An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department

Farah E. Shamout[1], Yiqiu Shen[2], Nan Wu[2], Aakash Kaku[2], Jungkyu Park[3], Taro Makino[3], Stanisław Jastrzębski[2], Duo Wang[3], Ben Zhang[3], Siddhant Dogra[3], Meng Cao[3], Narges Razavian[3], David Kudlowitz[3], Lea Azour[3], William Moore[3], Yvonne W. Lui[3], Yindalon Aphinyanaphongs[3], Carlos Fernandez-Granda[2], Krzysztof J. Geras[3]

[1] NYU Abu Dhabi, Abu Dhabi, UAE
[2] New York University, NY, USA
[3] NYU Langone Health, NY, USA

Abstract: We propose a data-driven approach for the automatic prediction of deterioration risk of COVID-19 patients using a deep neural network that learns from chest X-ray images, and a gradient boosting model that learns from routine clinical variables. Our AI prognosis system, trained using data from 3,661 patients, achieves an AUC of 0.786 (95% CI: 0.742-0.827) when predicting deterioration within 96 hours.

Methods

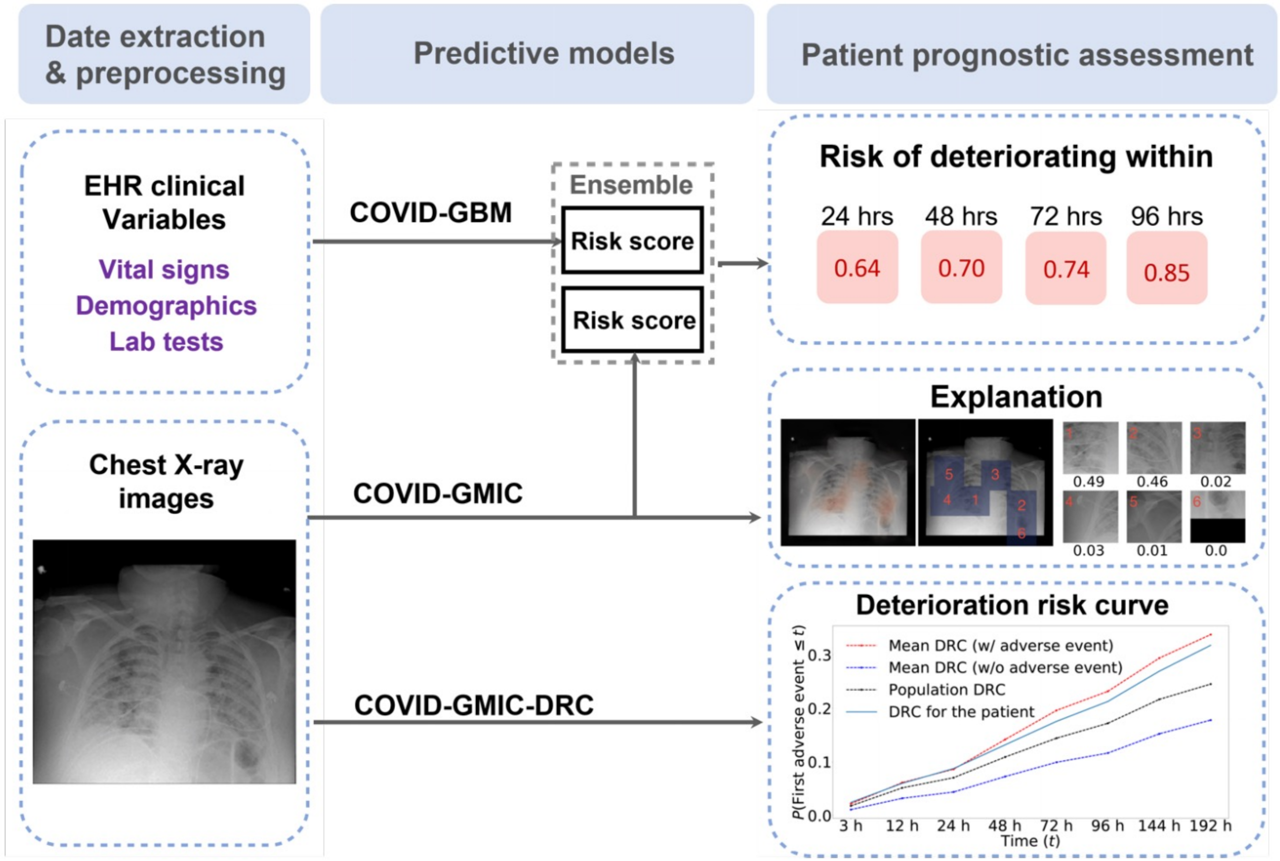


Fig 1. Overview of the AI system that assesses the patient's risk of deterioration every time a chest X-ray image is collected in the ED. We design two different models to process the chest X-ray images, both based on the GMIC neural network architecture. The first model, COVID-GMIC, predicts the overall risk of deterioration within 24, 48, 72, and 96 hours, and computes saliency maps that highlight the regions of the image that most informed its predictions. The predictions of COVID-GMIC are combined with predictions of a gradient boosting model that learns from routinely collected clinical variables, referred to as COVID-GBM. The second model, COVID-GMIC-DRC, predicts how the patient's risk of deterioration evolves over time in the form of deterioration risk curves.

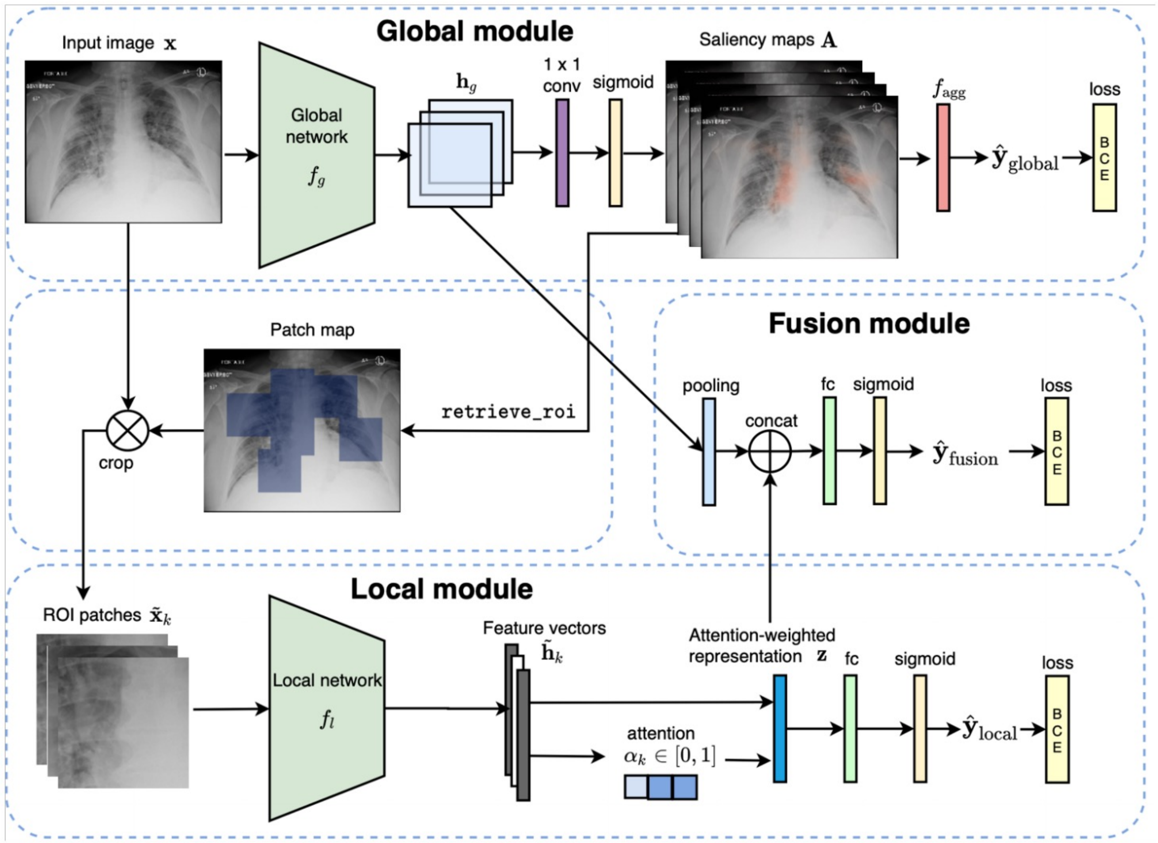


Fig 2. Architecture of COVID-GMIC. First, COVID-GMIC utilizes the global network to generate four saliency maps that highlight the regions on the X-ray image that are predictive of the onset of adverse events within 24, 48, 72, and 96 hours respectively. COVID-GMIC then applies a local network to extract fine-grained visual details from these regions. Finally, it employs a fusion module that aggregates information from both the global context and local details to make a holistic diagnosis.

Results

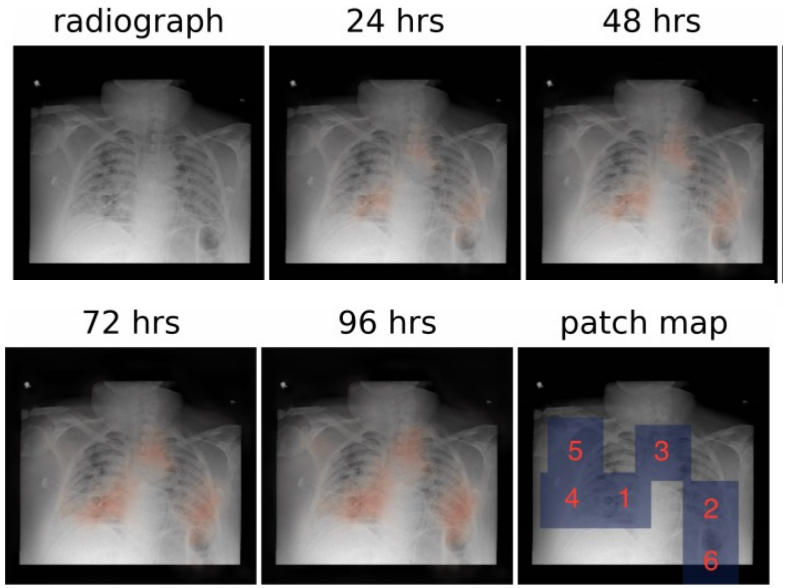
Test set (n=832)

	AUC				PR AUC			
	24 hours	48 hours	72 hours	96 hours	24 hours	48 hours	72 hours	96 hours
COVID-GBM	0.747 (0.692, 0.796)	0.739 (0.683, 0.788)	0.750 (0.701, 0.797)	0.770 (0.727, 0.813)	0.230 (0.164, 0.321)	0.325 (0.254, 0.421)	0.408 (0.337, 0.499)	**0.523** (0.446, 0.613)
COVID-GMIC	0.695 (0.627, 0.754)	0.716 (0.661, 0.766)	0.717 (0.661, 0.766)	0.738 (0.691, 0.781)	0.200 (0.140, 0.281)	0.302 (0.225, 0.395)	0.374 (0.296, 0.465)	0.439 (0.363, 0.532)
COVID-GBM + COVID-GMIC	**0.765** (0.713, 0.818)	**0.749** (0.700, 0.798)	**0.769** (0.720, 0.814)	**0.786** (0.742, 0.827)	**0.243** (0.187, 0.336)	**0.332** (0.254, 0.427)	**0.439** (0.351, 0.533)	0.517 (0.434, 0.605)

Reader study dataset (n=200)

	AUC				PR AUC			
	24 hours	48 hours	72 hours	96 hours	24 hours	48 hours	72 hours	96 hours
Radiologist A	0.613 (0.521, 0.707)	0.645 (0.559, 0.719)	0.691 (0.612, 0.764)	0.740 (0.666, 0.806)	0.346 (0.251, 0.475)	0.490 (0.381, 0.613)	0.640 (0.535, 0.744)	0.742 (0.650, 0.827)
Radiologist B	0.637 (0.544, 0.727)	0.636 (0.556, 0.720)	0.658 (0.578, 0.728)	0.713 (0.640, 0.777)	0.365 (0.268, 0.501)	0.460 (0.360, 0.585)	0.590 (0.479, 0.688)	0.704 (0.603, 0.792)
Radiologist A + Radiologist B	**0.642** (0.555, 0.729)	0.663 (0.580, 0.737)	0.692 (0.618, 0.763)	0.741 (0.673, 0.804)	**0.403** (0.286, 0.534)	0.499 (0.385, 0.618)	0.609 (0.507, 0.726)	0.740 (0.649, 0.830)
COVID-GMIC	**0.642** (0.550, 0.730)	**0.701** (0.621, 0.775)	**0.751** (0.681, 0.817)	**0.808** (0.746, 0.866)	0.381 (0.282, 0.527)	**0.546** (0.435, 0.671)	**0.676** (0.572, 0.788)	**0.789** (0.698, 0.879)

The optimal weights assigned to the COVID-GMIC prediction in the COVID-GMIC and COVID-GBM ensemble were derived through optimizing the AUC on the validation set. The ensemble of COVID-GMIC and COVIDGBM, denoted as 'COVID-GMIC + COVID-GBM', achieves the best performance across all time windows in terms of the AUC and PRAUC, except for the PR AUC in the 96 hours task. In the reader study, our main finding is that COVID-GMIC outperforms radiologists A & B across time windows longer than 24 hours, with 3 and 17 years of experience, respectively.



Saliency maps for a patient who was admitted to the intensive care unit and intubated within 48 hours. The saliency map highlights the medial right basilar and peripheral left basilar opacities.