Calibrating AI Models for Few-Shot Demodulation via Conformal Prediction

Kfir M. Cohen, Sangwoo Park, Osvaldo Simeone, and Shlomo Shamai (Shitz)

King's College London

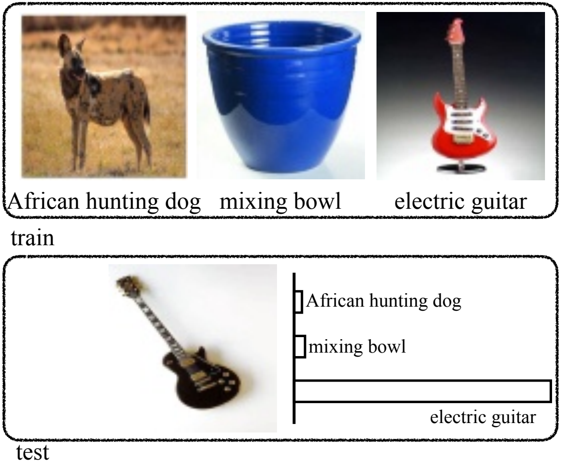
ICASSP 2023, Rhodes, Greece
June 2023

Overview

- AI tools are one of the main driving forces behind 6G
- They are capable of producing accurate, but not trustworthy, models
- In this work, we leverage Conformal Prediction (CP)[1] to ensure formal guarantees on reliability
- Application to demodulation

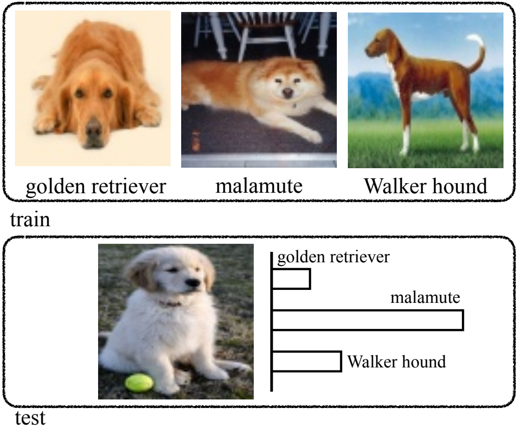
[1] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic Learning in a Random World, Springer, 2005.

Calibration of AI



- AI models typically output a hard decision, along with a **confidence level** (or, conversely, an **uncertainty level**).

Calibration of AI



- When failing, conventional **deep learning**-based AI systems tend to make **incorrect** decisions **confidently**.[2]

[2] G. Guo, et al, "On calibration of modern neural networks," in Proc. International conference on machine learning (ICML), 2017.

Calibration of AI

- Bayesian learning [3,4]
  - increases **computational complexity** as compared to conventional learning (by ensembling)
  - does not provide formal **finite-sample** calibration **guarantees**
- Post-hoc calibration schemes
  - address complexity by operating on a **pre-trained model**
  - can provide formal **finite-sample** calibration guarantees (**conformal prediction**[5,6])

.

[3] E. Angelino, et al, "Patterns of Scalable Bayesian Inference," Foundations and Trends in Machine Learning, 2016.

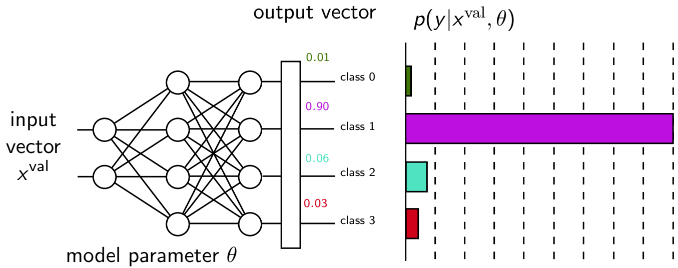
[4] O. Simeone, et al, "Machine Learning for Engineers," Cambridge University Press, 2022.

[5] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic Learning in a Random World, Springer, 2005.

[6] J. Cherian and L. Bronner, "How the Washington Post estimates outstanding votes for the 2020 presidential election".

Post-Hoc Calibration

- Some post-hoc calibration algorithms recalibrate a probabilistic model by matching accuracy estimated on a **validation set**.



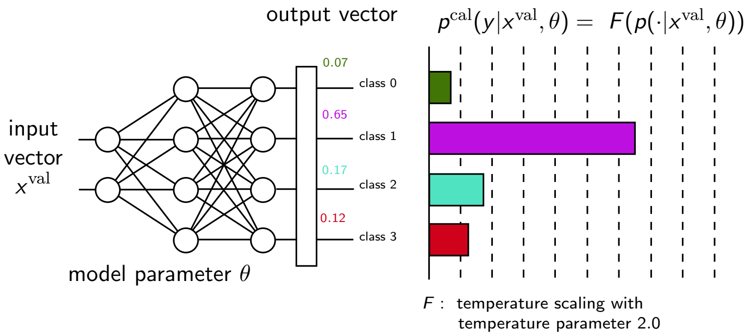
  - Temperature scaling, Platt scaling, isotonic regression
  - No guarantee of calibration: may **overfit the validation set**[7,8]

[7] A. Kumar, et al, "Verified Uncertainty Calibration," NeurIPS 2019.

[8] X. Ma and M. B. Blaschko, "Meta-Cal: Well-controlled Post-hoc Calibration by Ranking," ICML 2021.

Post-Hoc Calibration

- Some post-hoc calibration algorithms recalibrate a probabilistic model by matching accuracy estimated on a **validation set**.



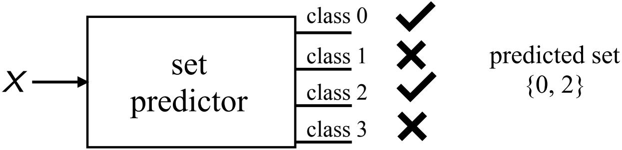
  - Temperature scaling, Platt scaling, isotonic regression
  - No guarantee of calibration: may **overfit the validation set**[7,8]

[7] A. Kumar, et al, "Verified Uncertainty Calibration," NeurIPS 2019.

[8] X. Ma and M. B. Blaschko, "Meta-Cal: Well-controlled Post-hoc Calibration by Ranking," ICML 2021.

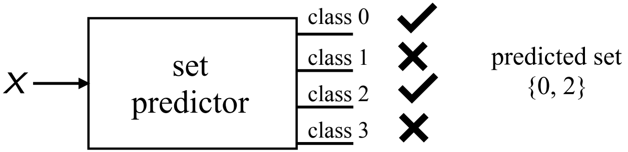
Conformal Prediction

- **Conformal prediction** produces set predictors.
- A **set predictor** is less informative than a probabilistic predictor:
  - Coarser, but easily **interpretable**, measure of **uncertainty** via **set size**
- **Conformal prediction** aims at extracting well-calibrated **set predictors** from probabilistic predictors.[9]



[9] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic Learning in a Random World, Springer, 2005.

Calibration of Set Predictors



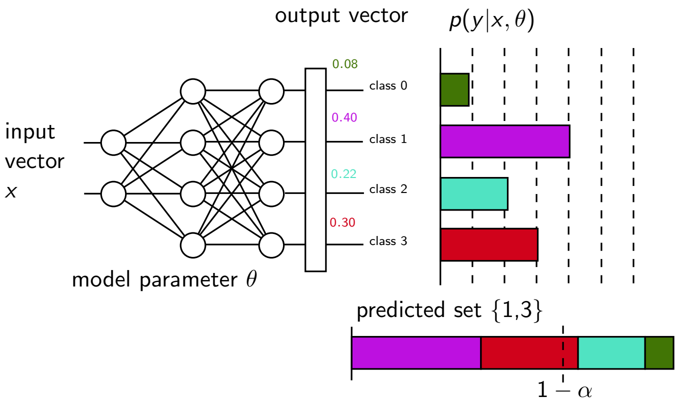
- A set predictor is **well calibrated** if

$$\mathbb{P}\big(\text{true label} \in \text{predicted set}\big) \geq 1 - \alpha$$

for some desired **coverage** probability $1 - \alpha$.

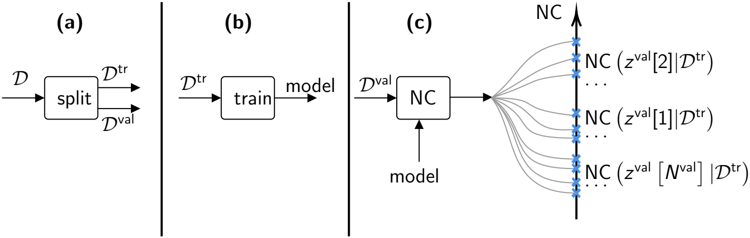
- Alternatively, we say it is $(1 - \alpha)$-**valid**.
- **Inefficiency** of a set predictor is the average predicted set sizes.

Set Predictors from Probabilistic Predictors



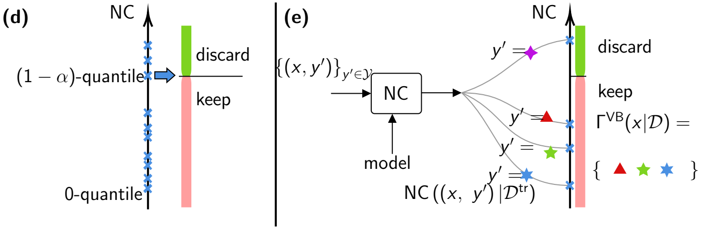
- Well-calibrated probabilistic predictor $\Longrightarrow$ well-calibrated set
- When $p(y|x,\theta) \neq p(y|x)$, this approach is invalid

Validation-based Conformal Prediction (VB-CP)



- **Nonconformity score**:
  - High when $z = (x, y)$ conforms poorly to $\mathcal{D}^{\text{tr}}$
  - E.g., for classification, $\mathrm{NC}\big((x,y)\big|\mathcal{D}^{\text{tr}}\big) = -\log p(y|x,\mathcal{D}^{\text{tr}})$
- Split data set into training and validation

Validation-based Conformal Prediction (VB-CP)



- Quantile analysis on the validation set
- VB-CP[10] is known to be $(1-\alpha)$-valid
- Assumption: test sample (x, y) and available data $\mathcal{D}$ are exchangeable

[10] V. Vovk, et al, "Algorithmic Learning in a Random World," Springer 2005.

K-cross-validation-based conformal prediction (K-CV-CP)

- Split the data into K folds
- For each of the folds $k = 1, \ldots, K$
  - A model is trained using the leave-fold-out
  - The fold is later used as a calibration set for that trained model
- Combine together the K predictions via quantile analysis[11]
- K-CV-CP is guaranteed to be $\approx (1 - 2\alpha)$-valid under the same exchangeability assumption

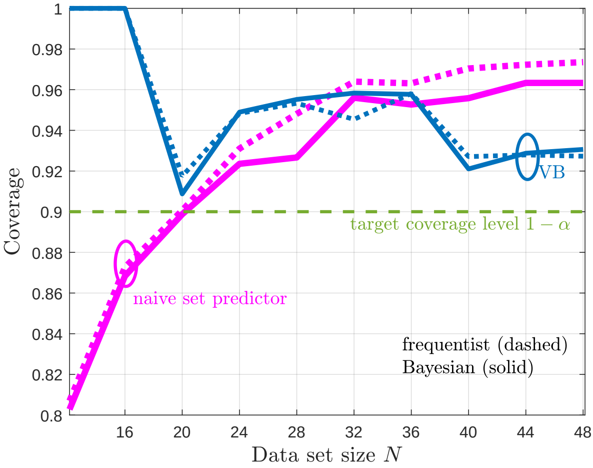
[11] R. F. Barber, et al, "Predictive inference with the jackknife+," The Annals of Statistics, 2021.

Experiments

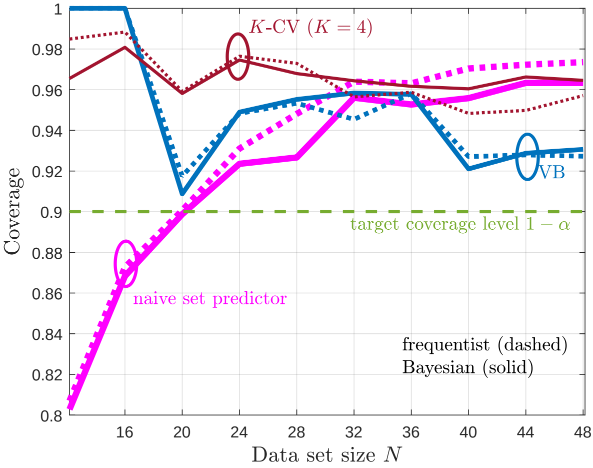
- Demodulation of 8-QAM constellation[12] ($|\mathcal{Y}| = 8$)
- Channel
  - Transmitter I/Q distortion
  - Random phase channel
  - AWGN
- Target miscoverage rate is set to $\alpha = 0.1$

[12] Z. Demeng, et al, "A Two-Stage Coded Modulation Scheme Based on the 8-QAM Signal for Optical Transmission Systems," Procedia Computer Science, 2018.

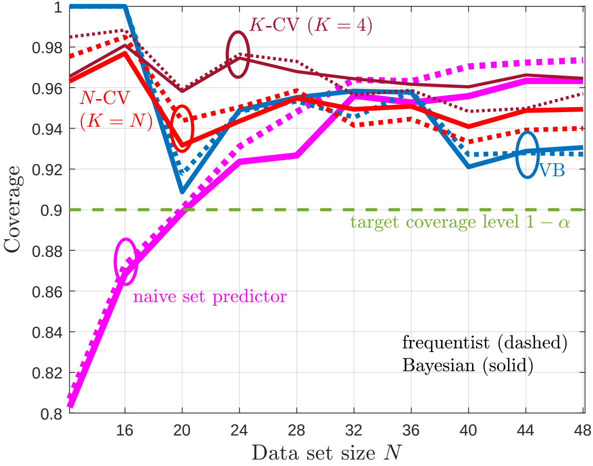
Experiments



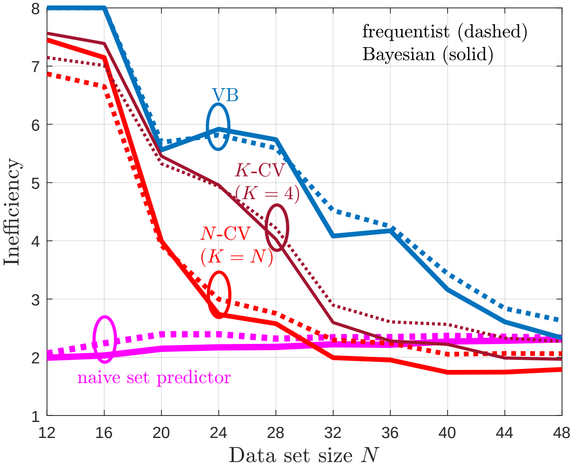
Experiments



Experiments



Experiments



Conclusions

- Forming set predictors directly from probabilistic predictors do not provide formal guarantees
- VB-CP provides calibration guarantees
  - $(1-\alpha)$-valid, even for misspecified models
- K-CV-CP better utilizes the available data, and is
  - $(1-2\alpha)$-valid de jure
  - $(1-\alpha)$-valid de facto[13,14]
  - More efficient, in the cost of training more models
- Gain of CP is prominent in the few-data regime

[13] R. F. Barber, et al, "Predictive inference with the jackknife+," The Annals of Statistics, 2021.

[14] Y. Romano, et al, "Classification with Valid and Adaptive Coverage," NeurIPS, 2020.

Acknowledgment

- The work of K. M. Cohen, S. Park and O. Simeone has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 725731.
- The work of O. Simeone has also been supported by an Open Fellowship of the EPSRC with reference EP/W024101/1, and by the European Union's Horizon Europe project CENTRIC (101096379).
- The work of S. Shamai has been supported by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement No. 694630.