Designing an Interactive Dashboard for Risk Prediction with Uncertainty: A Dual-Dataset Case Study on Heart Disease and Diabetes

Fnu Gaurav
Northeastern University, The Roux Institute
Portland, Maine, USA
yadav.gaurav2905@gmail.com

Abstract

Uncertainty visualization and explainability are essential to human-centered AI in clinical decision support. This paper presents an interactive dashboard that communicates model risk predictions with uncertainty and explainability across two classic medical datasets: UCI Heart Disease and Pima Indians Diabetes. We trained Logistic Regression and Random Forest models, applied bootstrap-based uncertainty estimation, and visualized probability intervals and quantile dotplots as dual encodings. The dashboard integrates calibrated probabilities, local and global SHAP explanations, and adjustable decision thresholds. Through six representative case studies (high, medium, low risk per dataset), we illustrate how uncertainty awareness changes interpretation and trust in model outputs. The study concludes with reflections on design transparency, ethical communication of uncertainty, and implications for explainable medical AI.

Keywords

Uncertainty visualization, explainable AI, calibration, SHAP, clinical decision support, human-data interaction

1 Motivation

Predictive models in healthcare often present risk as a single number, which can be misleading in high-uncertainty, high-stakes settings such as sepsis diagnosis [Zhang et al. 2024]. Clinicians and patients need more than a point estimate; they need to understand how confident the model is. This project addresses that gap by designing an interactive dashboard that explicitly visualizes uncertainty, calibrated probability, and explanations for each prediction. Building on recent Human-Data Interaction (HDI) research on human–AI collaboration in medical decision making [Zhang et al. 2024], our goal is to promote transparency and interpretability without requiring clinical deployment.

We use two interpretable medical datasets related to Heart Disease and Diabetes to compare how uncertainty behaves across different feature sets. Rather than user studies, this work focuses on creating a well-engineered, reproducible artifact that can ground future research and discussion in uncertainty communication design.

2 Related Work

Prior research on human–AI collaboration for medical decision making highlights challenges of presenting uncertain risk scores and the need for interfaces that support clinicians' reasoning rather than only final predictions [Zhang et al. 2024]. In explainable AI, SHAP values [Lundberg and Lee 2017] have become a standard approach to visualizing local and global feature attributions. Calibration research [Niculescu-Mizil and Caruana 2005] highlights the mismatch between predicted probabilities and real-world outcomes. Our work integrates these streams calibration, explainability, and uncertainty encoding into a unified, interactive system that visualizes model reliability and interpretability together.

3 Datasets

3.1 Heart Disease (Cleveland Subset)

The UCI Heart Disease dataset [Cleveland Clinic Foundation 1988] contains 303 samples and 13 interpretable features, including chest pain type (`cp`), resting blood pressure (`trestbps`), cholesterol (`chol`), and maximum heart rate (`thalach`). The binary target represents presence or absence of heart disease. This small but high-signal dataset supports fast experimentation and clear visual explanations.

3.2 Pima Indians Diabetes

The Pima dataset [National Institute of Diabetes and Digestive and Kidney Diseases 1990] includes 768 female patient records with 8 features (e.g., `Glucose`, `BMI`, `Age`, `Insulin`). The target predicts diabetes onset. Compared to Heart, this dataset shows class imbalance and greater metabolic feature variability, enabling generalization tests across domains.

4 Models

We trained two models per dataset:

- **Logistic Regression (LR)** — interpretable, smooth, and well-suited for calibrated probabilities.
- **Random Forest (RF)** — non-linear, robust to feature interactions, used for comparative calibration and uncertainty behavior.

In the dashboard, LR serves as the primary model for demonstration because its monotonic probability outputs align with clinical interpretability and SHAP explanations.

5 Calibration and Model Quality

We evaluated performance using ROC, Precision–Recall, and reliability curves. Figure 1 shows that the Diabetes model achieves substantially higher discriminative performance (ROC AUC ≈ 0.98, PR AUC ≈ 0.98) than the Heart model (ROC AUC ≈ 0.81, PR AUC ≈ 0.66). This indicates that Pima features provide a clearer separation between classes, whereas the Heart dataset produces more overlapping patterns.

Reliability diagrams (Figure 2) highlight important calibration differences. The Heart model shows underconfidence at low predicted probabilities and inconsistent calibration in the mid-range, with points fluctuating above and below the diagonal. This reflects a more complex decision boundary and limited data.

In contrast, the Diabetes model is close to well-calibrated overall, with predicted probabilities tracking the diagonal more closely. The model is slightly overconfident for high-probability predictions (0.6–0.9 bins), where observed positivity reaches 1.0, but remains substantially better calibrated than the Heart model.

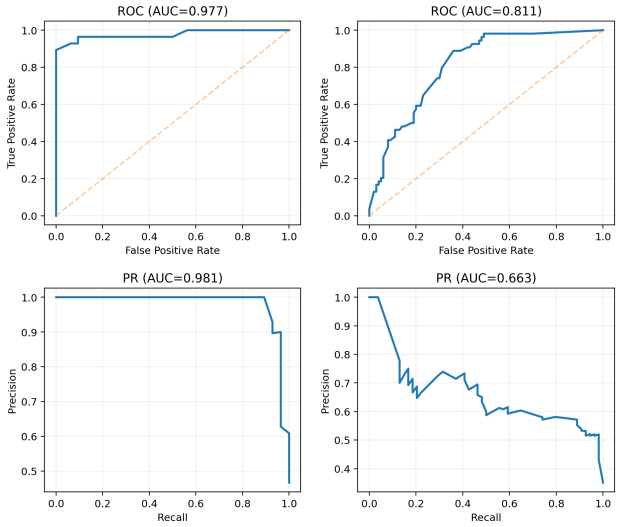

Figure 1: ROC and PR Curves for Heart (left) and Diabetes (right) datasets.

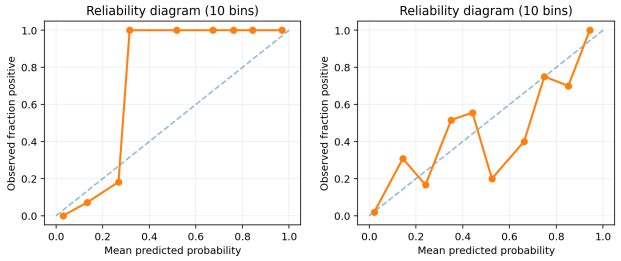

Figure 2: Reliability (calibration) curves. Heart model slightly underconfident; Diabetes well-aligned.

6 Uncertainty Estimation

To quantify predictive uncertainty, we used bootstrap resampling (B=100) [Efron and Tibshirani 1993]. Each model was refit on sampled subsets of the training data, producing 100 predicted probabilities per patient. We computed the median and 95% interval to capture both epistemic and aleatoric components.

Two uncertainty encodings were designed:

- **Encoding A: Interval Band.** Displays median probability and a 95% confidence interval. Narrow bands indicate model certainty.
- **Encoding B: Quantile Dotplot.** Represents the full bootstrap distribution as discrete quantile dots, revealing skew and disagreement across resamples.

These designs align with HDI and visualization principles, offering complementary trade-offs between clarity and richness.

7 Global Explainability (SHAP)

Global SHAP summaries reveal which features most strongly drive model predictions (Figure 3). For the Heart Disease dataset, features such as `ca`, `cp`, and `oldpeak` dominate: `ca` represents the number of major vessels (0–3) colored by fluoroscopy, `cp` denotes chest pain type (ranging from typical angina to asymptomatic), and `oldpeak` measures ST depression induced by exercise relative to rest—a marker of cardiac stress. Higher values of these features generally correspond to elevated heart disease risk. In contrast, for the Diabetes dataset, `Glucose` and `BMI` are most influential, reflecting well established clinical associations between blood glucose levels, body mass index, and diabetes onset. These results align with both domain intuition and prior medical literature, supporting the interpretability and face validity of the models.

Which features matter overall? (Global SHAP)

	feature	mean_abs_shap
10	ca_0	0.687
3	cp_3	0.4867
22	sex	0.4648
21	oldpeak	0.4389
20	thalach	0.3895
16	thal_2	0.3519
24	exang	0.3474
12	ca_2	0.2636
8	slope_1	0.2417
18	trestbps	0.2049

Which features matter overall? (Global SHAP)

	feature	mean_abs_shap
1	Glucose	0.9389
5	BMI	0.5748
0	Pregnancies	0.3079
6	DiabetesPedigreeFunction	0.2197
7	Age	0.151
4	Insulin	0.0241
2	BloodPressure	0.0114
3	SkinThickness	0.0095

Figure 3: Global SHAP feature importance for Heart (top) and Diabetes (bottom).

8 Case Studies

We illustrate six representative cases, each showing predicted probability, uncertainty interval, and local SHAP explanation.

8.1 Heart Disease: High Risk

A 51-year-old male (`cp`=3, `chol`=298, `oldpeak`=4.2) received a predicted disease probability of **0.971** (median **0.993**, 95% interval

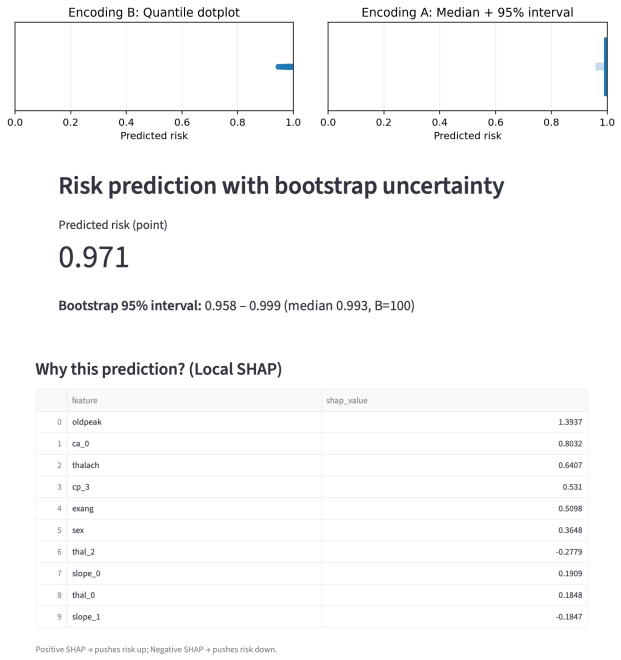

Figure 4: High-risk Heart Disease case: uncertainty encodings and local SHAP explanation.

[0.958, 0.999]). Both the interval band and quantile dotplot show a tight cluster near 1.0, indicating high model confidence and minimal epistemic uncertainty. Local SHAP analysis highlights `oldpeak`, `ca_0`, and `thalach` as the strongest contributors, all consistent with clinical patterns of exercise-induced cardiac stress. Within the dashboard, this case appears with almost no divergence across bootstrap samples, making the system's confidence immediately visible to the user.

8.2 Heart Disease: Medium Risk

A 46-year-old male (`cp`=3, `chol`=249, `oldpeak`=0.8) received a predicted disease probability of **0.353** (median **0.402**, 95% interval [0.175, 0.612]). Both the interval band and quantile dotplot display substantial spread, indicating high epistemic uncertainty and considerable model disagreement. Local SHAP analysis reveals competing influences: `cp_3` and `sex` increase risk, while `ca_0`, `thal_2`, and `exang` exert protective effects. This combination of mixed feature contributions aligns with the wide uncertainty range and reflects an ambiguous clinical profile where small variations in input can shift the prediction. In the dashboard, this case appears as a dispersed cloud of bootstrap samples, making its uncertainty immediately apparent to users.

8.3 Heart Disease: Low Risk

A 58-year-old female (`cp`=2, `thalach`=172, `oldpeak`=0.0) received a predicted disease probability of **0.038** (median **0.024**, 95% interval [0.002, 0.108]). The interval band and dotplot both cluster tightly near zero, indicating high model confidence in a low-risk

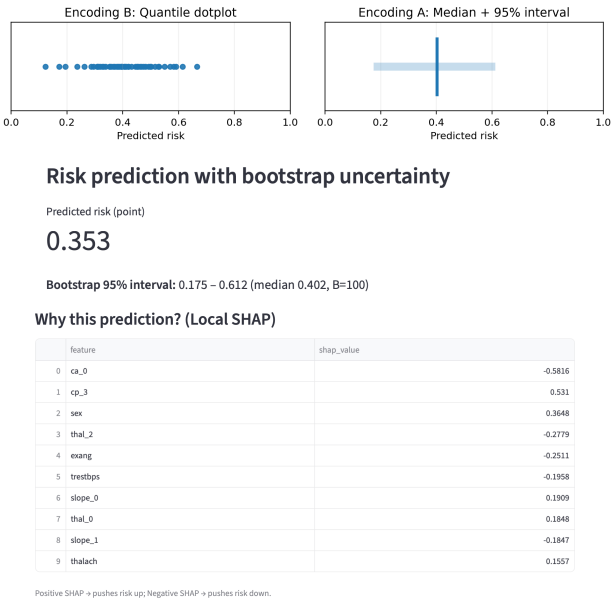

Figure 5: Medium-risk Heart Disease case: uncertainty encodings and local SHAP explanation.

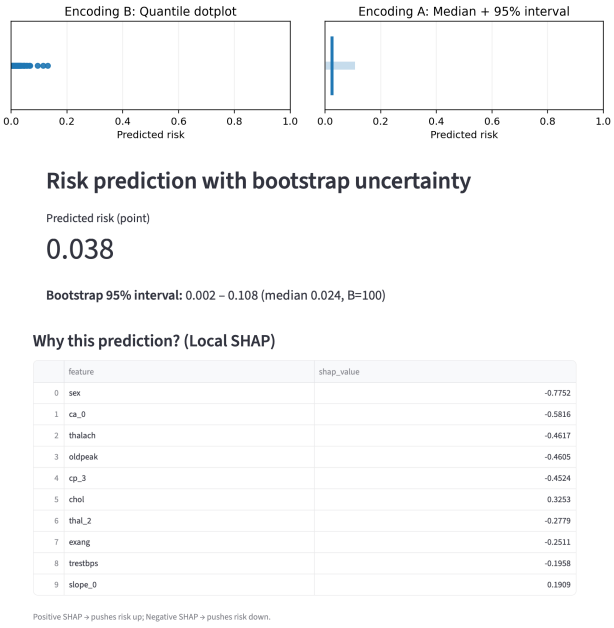

Figure 6: Low-risk Heart Disease case: uncertainty encodings and local SHAP explanation.

assessment. Local SHAP analysis shows that being female, having no major blocked vessels (`ca_0`), a high maximum heart rate (`thalach`), and no ST-segment depression (`oldpeak`=0) are the dominant protective factors. In the dashboard, this case appears with minimal dispersion across bootstrap samples, clearly signaling a stable low-risk profile to the user.

8.4 Diabetes: High Risk

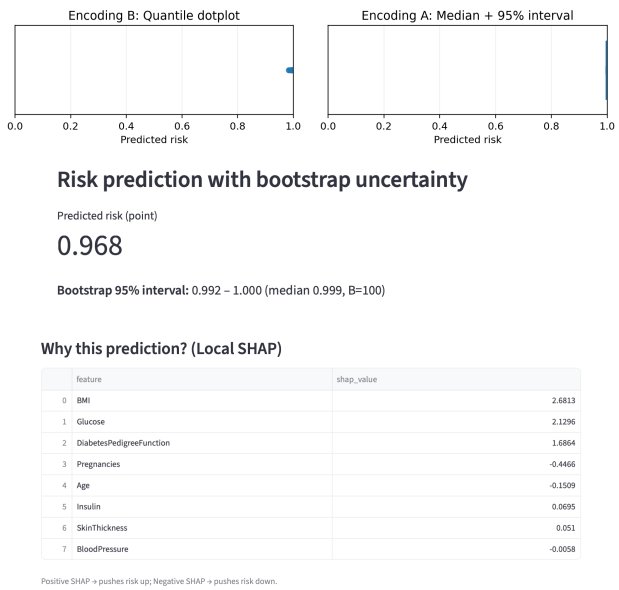

Figure 7: High-risk Diabetes case: uncertainty encodings and local SHAP explanation.

A 25-year-old patient (Glucose=180, BMI=59.4, DiabetesPedigreeFunction=2.42) received a predicted diabetes risk of **0.968** (median **0.999**, 95% interval [0.992, 1.000]). Both the interval band and quantile dotplot show an extremely tight cluster at the upper boundary, indicating very high model confidence. Local SHAP values highlight `BMI`, `Glucose`, and `DiabetesPedigreeFunction` as the dominant contributors, each exerting a strong positive influence consistent with clinical expectations. The absence of competing negative factors results in minimal uncertainty, and in the dashboard this case appears as a solid, undispersed block of bootstrap samples, visually signaling a stable high-risk profile.

8.5 Diabetes: Medium Risk

A 45-year-old patient (`BMI`=43.3, `Pregnancies`=8, `Glucose`=105) received a predicted diabetes risk of **0.519** (median **0.521**, 95% interval [0.386, 0.738]). The interval band and quantile dotplot show broad dispersion, indicating substantial epistemic uncertainty and strong model disagreement. Local SHAP values reveal competing influences: high BMI and a large number of pregnancies push risk upward, while moderate glucose and a low hereditary score exert protective effects. This combination of opposing factors creates an ambiguous clinical profile, naturally leading to a wide bootstrap distribution. In the dashboard, this case appears as a widely scattered dotplot, helping users recognize the model's uncertainty in borderline predictions.

8.6 Diabetes: Low Risk

A 21-year-old patient (Glucose=111, BMI=22.6, DiabetesPedigreeFunction=0.142) received a predicted diabetes risk of **0.000** (median

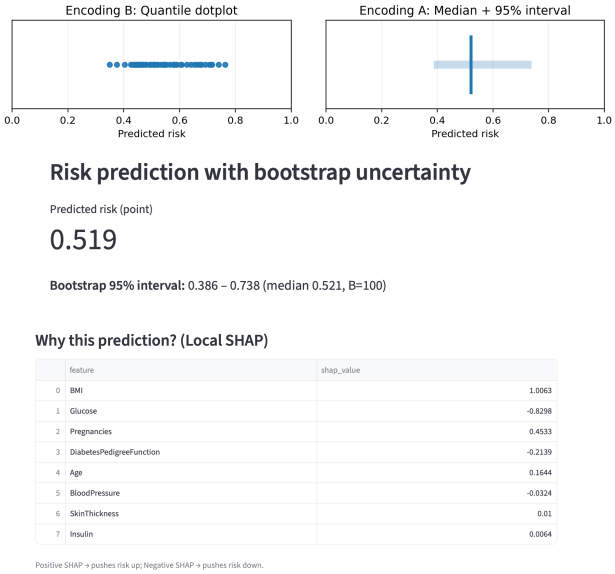

Figure 8: Medium-risk Diabetes case: uncertainty encodings and local SHAP explanation.

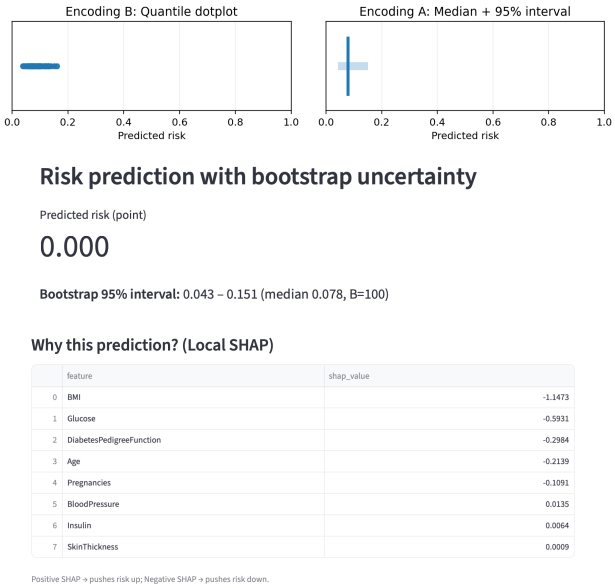

Figure 9: Low-risk Diabetes case: uncertainty encodings and local SHAP explanation.

0.078, 95% interval [0.043, 0.151]). Both the interval band and quantile dotplot concentrate near zero, indicating a confident low-risk assessment. Local SHAP values show that a healthy BMI, normal glucose levels, and low hereditary risk are the dominant protective factors. Minor contributions from other variables do not counteract these strong negative influences, resulting in a stable low-risk prediction. In the dashboard, this case appears as a compact cluster

of bootstrap samples near zero, visually reinforcing the model's confidence.

9 Design Reflections

Three insights emerged:

(1) **Uncertainty communication:** Intervals are intuitive for experts, while dotplots reveal richer nuance for ambiguous cases.
(2) **Transparency vs. overload:** Presenting uncertainty alongside SHAP explanations helps users contextualize confidence, but excessive visual detail can overwhelm.
(3) **Ethical implications:** Calibrated probabilities and honest uncertainty prevent over-trust in single-number risk estimates, echoing concerns from sepsis decision-support systems about overreliance on opaque alerts [Zhang et al. 2024]. Avoiding truncated axes and overly suggestive color schemes is also important to maintain fairness in visual communication. Bootstrap uncertainty may amplify underrepresented subgroup variability, reinforcing the need for fairness-aware interpretation.Beyond visual clarity, uncertainty estimates can also surface fairness concerns. Bootstrap variability tends to be higher for under-represented subgroups, meaning models may appear more stable for majority populations while silently amplifying uncertainty for minority groups. Making this variability visible is essential for equitable interpretation of clinical risk predictions.

10 Limitations

This work uses small, static datasets; uncertainty estimates may be noisy. Logistic regression assumes linear separability. We did not conduct a user study. Future work will integrate Bayesian models, MC dropout, and longitudinal data.

11 Conclusion and Future Work

We built a reproducible dashboard demonstrating how calibrated models, bootstrap uncertainty, and SHAP explanations can jointly improve trust in medical AI. Across two datasets, we observed distinct uncertainty behaviors reflecting data structure and model calibration. In addition, a natural next step is to study how different uncertainty encodings (e.g., interval bands versus quantile dotplots) influence users' confidence calibration and decision-making. Comparing how clinicians and non-experts interpret these visual encodings would extend this work toward human-centered evaluation of uncertainty communication, a key direction in medical HDI research.

Acknowledgments

The author thanks Professor Melanie Tory for her inspiring work on uncertainty communication, and the Human-Data Interaction Lab at Northeastern University for shaping this design perspective.

References

Cleveland Clinic Foundation. 1988. UCI Heart Disease Dataset. UCI Machine Learning Repository.

Bradley Efron and Robert Tibshirani. 1993. *An Introduction to the Bootstrap.* Chapman and Hall.

Scott Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems.*

National Institute of Diabetes and Digestive and Kidney Diseases. 1990. Pima Indians Diabetes Dataset. UCI Machine Learning Repository.

Alexandru Niculescu-Mizil and Rich Caruana. 2005. Predicting good probabilities with supervised learning. In *Proceedings of the International Conference on Machine Learning.*

Shao Zhang, Jianing Yu, Xuhai Xu, Changchang Yin, Yuxuan Lu, Bingsheng Yao, Melanie Tory, Lace M Padilla, Jeffrey Caterino, Ping Zhang, and Dakuo Wang. 2024. Rethinking human-AI collaboration in complex medical decision making: A case study in sepsis diagnosis. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.*