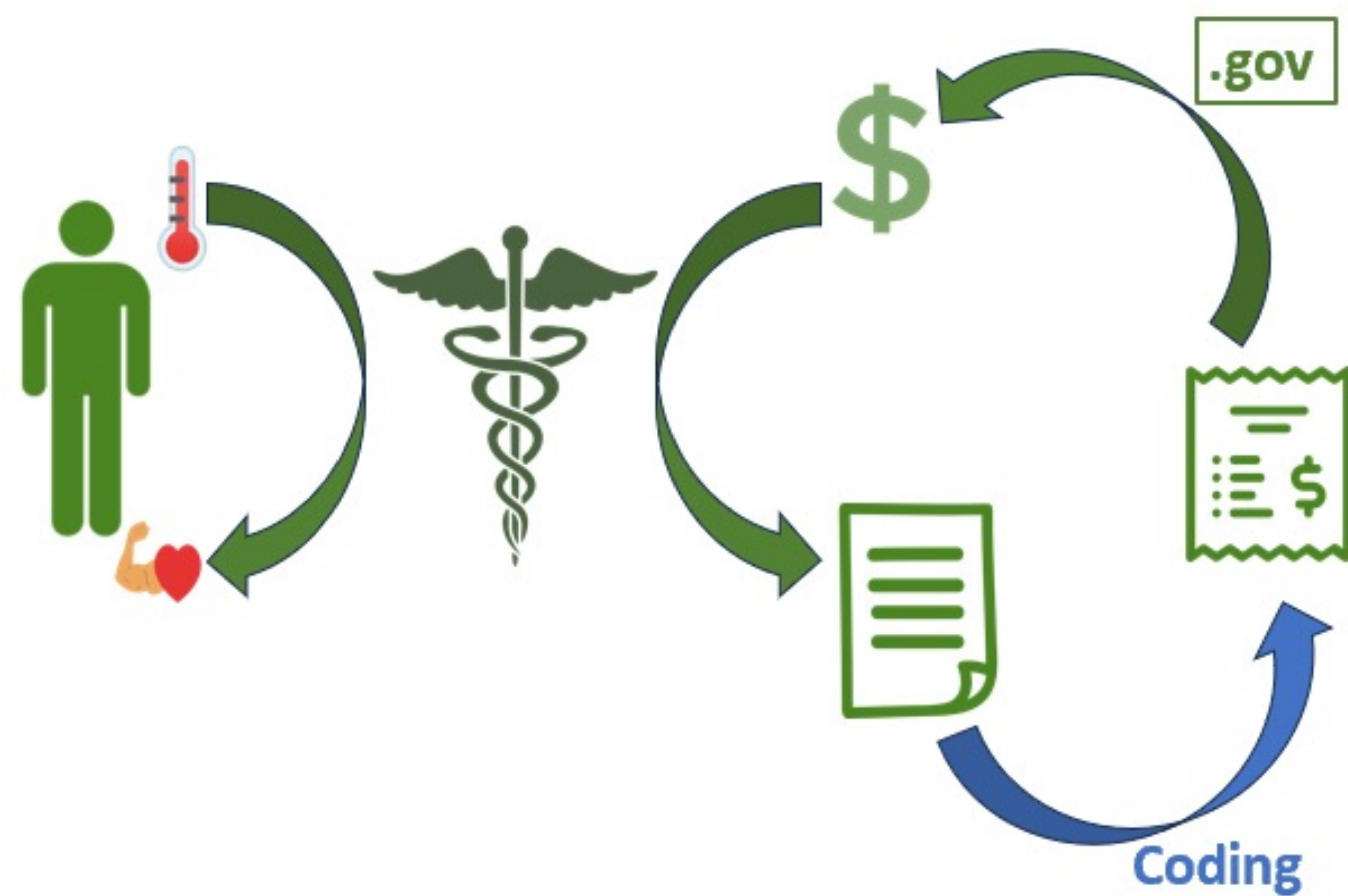




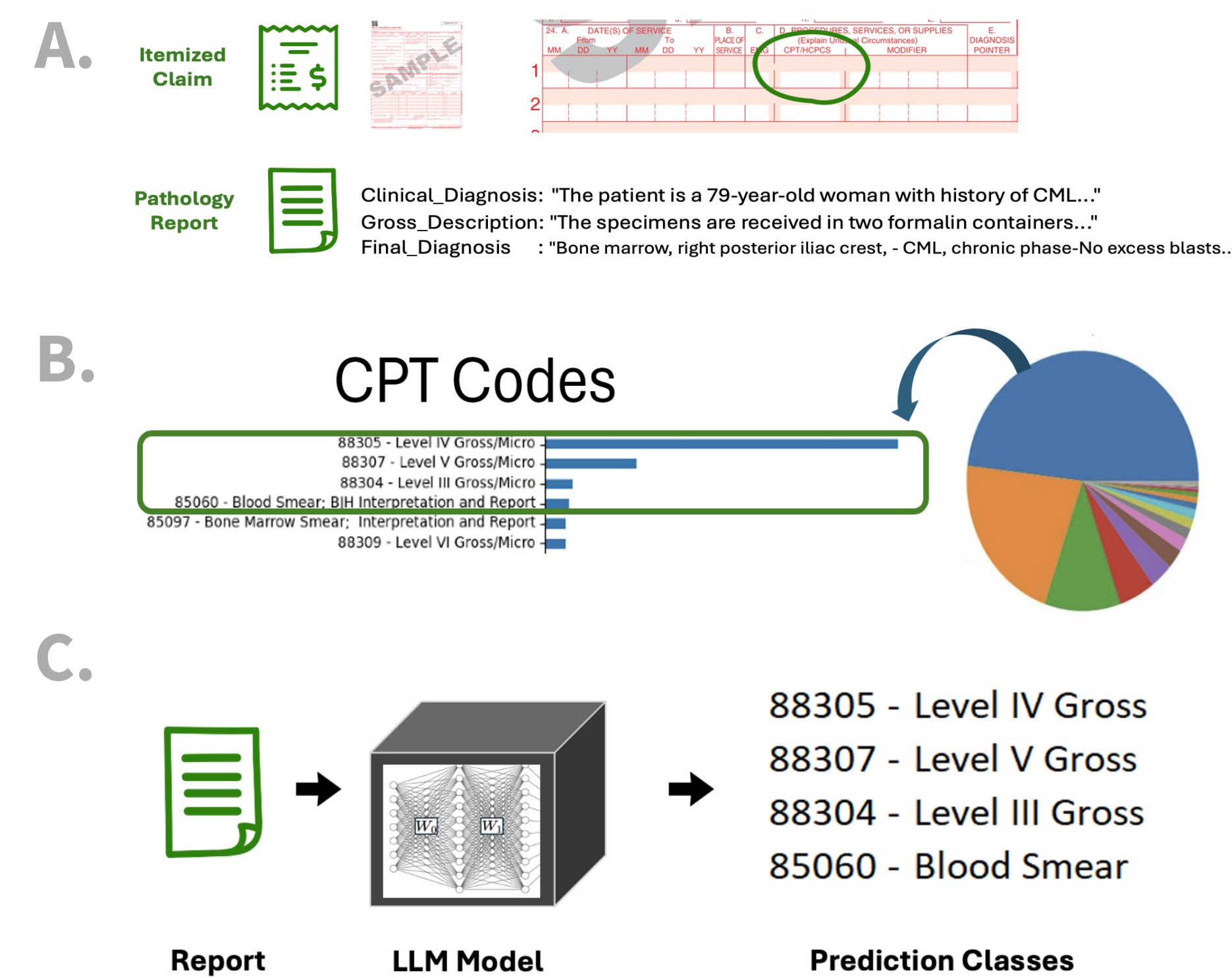
Conformal Prediction and Large Language Model for Medical Coding

Christopher Snyder, MD PhD
Department of Pathology & Immunology
Washington University in St. Louis School of Medicine, St. Louis, MO

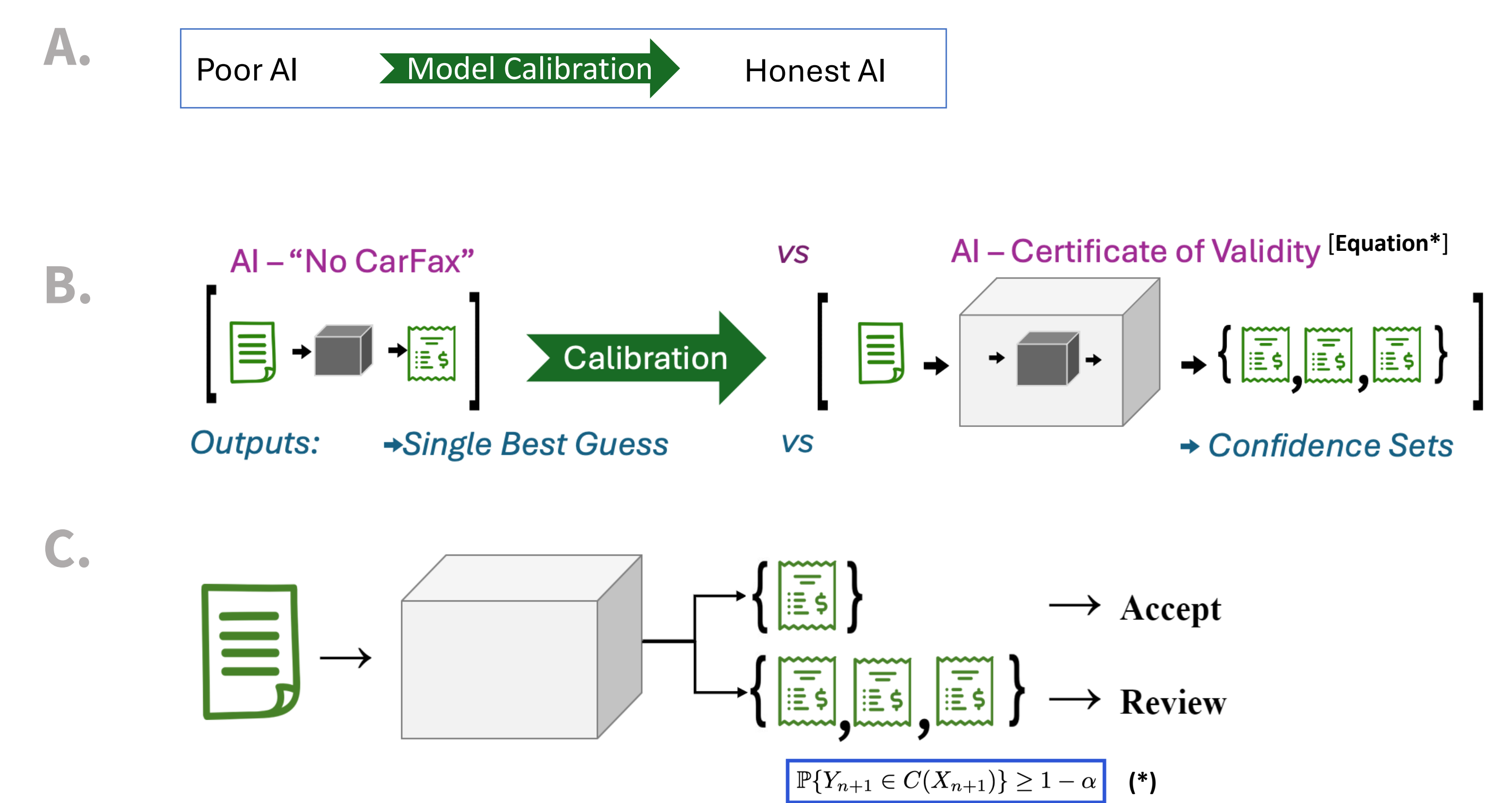
- Large Language Models (LLMs) promise efficiencies, but many applications are high-stakes, such as CPT procedure coding from Pathology reports.
- We introduce a calibration procedure called “Conformal Prediction (CP)” that measures experimental uncertainty in the classification process.
- We use holdout data to adapt a 95% accurate LLM CPT coder into a selective 99% accurate coder (without modifying any weights).



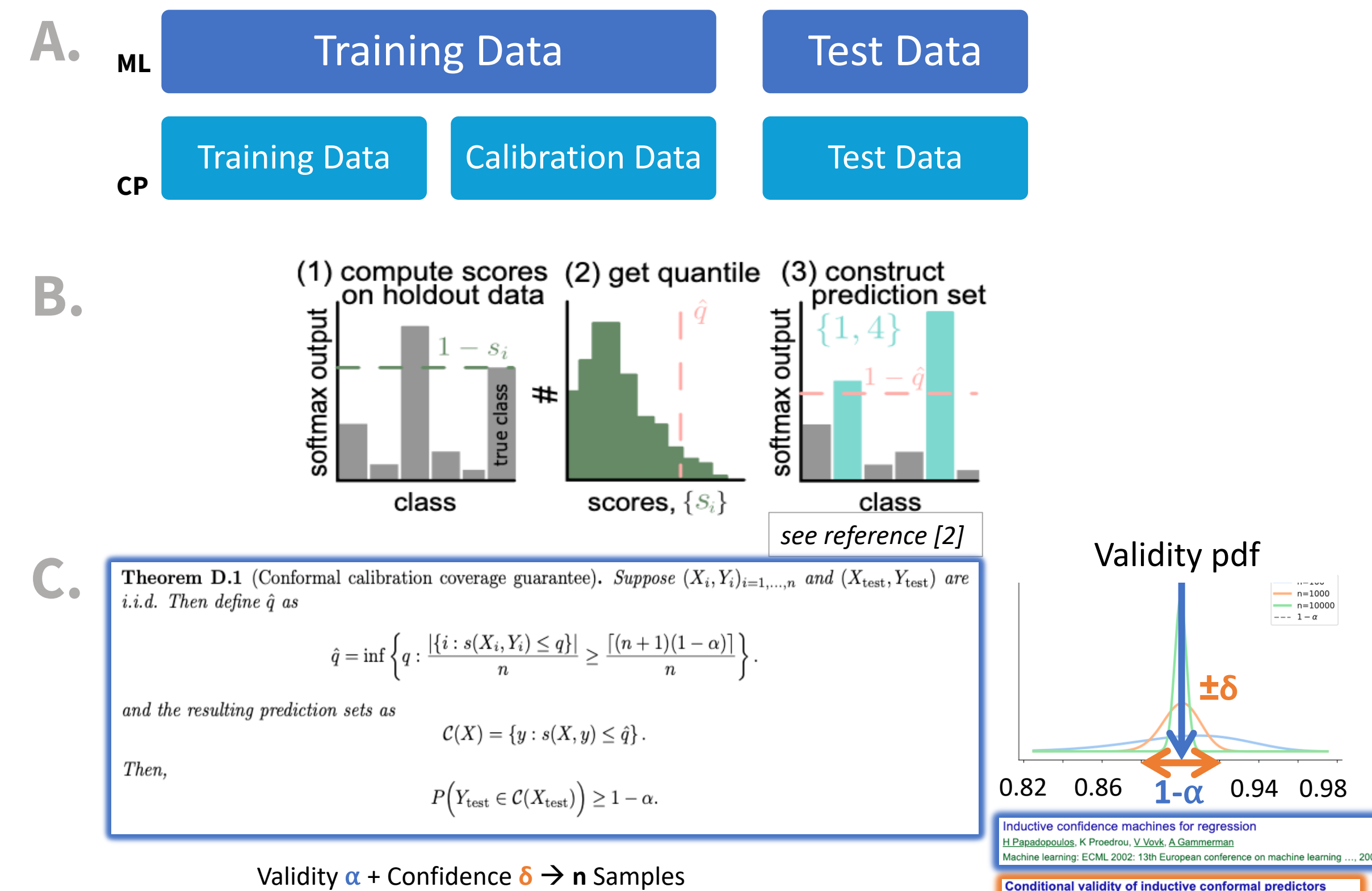
1 Large Language Models: as Medical Coders^[1]



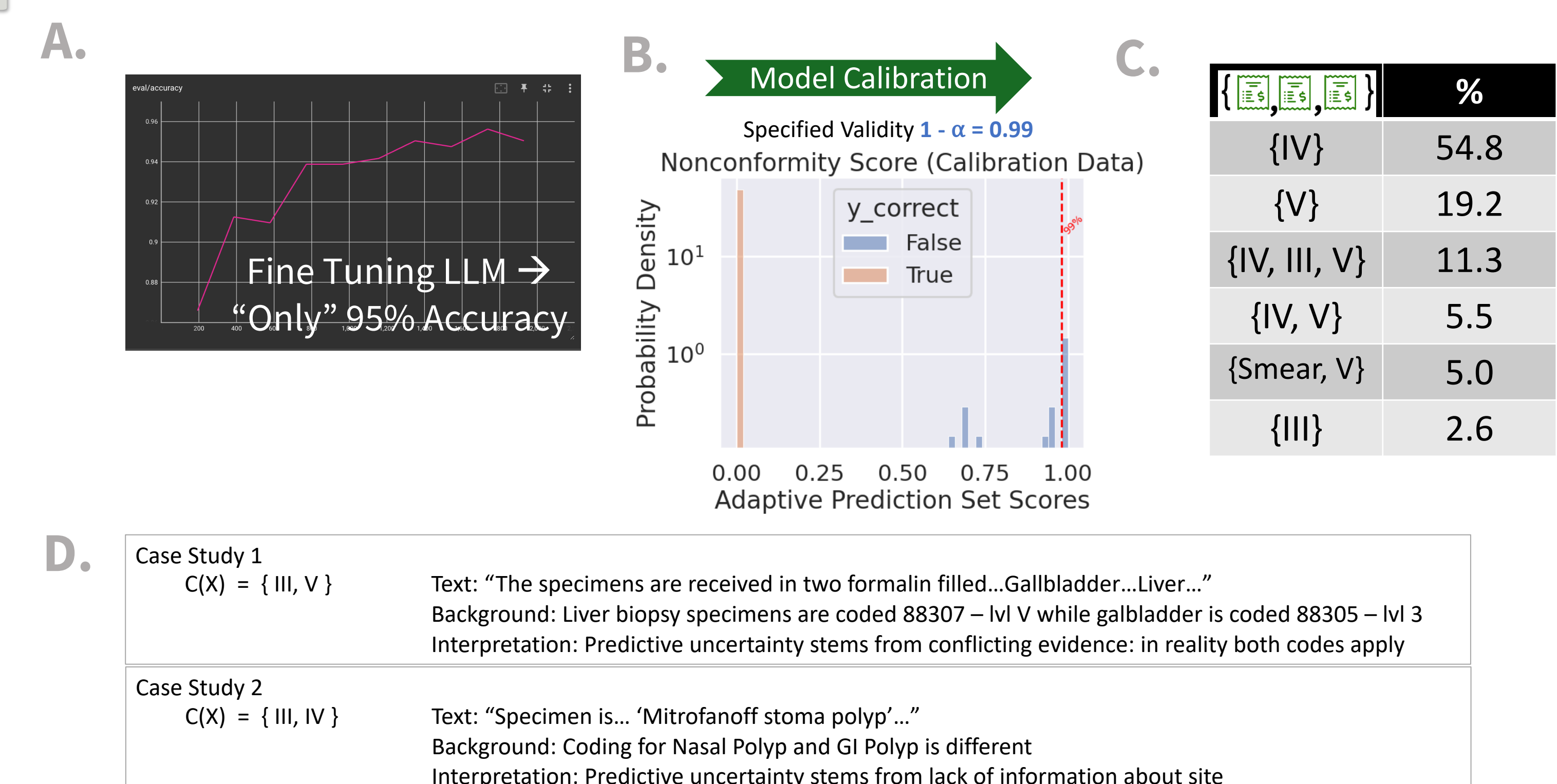
2 Big Picture of Conformal Prediction: Creating statistically rigorous uncertainty sets.



3 Theory of Conformal Prediction



4 Putting it together: Conformal Coding with an LLM



Abstract

Through fine tuning a language model on pathology report text alone, we can achieve 95% prediction accuracy of the top 5 most common CPT codes on our dataset. We also show the utility of conformal prediction, which allows us to raise our accuracy to 99.0% when we allow the model to abstain on making a prediction on 24% of the data. This is a value that is determined by a separate threshold on the scalar value of the predictor on the aforementioned validation set.

1. Allowing models to abstain allows differential automation of easy tasks.
2. LLMs with CP are a promising salve for the complexity of CPT coding.
3. No threshold effect of AI performance on AI utility in applications

[1] Ali Soroush et al. Large Language Models are Poor Medical Coders. NEJM 2024.

[2] Angelopoulos and Bates: Conformal Prediction – A Gentle Introduction. 2023.