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

Unveiling Bias in AI Model Training Data: Exploring the Impact of Intrinsic Data Variability on Lung Ultrasound Video Classification Models

Saunak Bhattacharjee, Boston University; Umair Khan, PhD; Russell Thompson, PhD; Lauren P. Etter; Ingrid Camelo, MD; Rachel C. Pieciak, MS; Ilse Castro-Aragon, MD; Bindu Setty, MD; Christopher C. Gill, MD Margrit Betke, PhD

Introduction/Background

Lung ultrasound (LUS) is an emerging tool for providing clinical support for patients with respiratory diseases (Fig. 1). However, its operator-dependent data acquisition and interpretation introduce potential variability in data collection and analysis. Factors such as variations in scanning techniques, duration of the recorded scan, and interpretation of visual patterns may introduce discrepancies. These discrepancies can affect data consistency and potentially bias the model training, impacting the development of generalizable artificial intelligence (AI)-based models.

Methods/Intervention

To investigate the inherent bias in LUS video data and its impact on AI model training, we employed a transformer-based video classification model aimed at identifying lung consolidations among pediatric patients (Fig. 2). This model was complemented by a frame-level transformer-based classification model that aggregates frame-level predictions to produce a video-level score. Both models were trained and validated on 2,400 videos collected in a sweep-acquisition fashion from 200 pediatric patients with pneumonia and were subsequently tested on an external dataset comprising another 2,400 LUS videos from 200 healthy individuals.

Results/Outcome

The analysis of the dataset revealed a correlation coefficient of 0.4039 between larger lung consolidations and longer video lengths, suggesting moderate operator bias in data collection (Fig. 3). The video classification model achieved a 100% accuracy on the external dataset. The frame-level model consistently predicted all frames from healthy individuals as lacking consolidations, with a confidence level above 0.73, demonstrating its ability to generalize to an external dataset.

Conclusion

This study highlights the need for careful consideration of data biases during AI model training to ensure accurate AI-aided diagnosis. Despite its high accuracy, caution is advised when generalizing these results, as the identified biases could affect the future performance of the models.

Statement of Impact

The findings underline the critical importance of addressing biases in LUS video datasets to develop reliable and generalizable AI-based diagnostic tools. Ensuring consistent data collection and interpretation practices is essential for the advancement of AI in medical diagnostics.

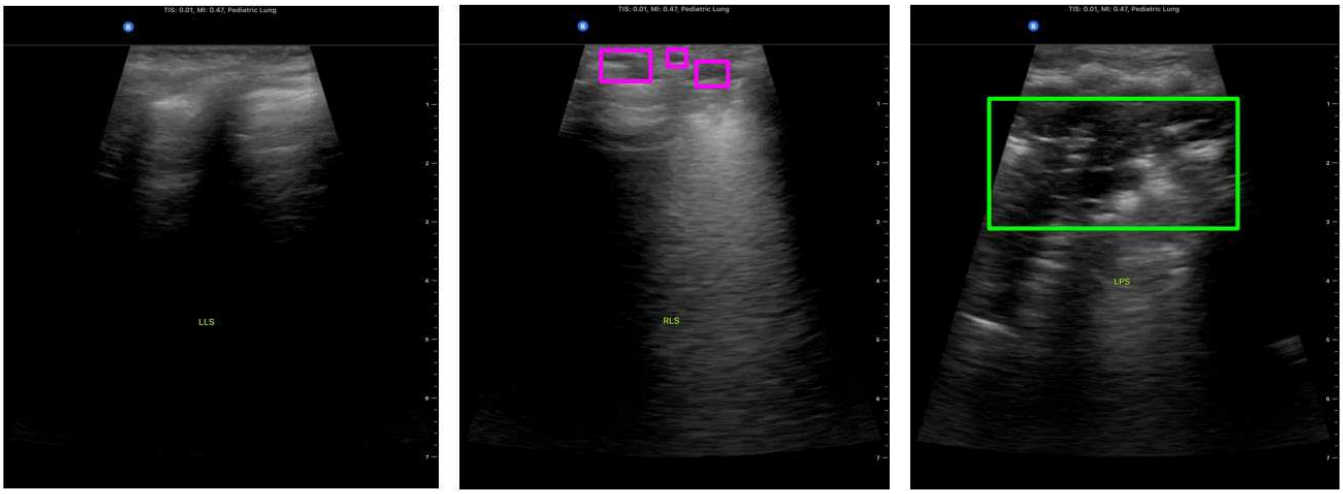

Fig. 1: Lung ultrasound video frames of pediatric patients without (left) and with findings (middle, right), the latter with small subpleural consolidation (pink bounding box) and a segmental/lobar consolidation (green bounding box).

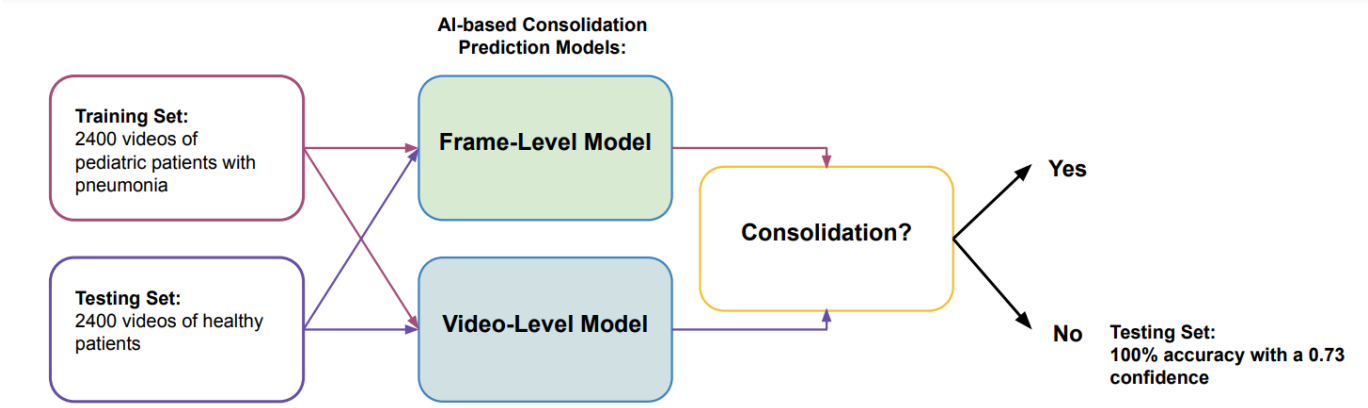

Fig. 2: A training set and a testing set (which is the external dataset in this case) are passed onto a video-level model and a frame-level model for the prediction of whether or not lung consolidations are present in the video.

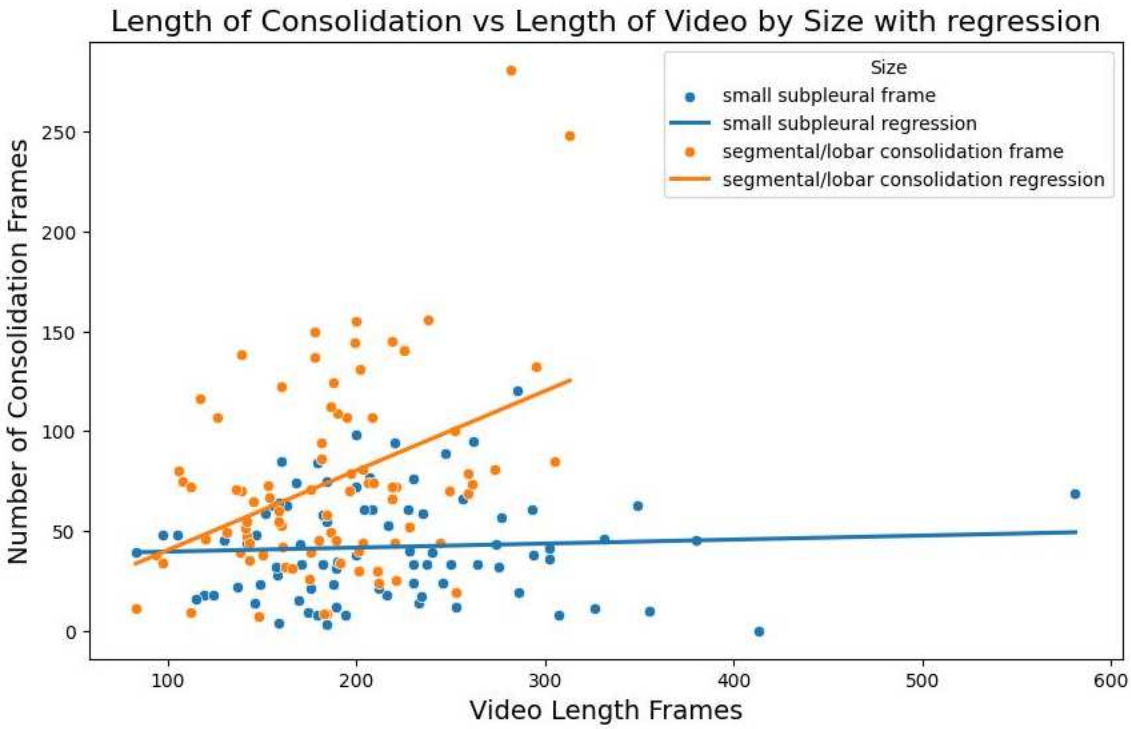

Fig. 3: Analysis of 2,400 lung ultrasound videos, comparing the number of consolidation frames within the videos and the length to the video for both small subpleural (blue) and larger segmental/lobar consolidation findings (orange). The plot shows the linear regression fit for the two variables. The R^2 value shows that 16.3% of the dependent data (number of consolidation frames) can be explained by the independent variable (video length frames). The longer the videos are, the more frames the videos contain with abnormalities, in particular, for videos showing larger consolidations, indicating that the operator sweeps the ultrasound probe slower across the region with findings.

Keywords

Lung Ultrasound (LUS)**;** Data Bias**;** Transformer-based Model**;** Lung Consolidation