ARTERYS

Retrospective Study X-ray Chest AI

Valenciennes Hospital France, Milvue and Arterys

Introduction

This retrospective study was conducted by Arterys and the Hospital of Valenciennes (CHV). Radiological data from CHV's emergency department from April to June 2020 was used to assess the standalone performance of Arterys' Chest AI module. Arterys' Chest AI module is an artificial intelligence (AI) algorithm developed by Milvue for the detection of anomalies on x-ray examinations

Objectives

The primary objective was to:

- Assess the clinical validity of Chest AI for chest pathologies

The secondary objective was:

- Assess the standalone performance of Chest AI across specific pathologies

Protocol

- 370 exams
- 1 anatomy
- 5 pathologies (nodules, pulmonary opacities, pleural effusion, pneumothorax, fractures)
- 4 senior radiologist readers
- 1 unaided (without AI) read

Material & Method

The 4 readers reviewed the 370 exams, without AI, and reported abnormalities as they would in their standard clinical workflow. The reads were completed in the Arterys platform in an environment similar to a clinical reading room. The readers had access to standard viewing tools (e.g., pan, zoom, WW/WL) and were given an unlimited amount of time to complete the review. The readers recorded their findings in a standardized excel file.

The assessment's exams were randomly selected using the predefined targets as shown in Figure 1. Timing data was not collected to avoid reading bias. The Milvue AI was assessed based on its standalone performance, i.e., no clinician interacted/reviewed the Milvue produced results prior to analysis.

Case Selection

Arterys defined the minimum number of exams per pathology to ensure the assessment's dataset was sufficiently enriched for the validation of the model (see Figure 1's Pre Defined Case Target). The exam selection for the assessment was established according to a keyword search of the exams' clinical radiological reports. (Figure 1: Count of reports parsed column).

Indications	Pre Defined Case Target	Count of Report Parsed	Count of 3/4 GT
Nodules	37	34	24
Opacities	37	128	105
Pleural effusion	37	91	89
Pneumothorax	37	20	22
Fracture	25	21	4

Figure 1 : Study dataset

Gold Standard Definition

Due to known clinical diagnostic variability, the Gold Standard was defined based on three fourths (3/4) consensus of the readers. To confirm this variability, Figure 2 shows the standalone Chest AI's sensitivity and specificity compared to the Gold Standard (GS) and individual readers (FB, L, S,P).

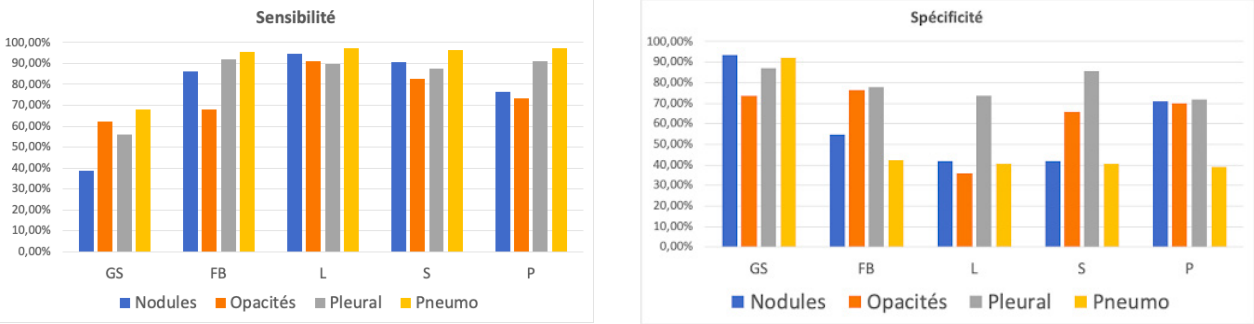

Figure 2 : Variation in sensitivity and specificity depending on pathologies and readers

Results

ROC Curves

The ROC curves were calculated using a trapezoidal estimate based on the Gold Standard of consensus of 3/4 of readers. ROC curves were performed for all pathologies independently and are shown in Figure 3. The area under the curve (AUC) for the indications are fracture: 0.817, nodule: 0.773, opacity: 0.876, pleural effusion: 0.923, pneumothorax 0.991.

Figure 3 shows a limited number of operating points on the curves for nodules, n = 24, and fractures, n = 4, due to the underrepresentation of these subpopulations in the assessment's exams. Figure 1 shows the 37 exam target for these subpopulations was not met, meaning the clinical prevalence of these subpopulations is low (nodules are also often reported as opacities by readers and fractures were treated in more detail in the retrospective bone and joint study). The number of pneumothorax is also low, but representative of its clinical prevalence.

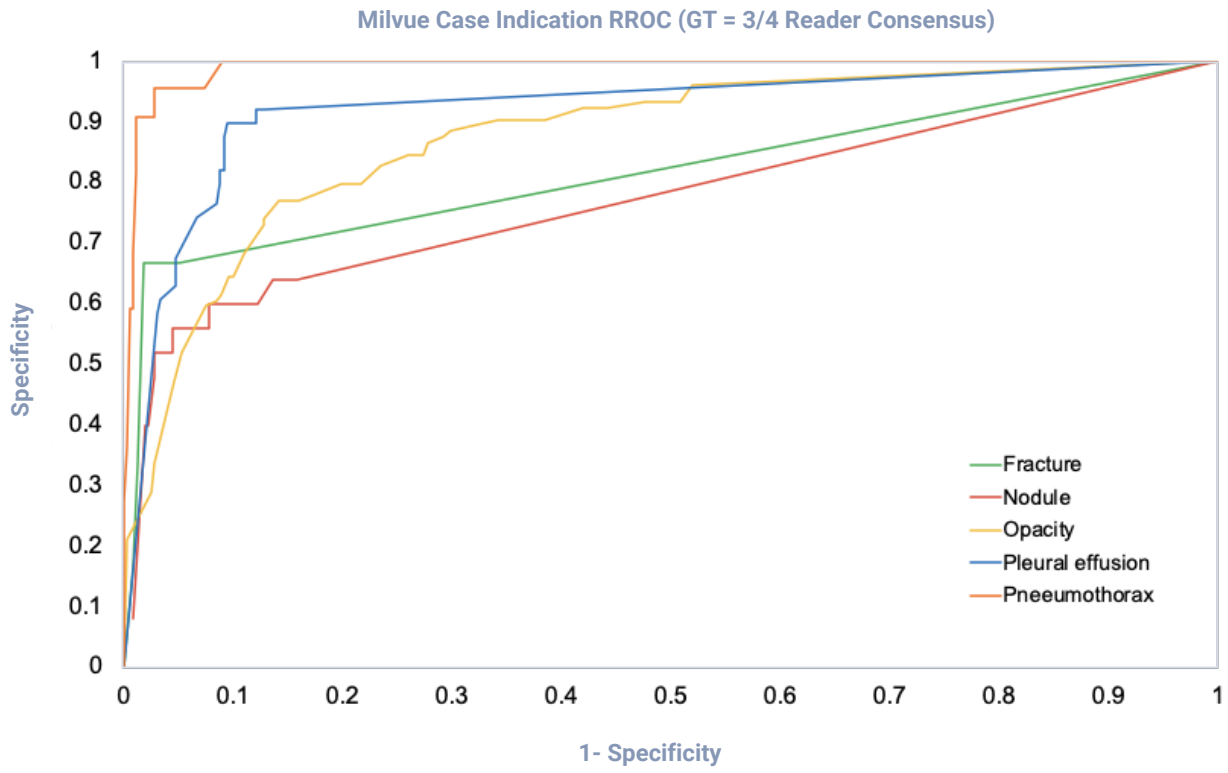

Figure 3 : AUC curve per pathologies (Fracture, nodules, opacities, pleural effusion, pneumothorax)

Performances

For the reasons discussed above, we only confidently considered opacities, pleural effusions, and pneumothorax in the analysis. For these latter pathologies, we observe a very good sensitivity and a good specificity.

Consensus = 3/4	Sensitivity	Specificity	VPP	VPN
All cases	98.80%	43.84%	57.14%	97.96%
Fractures	66.67%	94.99%	17.39%	99.45%
Nodules	64.00%	84.17%	21.92%	97.12%
Opacities	96.15%	48.04%	40.65%	97.12%
Pleural effusion	92.13%	87.16%	68.33%	97.36%
Pneumothorax	100.00%	75.76%	20.00%	100.00%

Figure 4 : Table of specificity, sensitivity, PPV (Positive Predictive Value), NPV (Negative Predictive Value) of the Chest AI model compared to the consensus of ¾ of readers.

Conclusion

Due to the limitation in exam selection, this assessment was used to identify areas of excellence for Arterys' Chest AI and its recommendations for use. Chest AI had very good sensitivity of the algorithm (> 92%), which provides a compelling use case as a triage tool to minimize the number of false negatives for opacities, pleural effusions and pneumothorax. The behavior of Chest AI also demonstrates that it is a perfect tool for homogenizing the divergent opinions of several practitioners. Finally, Chest AI performed well in the detection of both negative and positive pleural effusions.

Correlated with the Osteo-articular (retrospective) and Osteo-articular (prospective) studies in progress at CHV, Arterys will continue to assess the impact and performance of Chest AI and MSK AI in radiologic and emergency departments

ARTERYS

info@arterys.com
51 Federal St. Suite 305, San Francisco, CA 94107
www.arterys.com