ORIGINAL RESEARCH **Open access**

Neural Network with Adaptive Conformal Prediction for Providing Calibrated Uncertainty Intervals around Individualized Chemotherapy Toxicity Risk Predictions

Beatriz Romero[1], Ines Castillo[1*], Arturo Medina[2], Lucia Vega[1]

Abstract

Chemotherapy remains a cornerstone of cancer treatment, but it is frequently associated with severe toxicities, with 30–80% of patients experiencing grade 3–4 adverse events that may require dose reduction, treatment delays, or hospitalization. While machine learning models have shown strong potential in predicting chemotherapy-related toxicities using electronic health records, genomic data, and clinical variables, most existing approaches generate only point estimates (e.g., a single risk probability) without quantifying uncertainty, limiting their clinical reliability. Such miscalibrated predictions can lead to overconfident risk underestimation or excessive caution, both of which may negatively impact treatment decisions and patient outcomes. This manuscript proposes a conceptual framework that integrates neural network-based toxicity prediction with adaptive conformal prediction to produce calibrated, patient-specific prediction intervals with formal coverage guarantees. The framework combines a feedforward neural network for risk estimation, a non-conformity score to measure how atypical a patient is relative to the training data, and an adaptive calibration mechanism that updates interval thresholds over time to reflect shifts in patient populations and clinical practice. This design enables narrower intervals for well-represented, predictable cases and wider intervals for atypical or high-uncertainty patients, thereby making prediction reliability explicit. Importantly, the method provides finite-sample coverage guarantees without requiring distributional assumptions, ensuring that true toxicity outcomes fall within the predicted intervals at a user-specified confidence level. By transforming point predictions into uncertainty-aware, clinically interpretable intervals, the framework supports more robust, risk-stratified decision-making in chemotherapy planning and moves toward safer, more trustworthy AI-assisted oncology care.

Keywords Clinical decision support, Uncertainty quantification, Conformal prediction, Chemotherapy toxicity, Neural networks, Adaptive calibration

*Correspondence:
Ines Castillo
ines.castillo@gmail.com
[1] Department of AI Healthcare Systems, University of Zaragoza, Zaragoza, Spain
[2] Department of Clinical Intelligence Engineering, University of Malaga, Malaga, Spain

Introduction

Chemotherapy-induced toxicities represent a major source of morbidity and mortality in cancer care, with severe adverse events including febrile neutropenia, cardiotoxicity, peripheral neuropathy, chemotherapy-induced nausea and vomiting, and acute kidney injury occurring frequently across treatment regimens [1, 2]. These toxicities not only compromise patient quality of life but also force oncologists to reduce chemotherapy doses, delay treatment cycles, or discontinue curative-intent therapy altogether, potentially compromising long-term oncologic outcomes. The inter-

individual variability in toxicity susceptibility—driven by genetic polymorphisms, comorbidities, organ function, and concomitant medications—makes accurate, personalized risk prediction essential for optimizing the therapeutic index of chemotherapy. Existing clinical risk scores such as the Chemotherapy Assessment Score (CASS), CRASH score, and SAFE score provide population-level risk stratification but lack the granularity needed for truly individualized decision-making [3].

Machine learning approaches have demonstrated substantial improvements over traditional risk scores in predicting chemotherapy toxicities, with models based on logistic regression, random forests, gradient boosting, and neural networks achieving strong discriminative performance for neutropenia, neuropathy, and cardiotoxicity prediction [4, 5]. However, these models overwhelmingly produce point predictions—a single number representing the estimated probability of toxicity—without quantifying the uncertainty associated with that estimate. In clinical practice, a prediction of 25% risk of febrile neutropenia carries fundamentally different implications depending on whether the model is highly confident (e.g., the prediction derives from patients with near-identical clinical profiles in the training data) or profoundly uncertain (e.g., the patient has a rare combination of risk factors unseen during model development). The inability to communicate this distinction undermines clinician trust in machine learning predictions and limits their clinical utility for high-stakes decisions such as prophylactic growth factor administration or chemotherapy dose modification.

Table 1 clarifies how the proposed framework changes chemotherapy toxicity prediction from a point-estimation task into an uncertainty-aware decision-support process.

Table 1. Conceptual Distinction between Point Prediction and Adaptive Conformal Interval-Based Toxicity Prediction

Analytical dimension	Conventional neural network toxicity prediction	Adaptive conformal neural network framework	Added clinical value
Primary output	Single estimated probability of toxicity	Point estimate plus calibrated lower and upper uncertainty bounds	Prevent the mod output fro appearin more precise th clinically justified
Treatment of uncertainty	Usually implicit, unreported, or limited to population-level calibration	Explicit patient-specific risk interval with nominal coverage target	Allows clinicians distinguis confider from uncertai predictio
Response to atypical patients	May remain overconfident for patients unlike the training population	Produces wider intervals when non-conformity scores indicate prediction difficulty	Flags rar borderlin or poorl represent cases fo caution
Statistical guarantee	No finite-sample coverage guarantee for individual deployment settings	Distribution-free finite-sample marginal coverage under exchangeability	Provides stronge safety rationale clinical u
Clinical interpretability	Requires clinicians to act on a single probability	Communicates risk as a plausible calibrated range	Aligns model output wi clinical reasonin under uncertain
Main failure mode	Overconfident incorrect risk estimates may drive inappropriate treatment decisions	Intervals may be too wide or may not ensure conditional subgroup coverage	Makes uncertain visible rather tha hidden
Decision-support implication	Encourages binary threshold-based action	Supports tiered decisions based on interval position and width	Enables interventi monitorin or deferr based o uncertain structur

Standard neural networks, despite their expressive power and predictive accuracy, are particularly susceptible to producing miscalibrated probability estimates, often exhibiting overconfidence in incorrect predictions when applied to patients whose clinical features deviate from the training distribution [6, 7]. Post-hoc calibration methods such as Platt scaling or isotonic regression can improve overall calibration but do not provide patient-specific uncertainty estimates that reflect the difficulty of individual predictions. Bayesian neural networks and Monte Carlo dropout offer principled approaches to uncertainty quantification by modeling distributions over network weights, producing credible intervals that capture both aleatoric and epistemic uncertainty [5, 8]. However, these methods rely on distributional assumptions and approximations that may not hold in the heterogeneous, high-dimensional feature spaces characteristic of clinical prediction tasks, and their interval coverage is not guaranteed in finite samples.

This manuscript proposes a framework that combines neural network toxicity prediction with adaptive conformal prediction, a distribution-free method that provides finite-sample coverage guarantees without imposing parametric assumptions on the data or the model [9, 10]. Conformal prediction generates prediction intervals that are guaranteed to contain the true outcome with a user-specified probability, and the adaptive variant adjusts interval widths based on the difficulty of each prediction case, producing wider intervals for atypical patients and narrower intervals for routine cases. Applied to chemotherapy toxicity prediction, this framework enables oncologists to interpret each toxicity risk estimate within a calibrated uncertainty interval, distinguishing between predictions that are sufficiently reliable for decisive clinical action and those that warrant additional caution, monitoring, or diagnostic investigation. The following sections detail the background, architecture, and clinical applications of this uncertainty-aware predictive framework.

Background

Chemotherapy toxicity prediction

Chemotherapy-induced toxicities encompass a broad spectrum of adverse events affecting multiple organ systems, with neutropenia and febrile neutropenia representing the most common dose-limiting toxicities across cytotoxic regimens [3, 11]. Additional clinically significant toxicities include anthracycline-induced cardiotoxicity, platinum-induced peripheral neuropathy, chemotherapy-induced nausea and vomiting, mucositis, hepatotoxicity, and nephrotoxicity, each with distinct pathophysiological mechanisms and risk factors. The incidence of severe chemotherapy toxicity varies substantially across patient populations, with reported rates of grade 3–4 neutropenia ranging from 20% to 60% depending on the chemotherapy regimen, and rates of chemotherapy-induced peripheral neuropathy exceeding 70% in patients receiving platinum or taxane-based regimens [12, 13]. Machine learning models have been developed to predict individual patient risk for many of these toxicities, incorporating demographic variables, laboratory values, genetic polymorphisms, comorbidity indices, and treatment characteristics as predictive features [14, 15]. These prediction tools hold the potential to enable pre-emptive dose adjustment, targeted supportive care, and personalized chemotherapy selection, thereby reducing toxicity-related morbidity while maintaining therapeutic efficacy.

Existing prediction models

Current approaches to chemotherapy toxicity prediction span a methodological spectrum from clinical risk scores to advanced machine learning algorithms, each with distinct strengths and limitations for clinical deployment [4, 5]. Traditional clinical risk scores such as the CASS, CRASH, and SAFE models rely on a small number of hand-selected variables—typically age, baseline blood counts, liver function, and chemotherapy regimen intensity—to stratify patients into risk categories. Machine learning models including logistic regression, random forests, gradient-boosted trees, and neural networks have demonstrated improved discriminative performance by learning complex, non-linear interactions among larger sets of predictor variables from electronic health records [15-18]. However, the predominant output of these models is a point prediction—a single probability of toxicity—without accompanying quantification of the uncertainty surrounding that estimate, which limits their clinical interpretability and actionability for individual patients [19]. Furthermore, the calibration of these models, which reflects the alignment between predicted and observed toxicity rates, is frequently assessed only at the population level and may degrade substantially when applied to patient subgroups underrepresented in training data.

Conformal prediction

Conformal prediction provides a rigorous statistical framework for constructing prediction intervals with finite-sample coverage guarantees, requiring no assumptions about the underlying data distribution or the model used to generate predictions [1, 12]. In the split conformal prediction framework, the available data are partitioned into a training set used to fit the prediction model and a disjoint calibration set used to compute non-conformity scores—measures of how atypical each calibration example is relative to the model's predictions. For a new patient, the prediction interval is constructed by combining the model's point prediction with a quantile of the calibration non-conformity scores, yielding an interval that is guaranteed to contain the true outcome with probability at least 1−α for any user-specified confidence level [20, 21]. This guarantee holds marginally over the calibration and test data distributions under the assumption of exchangeability, making conformal prediction particularly attractive for clinical applications where distributional assumptions may be violated and rigorous uncertainty quantification is essential for patient safety. The resulting intervals are adaptive to the difficulty of each prediction when an appropriate non-conformity score is employed, producing wider intervals for patients whose features diverge from the training distribution.

Adaptive conformal prediction

Adaptive conformal prediction extends the standard conformal framework to accommodate settings where the data distribution may shift over time or where prediction difficulty varies systematically across patient subgroups, conditions that are commonplace in clinical oncology practice [13, 22]. The adaptive conformal inference algorithm dynamically adjusts the calibration threshold in response to observed patterns of prediction errors, enabling the framework to maintain valid coverage even when the exchangeability assumption is violated by temporal trends in chemotherapy regimens, changes in supportive care protocols, or evolving patient demographics. For chemotherapy toxicity prediction, this adaptivity is clinically important because toxicity risk profiles may shift with the introduction of new chemotherapy agents, changes in dosing guidelines, or seasonal variations in infection-related complications [14, 23]. The adaptive mechanism recalibrates the non-conformity threshold based on recent prediction errors, ensuring that prediction intervals remain valid for current patients while still leveraging the full calibration dataset for stable threshold estimation. This approach maintains the distribution-free coverage guarantees of conformal prediction while enhancing robustness to the non-stationarity inherent in real-world clinical data.

Framework Overview

High-level architecture

The proposed framework follows a sequential pipeline that transforms raw patient features into calibrated uncertainty intervals suitable for clinical decision support [15, 24]. Patient-level data—including demographics, baseline laboratory values, chemotherapy regimen details, comorbidity indices, and genetic markers—are first processed through a feedforward neural network trained to predict the probability of a specified chemotherapy toxicity outcome, such as febrile neutropenia or peripheral neuropathy. The neural network outputs a point prediction representing the estimated toxicity risk, which is then paired with a non-conformity score that quantifies how unusual or difficult-to-predict the patient's clinical profile is relative to the calibration dataset. An adaptive calibration mechanism adjusts the non-conformity threshold based on recent prediction performance, and the final prediction interval is constructed by combining the point prediction with this calibrated threshold to produce an upper and lower bound on the toxicity risk estimate [6, 25]. This architecture preserves the predictive flexibility of neural networks while augmenting their outputs with rigorous, distribution-free uncertainty quantification that directly supports clinical interpretation and decision-making.

Figure 1 presents the proposed left-to-right framework through which patient-level chemotherapy features are transformed into calibrated toxicity risk intervals and translated into uncertainty-aware clinical decision support.

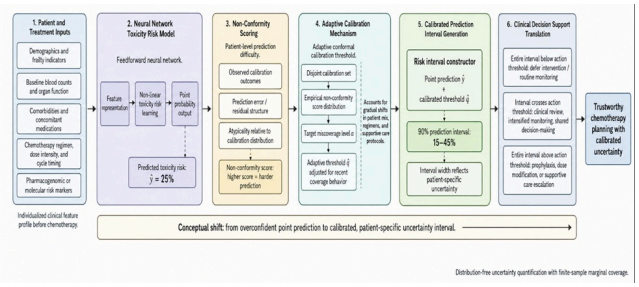

Figure 1. Adaptive conformal neural network architecture for calibrated chemotherapy toxicity risk intervals.

Core assumptions

The framework operates under a set of assumptions that determine the validity and scope of its coverage guarantees, with the principal requirement being exchangeability of the calibration and test data [16, 26]. Exchangeability—the condition that the joint distribution of the data is invariant under permutation—is approximately satisfied when calibration and test patients are drawn from the same underlying population and there are no systematic differences in data collection, patient characteristics, or outcome ascertainment between the two sets. For the adaptive variant of the framework, this assumption is relaxed to allow for gradual distribution shifts, with the adaptive threshold updating mechanism maintaining approximate coverage validity as long as shifts occur slowly relative to the rate of threshold adaptation [13, 22]. Additional assumptions include the availability of a representative calibration dataset of sufficient size to estimate non-conformity score quantiles with adequate precision, and the requirement that the neural network architecture and training procedure are held fixed after calibration to preserve the validity of the coverage guarantee. These assumptions are generally consistent with the design of clinical prediction model development and validation studies in oncology.

Design principles

The framework is guided by four design principles that collectively ensure its suitability for high-stakes clinical applications in chemotherapy toxicity prediction [17, 27]. The first principle is distribution-free validity, meaning that the coverage guarantee of the prediction intervals holds regardless of the underlying data distribution, the neural network architecture, or the training procedure employed, distinguishing conformal prediction from parametric uncertainty quantification methods that rely on distributional assumptions. The second principle is finite-sample coverage, ensuring that the prediction intervals achieve the nominal coverage level in finite samples rather than asymptotically, which is critical for clinical settings where large sample sizes cannot be assumed for rare toxicities or specific patient subgroups. The third principle is adaptivity to prediction difficulty, enabling the interval width to reflect the inherent uncertainty in each individual prediction, with wider intervals flagging patients for whom the model's toxicity risk estimate is less reliable. The fourth principle is interpretability of the resulting intervals, with upper and lower bounds on toxicity probability directly supporting clinical risk stratification and shared decision-making without requiring clinicians to understand the underlying statistical machinery.

Neural Network Architectur

Risk prediction model

The neural network component of the framework comprises a feedforward architecture with two to four hidden layers employing rectified linear unit activations, trained via stochastic gradient descent with binary cross-entropy loss to predict the probability of chemotherapy toxicity occurrence [18, 28]. Input features are drawn from domains known to influence chemotherapy toxicity risk, including patient demographics, baseline hematologic and biochemical laboratory values, estimated glomerular filtration rate and hepatic function markers, body surface area and body mass index, chemotherapy agent types and planned dose intensity, prior treatment history, comorbidity indices, concomitant medications, and germline genetic variants in pharmacokinetic and pharmacodynamic genes implicated in drug metabolism and transport. The output layer employs a sigmoid activation function to produce a predicted probability between zero and one, representing the estimated risk of the specified toxicity outcome, with model complexity and regularization strength selected to optimize discriminative performance while mitigating overfitting. This neural network architecture has been widely adopted in chemotherapy toxicity prediction studies, where it has demonstrated the capacity to capture complex, non-linear relationships between patient features and toxicity outcomes that simpler linear models may fail to identify [24, 26].

Non-conformity score

The non-conformity score function serves as the bridge between the neural network's point prediction and the conformal prediction interval, quantifying how atypical or difficult-to-predict each patient is relative to the patterns learned by the model [19, 29]. For binary toxicity outcomes, a natural non-conformity score is one minus the predicted probability for patients who experience toxicity, and the predicted probability itself for patients who do not, yielding larger scores when the model assigns low probability to observed toxicities or high probability to patients who remain toxicity-free. Alternative non-conformity scores, such as the absolute residual between the observed binary outcome and the predicted probability or more

sophisticated scores based on conditional density estimation, can be substituted depending on the desired sensitivity of the resulting intervals to specific types of prediction errors. The choice of non-conformity score directly influences the adaptivity of the prediction intervals, with scores that capture the magnitude and direction of prediction errors producing intervals that widen appropriately for patients whose clinical profiles diverge from well-represented training examples. This adaptive property is essential for clinical utility, as it ensures that patients with rare comorbidity patterns or unusual chemotherapy regimens receive appropriately wide intervals that reflect the model's increased uncertainty.

Adaptive Conformal Prediction

Split conformal setup

The split conformal prediction procedure partitions the available patient data into three disjoint sets: a training set for fitting the neural network model, a calibration set comprising 20–30% of the data for computing non-conformity scores and estimating quantile thresholds, and a held-out test set for evaluating prediction interval performance [2, 21]. The calibration set patients are processed through the trained neural network to obtain point predictions, and non-conformity scores are computed for each calibration patient using the chosen score function, producing an empirical distribution of scores that captures the range of prediction errors expected under the model. A critical requirement of the split conformal approach is that the calibration data remain entirely unused during model training, ensuring that the non-conformity scores provide an unbiased estimate of the model's prediction error on new patients drawn from the same distribution [12, 20]. This data-splitting strategy is straightforward to implement in typical clinical prediction model development workflows and aligns with established practices for separating model development from validation in prognostic research.

Adaptive calibration threshold

The adaptive calibration threshold extends the standard conformal quantile to accommodate non-stationarity in the patient population or toxicity outcome distribution, updating the threshold in response to observed prediction errors over time [22, 23]. The threshold is defined as the (1−α)(1+1/|D_cal|)-th empirical quantile of the calibration non-conformity scores, where α is the desired miscoverage rate and the finite-sample correction factor 1/|D_cal| accounts for the uncertainty in quantile estimation from a finite calibration set. In the adaptive variant, this threshold is dynamically adjusted using an online learning mechanism that increases the threshold when recent empirical coverage falls below the target level and decreases it when coverage exceeds the target, maintaining valid coverage under gradual distribution shifts while preventing interval widths from growing unnecessarily large [13, 14]. This adaptivity is particularly relevant in oncology settings where chemotherapy protocols, supportive care guidelines, and patient referral patterns may evolve over the course of a clinical prediction model's deployment, potentially degrading the performance of static calibration thresholds.

Prediction interval

For a new patient with neural network point prediction $\hat{y}$ and adaptive calibration threshold $\hat{q}$, the conformal prediction interval is constructed to contain the true toxicity outcome with probability at least 1−α [1, 6]. In the regression setting where the toxicity outcome is modeled as a continuous severity score, the interval takes the form $[\hat{y} - \hat{q}, \hat{y} + \hat{q}]$, producing symmetric bounds around the point prediction whose width reflects the calibration threshold and, through the threshold's adaptivity, the difficulty of the individual prediction case. For the binary classification setting more typical of chemotherapy toxicity prediction, the interval is interpreted as covering the predicted probability distribution, with the non-conformity score determining which probability values are considered sufficiently consistent with the model's predictions to be included [8, 9]. The resulting interval can be communicated to clinicians as a calibrated toxicity risk range—for example, a prediction of 25% risk with a 90% prediction interval of 15–45%—which conveys both the point estimate and the model's confidence in that estimate in a clinically interpretable format. This transformation of point predictions into calibrated intervals represents the core contribution of the framework to uncertainty-aware chemotherapy toxicity prediction.

Uncertainty Interval Generation

Point prediction versus interval

The transition from a point prediction to a conformal prediction interval fundamentally transforms how chemotherapy toxicity risk is communicated to the clinical team, replacing a single number that conveys an illusion of precision with a calibrated range that transparently represents predictive uncertainty [25, 28]. A clinician receiving a point prediction of "25% risk of febrile neutropenia" has no information about whether this estimate derives from a robust model trained on thousands of similar patients or from a model extrapolating uncertainly from a sparse region of the feature space, and this ambiguity can lead to either excessive or insufficient clinical intervention depending on the clinician's subjective risk tolerance. In contrast, the conformal prediction interval presents the same underlying prediction as "25% risk (90% PI: 15–45%)," making explicit that the model's best estimate is 25% but that values between 15% and 45% are all plausibly consistent with the patient's clinical profile and the model's uncertainty. This additional information allows the clinician to calibrate their response to the prediction, distinguishing between confident high-risk predictions that unambiguously warrant aggressive supportive care and uncertain intermediate-risk predictions that justify a more nuanced, monitoring-intensive approach [9, 10]. The prediction interval thus serves as both a quantitative uncertainty estimate and a communication tool that aligns machine learning outputs with the clinical reality that risk prediction is inherently imprecise for individual patients.

Adaptive width interpretation

The width of the conformal prediction interval provides a direct, interpretable measure of predictive confidence that varies across patients in clinically meaningful ways, with narrower intervals indicating higher model confidence and wider intervals signaling greater uncertainty [19, 26]. A narrow interval such as 25% risk (90% PI: 20–32%) suggests that the patient's clinical features align closely with well-represented training examples, that the neural network has learned a stable relationship between those features and the toxicity outcome, and that the point prediction can be used with reasonable confidence for clinical decision-making. A wide interval such as 25% risk (90% PI: 5–60%) indicates that the patient's profile is atypical or lies near a decision boundary where small changes in features would substantially alter the prediction, and this width itself becomes clinically actionable by flagging the patient as a candidate for intensified monitoring, additional diagnostic testing, or specialist consultation [7, 27]. The adaptive property of the conformal framework ensures that interval widths respond appropriately to patient-specific difficulty, with patients harboring rare genetic variants, unusual comorbidity combinations, or atypical chemotherapy regimens receiving appropriately wide intervals that reflect the model's epistemic uncertainty. This direct mapping from interval width to clinical action—narrow intervals supporting decisive treatment, wide intervals triggering caution—provides a principled basis for integrating machine learning predictions into chemotherapy toxicity management workflows.

Clinical Decision Support

Risk-based action thresholds

Prediction intervals map onto clinical action thresholds through a tiered decision strategy: when the entire interval exceeds the threshold, intervention is clearly indicated; when it falls entirely below, intervention can be safely deferred; and when the interval straddles the threshold, clinical judgment with intensified monitoring is warranted [3, 25]. This prevents both under-treatment driven by overconfident low predictions and over-treatment driven by uncertain elevated predictions. For toxicities without established thresholds, interval width alone can guide monitoring frequency [15].

Table 2 translates calibrated toxicity risk intervals into clinically interpretable action categories for chemotherapy planning.

Table 2. Decision Logic for Translating Calibrated Toxicity Risk Intervals Into Chemotherapy Management Actions

Interval relationship to clinical action threshold	Interpretation of model output	Recommended clinical response	Rationa
Entire interval below threshold	Toxicity risk is consistently estimated as low, even after accounting for uncertainty	Continue planned chemotherapy with routine monitoring	Both poi estimate a uncertair bounds support lo risk classificat

Point estimate below threshold but upper bound exceeds threshold	Apparent low risk is uncertain because plausible risk values cross the action boundary	Consider closer monitoring, additional laboratory review, or patient counseling	Prevent under-treatme caused l overconfic low poir predictio
Interval straddles threshold widely	Prediction is clinically ambiguous and model uncertainty is substantial	Escalate to clinician review, shared decision-making, or specialist input	Wide uncertair indicates t automate risk stratificat alone is insufficie
Point estimate above threshold but lower bound falls below threshold	Elevated risk signal exists but is not decisively supported across the full interval	Consider individualized supportive care based on patient preference and toxicity severity	Avoids unnecess interventi when hig risk classificat is uncerta
Entire interval above threshold	Toxicity risk remains high even after uncertainty adjustment	Initiate prophylaxis, dose modification, intensified monitoring, or supportive care escalation	Interva positior support decisiv action un calibrate uncertair
Extremely wide interval regardless of threshold position	Patient profile is atypical or poorly represented in calibration data	Treat uncertainty itself as clinically actionable; request additional assessment or data review	Interval w identifie cases wh the mod should no used as sole decis basis

Shared decision-making

Prediction intervals enable transparent communication of uncertainty during shared decision-making, with oncologists presenting both the point estimate and its range to convey model confidence [4, 17]. Wide intervals prompt incorporation of closer monitoring and adaptive treatment modification into the care plan, converting uncertainty into an actionable component rather than a source of anxiety. This aligns with precision oncology's emphasis on individualized treatment tailored to patient-specific characteristics and preferences.

Evaluation Strategy

Coverage metrics

Empirical coverage—the proportion of test patients whose observed toxicity falls within the prediction interval—is the primary evaluation metric and should approximate the nominal 90% level under exchangeability [1, 20]. Coverage is assessed marginally across the population and within subgroups defined by age, chemotherapy regimen, and comorbidity burden to detect systematic undercoverage in specific populations [21]. Conditional coverage evaluation, while not guaranteed by the marginal framework, identifies patient subgroups for whom intervals may be misleadingly narrow or wide.

Efficiency metrics

Average interval width quantifies precision, with narrow intervals indicating low uncertainty and wide intervals signaling limited clinical actionability [6, 19]. The adaptive width ratio—comparing the widest to narrowest patient quartiles—measures how effectively the framework differentiates easy from difficult prediction cases. These metrics ensure the framework achieves valid coverage without producing intervals too wide to be clinically useful.

Baseline comparisons

Comparisons to fixed-width intervals, Bayesian neural network credible intervals, and Monte Carlo dropout prediction intervals contextualize the framework's performance [5, 8]. Fixed-width intervals provide a non-adaptive benchmark, while Bayesian and dropout methods offer parametric alternatives that capture uncertainty but lack finite-sample coverage guarantees. Comparative evaluation reveals whether guaranteed coverage incurs a meaningful width penalty relative to these alternatives.

Limitations

Technical limitations

The marginal coverage guarantee does not ensure conditional coverage, meaning intervals may systematically undercover for patients with rare chemotherapy regimens, uncommon genetic variants, or underrepresented demographic groups, raising health equity concerns [1, 22]. Calibration set size trades off against quantile precision, and rare toxicities with low event rates pose particular challenges for single-institution datasets. Exchangeability violations due to evolving chemotherapy protocols or referral patterns further limit the guarantee to approximate validity under the adaptive mechanism.

Clinical limitations

Clinicians may misinterpret prediction intervals as Bayesian credible intervals rather than frequentist coverage statements, potentially leading to inappropriate decisions if the distinction is not clearly communicated [9, 17]. Statistically appropriate intervals may be too wide for clinical actionability in complex patients, providing little guidance beyond acknowledging uncertainty. Prospective validation and clinician training are essential before implementation to establish whether uncertainty-aware predictions improve toxicity outcomes compared to standard decision support.

Conclusion

This manuscript presented a framework integrating neural network toxicity prediction with adaptive conformal prediction to generate calibrated, patient-specific uncertainty intervals. The approach transforms point predictions into intervals with finite-sample coverage guarantees, maintaining neural network flexibility while providing distribution-free uncertainty quantification.

Key advantages include adaptivity to patient difficulty, producing wider intervals for atypical cases, and compatibility with any neural network architecture without model modification. These properties distinguish conformal prediction from parametric alternatives that lack guaranteed coverage.

Limitations include the gap between marginal and conditional coverage, raising equity concerns, and calibration set requirements that challenge single-institution deployment. Clinician interpretation and workflow integration represent additional barriers requiring implementation research.

Future work should apply this framework to real-world chemotherapy datasets and conduct prospective studies comparing uncertainty-aware decision support to standard tools. Embedding rigorous uncertainty quantification into clinical prediction models will be essential for building trust and improving the safety of machine learning in oncology.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 27 Dec 2024 Revised: 16 Feb 2025 Accepted: 26 Apr 2025
Published online: 20 July 2025

References

Lei J, G'Sell M, Rinaldo A, Tibshirani RJ, Wasserman L. Distribution-free predictive inference for regression. J Am Stat Assoc. 2018;113(523):1094-111.

Vovk V, Shen J, Manokhin V, Xie MG. Nonparametric predictive distributions based on conformal prediction. In: Conformal and probabilistic prediction and applications. PMLR; 2017. p. 82-102.

Shashikumar SP, Wardi G, Malhotra A, Nemati S. Artificial intelligence sepsis prediction algorithm learns to say "I don't know". NPJ Digit Med. 2021;4(1):134.

Sreenivasan AP, Vaivade A, Noui Y, Khoonsari PE, Burman J, Spjuth O, et al. Conformal prediction enables disease course prediction and allows individualized diagnostic uncertainty in multiple sclerosis. NPJ Digit Med. 2025;8(1):224.

Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. Adv Neural Inf Process Syst. 2017;30.

Angelopoulos AN, Bates S. Conformal prediction: A gentle introduction. Found Trends Mach Learn. 2023;16(4):494-591.

Kath C, Ziel F. Conformal prediction interval estimation and applications to day-ahead and intraday power markets. Int J Forecast. 2021;37(2):777-99.

Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? Adv Neural Inf Process Syst. 2017;30.

Vazquez J, Facelli JC. Conformal prediction in clinical medical sciences. J Healthc Inform Res. 2022;6(3):241-52.

Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med. 2021;4(1):4.

Sun X, Nakashima M, Nguyen C, Chen PH, Tang WW, Kwon D, et al. FairICP: identifying biases and increasing transparency at the point of care in post-implementation clinical decision support using inductive conformal prediction. J Am Med Inform Assoc. 2025;32(8):1299-309.

Bloomingdale P, Mager DE. Machine learning models for the prediction of chemotherapy-induced peripheral neuropathy. Pharm Res. 2019;36(2):35.

Cho BJ, Kim KM, Bilegsaikhan SE, Suh YJ. Machine learning improves the prediction of febrile neutropenia in Korean inpatients undergoing chemotherapy for breast cancer. Sci Rep. 2020;10(1):14803.

Cuplov V, André N. Machine learning approach to forecast chemotherapy-induced haematological toxicities in patients with rhabdomyosarcoma. Cancers (Basel). 2020;12(7):1944.

Venäläinen MS, Heervä E, Hirvonen O, Saraei S, Suomi T, Mikkola T, et al. Improved risk prediction of chemotherapy-induced neutropenia—model development and validation with real-world data. Cancer Med. 2022;11(3):654-63.

Hughes JH, Tong DM, Burns V, Daly B, Razavi P, Boelens JJ, et al. Clinical decision support for chemotherapy-induced neutropenia using a hybrid pharmacodynamic/machine learning model. CPT Pharmacometrics Syst Pharmacol. 2023;12(11):1764-76.

Guo L, Wang W, Xie X, Wang S, Zhang Y. Machine learning for genetic prediction of chemotherapy toxicity in cervical cancer. Biomed Pharmacother. 2023;161:114518.

Cai L, Deutsch TM, Sidey-Gibbons C, Kobel M, Riedel F, Smetanay K, et al. Machine learning to predict the individual risk of treatment-relevant toxicity for patients with breast cancer undergoing neoadjuvant systemic treatment. JCO Clin Cancer Inform. 2024;8:e2400010.

Kim S. Predicting chemotherapy-induced peripheral neuropathy using transformer-based multimodal deep learning. Research. 2025;8:0795.

Barber RF, Candes EJ, Ramdas A, Tibshirani RJ. Predictive inference with the jackknife+. Ann Stat. 2021;49(1):486-507.

Wu Y, Zhao W, Zhang L, Wang Y, Wen Y, Liu L. Machine learning models for predicting chemotherapy-induced adverse drug reactions in colorectal cancer patients. Dig Liver Dis. 2025; (in press, no volume/issue given as per original).

Gibbs I, Candes E. Adaptive conformal inference under distribution shift. Adv Neural Inf Process Syst. 2021;34:1660-72.

Isaksson LJ, Pepa M, Zaffaroni M, Marvaso G, Alterio D, Volpe S, et al. Machine learning-based models for prediction of toxicity outcomes in radiotherapy. Front Oncol. 2020;10:790.

Froicu EM, Oniciuc OM, Afrăsânie VA, Marinca MV, Riondino S, Dumitrescu EA, et al. The use of artificial intelligence in predicting chemotherapy-induced toxicities in metastatic colorectal cancer: a data-driven approach for personalized oncology. Diagnostics (Basel). 2024;14(18):2074.

Choo H, Yoo SY, Moon S, Park M, Lee J, Sung KW, et al. Deep-learning-based personalized prediction of absolute neutrophil count recovery and comparison with clinicians for validation. J Biomed Inform. 2023;137:104268.

Amouheidari A, Alirezaei Z, Rauh S, Hassanpour M. PrACTiC: a predictive algorithm for chemoradiotherapy-induced cytopenia in glioblastoma patients. J Oncol. 2022;2022:1438190.

Deng J, Lou H, Liu L, Zhong M, Wu S, Zeng Y, et al. Machine learning-based prediction model for delayed chemotherapy-induced nausea and vomiting in pediatric cancer: a prospective cohort study. Pediatr Blood Cancer. 2026; (e70123, online ahead of print; no volume/issue given).

Riley RD, Collins GS, Kirton L, Snell KI, Ensor J, Whittle R, et al. Uncertainty of risk estimates from clinical prediction models: rationale, challenges, and approaches. BMJ. 2025;388.

Lu J, Zhao S, Ma W, Shao H, Hu X, Xi Y, et al. Uncertainty-aware pre-trained foundation models for patient risk prediction via Gaussian process. In: Companion Proceedings of the ACM Web Conference 2024. 2024. p. 1162-5.