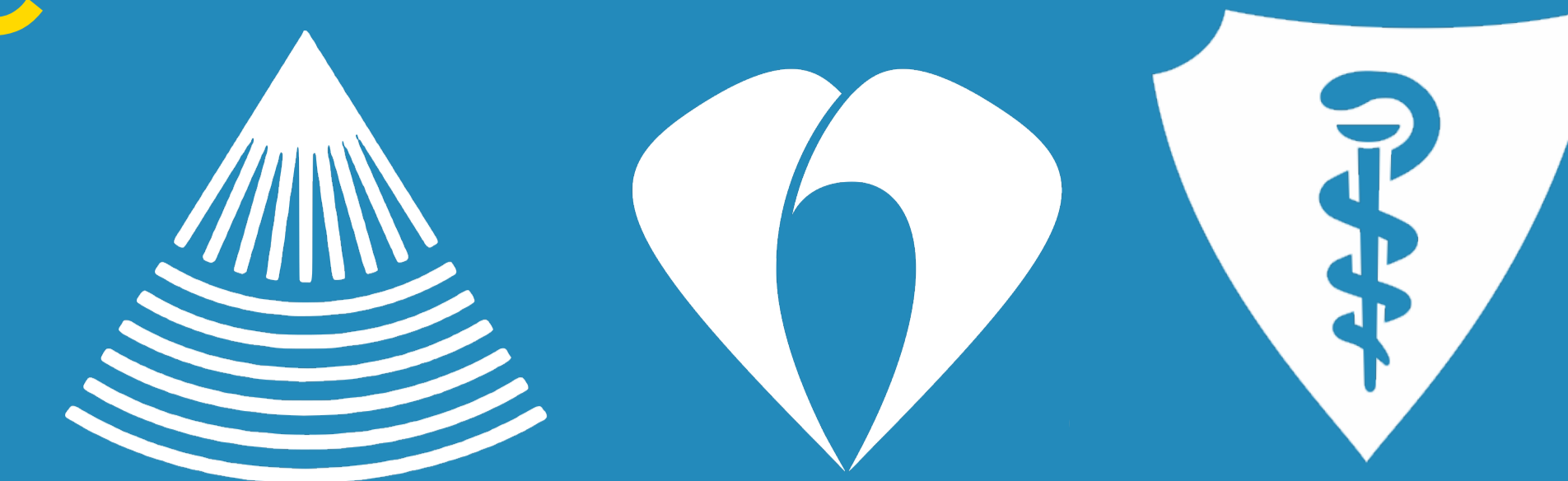


Možnosti hodnocení snímků hrudníku pomocí umělé inteligence

Autor: Jakub Olejko

Školitel: Filip Čtvrtlík, doc. MUDr. Ph.D.

Radiologická klinika LF UP v Olomouci a FNOL



Prezentováno na studentské konferenci SVOČ, Teoretické ústavy LF UP v Olomouci, 30.04. 2024

Úvod

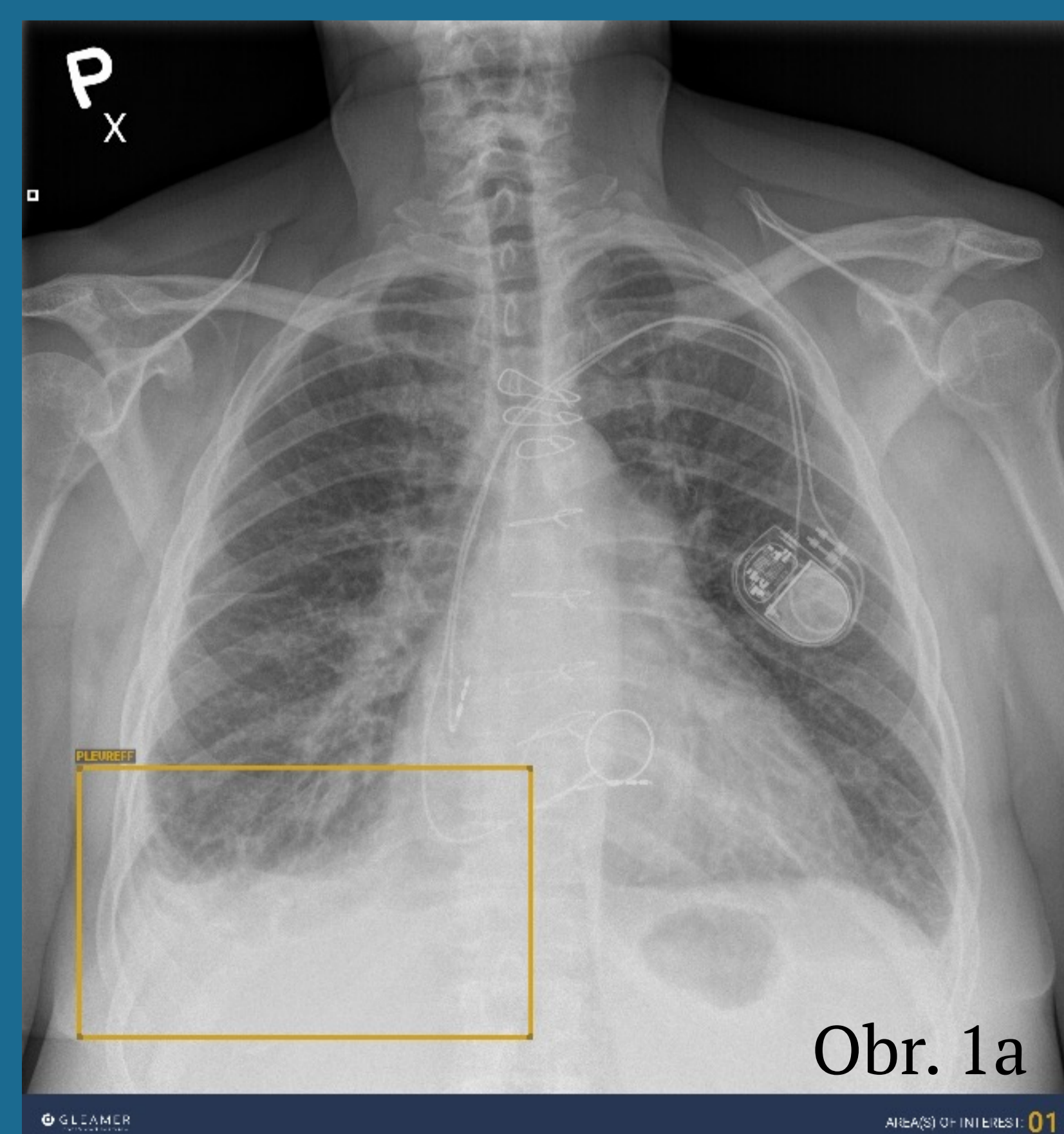
V oblasti radiologické diagnostiky si umělá inteligence (AI) získala značnou pozornost díky svému potenciálu zvýšit přesnost a efektivitu. Jednou z konkrétních oblastí zájmu je porovnání diagnostické úspěšnosti umělé inteligence a lékařů při hodnocení rentgenových snímků hrudníku ve vztahu ke zlatému standardu diagnostiky výpočetní tomografií (CT).

Cíl

Cílem této studie je srovnat schopnost systému umělé inteligence (ChestView; Gleamer) a lékařů správně a efektivně popsat nálezy na rentgenových snímcích hrudníku, přičemž jako měřítko přesného nálezu slouží CT vyšetření hrudníku.

Metodika

U dvou skupin (skupina A 197 pacientů a skupina B 194 pacientů) jsme měli k dispozici rentgenové snímky hrudníku, které byly čteny modelem umělé inteligence (AI). Ve skupině A měl radiolog k dispozici výsledek z modelu AI, rentgenový snímek a pak prováděl diagnostiku. Ve skupině B měl radiolog k dispozici pouze rentgenový snímek, bez hodnocení AI. Umělá inteligence byla schopná sledovat jen 5 specifických nálezů (pneumotorax, fluidothorax, konsolidace, mediastinální masa a nodulace). Úspěšnost diagnostiky byla vyhodnocována na základě CT skenů (výsledek CT vyšetření považujeme za správný). K testování hypotézy, byl použit párový t-test.



Obr. 1a



Obr. 1b



Obr. 2

Obr. 1: RTG S+P zpracován pomocí AI (1a) s hodnocením AI (1b)

Obr. 2: Fluidothorax na CT

Výsledky

Ve skupině A je rozdíl mezi pravděpodobností správného popisu pro AI (odhad úspěšnosti 0.58) a lékaře, který měl k dispozici výsledek AI (odhad úspěšnosti 0.71), statisticky významný (p-hodnota <0.001) (viz. tab. 1). Ve skupině B je rozdíl mezi pravděpodobností správného popisu pro AI (odhad úspěšnosti 0.63) a lékaře, který neměl k dispozici výsledek AI (odhad úspěšnosti 0.73), rovněž statisticky významný (p-hodnota <0.001) (viz. tab. 2). Úspěšnost diagnostiky se zdá být při diagnostice fluidothoraxu a nodulace u obou skupin stejná, a proto se vliv AI na diagnostiku lékaře nezdá být významný.

Tab. 1 Skupina A	AI				L+AI				L+AI - AI	
	Spr.	Nespr.	Neroz.	Úspěšnost	Spr.	Nespr.	Neroz.	Úspěšnost	Rozdíl	P-hodnota
Pneumothorax	185	6	6	0.94	195	2	0	0.99	0.05	0.001
Fluidothorax	172	25	0	0.87	171	16	10	0.87	0.01	0.809
Konsolidace	160	18	19	0.81	167	19	11	0.85	0.04	0.034
Mediastinální masa	182	8	7	0.92	195	2	0	0.99	0.07	0.001
Nodulus	167	12	18	0.85	184	12	1	0.93	0.09	< 0.001
Celkem	114	41	42	0.58	140	37	20	0.71	0.13	< 0.001

Tab. 2 Skupina B	AI				L				L - AI	
	Spr.	Nespr.	Neroz.	Úspěšnost	Spr.	Nespr.	Neroz.	Úspěšnost	Rozdíl	P-hodnota
Pneumothorax	188	4	2	0.97	194	0	0	1.00	0.03	0.014
Fluidothorax	173	21	0	0.89	169	16	9	0.87	0.02	0.249
Konsolidace	166	16	12	0.86	173	16	5	0.89	0.04	0.052
Mediastinální masa	181	2	11	0.93	187	2	5	0.96	0.034	0.03
Nodulus	165	12	17	0.85	181	12	1	0.93	0.08	< 0.001
Celkem	122	33	39	0.63	141	35	18	0.73	0.10	< 0.001

Tab. 1 a 2: Tabulky uvádí počty správně (Spr.) a nesprávně (Nespr.) diagnostikovaných pacientů a také počty případů, kdy bylo hodnocení nerozhodné (Neroz.). Úspěšnost vyjadřuje poměr počtu správně diagnostikovaných pacientů k celkovému počtu pacientů. Uvedeny jsou také odpovídající p-hodnoty.

Závěr

Zkoumaný program využívající AI (zatím) umí na snímku hrudníku hodnotit jen omezené spektrum patologických změn a neumí popsat klinicky významné vedlejší nálezy, jako je například pneumoperitoneum nebo dilatace kliček střev. Prokázali jsme, že při jejich hodnocení byla úspěšnost popisu nálezů hodnocených radiologem lepší než samotné AI (skupina B). Také se zdá, že vliv AI na lékaře se jeví být nevýznamným (diagnostika nodulací a fluidothoraxu).