On Subset Selection of Multiple Humans To Improve Human-AI Team Accuracy

Sagalpreet Singh, Shweta Jain, Shashi Shekhar Jha
Indian Institute of Technology Ropar

Introduction

Classification tasks are often hindered by the limitations of both human and AI models, leading to inaccuracies. Our work explores the use of multiple human labels in combination with the AI model's output, drawing inspiration from the crowdsourcing literature, which combines labels from multiple humans.

Our key **contributions** are:

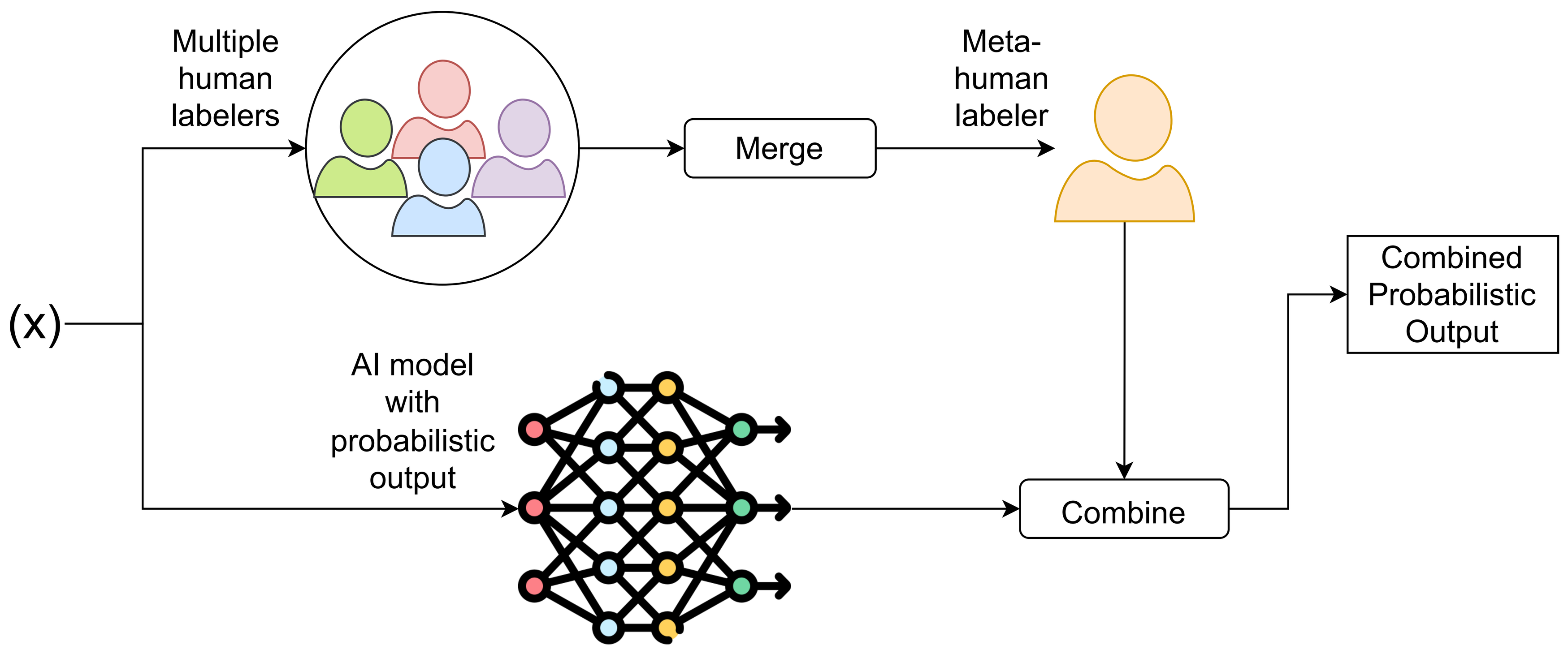
1. Approach to combine the predicted class labels from multiple humans with the probabilistic output of an AI model
2. Empirical validation of our approach on the CIFAR-10H image classification dataset
3. Efficient algorithm to select a subset of humans whose class-level outputs combined with probabilistic output of the AI model leads to maximum accuracy

Notation

Notation	Definition
k or $\mathcal{K}$	number of classes
$\mathcal{Y} = \{1, \ldots, k\}$	set of labels
x	input image
n	number of humans
$c(x)$	prediction made by the combined model
$\varphi^{[i]}$	confusion matrix corresponding to h_i

Notation	Definition
$\mathcal{M}$	AI model
$y(x)$	true label for x
$m(x)$	output of $\mathcal{M}$
h_i	i^{th} human
$h_i(x)$	hard label predicted by h_i for x
$h(x)$	collection $\{h_1(x), h_2(x), \ldots, h_n(x)\}$

Single Label from Multiple Humans



The following strategies have been tested for getting the meta-human labeler:

Best Human: The most accurate human labeler

Best Majority Human: The most accurate human whose predicted class is same as the majority prediction label

Best Weighted-Majority Human: The most accurate human whose predicted class is same as the accuracy-weighted majority prediction label

ComHAI: Combining Humans and AI

We can generalize the combination method from [1] as follows:

$$p(y(x)=j|h(x)=\{l_1,l_2,\ldots,l_n\},m(x)) = \frac{m_j(x)\prod_{i\in[n]}\varphi^{[i]}_{l_ij}}{\sum_{k=1}^{\mathcal{K}} m_k(x)\prod_{i\in[n]}\varphi^{[i]}_{l_ik}} \quad (1)$$

Lower bound on accuracy of the combined prediction is given as follows:

$$\mathbb{E}[\mathbb{1}(c(x)=y(x))] \ge \mathbb{P}\left\{\prod_{i\in[n]}\frac{\varphi^{[i]}_{h_i(x)y(x)}}{1-\varphi^{[i]}_{h_i(x)y(x)}} > \frac{1-m^\theta_{y(x)}(x)}{m^\theta_{y(x)}(x)}\right\} \quad (2)$$

Instead of combining predictions from all humans, we can generalize 1 and 2 by allowing combination from a subset of humans as follows:

$$\mathbb{E}[\mathbb{1}(c(x)=y(x))] \ge \mathbb{P}\left\{\prod_{i\in S}\frac{\varphi^{[i]}_{h_i(x)y(x)}}{1-\varphi^{[i]}_{h_i(x)y(x)}} > \frac{1-m^\theta_{y(x)}(x)}{m^\theta_{y(x)}(x)}\right\} \quad (3)$$

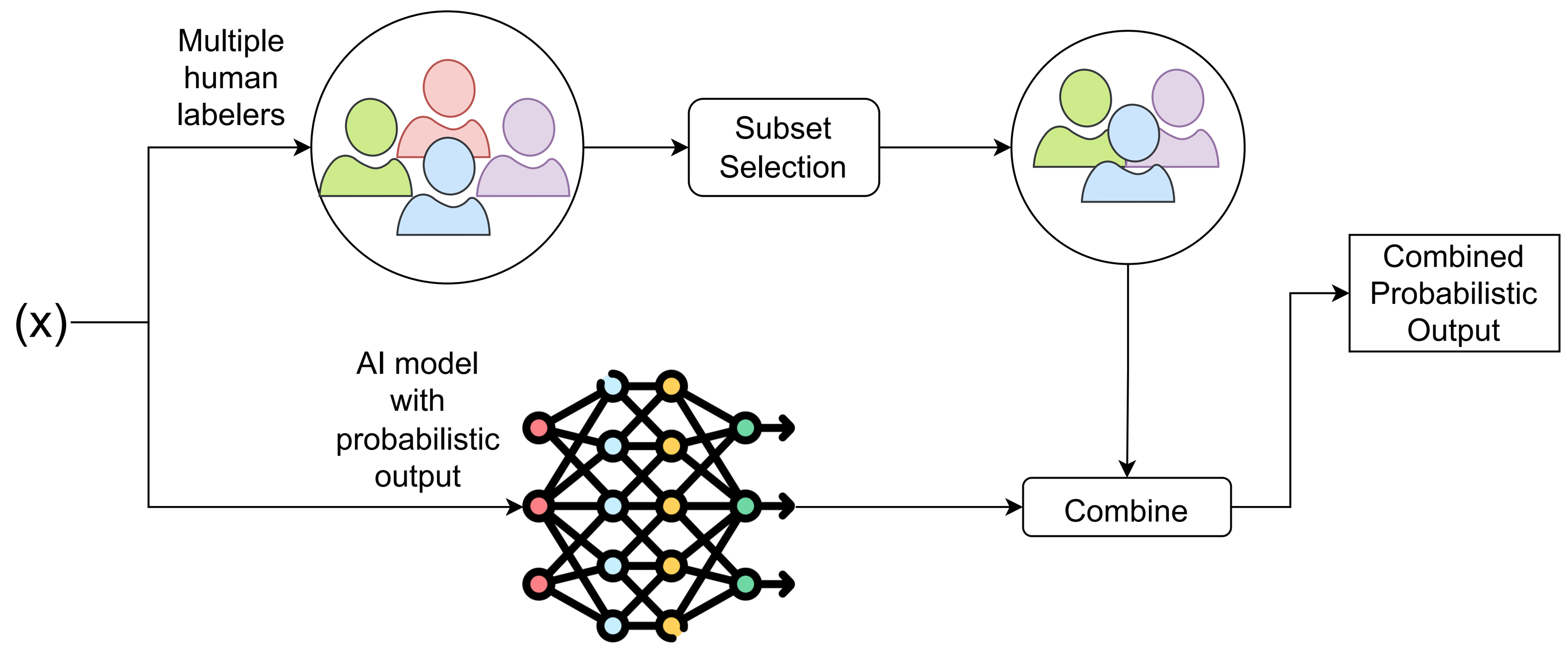
For combination, choose the subset which maximizes the lower bound as follows:

$$S^* = \arg\max_{S}\left(\prod_{i\in S}\frac{\varphi^{[i]}_{h_i(x)y(x)}}{1-\varphi^{[i]}_{h_i(x)y(x)}}\right) \quad (4)$$

Since, $y(x)$ is unknown, pseudo-optimal subset is selected by max-max optimization as follows:

$$S^{**} = \arg\max_{S}\max_{1\le j\le\mathcal{K}}\left(\prod_{i\in S}\frac{\varphi^{[i]}_{h_i(x)j}}{1-\varphi^{[i]}_{h_i(x)j}}\right) \quad (5)$$

Subset Selection & Combination Methods



All Humans Selected: Use predictions from all humans to combine with model probabilities using the Equation 1 with calibrated model probabilities.

Random Subset Selection: Select a subset of humans randomly (a human is selected or not with 0.5 probability) and use predictions from those humans to combine with model probabilities using the Equation 1 with calibrated model probabilities.

True LB Subset Selection: Select the optimal subset of humans using Equation 4. Although this method is impractical (as it requires knowledge of true label), we evaluate this combination method as it maximizes the true lower bound.

Pseudo LB Subset Selection: Select a pseudo-optimal subset of humans using Equation 5 and use their predictions humans to combine with model probabilities.

Algorithm

GreedySubsetSelection: Pseudo LB Subset Selection

Input : $n, \mathcal{K} \in \mathbb{N}$
$\mathcal{K} \times \mathcal{K}$ dimension matrices $\varphi^{[i]}, 1 \le i \le n$
$h_i(x), 1 \le i \le n$

Output: S

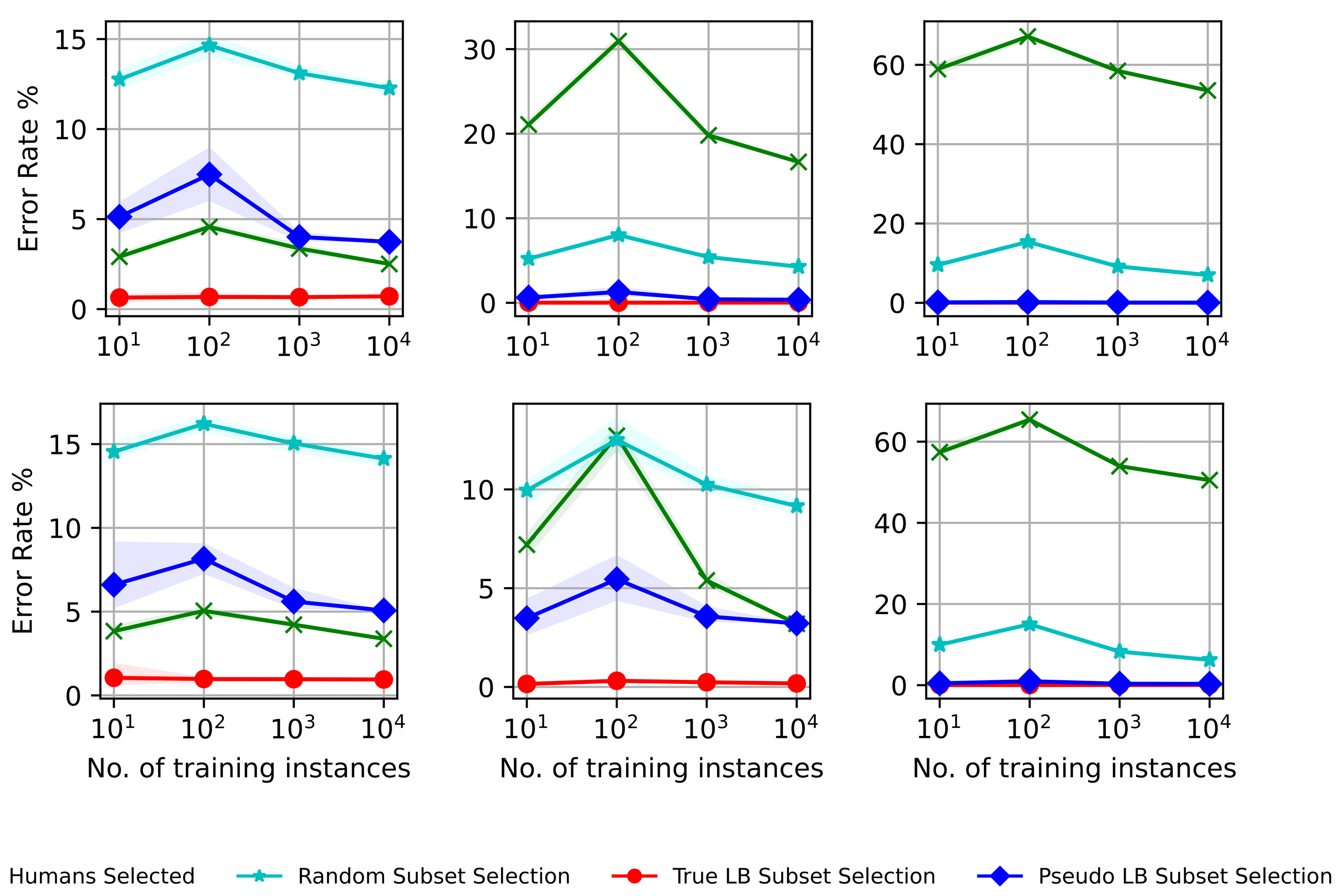
for $i \leftarrow 1$ **to** n **do**
  for $j \leftarrow 1$ **to** $\mathcal{K}$ **do**
    $M[i][j] \leftarrow \frac{\varphi^{[i]}_{h_i(x)j}}{1-\varphi^{[i]}_{h_i(x)j}}$

for $j \leftarrow 1$ **to** $\mathcal{K}$ **do**
  $C[j] \leftarrow 1$
  for $i \leftarrow 1$ **to** n **do**
    if $M[i][j] > 1$ **then**
      $C[j] \leftarrow C[j] \times M[i][j]$;

$y^* \leftarrow \arg\max_{1\le j\le\mathcal{K}} C[j]$
$S^{**} \leftarrow \{1 \le i \le n \mid M[i][y^*] > 1\}$
return S^{**}

Experiments & Results

Learning curves for classification on CIFAR-10h dataset with custom CNN (56.74% accuracy) as the AI model and different set of human labelers



Learning curve on CIFAR-10H using subset selection based methods. The number of humans considered for plots in the first row is 5, 10, and 15, respectively, all having an accuracy 70%. Plots in the second row correspond to 4, 7, and 13 humans, with accuracies ranging from 0.5 to 0.8.

[1] Gavin Kerrigan, Padhraic Smyth, and Mark Steyvers.
Combining human predictions with model probabilities via confusion matrices and calibration.
Advances in Neural Information Processing Systems, 34:4421–4434, 2021.