SIIM CMIMI24
Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

Automated Identification of Challenging Samples in Medical Imaging for Unbiased AI Model Training

Frank Li, PhD, MS, Emory University; Theo Dapamede, MD, PhD; Bardia Khosravi, MD, MPH, MHPE; Mohammadreza Chavoshi, MD; Saptarshi Purkayastha, PhD; Hari Trivedi, MD; Judy Gichoya, MD, MS, FSIIM

Introduction/Background

In medical imaging datasets, "shortcuts" or spurious correlations can cause AI models to unintentionally depend on irrelevant features when making decisions. For instance, presence of support devices like chest tubes act as shortcuts when predicting pneumothorax (easy cases), and pneumothorax cases without chest tubes are harder for the model to learn (hard-to-learn cases). However, hard-to-learn cases are not always known a priori. In this study, we aim to establish a pipeline to automatically differentiate easy- and hard-to-learn cases during model development.

Methods/Intervention

We used a bias amplification (BAM) technique during model training to identify hard-to-earn samples within the SIIM-ACR pneumothorax dataset. BAM incorporates a trainable auxiliary variable (b) to track errors made by the model during training to identify hard-to-learn samples and amplifies them during training process. For our experiments, predicted probabilities below 0.25 for positive samples (false negatives, FN) and above 0.75 for negative samples (false positives, FP) were designated as hard-to-learn samples. Conversely, predicted probabilities above 0.75 for positive samples (true positives, TP) and below 0.25 for negative samples (true negatives, TN) were regarded as easy-to-learn samples. Grad-CAM++ was used generate saliency maps and images reviewed by two radiologists.

Results/Outcome

The magnitude of the auxiliary variable (b) increased with the level of learning difficulty, implying that the model leaned more heavily on b when facing challenging samples. As expected, FP and TP examples had a higher presence of support devices and FN were often missing support devices. Saliency maps and radiologist review revealed that the model focused more on support devices, further supporting this observation.

Conclusion

Our findings validate the hypothesis that images containing both pneumothorax and support devices are easier for the model to learn from. The proposed pipeline may serve as an automated tool to identify hard-to-learn samples in medical imaging datasets, facilitating the training of unbiased AI models and reducing the reliance on labor-intensive human labeling.

Statement of Impact

This study offers an automated method of narrowing down datasets for AI training and validation to alleviate the need for extensive human labeling of granular labels in medical imaging datasets.

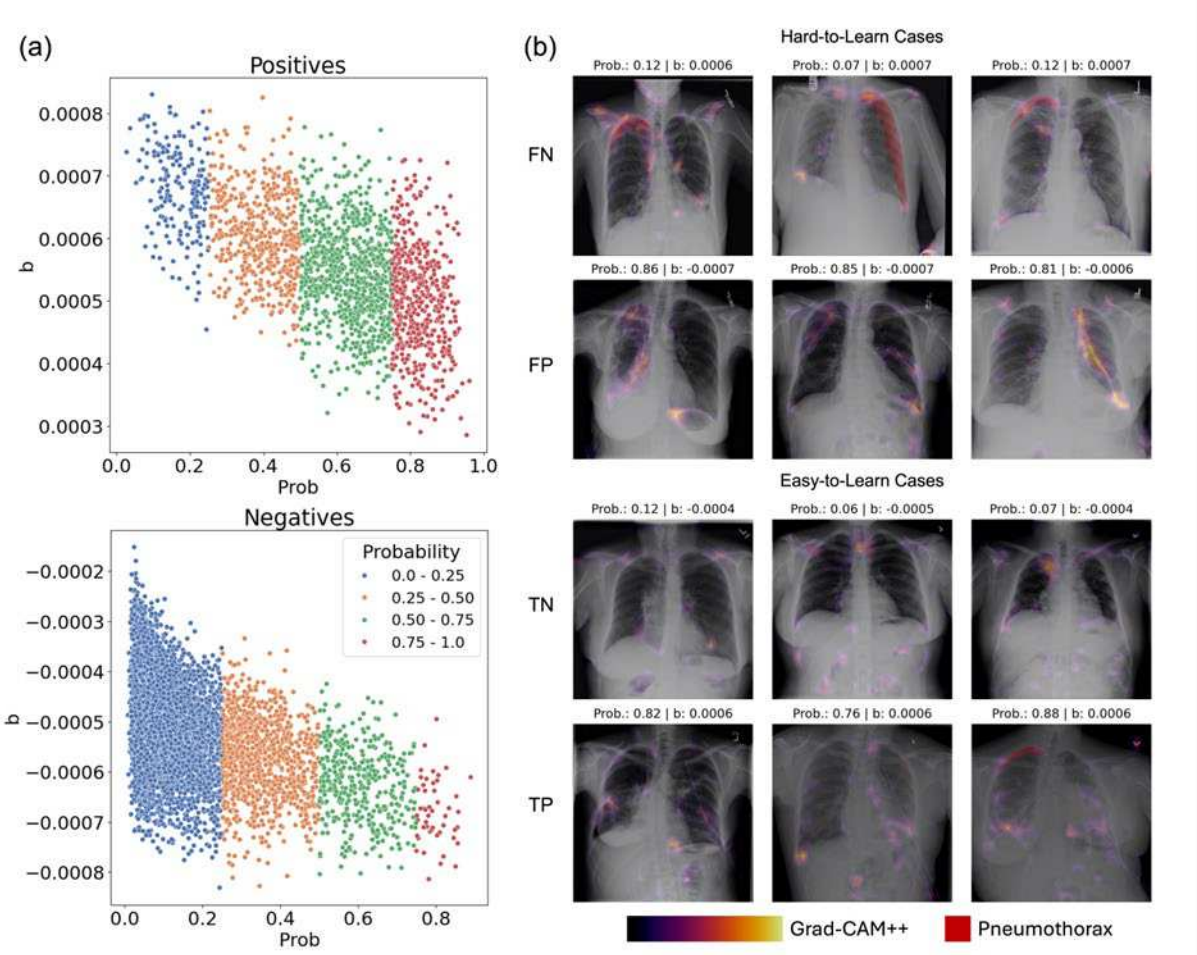

Figure 1. (a) The magnitude of the auxiliary variable b increased with the level of learning difficulty, indicating that the model leaned more heavily on b when facing challenging samples. (b) Fewer support devices were observed in FN and TN, while FP and TP had a higher presence of support devices. Saliency maps revealed that the model focused more on support devices.

Keywords

AI Bias; Shortcut Learning; Medical Imaging; AI; pneumothorax; chest tubes