What If Without the Conformal Prediction?

Jing LIANG
Industrial Engineering and Decision Analytics
Hong Kong University of Science and Technology
jliangcd@connect.ust.hk

Abstract

To be continued...

1 Introduction

With the growing use of machine learning in risk-sensitive domains such as medical diagnosis and financial risk management, aggregate metrics such as average accuracy are no longer sufficient. Conventional learning algorithms such as neural networks, support vector machines, and decision trees typically provide only point predictions and do not offer calibrated measures of uncertainty. In practice, decision-makers require not only accurate predictions but also a principled assessment of their reliability. Conformal prediction meets this requirement by providing instance-wise, distribution-free validity guarantees that quantify the uncertainty associated with each prediction.

If conformal prediction did not exist, what we would lose is not merely an algorithm but a critical meta-methodology: a systematic framework that decouples uncertainty quantification from model design through the construction of nonconformity scores. The impact would extend beyond the disappearance of a single technique, representing a fundamental shift in the paradigm of algorithm design.

2 Background and Need

With the rapid development of artificial intelligence and its widespread deployment across society, both industry and government agencies are increasingly adopting AI-based decision-support systems and agentic task-execution systems. These systems are typically built upon predictive machine learning models, which generate predictions or recommended actions through supervised learning tasks[1].

At the same time, AI technologies are accelerating their penetration into high-risk domains such as medical diagnosis, biometrics, face recognition, nuclear fusion, and industrial diagnostics[2]. In these settings, model failures can lead to severe medical, ethical, economic, or safety consequences. As a result, society and regulatory bodies have raised expectations regarding the reliability, fairness, transparency, and accountability of AI systems, thereby driving the development of Trustworthy AI (TAI).

Within the TAI framework, uncertainty quantification (UQ) plays a central role in ensuring reliability. The goal of UQ is to characterize and evaluate various sources of uncertainty in predictive models, providing reliable and calibrated confidence information[3]. This enables decision-makers to understand the trust boundaries of model outputs and make robust decisions accordingly.

Traditional UQ approaches face structural limitations that constrain their effectiveness in high-risk applications. Bayesian methods, such as Bayesian neural networks, provide probabilistic predictions but rely heavily on strong prior assumptions and the correctness of the model specification; inference is also computationally expensive for large-scale models. Frequentist methods, such as bootstrapping

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

or dropout-based UQ, estimate uncertainty through resampling or approximate inference but typically lack distribution-free, finite-sample guarantees. They also become unstable under distribution shift and incur high computational costs.

Consequently, traditional UQ methods generally fail to provide instance-wise, distribution-free guarantees of validity.

Conformal prediction, however, offers a systematic solution. It provides verifiable, finite-sample validity guarantees for each individual instance without requiring any distributional assumptions, enabling precise quantification of predictive uncertainty.

3 Conformal Prediction

3.1 Historical Development

The theoretical origins of conformal prediction were established in the late 1990s and early 2000s by Vladimir Vovk, Alex Gammerman, and Glenn Shafer[4, 5, 6]. Unlike traditional parametric statistics or Bayesian inference, the methodology arises from the frameworks of algorithmic randomness and game-theoretic probability[6, 7]. Its objective is to construct distribution-free predictive guarantees for machine learning that remain valid for finite samples. Within this paradigm, predictions produce sets of plausible labels rather than single point estimates, and the size of these sets adapts to model uncertainty, thereby providing a direct quantification of predictive reliability.

The foundational contribution to the field is the work that establishes the theoretical basis of conformal prediction [6]. A more recent overview is provided by the monograph "Conformal prediction for reliable machine learning: theory, adaptations and applications" [2].

Recent developments in conformal prediction have focused on three main areas: improving the validity concept, enhancing computational efficiency, and expanding application scope[8].

3.2 Conformal Prediction Framework

Vovk's work introduces full conformal prediction or transductive conformal prediction [6]. The central idea is that, given a training set and a new test instance, each possible label for the test instance is tentatively paired with the training data to compute a nonconformity score, and the label is included in the prediction set based on its rank among these scores. Because this procedure becomes computationally expensive as the dataset grows, subsequent research has proposed more efficient variants, most notably split conformal prediction[9], which is now widely used. This subsection will focus on split conformal prediction and its technical details, and will illustrate the overall conformal framework by comparing the associated forms of variance.

A simple introduction to the conformal prediction algorithm

1. Core insight: a revolutionary abstraction of the "score" definition Your observation that "the definition of the score removes the dependence on the model's internal structure" captures the essence of conformal prediction.

4 Key Insights and Benefit of Conformal Prediction

(1) The key step: the nonconformity score • Traditional view: model outputs (such as softmax probabilities or SVM decision values) are used directly as proxies for confidence. These are tightly coupled to the model's internal structure. A neural network probability and a random forest probability do not have the same meaning and cannot be compared directly. • CP's abstraction: it introduces a nonconformity score A(x, y). The only requirement is that for a data point (x, y), a higher score means the pair looks more "strange" or "nonconforming" relative to the model or data distribution. • Structure removed: this score A can be anything—prediction error |y f(x)|, negative log-likelihood, distance to a decision boundary, or variance across an ensemble. CP does not care how A is constructed; it only uses the ranking of these scores. It compresses any model into a scalar score generator.

(2) Significance: this abstraction achieves model-agnosticism. Whether using a ResNet, a transformer, or logistic regression, as long as a "strangeness score" can be computed for each candidate label, CP can provide coverage guarantees. This decouples uncertainty quantification from model design.

2. Theoretical foundation: trading "exchangeability" for mathematical guarantees Your point that CP "uses exchangeability to obtain mathematical guarantees" touches its statistical core.

(1) Mild assumption: CP does not require the data to be i.i.d.; it only needs exchangeability—the joint distribution remains unchanged under index permutations. This is weaker than i.i.d. and closer to practical settings (e.g., time series can often be made approximately exchangeable after proper preprocessing). (2) Strong guarantee: under this mild assumption, CP yields its key validity guarantee: for any new test point, the probability that the true label falls outside the prediction set is at most , the preset error rate. This is a finite-sample, distribution-free guarantee independent of model complexity or data distribution. (3) Insight: CP demonstrates a path where strong, practical statistical guarantees arise not from assuming a specific data-generating distribution, but from exploiting symmetry properties of data order. This is foundational for statistical inference in modern high-dimensional, complex data settings.

Therefore, the true insight of conformal prediction is that it offers a meta-methodology:

1. Decoupling: It separates model performance (implicitly reflected through the score function A) from output guarantees (derived from rank-based calibration and exchangeability). 2. Guarantee-first design: It reverses the traditional approach. Instead of trying to make model outputs "look like" probabilities and then hoping for guarantees, CP first builds a framework that inherently provides guarantees (via calibration and set-valued prediction), and then embeds any model inside it. 3. Infrastructure-friendly: It provides a simple, unified, and theoretically sound reliability layer for the diverse and ever-changing machine-learning ecosystem. This is not only an algorithmic innovation but also an elevation in system-level engineering thinking.

5 Alternative Methods

A wide range of uncertainty quantification techniques have been developed across statistics and machine learning. These approaches differ in their underlying assumptions, the form of guarantees they provide, and their suitability for various application settings. Table 1 summarizes the main methodological families commonly used as competitors to conformal prediction.

Method Family	Key Idea	Representative Competitors
Classical parametric prediction intervals	Assume a fully specified parametric model, typically linear with Gaussian noise	Gaussian linear model (t-interval)
Semi- / Non-parametric prediction bands	Assume a functional form for estimation, such as conditional means or quantiles, without making distributional assumptions	Quantile regression bands
Bayesian framework	Characterize uncertainty through the posterior predictive distribution	Bayesian neural networks; Bayesian model averaging
PAC-learning	Provide worst-case upper bounds on prediction error over the hypothesis class	Vapnik–Chervonenkis bounds; Littlestone–Warmuth bounds
Hold-out and resampling-based methods	Estimate uncertainty empirically via data splitting or repeated resampling	Train–test hold-out; cross-validation; bootstrap

Table 1: Major competitor methods to conformal prediction

None of these approaches simultaneously achieve the combination of distribution-free operation, finite-sample validity, and instance-wise guarantees that conformal prediction provides. Nevertheless, each method family has its own advantages and suitable application contexts. The remainder of this section outlines the core ideas behind several representative competitors and contrasts them with conformal prediction to highlight its distinct value.

5.1 Bayesian framework

5.2 Boostrap resampling

5.3 quantile regression

5.4 Others

6 Implications of a World Without Conformal Prediction

TBC

7 Conclusion

TBC

References

[1] Yuzhou Qian, Keng L. Siau, and Fiona F. Nah. Societal impacts of artificial intelligence: Ethical, legal, and governance issues. *Societal Impacts*, 3:100040, 2024. ISSN 2949-6977. doi: https://doi.org/10.1016/j.socimp.2024.100040. URL `https://www.sciencedirect.com/science/article/pii/S2949697724000055`.

[2] Vineeth Balasubramanian, Shen-Shyang Ho, and Vladimir Vovk. *Conformal prediction for reliable machine learning: theory, adaptations and applications*. Newnes, 2014.

[3] Ralph C Smith. *Uncertainty quantification: theory, implementation, and applications*. SIAM, 2024.

[4] C. Saunders, A. Gammerman, and V. Vovk. Transduction with confidence and credibility. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI '99)*, pages 722–726, 1999.

[5] Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In Tapio Elomaa, Heikki Mannila, and Hannu Toivonen, editors, *Machine Learning: ECML 2002*, pages 345–356, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. ISBN 978-3-540-36755-0.

[6] Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

[7] Alexander Gammerman and Vladimir Vovk. Kolmogorov complexity: Sources, theory and applications. *Computer Journal*, 42(4), 1999.

[8] Matteo Fontana, Gianluca Zeni, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *Bernoulli*, 29(1):1–23, 2023.

[9] Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018.