RgeoJSD: robust noise-tolerant loss for cerebral emboli classification

Medical Imaging Research Laboratory

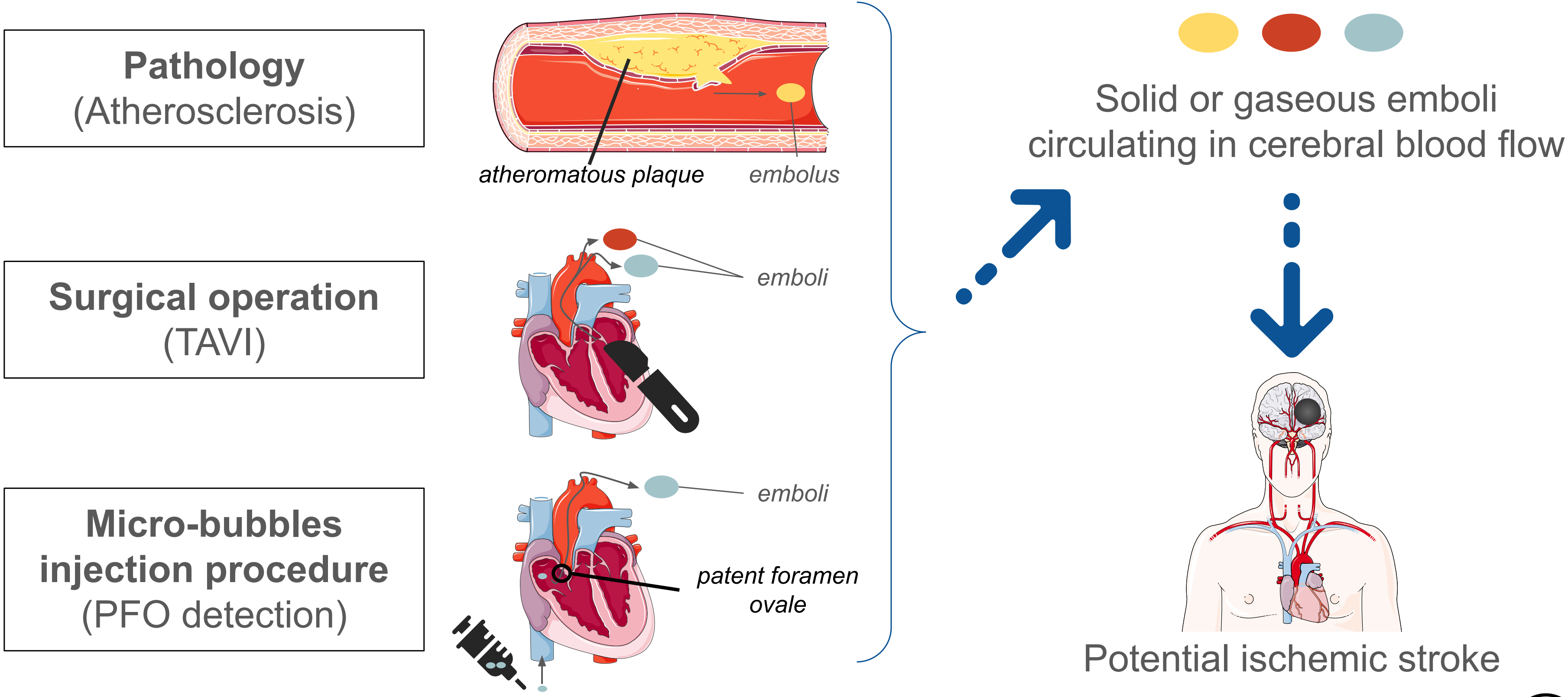
Mathilde Dupouy[1]
Yamil Vindas[2]
Thibaut Dambry[3]
Blaise Kévin Guépié[4]
Philippe Delachartre[1]

[1]CREATIS; CNRS (UMR 5220); INSERM (U1294); INSA Lyon; Université de Lyon, Lyon, France
[2]Data science for digital health, Geneva University, Geneva, Switzerland
[3]Atys Medical, Soucieu-en-Jarrest, France
[4]LIS2N, Université technologique de Troyes, Troyes, France

IEEE IUS 2025
International Ultrasonics Symposium
September 15 - 18, 2025 || Utrecht, Netherlands

Cerebral emboli

Cerebral emboli are solid or gaseous material in the cerebral blood flow, and are one of the main risk of stroke.



Pathology
(Atherosclerosis)

Surgical operation
(TAVI)

Micro-bubbles injection procedure
(PFO detection)

Solid or gaseous emboli circulating in cerebral blood flow

Potential ischemic stroke

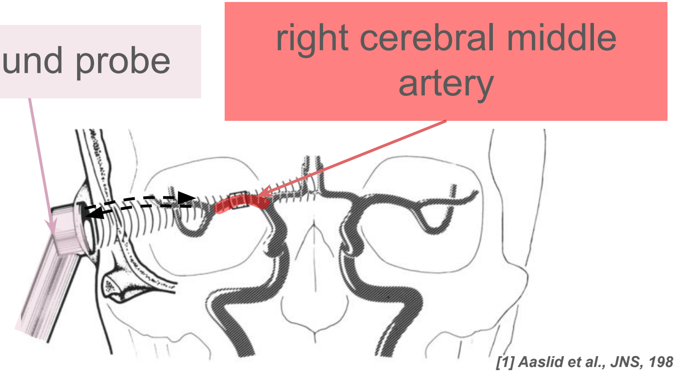
__PFO = Patent Foramen Ovale__ (hole between left and right atriums)
__TAVI = Transcatheter Aortic Valve Implantation__ (percutaneous endovascular technique of aortic valve replacement)

@Servier Medical Art

Emboli detection

Transcranial Doppler is a unique non-invasive modality to monitor emboli, detected as high intensity transient signals.

Transcranial Doppler (TCD)



ultrasound probe

right cerebral middle artery

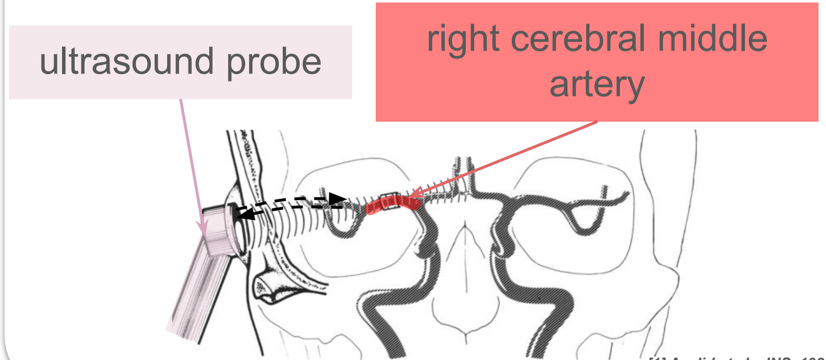
[1] Aaslid et al., JNS, 1982

[1] Aaslid R. et al.*, JNS, *1982 *Noninvasive transcranial Doppler ultrasound recording of flow velocity in basal cerebral arteries*
[2] Guépié B. K. et al.*, IEEE JBHI, *2018 *Sequential Emboli Detection from Ultrasound Outpatient Data*

Emboli detection

Transcranial Doppler is a unique non-invasive modality to monitor emboli, detected as high intensity transient signals.

Transcranial Doppler (TCD)



[1] Aaslid et al., JNS, 1982

Portable and robotized TCD

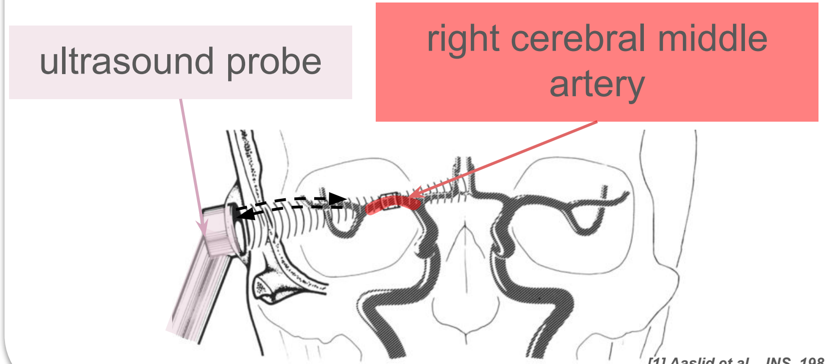
Atys medical

Longer recordings with **low constraints** for the patient **[2]**

*[1] **Aaslid R. et al.**, JNS, **1982** Noninvasive transcranial Doppler ultrasound recording of flow velocity in basal cerebral arteries*
*[2] **Guépié B. K. et al.**, IEEE JBHI, **2018** Sequential Emboli Detection from Ultrasound Outpatient Data*

Transcranial Doppler is a unique non-invasive modality to monitor emboli, detected as high intensity transient signals.

Transcranial Doppler (TCD)



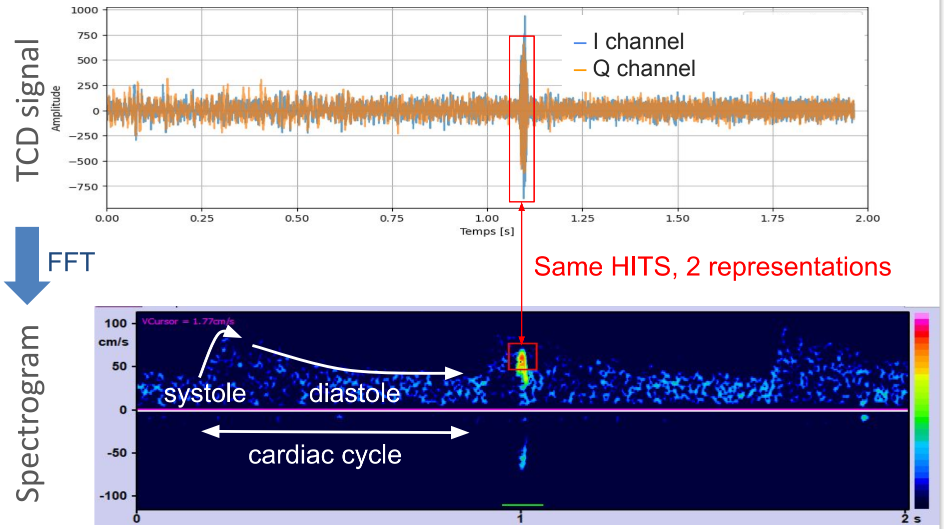
[1] Aaslid et al., JNS, 1982

Portable and robotized TCD

Atys medical

Longer recordings with **low constraints** for the patient **[2]**

High Intensity Transient Signal (HITS)

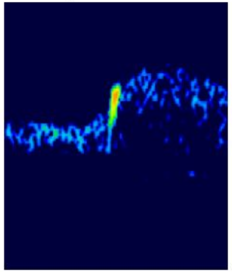
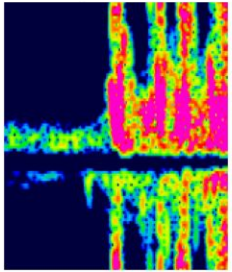
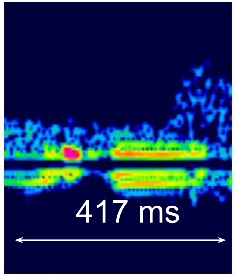
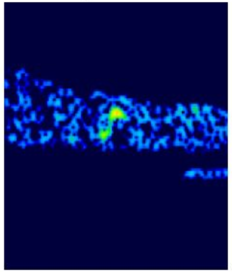
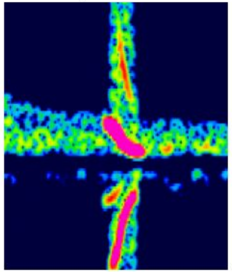
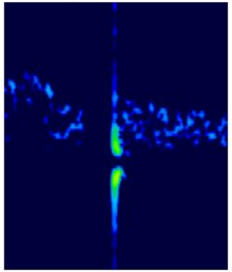


Same HITS, 2 representations

*[1] **Aaslid R. et al.**, **JNS, 1982** Noninvasive transcranial Doppler ultrasound recording of flow velocity in basal cerebral arteries*
*[2] **Guépié B. K. et al.**, **IEEE JBHI, 2018** Sequential Emboli Detection from Ultrasound Outpatient Data*

Emboli classes

Medical interest is to distinguish between artefact / emboli (counting) and solid emboli / gaseous emboli (risk, causes).

Solid Emboli (SE)	Gaseous Emboli (GE)	Artifacts (A)
 Carotid stenosis Emboli detection	 Aortic valve stenosis TAVI	 - Emboli detection
 - Emboli detection	 PFO PFO detection	 - Emboli detection

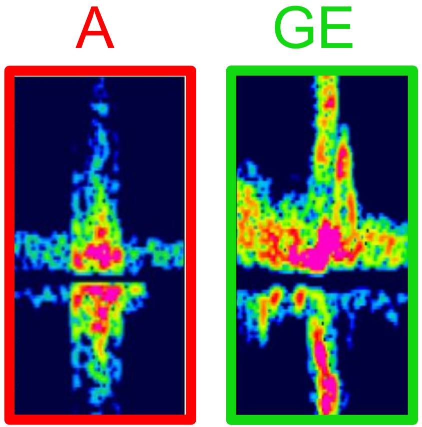
International consensus 1998
→ duration < 300 ms
→ intensity increase > 3 dB w.r.t blood flow signal
→ unidirectional
→ musical “chirp”, “snap” sound

Ringelstein E. B. et al., Stroke, 1998
Consensus on microemboli detection by TCD

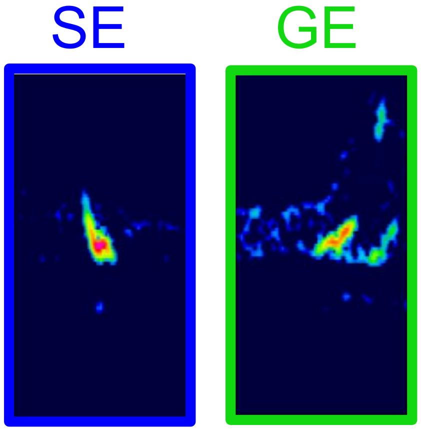
Our main objective: deep learning for HITS classification A/SE/GE

Ambiguous cases

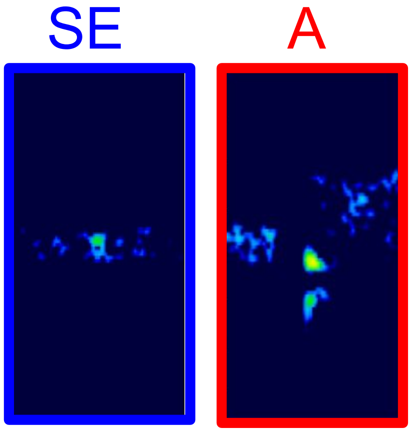
Annotation can be difficult because of ambiguous signal and annotator subjectivity.



scalpel noise vs. emboli shower



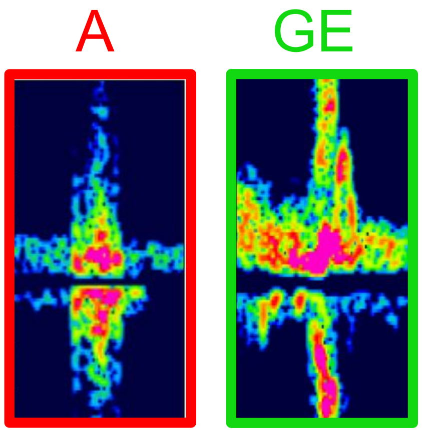
Speed modulation ambiguity



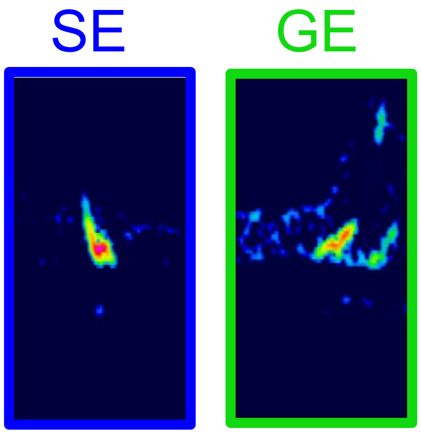
Small solid emboli vs. speckle and other artifacts

Ambiguous cases

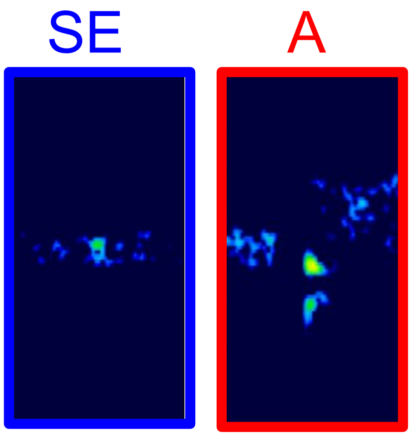
Annotation can be difficult because of ambiguous signal and annotator subjectivity.



scalpel noise vs. emboli shower



Speed modulation ambiguity



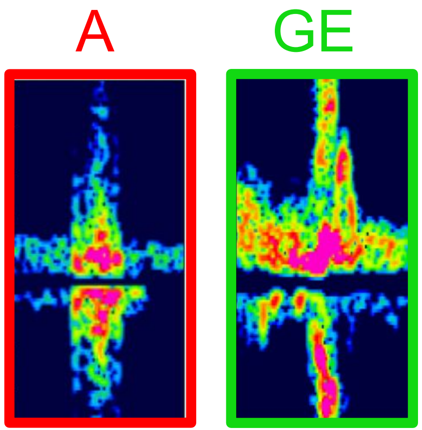
Small solid emboli vs. speckle and other artifacts

Challenge for supervised learning

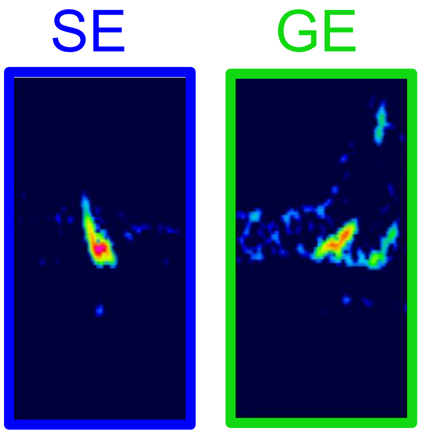
Designing methods that are robust to label noise

Ambiguous cases

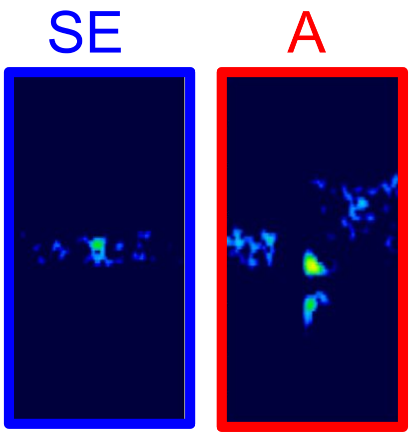
Annotation can be difficult because of ambiguous signal and annotator subjectivity.



scalpel noise vs. emboli shower



Speed modulation ambiguity



Small solid emboli vs. speckle and other artifacts

Challenge for supervised learning

Designing methods that are robust to label noise

Our proposition

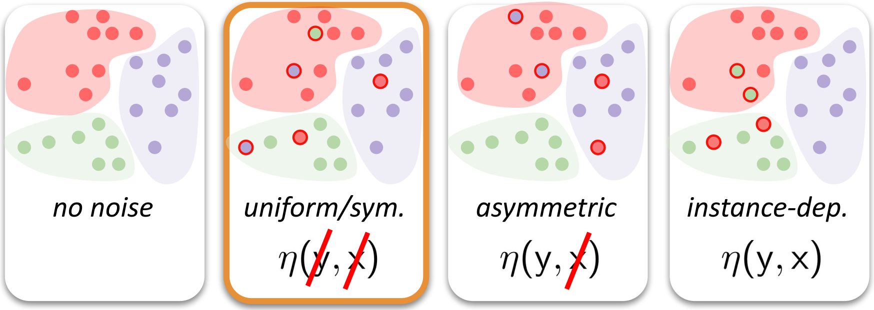
RgeoJSD, new label robust loss function for classification

Robust losses

Loss functions robust to noisy labels are a way to deal with noisy labels that is architecture-free and computationally light.

Noise rate η

$$\eta(y,x) = p(\overset{\text{noisy label}}{y_\eta} \neq i \mid \overset{\text{true label}}{y} = i, \underset{\text{sample}}{x})$$



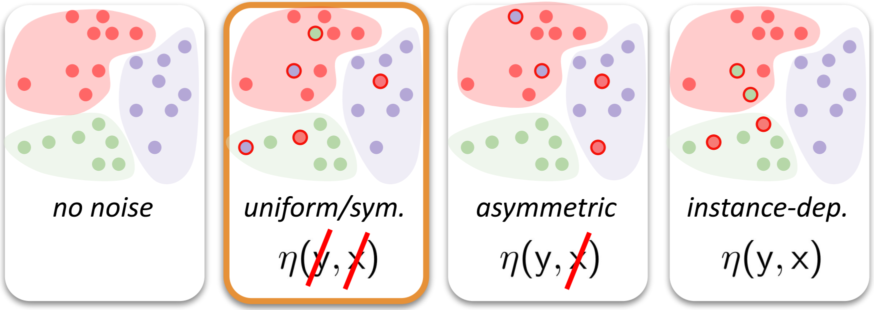
*[3] **Ghosh et al., AAAI, 2017** Robust Loss Functions under Label Noise for Deep Neural Networks*

Robust losses

Loss functions robust to noisy labels are a way to deal with noisy labels that is architecture-free and computationally light.

Noise rate η

$$\eta(y,x) = p(\overset{\text{noisy label}}{y_\eta} \neq i | \overset{\text{true label}}{y} = i, \underset{\text{sample}}{x})$$



Noise tolerant loss

$$\mathbb{P}_{\text{clean}}[\underset{\text{Minimizer of clean risk}}{f^*}(\mathbf{x}) = y_{\mathbf{x}}] = \mathbb{P}_{\text{clean}}[\underset{\text{Minimizer of noisy risk}}{f^*_\eta}(\mathbf{x}) = y_{\mathbf{x}}]$$

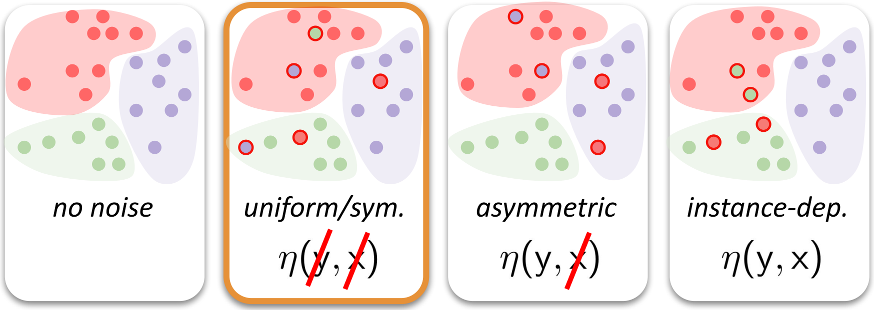
*[3] **Ghosh et al., AAAI, 2017** Robust Loss Functions under Label Noise for Deep Neural Networks*

Robust losses

Loss functions robust to noisy labels are a way to deal with noisy labels that is architecture-free and computationally light.

Noise rate η

$$\eta(y,x) = p(\overbrace{y_\eta}^{\text{noisy label}} \neq i \mid \overbrace{y}^{\text{true label}} = i, \underbrace{x}_{\text{sample}})$$



Noise tolerant loss

$$\mathbb{P}_{\text{clean}}[\underbrace{f^*}_{\text{Minimizer of clean risk}}(\mathbf{x}) = y_{\mathbf{x}}] = \mathbb{P}_{\text{clean}}[\underbrace{f^*_\eta}_{\text{Minimizer of noisy risk}}(\mathbf{x}) = y_{\mathbf{x}}]$$

Sufficient conditions [3]

$$\left\{\begin{array}{l} \textit{Symmetric loss } \text{L [3]} \\ \text{“Reasonable” noise rate} \quad \eta < \underbrace{\dfrac{K}{K-1}}_{\textbf{0.66} \text{ for K = 3 classes}} \end{array}\right. \Rightarrow \text{L is } \textit{noise tolerant} \text{ under } \textit{uniform} \text{ label noise}$$

*[3] **Ghosh et al., AAAI, 2017** Robust Loss Functions under Label Noise for Deep Neural Networks*

RgeoJSD

The proposed RgeoJSD is derived from geometric JSD by balancing convergence and tolerance.

Geometric Jensen-Shannon Divergence (GeoJSD) [4]

$$\mathrm{GeoJSD}(f(x), y) = (1 - \alpha)\mathrm{KL}(f(x)||m) + \alpha\mathrm{KL}(e_y||m)$$

$$m = f(x)^{1-\alpha} e_y^{\alpha}$$

[4] ***Nielsen F.*****, Entropy*****, 2019*** *On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means*

RgeoJSD

The proposed RgeoJSD is derived from geometric JSD by balancing convergence and tolerance.

Geometric Jensen-Shannon Divergence (GeoJSD) [4]

$$\mathrm{GeoJSD}(f(x), y) = (1-\alpha)\mathrm{KL}(f(x)||m) + \alpha\mathrm{KL}(e_y||m) \qquad m = f(x)^{1-\alpha} e_y^{\alpha}$$

$$= \alpha(1-\alpha)[H(e_y, f(x)) - H(f(x), f(x)) + H(f(x), e_y) - H(e_y, e_y)]$$

*[4] **Nielsen F., Entropy, 2019** On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means*

RgeoJSD

The proposed RgeoJSD is derived from geometric JSD by balancing convergence and tolerance.

Geometric Jensen-Shannon Divergence (GeoJSD) [4]

$$\mathrm{GeoJSD}(f(x), y) = (1-\alpha)\mathrm{KL}(f(x)||m) + \alpha \mathrm{KL}(e_y||m) \qquad m = f(x)^{1-\alpha} e_y^{\alpha}$$

$$= \alpha(1-\alpha)[\underbrace{H(e_y, f(x)) - H(f(x), f(x))}_{\substack{\propto \text{ cross-entropy} \\ \Rightarrow \text{ convergence properties}}} + \underbrace{H(f(x), e_y) - H(e_y, e_y)}_{\substack{\propto \text{ reverse cross-entropy} \\ \text{symmetric loss} \\ \Rightarrow \text{ label noise tolerant}}}]$$

*[4] **Nielsen F., Entropy, 2019** On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means*

RgeoJSD

The proposed RgeoJSD is derived from geometric JSD by balancing convergence and tolerance.

Geometric Jensen-Shannon Divergence (GeoJSD) [4]

$$\mathrm{GeoJSD}(f(x), y) = (1-\alpha)\mathrm{KL}(f(x)||m) + \alpha \mathrm{KL}(e_y||m) \qquad m = f(x)^{1-\alpha} e_y^{\alpha}$$

$$= \alpha(1-\alpha)[\underbrace{H(e_y, f(x)) - H(f(x), f(x))}_{\substack{\propto \text{ cross-entropy} \\ \Rightarrow \text{ convergence properties}}} + \underbrace{H(f(x), e_y) - H(e_y, e_y)}_{\substack{\propto \text{ reverse cross-entropy} \\ \underline{\text{symmetric}} \text{ loss} \\ \Rightarrow \text{ label noise robustness}}}]$$

Robust GeoJSD (RgeoJSD)

$$\mathrm{RgeoJSD}(f(x), y) = \underset{\text{NEW}}{\beta} \cdot L_C + \underset{\text{NEW}}{\gamma} \cdot L_R$$

*[4] **Nielsen F., Entropy, 2019** On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means*

We used a private dataset of spectrograms from TCD recordings in different hospitals.

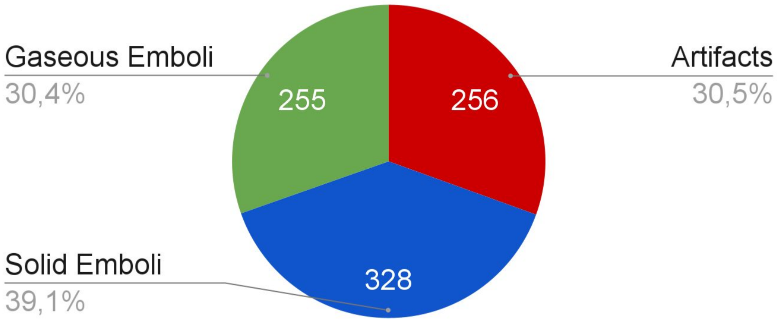
Dataset

Spectrograms (400 ms)
1232 RGB images (160×192×3)
👤 35 subjects (21-85 y.o.)
🏥 10 hospital sources 📍 5 FR, 5 EU
⚕️ various services
💉 various examinations
😷 various causes

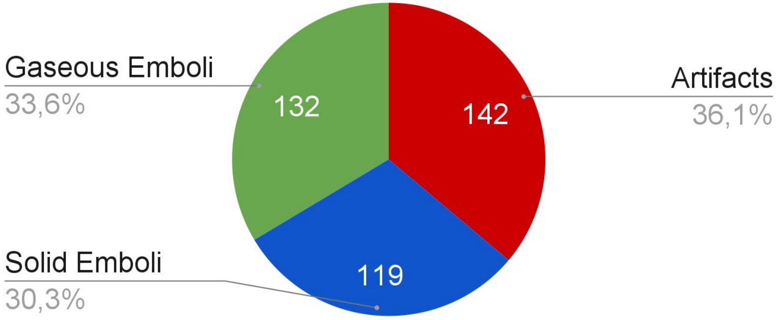
Acquisition

TCD-X Robotized Holter
1.5MHz, PRF ~6kHz
Volume 8-22 mm³
Depth 45-55mm
HBR detection [2]

Train set (839 HITS, 68%)



Test set (393 HITS, 32%)

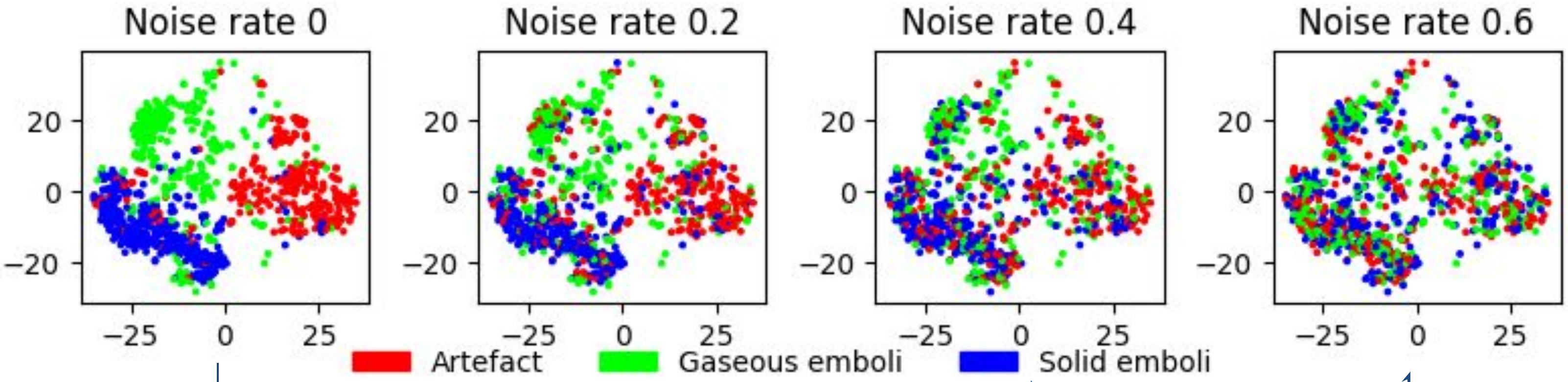


*[2] **Guépié B. K. et al.**, **IEEE JBHI, 2018** Sequential Emboli Detection from Ultrasound Outpatient Data*

Adding noise

Uniform label noise is synthetically added with a noise rate η.

$$p(\overset{\text{noisy}}{y_\eta} = i \,|\, \overset{\text{true}}{y} = j) = \begin{cases} 1 - \overset{\text{noise rate}}{\eta} & \text{if } i = j \\ \eta/2 & \text{if } i \neq j \end{cases}$$

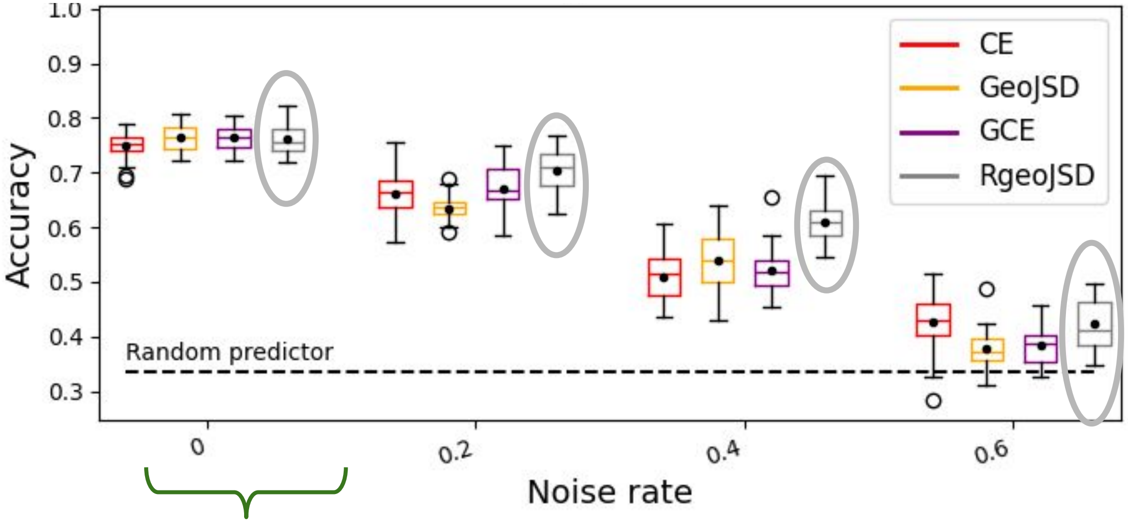


Point cloud using t-SNE on classifier latent space

Coloring points for different noise rates

Classification performance

Accuracy on test set (20 runs)



$$\mathrm{CE}(f(x), y) = H(f(x), y)$$

$$\mathrm{GeoJSD}(f(x), y) = \alpha' L_C + \alpha' L_R$$

$$\alpha' = 0.25$$

$$\mathrm{GCE}(f(x), y) = \frac{1 - f_y(x)^q}{q} \qquad q = 0.7$$

$$\mathrm{RgeoJSD}(f(x), y) = \beta \cdot L_C + \gamma \cdot L_R$$

$$\beta = 2.5 \cdot 10^{-4} \quad \gamma = 2.5 \cdot 10^{-2}$$

Similar performances under no noise

Best method for medium noise rates

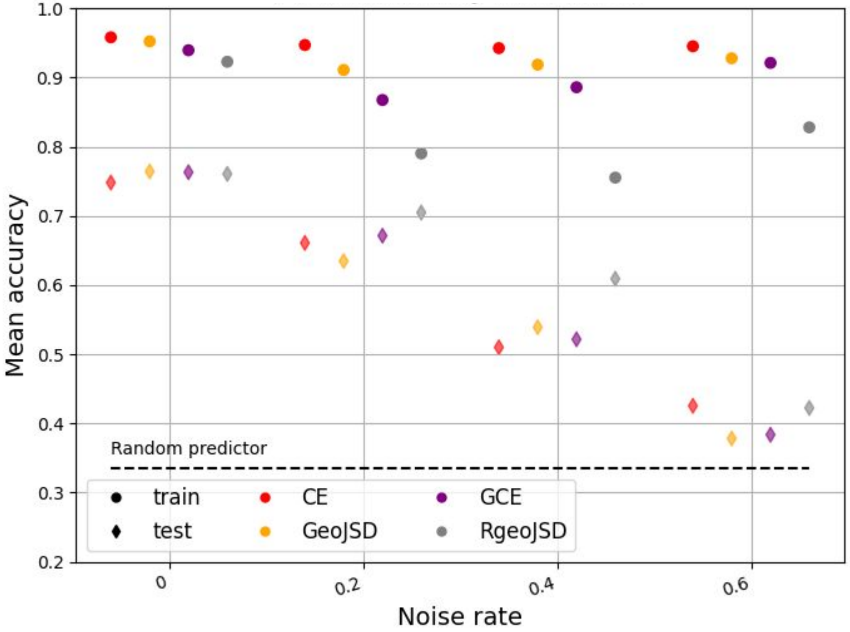
+ global performance drop
+ global dispersion

Performance drop for high noise rate

*[4] **Zhilu Z., NeurIPS, 2018** Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels*

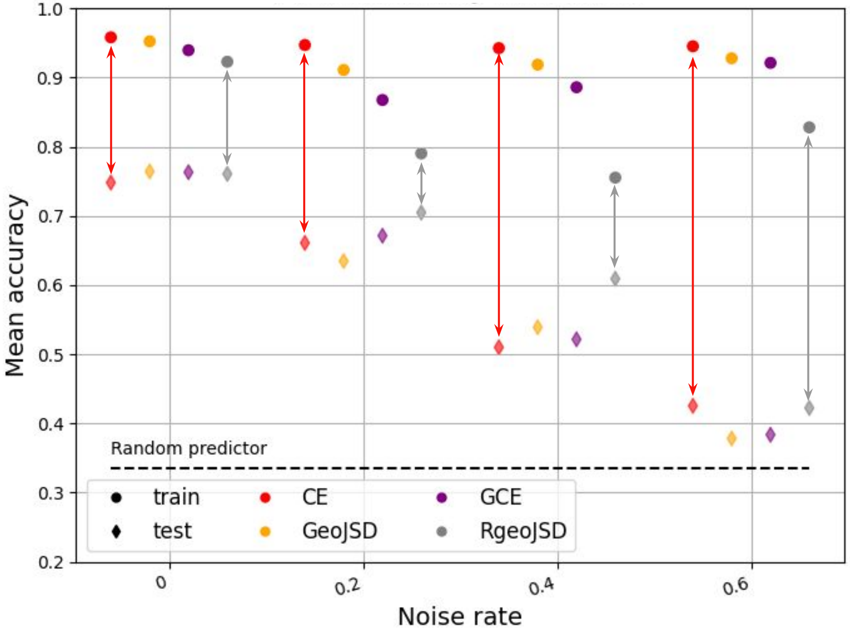
Training behavior

Mean accuracy on test set (20 runs)



Training behavior

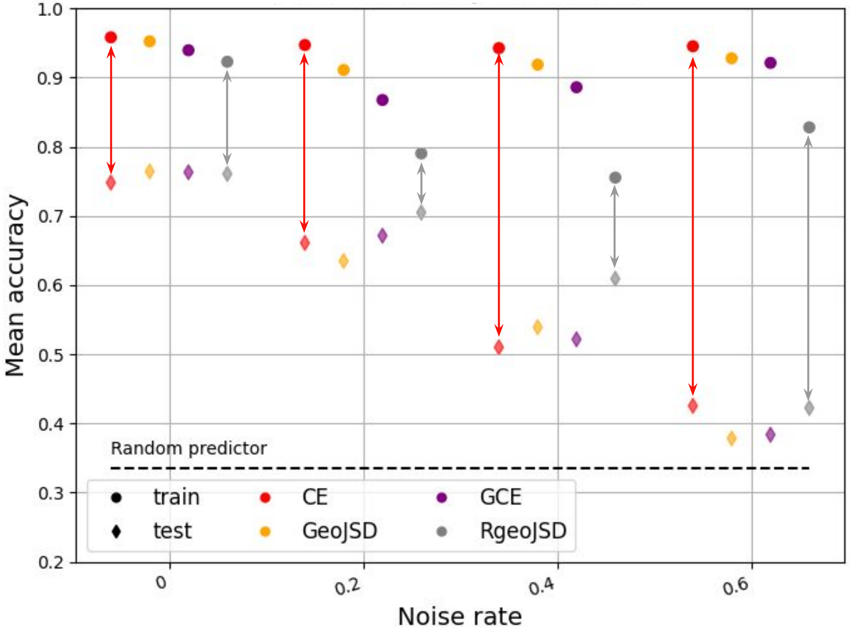
Mean accuracy on test set (20 runs)



⇒ Symptomatic of overfitting

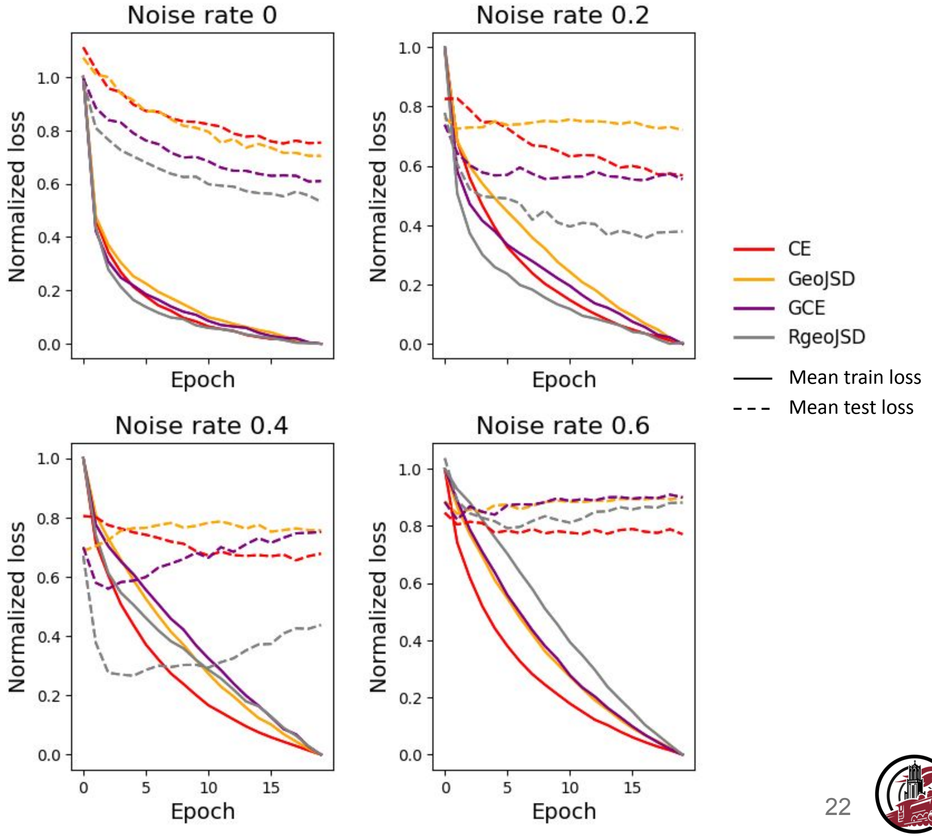
Training behavior

Mean accuracy on test set (20 runs)



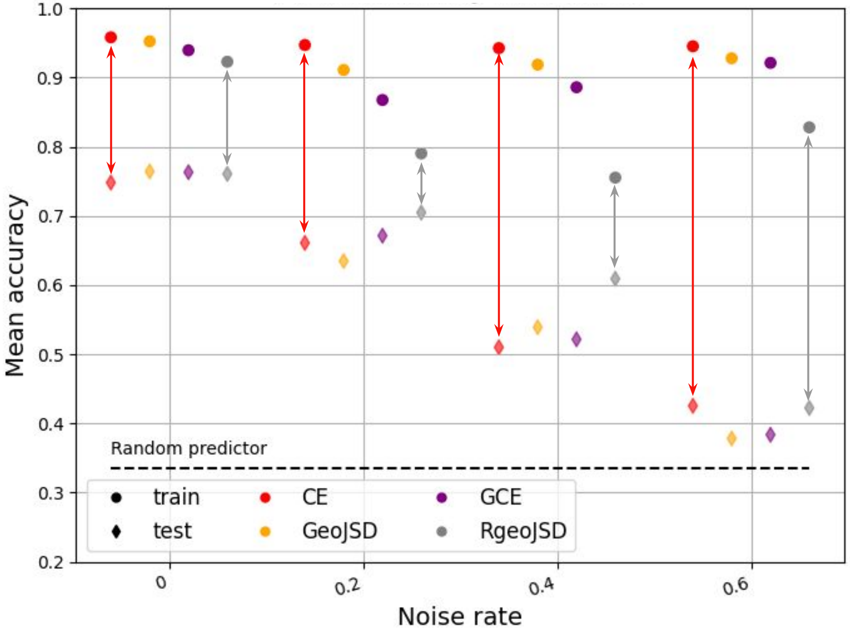
⇒ Symptomatic of overfitting

Mean train and test losses over 20 runs

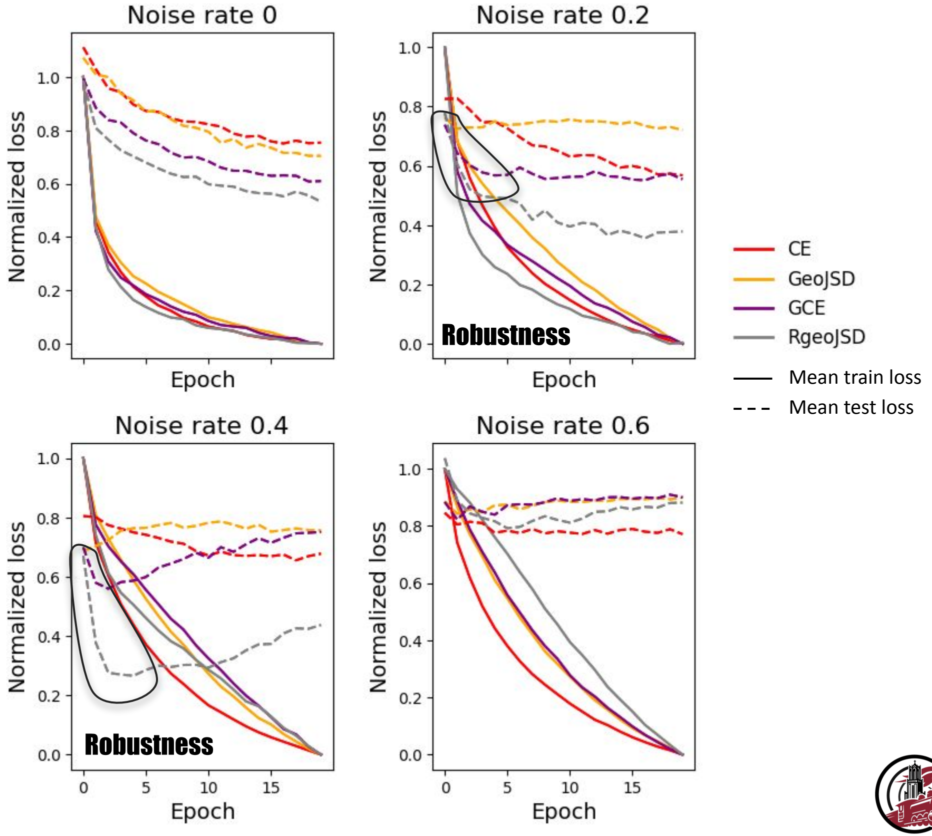


Training behavior

Mean accuracy on test set (20 runs)



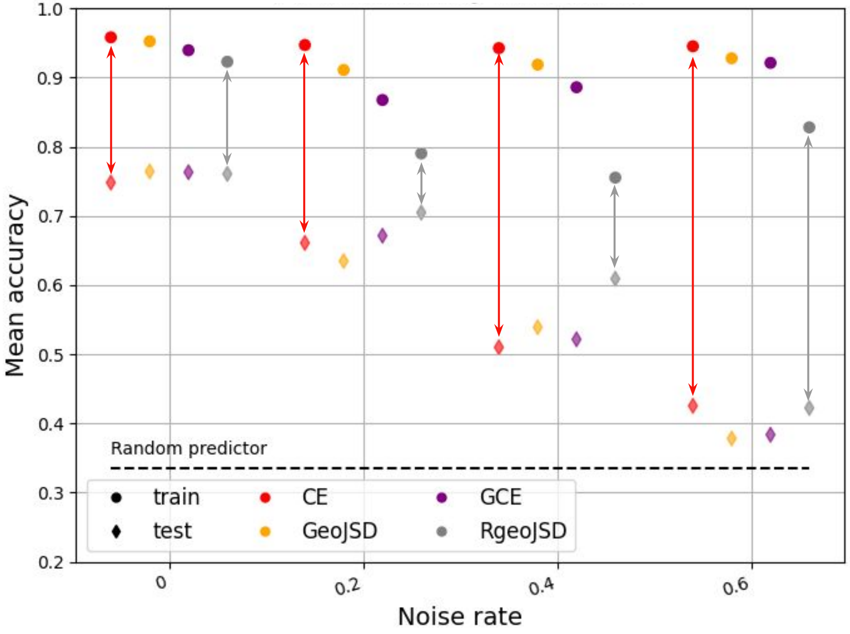
Mean train and test losses over 20 runs



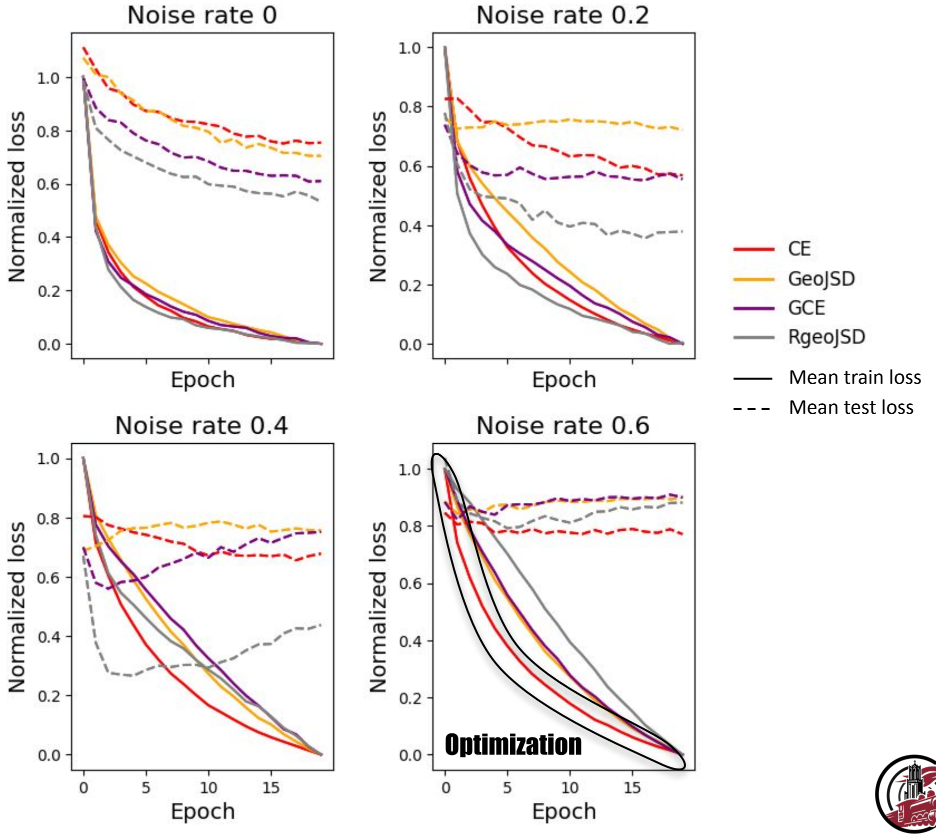
⇒ Symptomatic of overfitting

Training behavior

Mean accuracy on test set (20 runs)



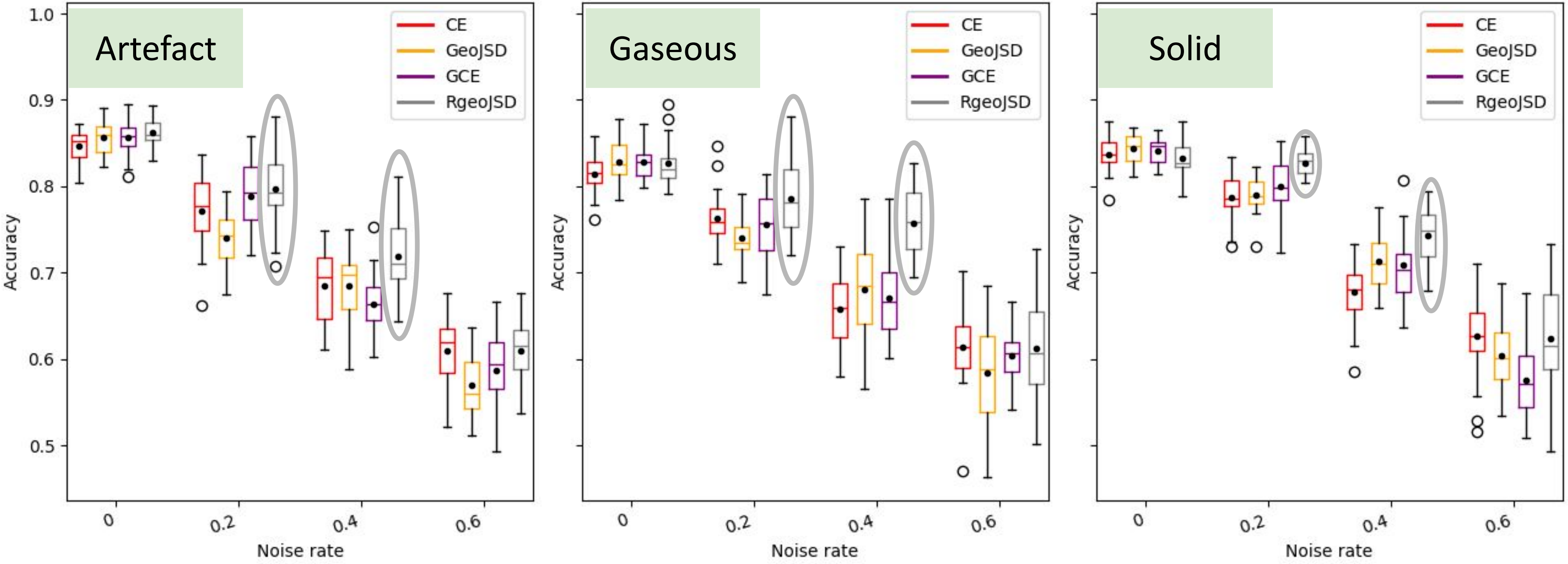
Mean train and test losses over 20 runs



⇒ Symptomatic of overfitting

Classification performance

Accuracy on test set (20 runs)



→ At medium noise rates, our method is better on all classes
→ Higher dispersion of accuracy for SE and GE at high noise rates

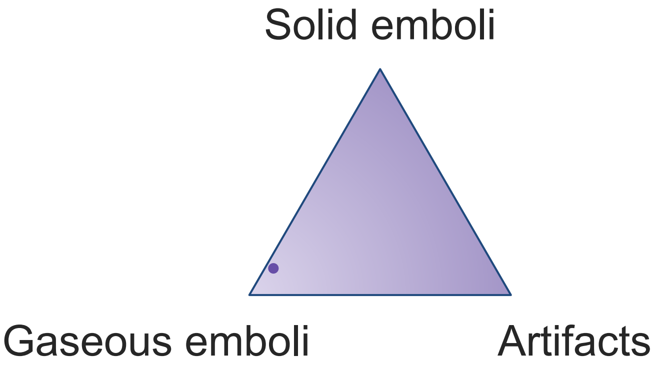
Conclusion

New loss tolerant to label noise
→ Easy and efficient implementation
→ Best performances on our TCD dataset for medium noise rates (uniform noise)
→ Same performance under no noise

$$\boxed{\mathrm{RgeoJSD}(f(x), y) = \beta \cdot L_C + \gamma \cdot L_R}$$

Perspectives

- Illustrate weights effect
- Test on other types of noise
- Study the impact on model uncertainty



Thanks for your attention, any question ?

🏠Take home messages

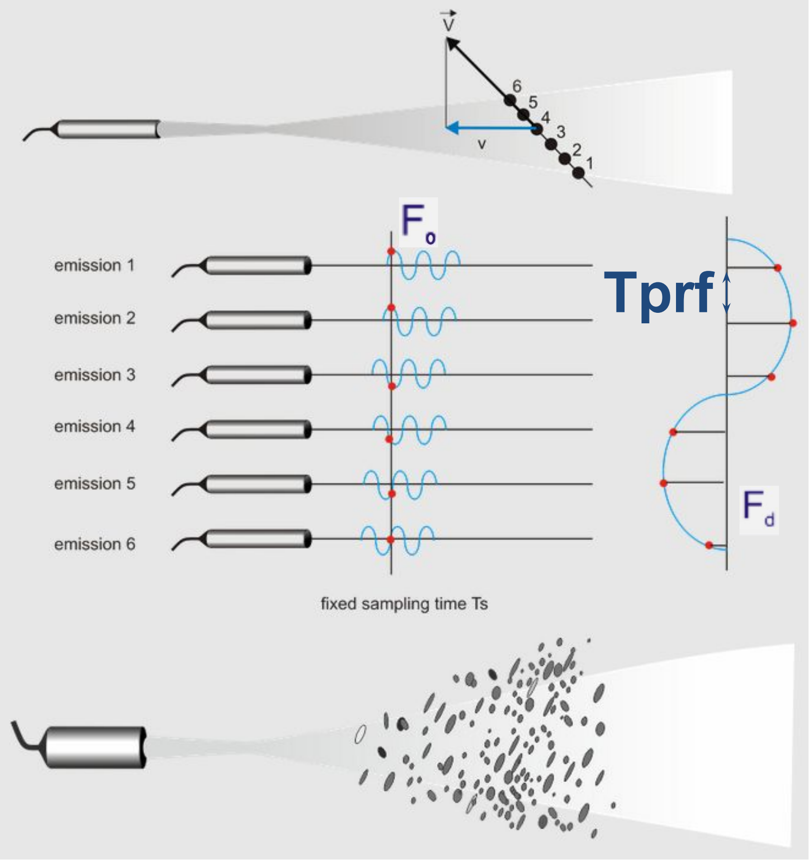
→ TCD data needs a particular care regarding label noise
→ Classification loss similar to cross-entropy while more robust to label noise

📜 Proceeding available

https://hal.science/hal-05234901

Contact: mathilde.dupouy@insa-lyon.fr

TCD acquisition



Doppler effect for pulsed echography

Speed wrt measured frequency

$$V = \frac{F_d * C}{2 * F_o * \cos \delta}$$

Max. speed wrt PRF

$$V_{max} = \frac{F_{prf} * C}{4 * F_o * \cos \delta}$$

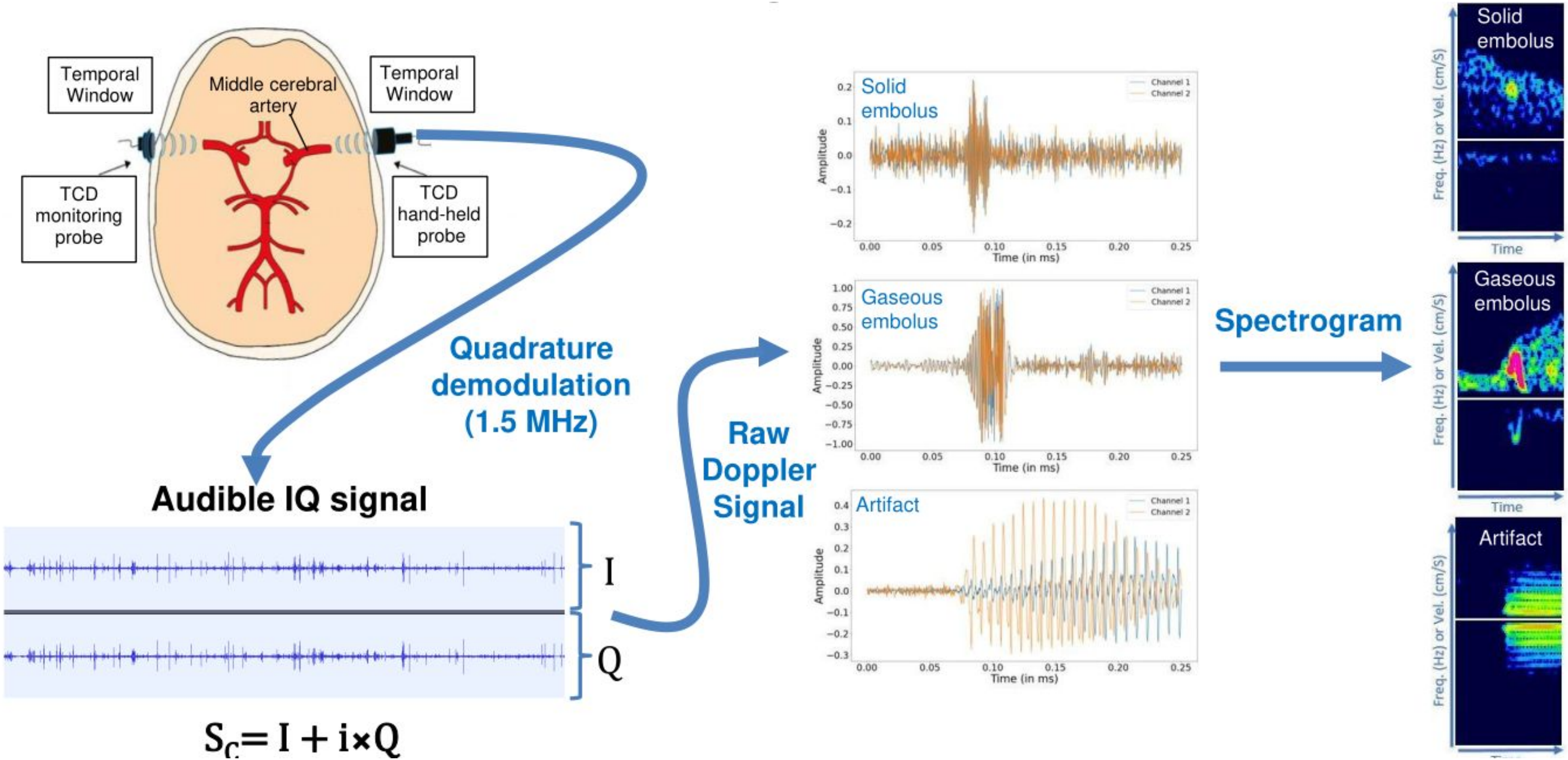
Depth wrt PRF

$$P_{max} = \frac{C}{2 * F_{prf}}$$

- Pulse repetition frequency: 4.4-6.2 kHz;
- Transmitted ultrasound frequency: 1.5 MHz;
- Insonation depth: $45 - 55\ mm$;
- Sample volume: $8 - 10\ mm^3$.

Credit: https://www.signal-processing.ch/background.php

TCD acquisition



Quadrature demodulation (1.5 MHz)

Audible IQ signal

$$S_C = I + i \times Q$$

Raw Doppler Signal

Spectrogram

Credit: Yamil Vindas PhD defense

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

How well Q approximates the true distribution P, the information gain achieved if P would be used instead of Q which is currently used.

$$H(p,q) = -\sum_{x \in \mathcal{X}} p(x) \log q(x).$$

Measures how well the distribution at hand q approximates the true distribution p.

$$H(p,q) = H(p) + D_{\mathrm{KL}}(p \parallel q)$$

$$[\mathrm{Geo}]\mathrm{JSD}(\mathrm{f}(\mathrm{x}), \mathrm{y}) = (1-\alpha)\mathrm{KL}(\mathrm{f}(\mathrm{x})||\textcircled{m}) + \alpha\mathrm{KL}(\mathrm{e}_{\mathrm{y}}||\textcircled{m})$$

Jensen-Shannon divergence is derived from KL and is symmetric (arguments invertible) with finite values. Its square root is a metric. Its value is lower when distributions are closer.
The geometric Jensen–Shannon divergence is a variation that yields a closed-form formula for divergence between two Gaussian distributions.

Risk and symmetric loss

Noise tolerant loss

$$\mathbb{P}_{\text{clean}}[f^*(x) = y_x] = \mathbb{P}_{\text{clean}}[f^*_\eta(x) = y_x]$$

Minimizer of clean risk

Minimizer of noisy risk

Risk

$$R_L(f) = \mathbb{E}_{\mathbf{x},y}[L(f(\mathbf{x}), y_{\mathbf{x}})]$$

Symmetric loss [3]

Symmetric loss L $\exists C, \forall f, \forall x, \sum_{k=1}^{K} L(f(x), k) = C$

number of classes

+ reasonable noise rate $\eta < \frac{K}{K-1}$

⇒ Robustness to uniform label noise

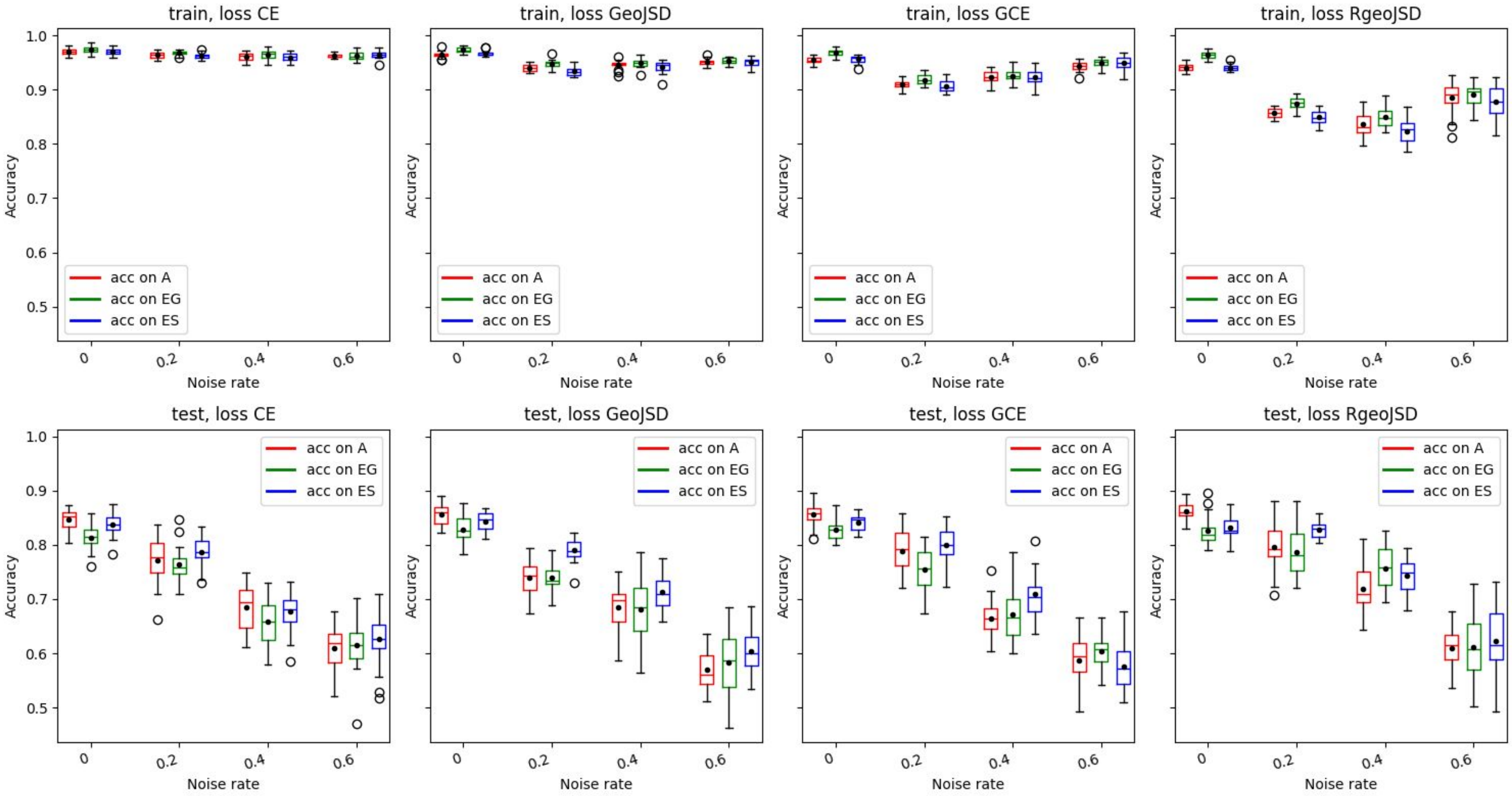
⇒ Bounded risk under simple non-uniform noise

+ true minimizer cancels out the risk

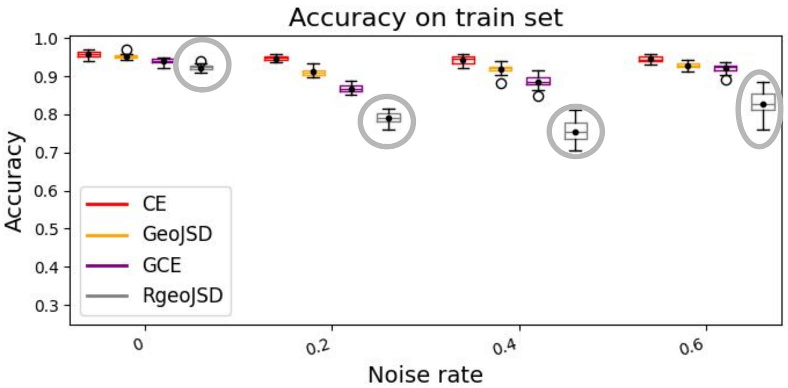
⇒ Robustness to simple non-uniform noise

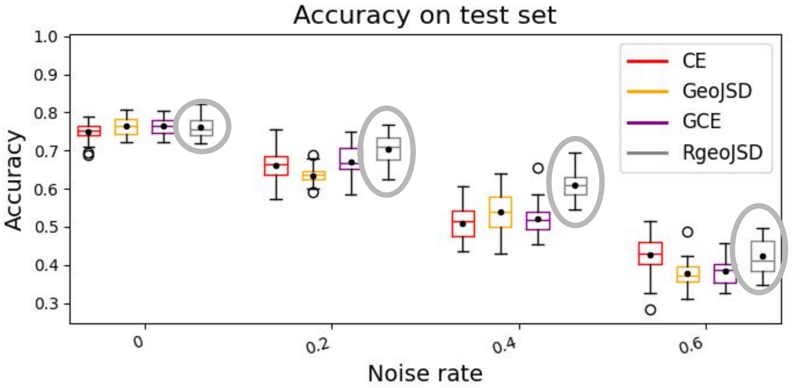
*[3] **Ghosh et al., AAAI, 2017** Robust Loss Functions under Label Noise for Deep Neural Networks*

Classification performance



Training behavior





⇒ Symptomatic of overfitting

Mean train and test losses over 20 runs