

# Transfer learning for accurate tissue of origin classification from cfDNA methylation

Shiva Farashahi, Dorna Kashef, Amirali Kia, Esther Brown, Feras Hantash, Kieran I Chacko  
Harbinger Health, Cambridge, MA USA

AACR: Liquid Biopsy November 2024 | Poster B055

## BACKGROUND

Accurately predicting the tissue of origin (TOO) is required for blood-based multi-cancer screening to guide follow-up imaging and care when cancer is detected. However, precisely determining the TOO using cell-free DNA (cfDNA) methylation remains challenging due to low and variable concentrations of circulating tumor DNA and cfDNA and the complex, poorly characterized nature of cancer type-specific methylation patterns. To address these challenges, we developed a transfer learning approach that leverages methylation profiles learned from tissue biopsy samples to predict the TOO of cfDNA cancer samples.

## METHODS

### Samples

Unmatched tissue biopsy (N=517) and blood (N=1,155, NCT05435066) samples were collected from treatment-naïve cancer patients across 21 cancer types consolidated into 17 cancer categories. All samples were analyzed using a custom targeted bisulfite sequencing hybrid capture assay (18.6 Mb). Five unique metrics were used to quantify informative methylation signal at each TOO region of interest. Samples from 10 cancer categories with N>25 were used for training (N=1,041). Results are reported for samples with tumor content greater than 0.1% (N=415) (Table 1).

### Model design

A single-layer, fully connected feed-forward neural network (NN) was initially trained on cancer biopsy methylation data, to extract informative features. Next, weights of the feature-extraction layer was transferred to a two-layers fully connected feed-forward NN. A second layer was trained, and the transferred layer was fine-tuned using plasma cfDNA methylation features (Figure 1).

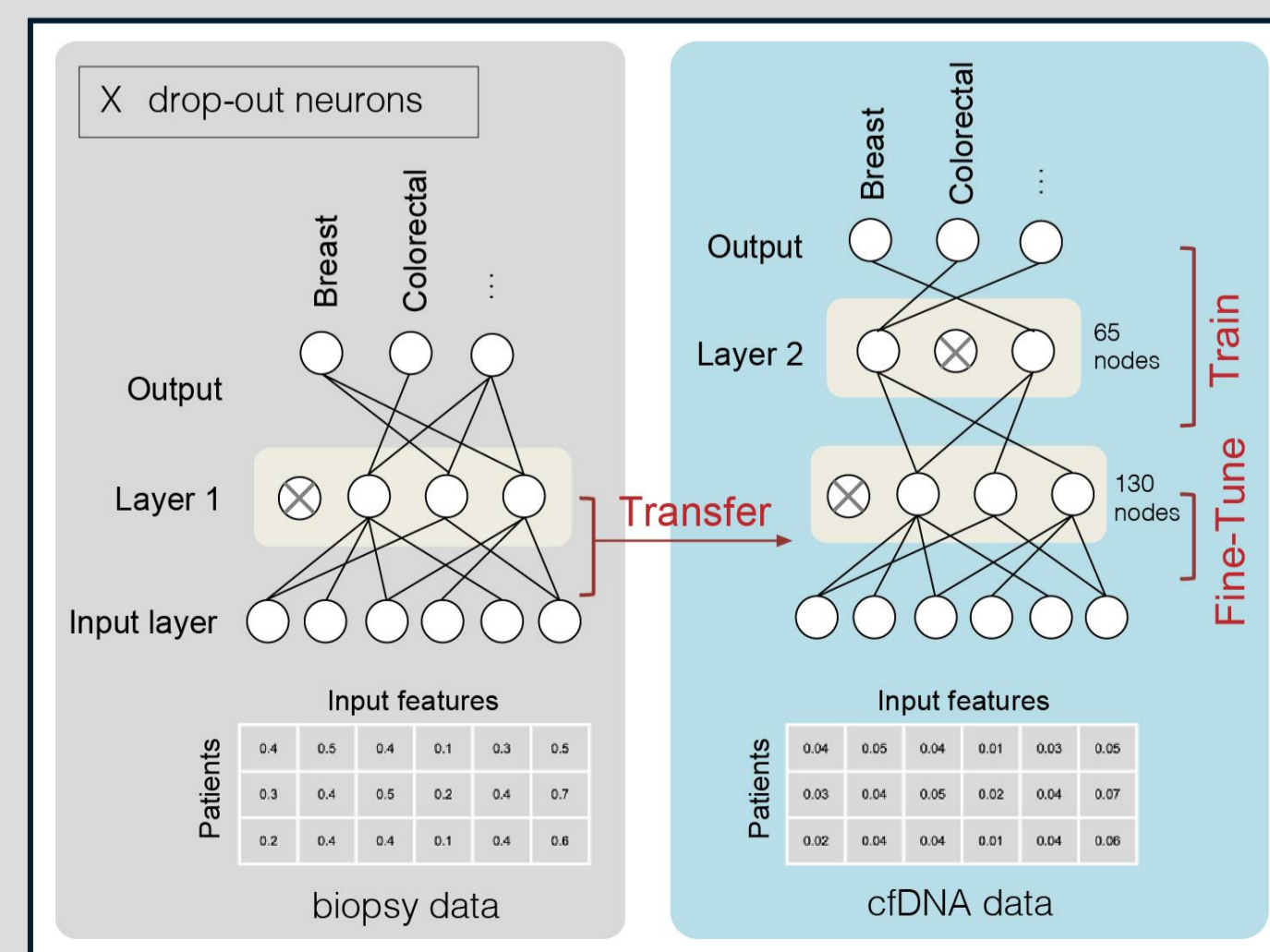


Figure 1. Schematic of transfer learning model.

## RESULTS

### Classification performance

	High tumor content, blood	Train & Test set, Blood	Cancer tissue biopsy
Breast	56	305	64
Colorectal	42	84	30
H&N	56	79	24
Liver,Biliary	21	28	9
Lung	84	14	63
Lymphoid	46	86	30
Panc., GB	54	92	20
Prostate	8	79	30
Upper GI	38	73	61
Uterine	10	70	31

Table 1. Breakdown of samples.

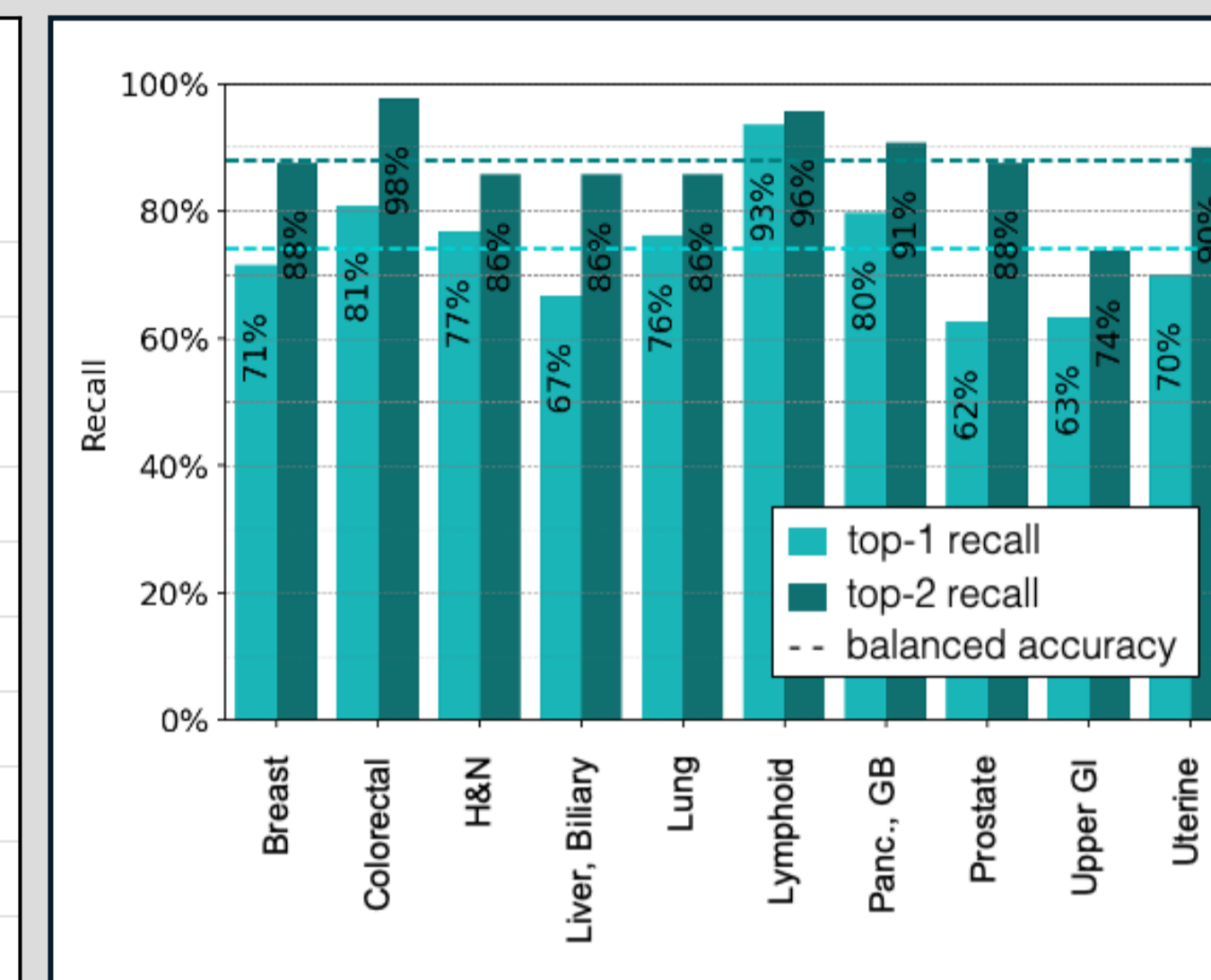


Figure 2. TOO top-1 and top-2 BA and recall.

Our TOO top-1 and top-2 performance was evaluated using 10-fold cross validation on plasma cfDNA samples. We achieved 73% and 89% balanced accuracy (BA) for top-1 and top-2 predictions, respectively (Figure 2). Top-1 prediction is the cancer category with highest probability predicted by the model and top-2 predictions are the cancer categories with highest and second highest probabilities predicted by the model. We observed minimal confusion between most cancer categories expect for Liver + Biliary and Uterine (Figure 3).

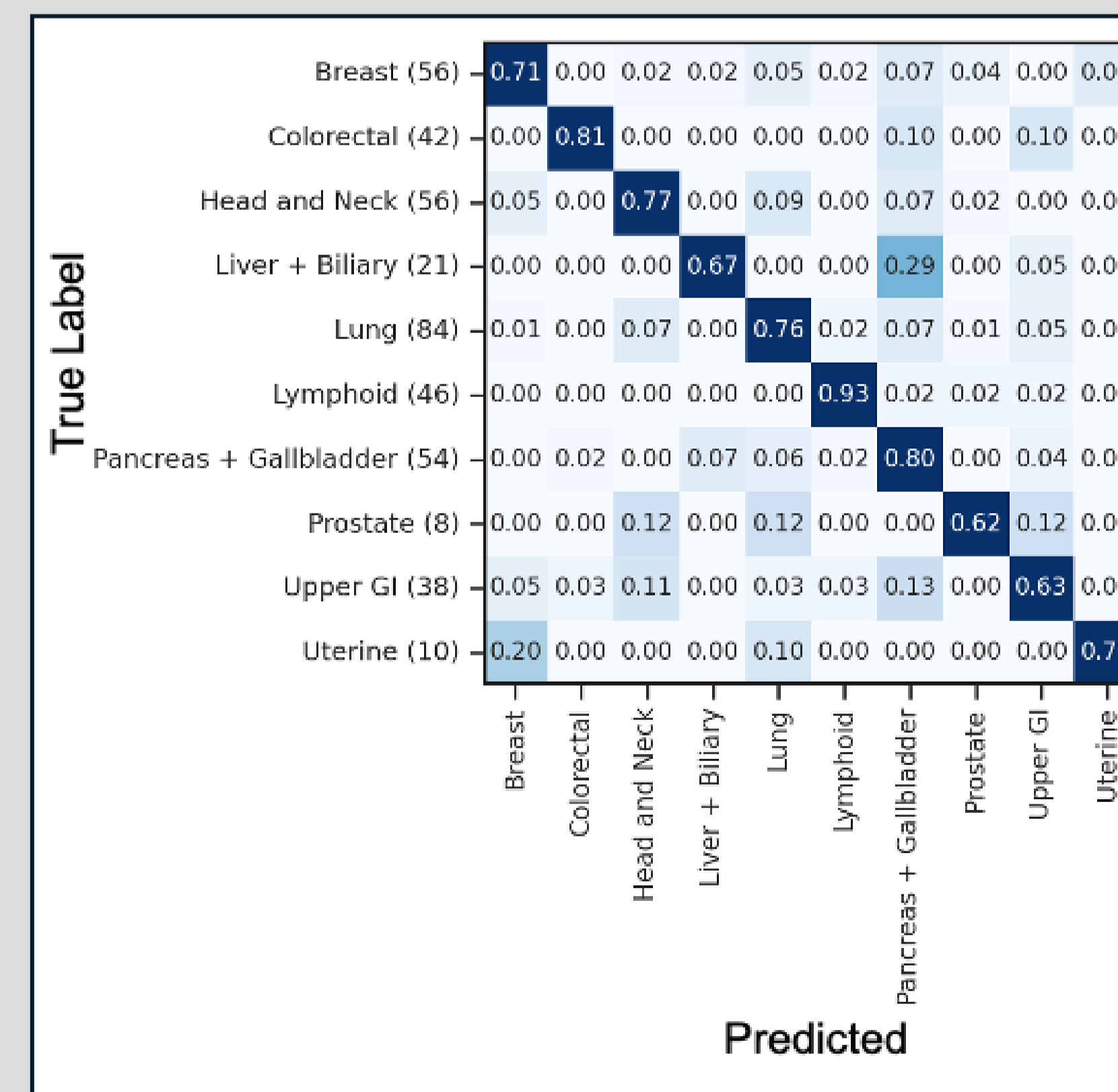


Figure 3. TOO confusion matrix.

## RESULTS

### Evaluating the effect of transfer learning

To quantify the impact of transfer learning, we trained a new model without transferring features learned from cancer tissue biopsy. TOO classification with and without transfer learning achieves 73% and 69% BA for top-1 predictions, respectively. (Figure 4). Accuracy per indication revealed that training on biopsy data significantly benefits the detection of tumor types that shed at low rates; Breast cancer +14% points, Prostate cancer +12% points, and Uterine cancer +20% points.

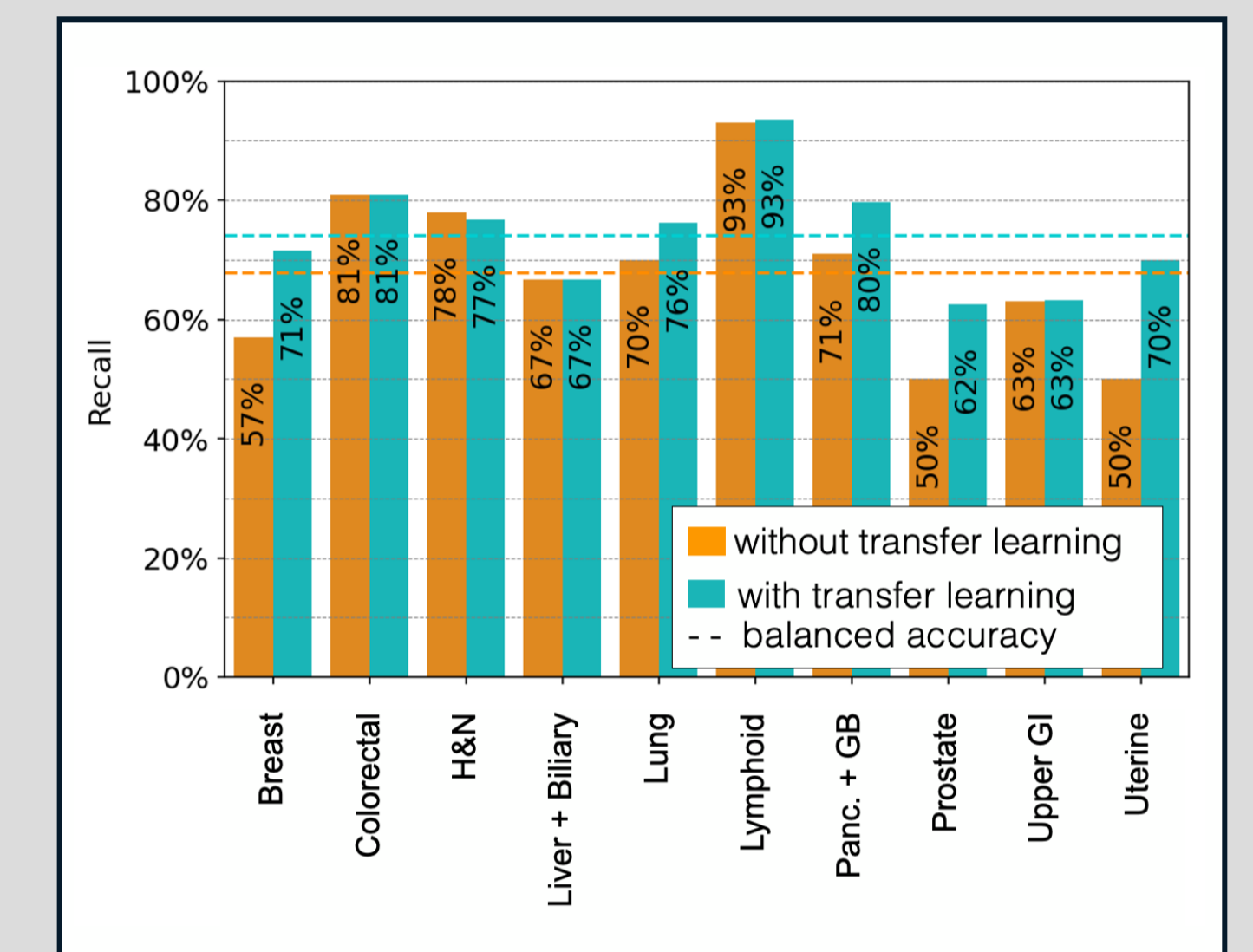


Figure 4. TOO top-1 BA and recall with and without transfer learning.

## CONCLUSIONS

- We developed a novel transfer learning model that extracts **cancer type-specific methylation features** from signal-rich tissue biopsy data, significantly enhancing tissue-of-origin (TOO) classification in blood based liquid biopsy.
- We achieved **89% top-2 balanced accuracy** across 10 cancer categories, outperforming models trained solely on plasma cfDNA, and demonstrating its robustness in detecting challenging cancers with low shedding rates into blood.
- Our approach leverages **automated feature extraction**, enabling a more scalable and adaptable solution compared to traditional manual feature extraction methods.

**Disclosure:** SF, DK, EB, FH and KC are full-time employees of Harbinger Health. AK is a former employee of Harbinger Health.