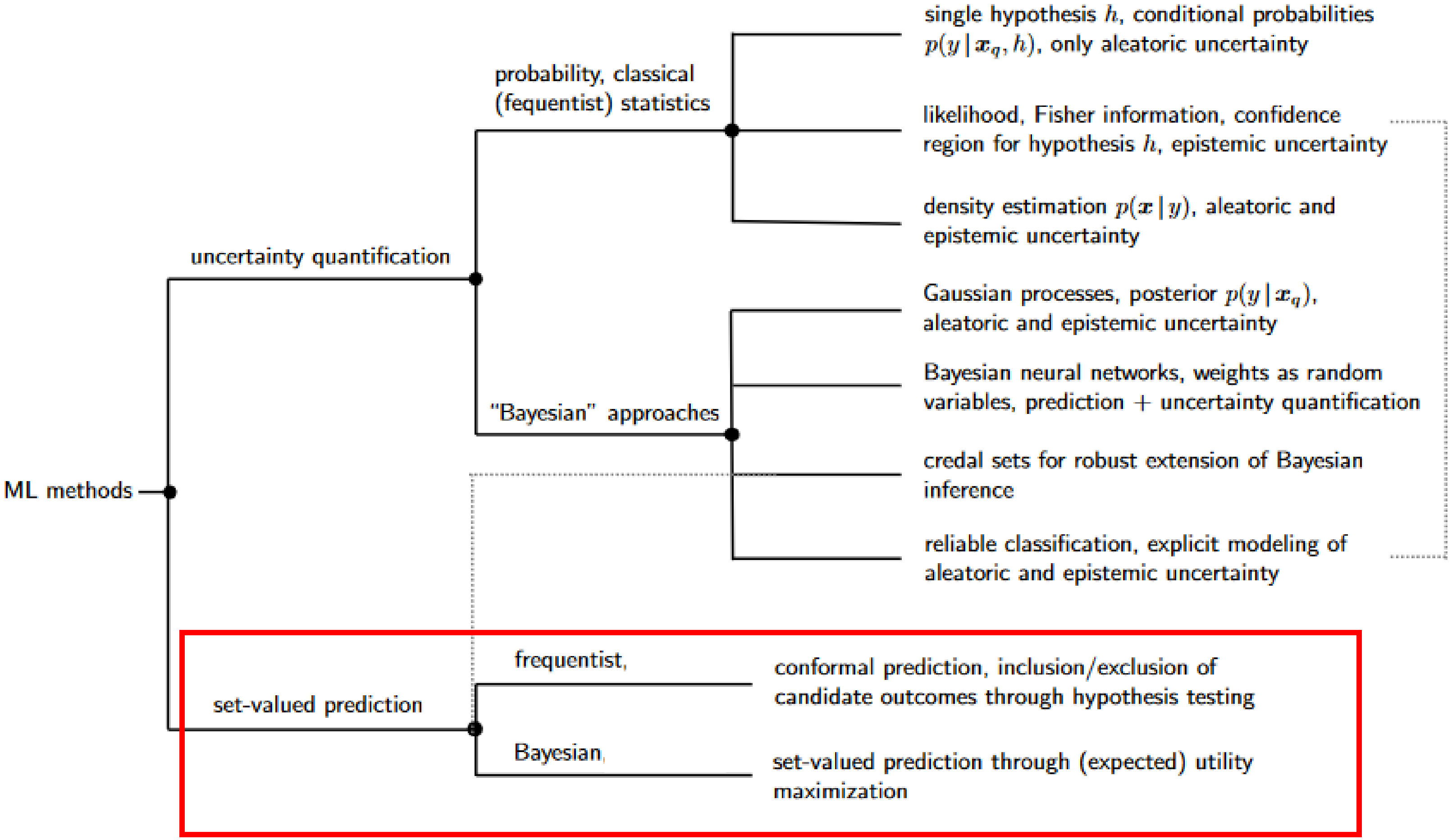
Journals

- **Campagner, A.** (2023). Learning from Fuzzy Labels: Theoretical Issues and Algorithmic Solutions. *International Journal of Approximate Reasoning* (Under Review)
- **Campagner, A.**, Famiglini, L., Carobene, A., Cabitza, F. (2023). Everything is Varied: The Surprising Impact of Individual Variation on ML Reliability in Medicine. *IEEE Transactions on Neural Networks and Learning Systems* (Under Review)
- **Campagner, A.**, Ciucci, D., Denœux, T. (2022). A General Framework for Evaluating and Comparing Soft Clusterings. *Information Sciences*, 623, 70-93
- **Campagner, A.**, Ciucci, D.,& Denœux, T. (2022). Belief functions and rough sets: Survey and new insights. *International Journal of Approximate Reasoning*, *143*, 192-215.
- **Campagner, A**., Ciucci, D., Svensson, C. M., et al. (2021). Ground truthing from multi-rater labeling with three-way decision and possibility theory. *Information Sciences*, 545, 771-790
- **Campagner, A.**, Ciucci, D., Hullermeier, E. (2021). Rough set-based feature selection for weakly labeled data. *International Journal of Approximate Reasoning*, 136, 150-167
- **Campagner, A**., Dorigatti, V., Ciucci, D. (2020). Entropy-based shadowed set approximation of intuitionistic fuzzy sets. *International Journal of Intelligent Systems*, *35*(12), 2117-2139.
- Seveso, A., **Campagner, A**., Ciucci, D., Cabitza, F. (2020). Ordinal labels in machine learning: a user-centered approach to improve data validity in medical settings. *BMC Medical Informatics and Decision Making*, 20(5), 1-14

Conferences

- Cabitza, F., **Campagner, A**., Basile, V. (2023). Toward a Perspectivist Turn in Ground Truthing for Predictive Computing. *AAAI 2023* (Accepted)
- **Campagner, A.**, Ciucci, D., Denœux, T. (2022). A Distributional Approach for Soft Clustering Comparison and Evaluation. *Lecture Notes in Computer Science*, vol 13506
- **Campagner, A.**, Lienen, J. Hullermeier, E., Ciucci, D. (2022). Scikit-Weak: A Python Library for Weakly Supervised Machine Learning. *Lecture Notes in Computer Science*, vol 13633
- **Campagner, A.**, Ciucci, D. (2022). Rough-set Based Genetic Algorithms for Weakly Supervised Feature Selection. C*ommunications in Computer and Information Science*, vol 1602
- **Campagner, A**. (2021). Learnability in "Learning from Fuzzy Labels". In *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-6).
- **Campagner, A.**, Ciucci, D. (2021). Feature selection and disambiguation in learning from fuzzy labels using rough sets. *Lecture Notes in Computer Science*, vol 12872

Cautious Inference

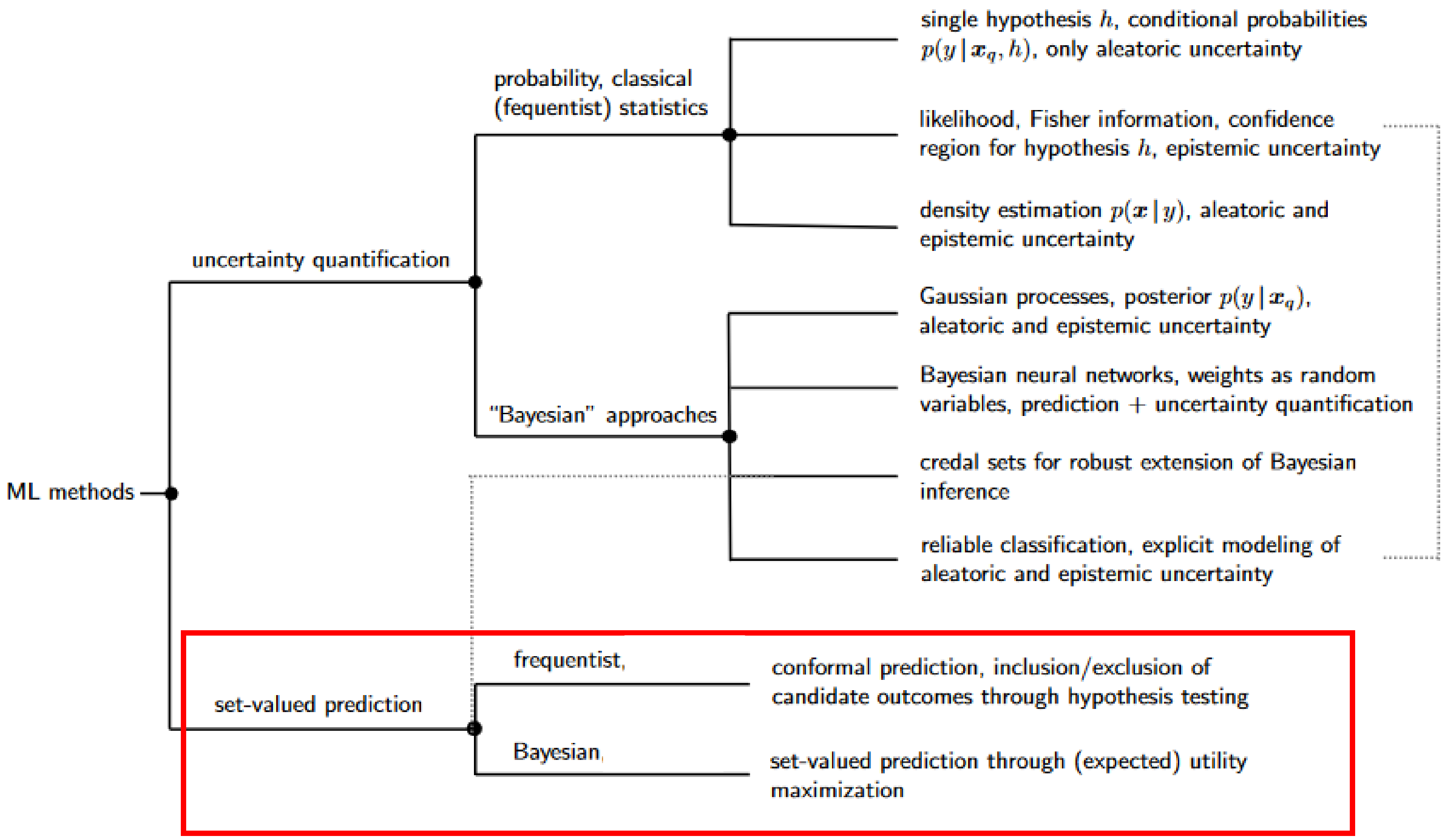


Hüllermeier, E., Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*

Cautious Inference: Why is it Important?

Uncertainty Quantification is critical in decision-making

- Increase tolerance to errors
- Reduce risk of biases…
- … particularly so in critical settings!



Hüllermeier, E., Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*

Cautious Inference: Why is it Important?

Uncertainty Quantification is important in decision-making

Advantages w.r.t. "probabilistic methods" (e.g. Bayesian methods, ensembles)

- Simpler to interpret (sets vs distributions)
- Not require priors/posteriors, approximate inference, ecc.
- Some techniques (selective prediction, conformal prediction) satisfy appealing theoretical guarantees

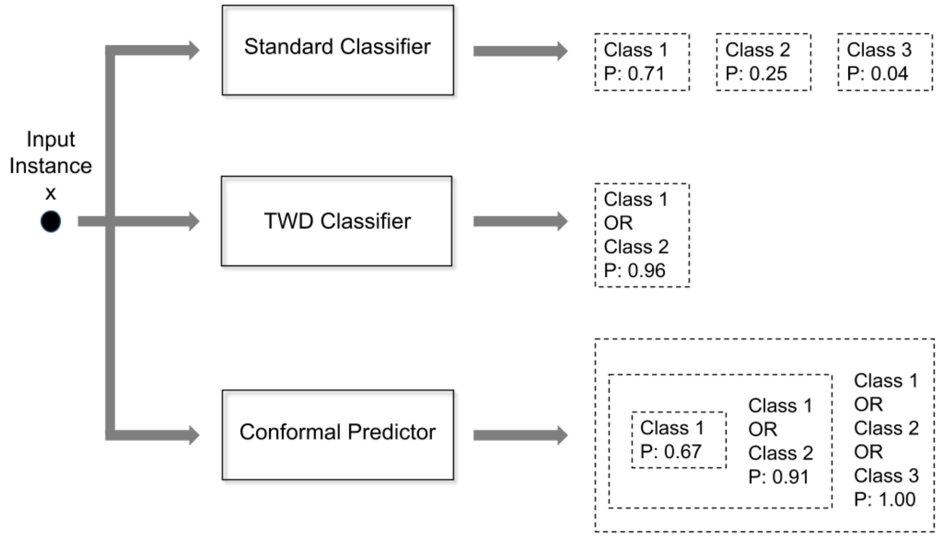
Hüllermeier, E., Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*

Cautious Inference: State of the Art

Most popular approaches:

- **Conformal Prediction**
- **"Decision-theoretic" Methods** (*three-way decision*)
- **(Selective prediction)**

Theory and practice more consolidated than imprecision in the input...

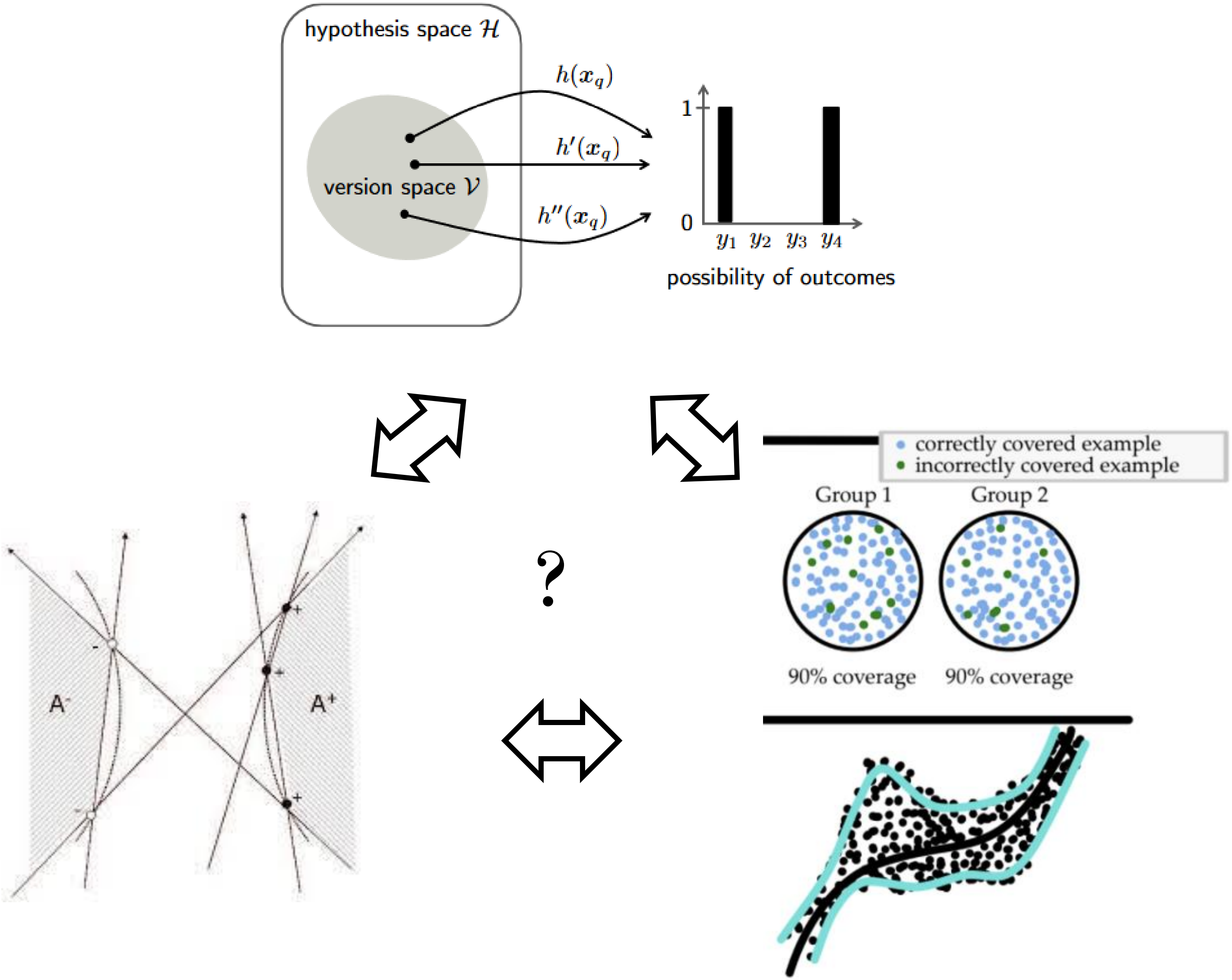


Campagner, A., Milella, F., Ciucci, D., Cabitza, F. (2023). Three-Way Decision in Machine Learning tasks: a Systematic Review. *Artificial Intelligence Review* (Under Review)
Hüllermeier, E., Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*
Campagner, A., Cabitza, F., Ciucci, D. (2020). Three-Way Decision for Handling Uncertainty in Machine Learning: A Narrative Review. *Lecture Notes in Computer Science, vol 12179*

Cautious Inference: Research Gaps (RQ1)

Limited **understanding of relationships**

Methods have been studied *"in isolation"* with different focus... yet they share similarities: **possible to translate results between different paradigms**?



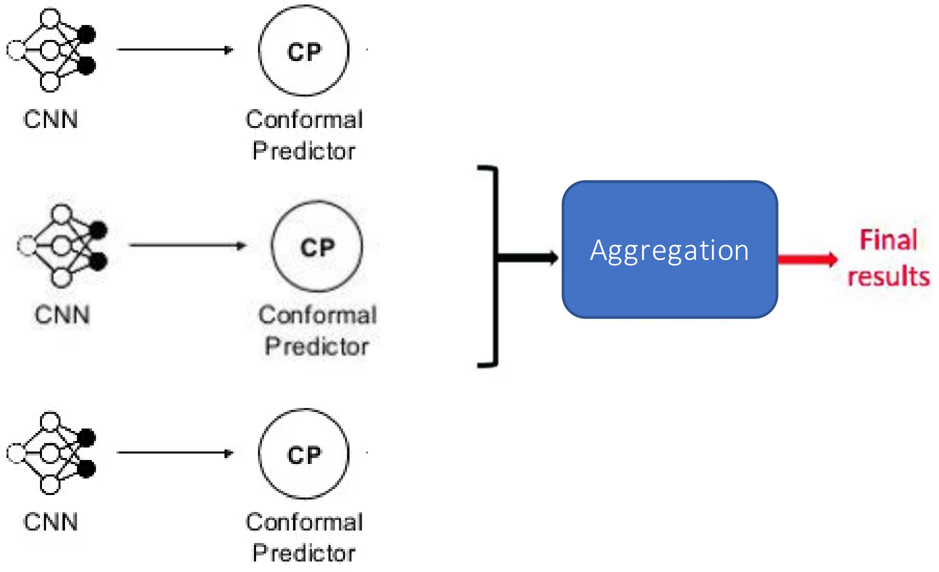
Campagner, A., Cabitza, F., Ciucci, D. (2020). Three-way decision for handling uncertainty in machine learning: A narrative review. In *International Joint Conference on Rough Sets* (pp. 137-152).
Hüllermeier, E., Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*

Cautious Inference: Research Gaps (RQ2)

Validity/Efficiency/Precision Trade-off

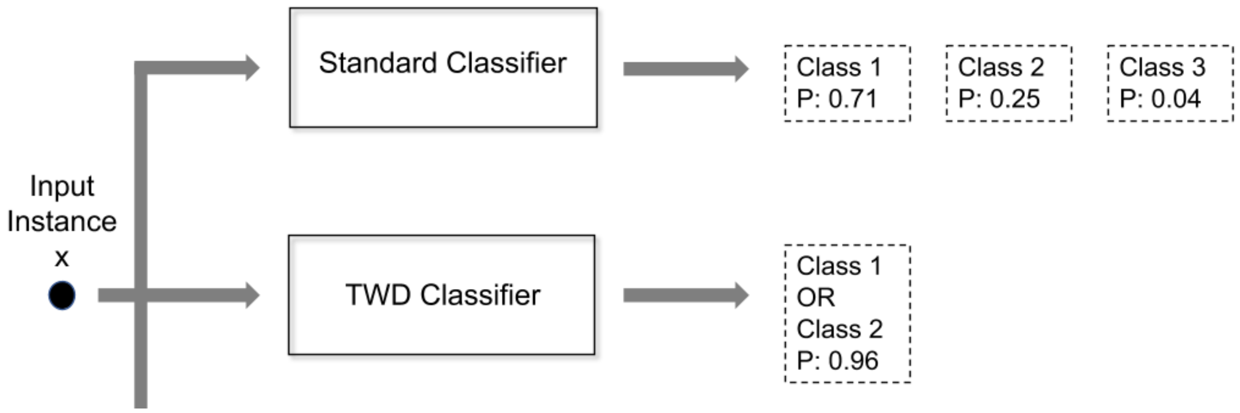
Find a compromise between *theoretical guarantees* (coverage), computational efficiency (esp. *sample efficiency*) and *precision*

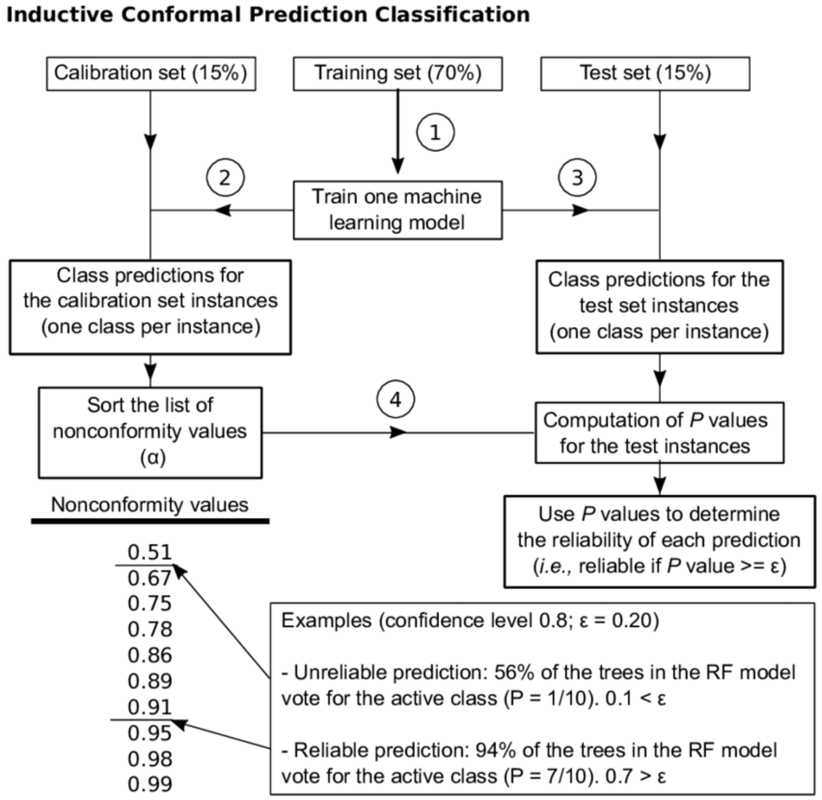
Ensemble methods have been proposed to address this problem… effective empirically, but **limited theoretical understanding**!



Balasubramanian, V. N., Chakraborty, S., & Panchanathan, S. (2015). Conformal predictions for information fusion. *Annals of Mathematics and Artificial Intelligence*, *74*(1), 45-65.
Carlsson, L., Eklund, M., Norinder, U. (2014). Aggregated conformal prediction. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 231-240)
Cherubin, G. (2019). Majority vote ensembles of conformal predictors. *Machine Learning*, *108*(3), 475-488.
Toccaceli, P., Gammerman, A. (2017). Combination of conformal predictors for classification. In *Conformal and Probabilistic Prediction and Applications* (pp. 39-61). PMLR.

$$T^*(x) = \arg\min_{T \in 2^Y} \sum_{y \in Y} \epsilon_{T,y} p_h^y(x) + \alpha(|T|)$$





Three-way Decision

Can be implemented efficiently both at *training time* and *post-hoc*

No study about theoretical properties...

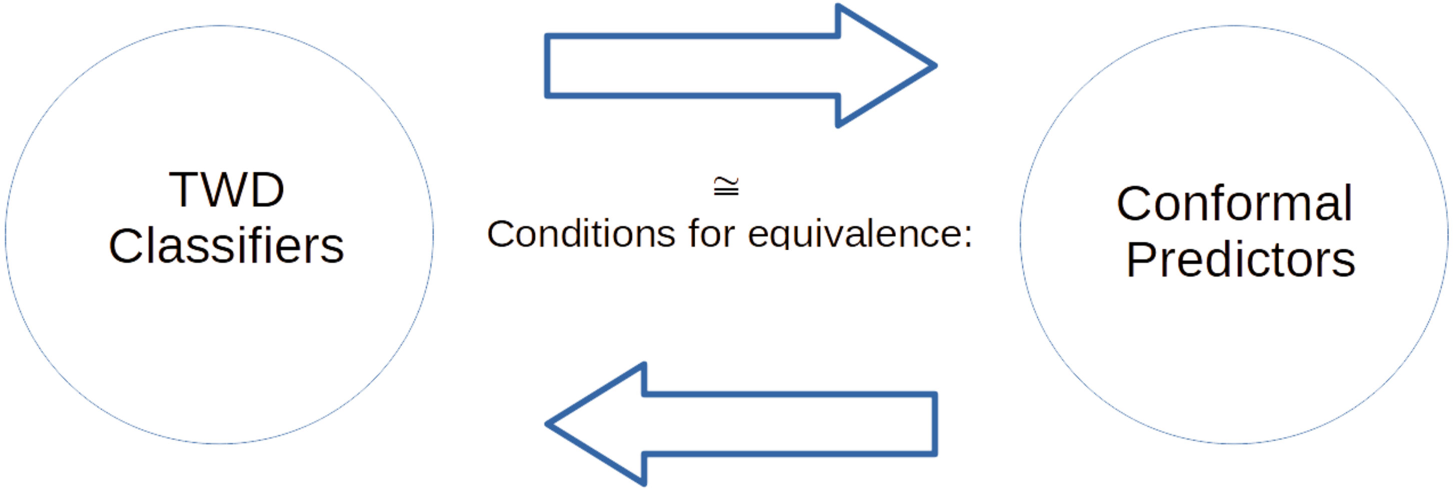
Conformal Prediction (CP)

Theoretical guarantee (Validity): interpret the imprecise predictions as frequentist confidence intervals

But not data efficient (cannot use all data for training) -> less precise

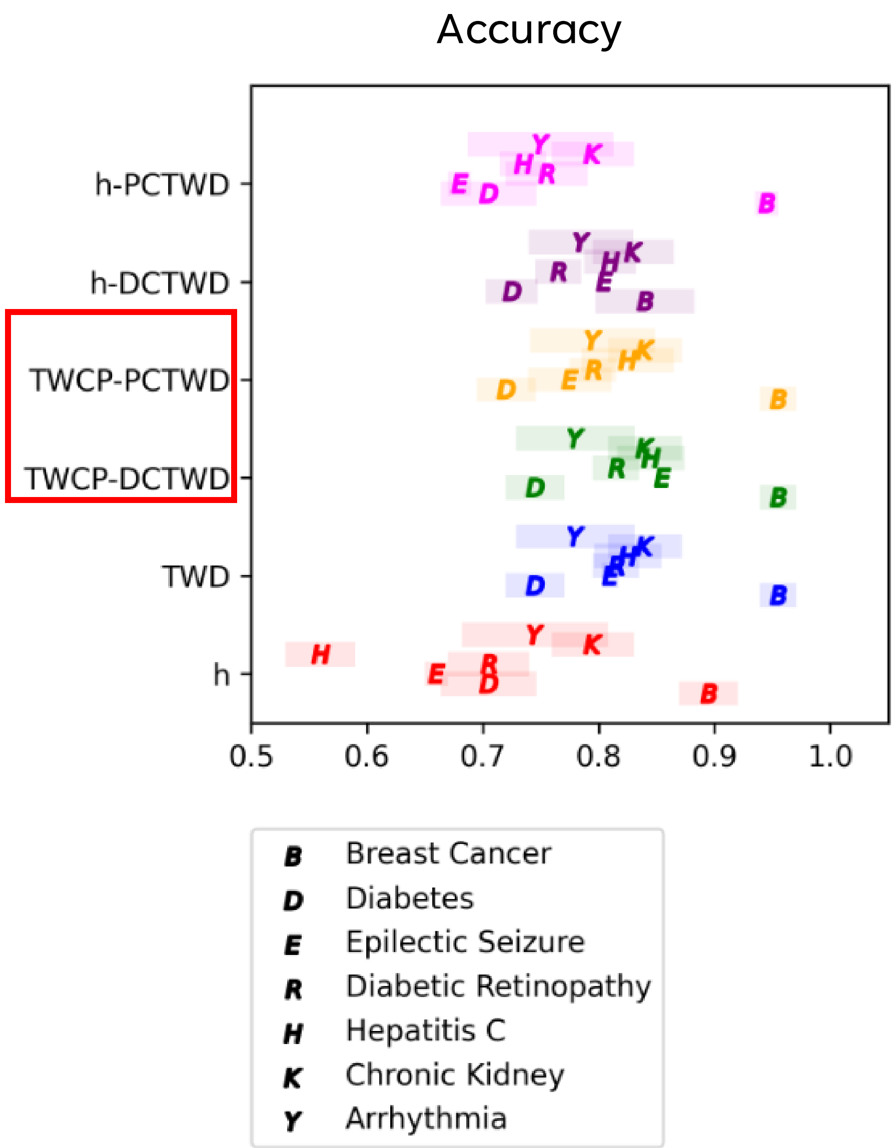
Campagner, A., Cabitza, F., Ciucci, D. (2020). Three-Way Decision for Handling Uncertainty in Machine Learning: A Narrative Review. *Lecture Notes in Computer Science, vol 12179*
Campagner, A., Cabitza, F., Ciucci, D. (2020). The three-way-in and three-way-out framework to treat and exploit ambiguity in data. *International Journal of Approximate Reasoning, 119*, 292-312.

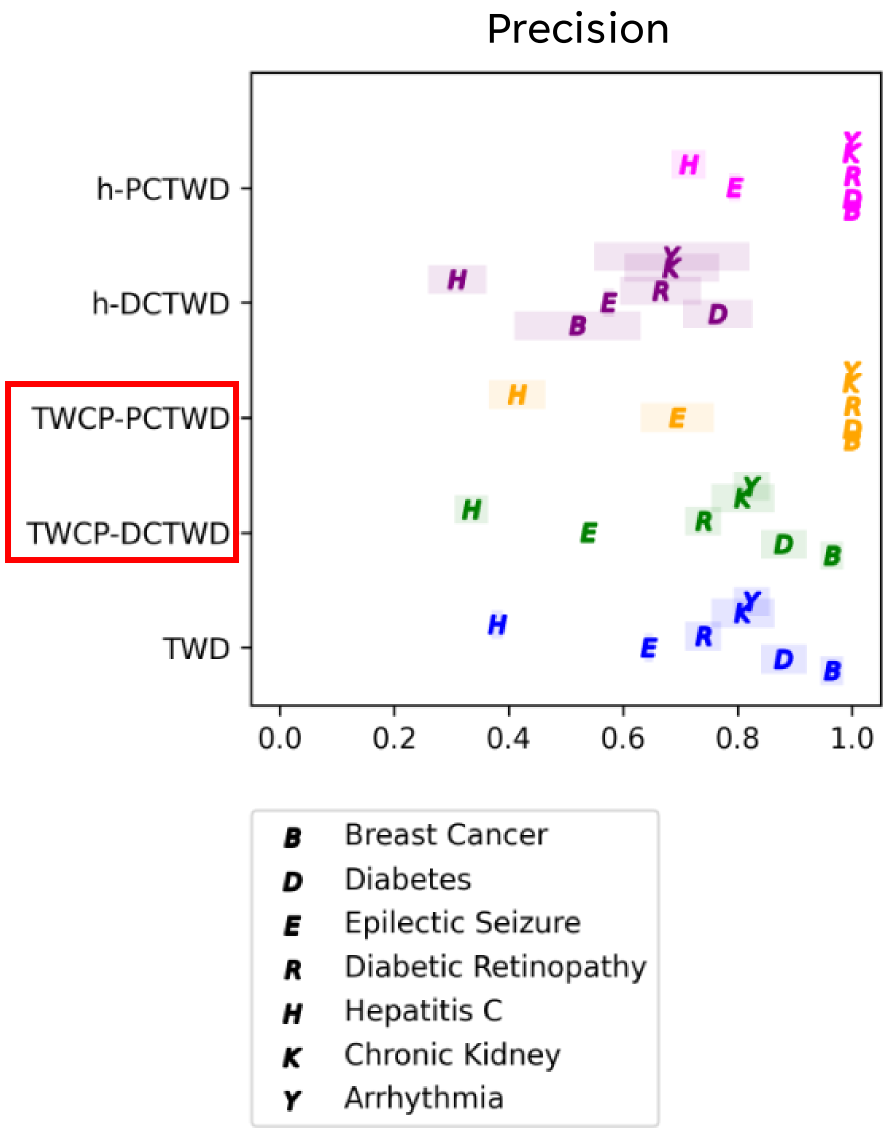
Three-way Conformal Predictor
- Improve validity
- Conditions for equivalence



Decision-Theoretic/Possibilistic Conformal TWD classifier
- Preserve validity
- Improve cost-sensitiveness

Campagner, A., Cabitza, F., Berjano, P., Ciucci, D. (2021). Three-way decision and conformal prediction: Isomorphisms, differences and theoretical properties of cautious learning approaches. Information Sciences, 579, 347-367





Campagner, A., Cabitza, F., Berjano, P., Ciucci, D. (2021). Three-way decision and conformal prediction: Isomorphisms, differences and theoretical properties of cautious learning approaches. Information Sciences, 579, 347-367

$$
\begin{aligned}
\Gamma_{\min}(x, \epsilon) &= \{y \in Y : \min\{p_i^y\} \geq \epsilon\} \\
\Gamma_{\max}(x, \epsilon) &= \{y \in Y : \max\{p_i^y\} \geq \epsilon\} \\
\Gamma_{\mathbf{w}}(x, \epsilon) &= \{y \in Y : \sum w_i * p_i^y \geq \epsilon\}
\end{aligned}
$$

Focus on CP framework: easy to study and strong guarantees

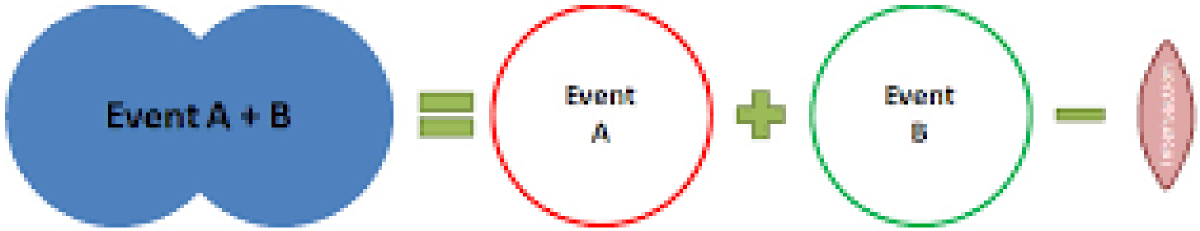
Different CP are trained on different datasets (may be related or not) and then ensembled

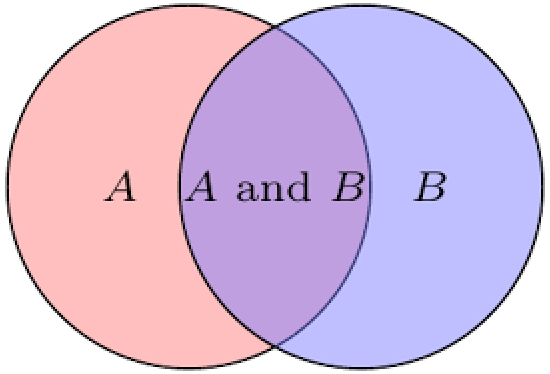
Approach based on aggregation theory and focus on 3 popular rules:

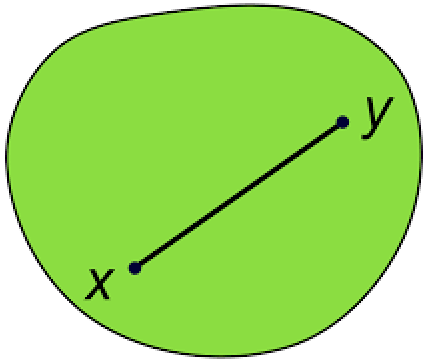
- *Min (t-norms)*
- *Max (t-conorms)*
- *Weighted mean*

Campagner, A., Barandas, M., Folgado, D., et al. (2023). Evidential Predictors: Evidential Combination of Conformal Predictors for Multivariate Time Series Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence (Under Review)

Destercke, S., & Antoine, V. (2013). Combining imprecise probability masses with maximal coherent subsets: Application to ensemble classification. In *Synergies of Soft Computing and Statistics for Intelligent Data Analysis* (pp. 27-35). Springer, Berlin, Heidelberg.







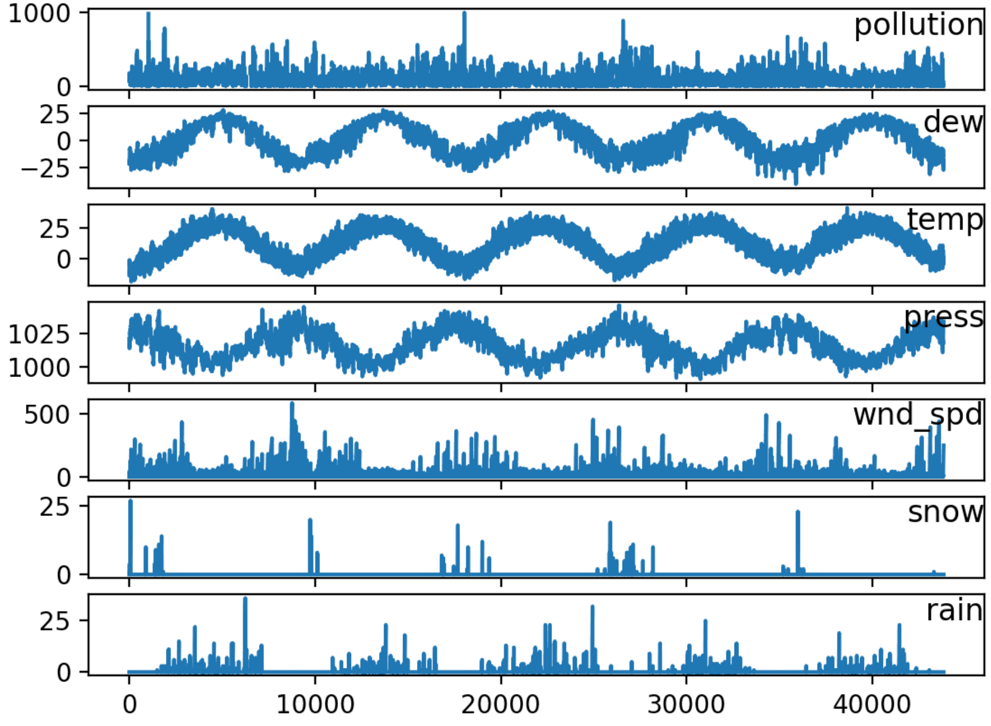
Results

T-norms are the most efficient but generally lose validity: min is the best!

T-conorms are always valid but not efficient: max is the best!

Weighted mean is a good compromise, improves efficiency and is valid in general conditions

Campagner, A., Barandas, M., Folgado, D., et al. (2023). Evidential Predictors: Evidential Combination of Conformal Predictors for Multivariate Time Series Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence (Under Review)
Carlsson, L., Eklund, M., Norinder, U. (2014). Aggregated conformal prediction. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 231-240)
Jaworski, P., Durante, F., Hardle, W. K., & Rychlik, T. (2010). *Copula theory and its applications* (Vol. 198). Berlin: Springer.



Example Application (others in the thesis!)

Multivariate Time Series Classification: multiple timed signals... categorize into some meaningful classes

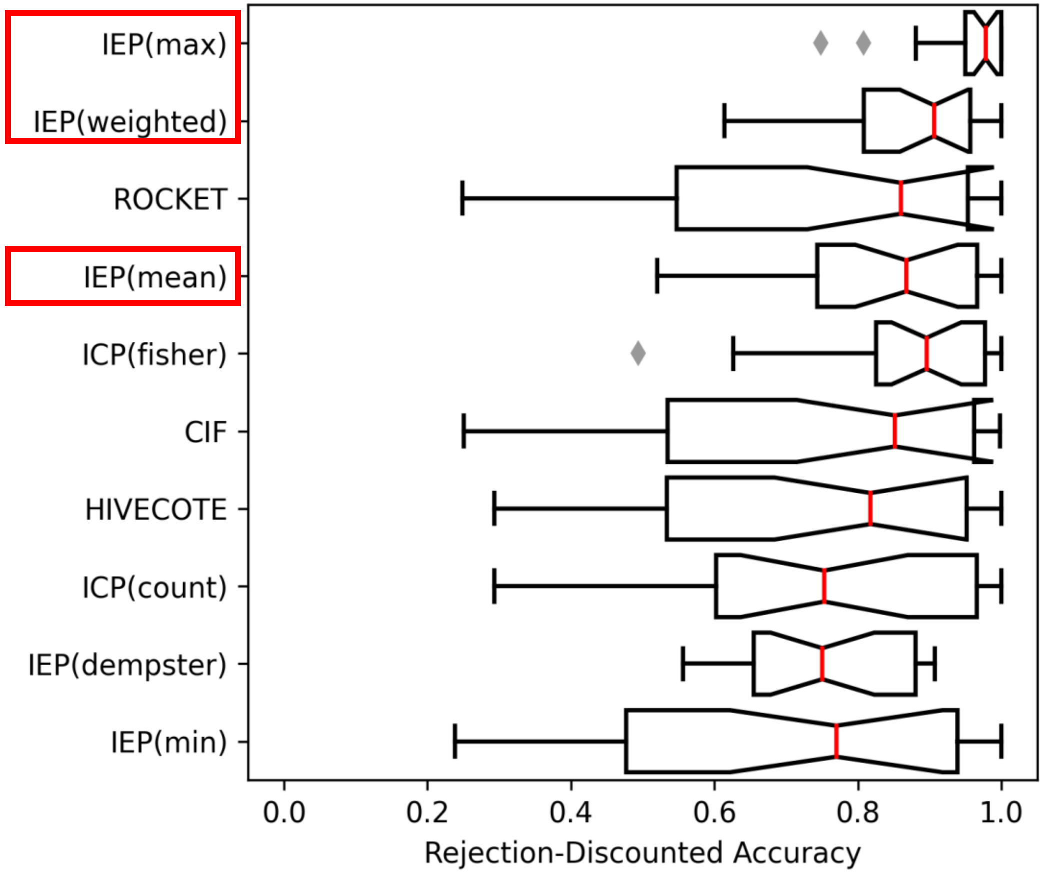
"Naive" strategy: one model for each dimension... then ensemble!

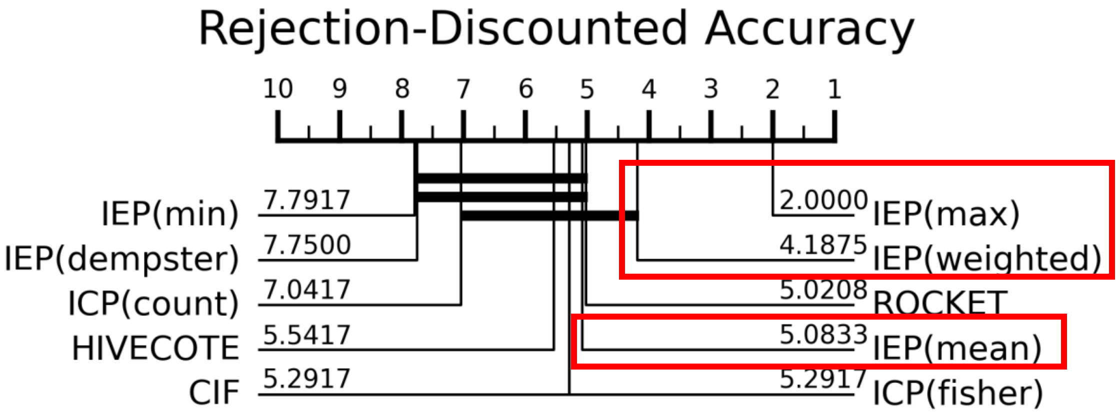
How does this approach fares when *base models are cautious inference* ones?

Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., & Bagnall, A. (2021). The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, *35*(2), 401-449.

Dataset	TrainSize	TestSize	NumDimensions	SeriesLength	NumClasses	Type
ArticularyWordRecognition	275	300	9	144	25	Coordinates
AtrialFibrillation	15	15	2	640	3	ElectricBiosignals
BasicMotions	40	40	6	100	4	AccelerometerGyroscope
Cricket	108	72	6	1197	12	AccelerometerGyroscope
EigenWorms	128	131	6	17984	5	Other
Epilepsy	137	138	3	206	4	AccelerometerGyroscope
EthanolConcentration	261	263	3	1751	4	Other
ERing	30	270	4	65	6	Other
FaceDetection	5890	3524	144	62	2	ElectricBiosignals
FingerMovements	316	100	28	50	2	ElectricBiosignals
HandMovementDirection	160	74	10	400	4	ElectricBiosignals
Handwriting	150	850	3	152	26	AccelerometerGyroscope
Heartbeat	204	205	61	405	2	Audio
Libras	180	180	2	45	15	Coordinates
LSST	2459	2466	6	36	14	Other
MotorImagery	278	100	64	3000	2	ElectricBiosignals
NATOPS	180	180	24	51	6	AccelerometerGyroscope
PenDigits	7494	3498	2	8	10	Coordinates
PEMS-SF	267	173	963	144	7	Other
RacketSports	151	152	6	30	4	AccelerometerGyroscope
SelfRegulationSCP1	268	293	6	896	2	ElectricBiosignals
SelfRegulationSCP2	200	180	7	1152	2	ElectricBiosignals
StandWalkJump	12	15	4	2500	3	ElectricBiosignals
UWaveGestureLibrary	120	320	3	315	8	AccelerometerGyroscope

Ruiz, A. P., Flynn, M., Large, J., Middlehurst, M., & Bagnall, A. (2021). The great multivariate time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, *35*(2), 401-449.





Campagner, A., Barandas, M., Folgado, D., et al. (2023). Evidential Predictors: Evidential Combination of Conformal Predictors for Multivariate Time Series Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence (Under Review)

Journals

- **Campagner, A., Milella, F., Ciucci, D., Cabitza, F. (2023).** Three-Way Decision in Machine Learning tasks: a Systematic Review. *Artificial Intelligence Review* (Under Review)
- **Campagner, A.**, Barandas, M., Folgado, D., et al. (2023). Evidential Predictors: Evidential Combination of Conformal Predictors for Multivariate Time Series Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Under Review)
- **Campagner, A.**, Cabitza, F., Berjano, P., Ciucci, D. (2021). Three-way decision and conformal prediction: Isomorphisms, differences and theoretical properties of cautious learning approaches. *Information Sciences*, 579, 347-367
- **Campagner, A.**, Cabitza, F., & Ciucci, D. (2020). The three-way-in and three-way-out framework to treat and exploit ambiguity in data. *International Journal of Approximate Reasoning*, *119*, 292-312.

Conferences

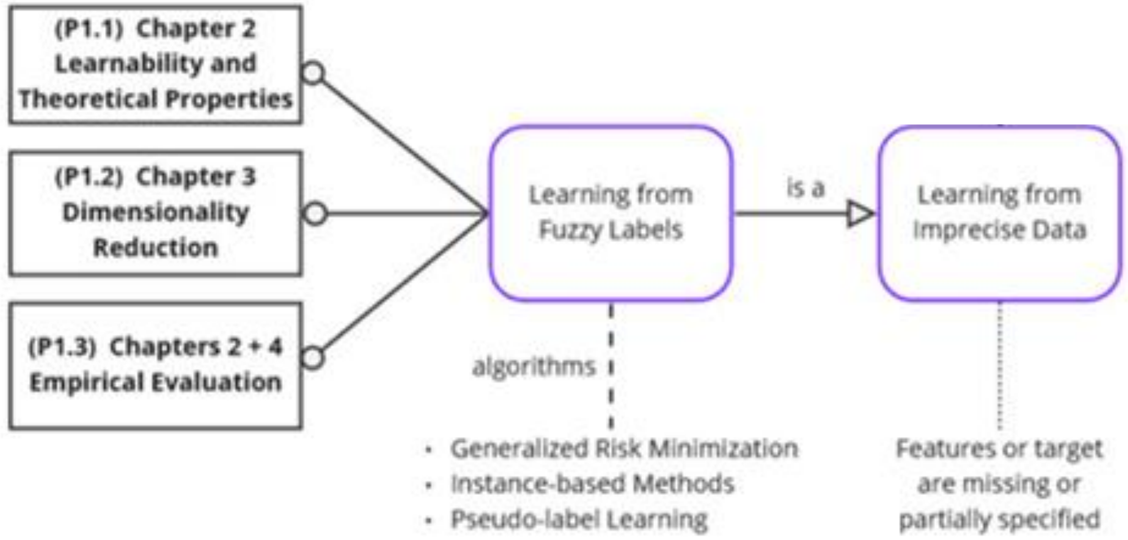
- **Campagner, A.**, Ciucci, D. (2022). Three-way Learnability: A Learning Theoretic Perspective on Three-way Decision. *17th Conference on Computer Science and Intelligence Systems* (pp. 243–246)
- **Campagner, A.**, Famiglini, L., Cabitza, F. (2022). Re-calibrating Machine Learning Models Using Confidence Interval Bounds. *International Conference on Modeling Decisions for Artificial Intelligence* (pp. 132-142)
- **Campagner, A.**, Cabitza, F., Ciucci, D. (2020). Three-Way Decision for Handling Uncertainty in Machine Learning: A Narrative Review. *Lecture Notes in Computer Science, vol 12179*
- **Campagner, A.**, Ciucci, D., & Cabitza, F. (2020). Ensemble Learning, Social Choice and Collective Intelligence. *International Conference on Modeling Decisions for Artificial Intelligence* (pp. 53-65)

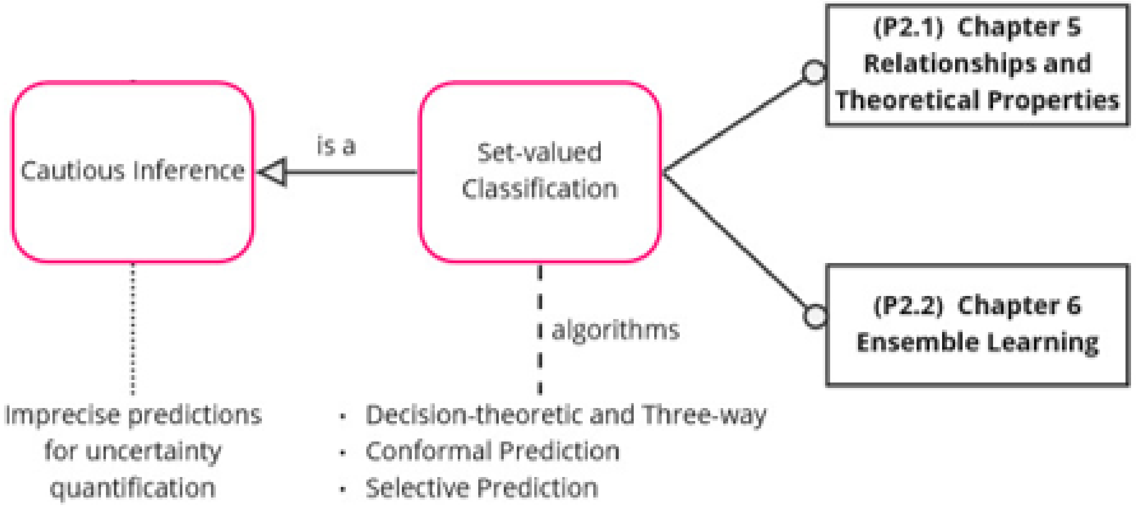
CONTRIBUTIONS

1. Characterization of sample and computational complexity
2. Proposal and analysis of a novel pseudo-label learning method (RRL)

1. Proposal and analysis of a novel Rough Set-based feature selection algorithm
2. Empirical analyis

1. Empirical analysis
2. Applications in medical problems





CONTRIBUTIONS

1. Characterization of sample complexity for three-way decision and relationships with selective prediction
2. Relationships between three-way decision and conformal prediction

1. Empirical analysis of cautious inference base models for ensemble learning
2. Theoretical study of validity and efficiency of ensemble conformal predictors
3. Applications in multi-variate time series classification

Open Problems

Results on learning from fuzzy labels represent an opportunity for interaction between uncertainty management and machine learning theory... can these ideas be carried over to more general formalisms, e.g. imprecise probabilities?

Cautious learning has a deep, rich theory with connections with active learning, adversarial learning and over-parameterized models... can we extend results to provide connections with these settings?

Cautious inference has been motivated in the literature as a way to reduce overreliance and optimize human-AI interaction... surprisingly few user studies on this topic!

Thank you

a.campagner@campus.unimib.it