Wise Instance Selection Algorithms Help Reduce Annotation Work in Multi-Task Multi-Class Urinary Tract Dilation Prediction

Yining Hua[1,2], MSc; Michael Lingzhi Li[1,3*], PhD; Hsin-Hsiao Scott Wang[1*], MD, MPH, MBAn*

[1]Department of Urology, Boston Children's Hospital; [2]Department of Epidemiology, Harvard T.H. Chan School of Public Health; [3]Harvard Business School;

Introduction

We explore methods to optimize annotation work efficiency for developing real-life clinical prediction models on the task of predicting urinary tract dilation (UTD) system classification using infant hydronephrosis ultrasound reports.

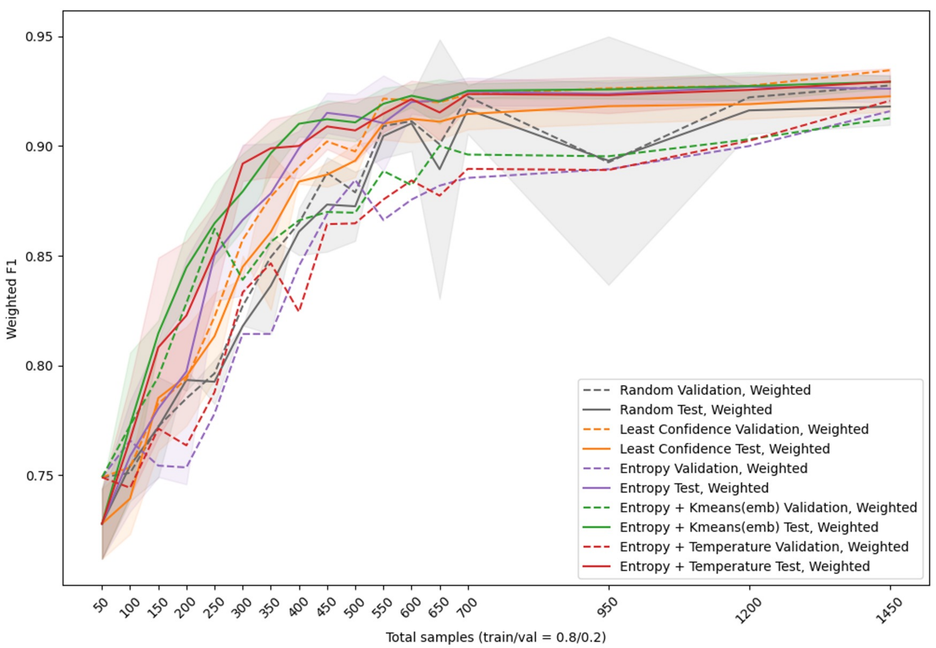
Methods

A **retrospective analysis** of Boston Children's Hospital urology radiology records identified a cohort of infants (n = 2479) aged 0-90 days with early ultrasounds. The research team analyzed the report and images to label **11 UTD classification outcomes** as i) *absence* or ii) *presence* of the UTD feature, or iii) *missing kidney*.

We study previously published and novel algorithms to aid in instance selection for training **multitask and multiclass prediction** models in a **semi-supervised learning** manner. We design (1) **Temperature Scaling + Entropy**: An optimal temperature parameter is calculated to scale the logits of the model after each round of fine-tuning using the logits and actual labels in the validation set. This temperature is used to scale prediction logits in the next iteration. The rest follows the Entropy method. (2) **K-Means + Entropy**: Unlabeled data points are transformed into feature vectors. The model computes their entropy score based on their task logits for each data point. K-Means clustering partitions the data into k clusters. Within each cluster, instances are ranked by their entropy scores. Top n/k instances with the highest entropy are chosen from each cluster for model retraining. In addition, we implement (3) **Least Confidence**, and (4) **Entropy** in a multitask and multiclass setting.

Experiment Settings We use the ClinicalBERT encoder (the classifier adds 11 linear layers to the encoder) and calculate the mean performance across ten runs for each task. For each experiment, 50 instances were selected per iteration for the first 14 iterations and 250 instances for the subsequent iterations. We compare the algorithms to (5) **Random sampling**.

Results



Entropy-based instance selection algorithms help **reduce** substantial **annotation workload** for Urinary Tract Dilation Prediction.

Scan to
read the abstract