# DIAGNOSTIC SENSITIVITY OF ARTIFICIAL INTELLIGENCE IN DETECTING ACUTE SURGICAL ABDOMEN ON PAEDIATRIC ABDOMINAL X-RAYS COMPARED WITH OPERATIVE FINDINGS

ABDUL MANNAN, ZAHRA AUQIL, BATOOL FATIMA, ZUHA ZAFAR, REHMAN WAHEED, MUHAMMAD KASHIF BASHIR.

DEPARTMENT OF PEDIATRIC SURGERY, MAYO HOSPITAL LAHORE, PAKISTAN

## INTRODUCTION

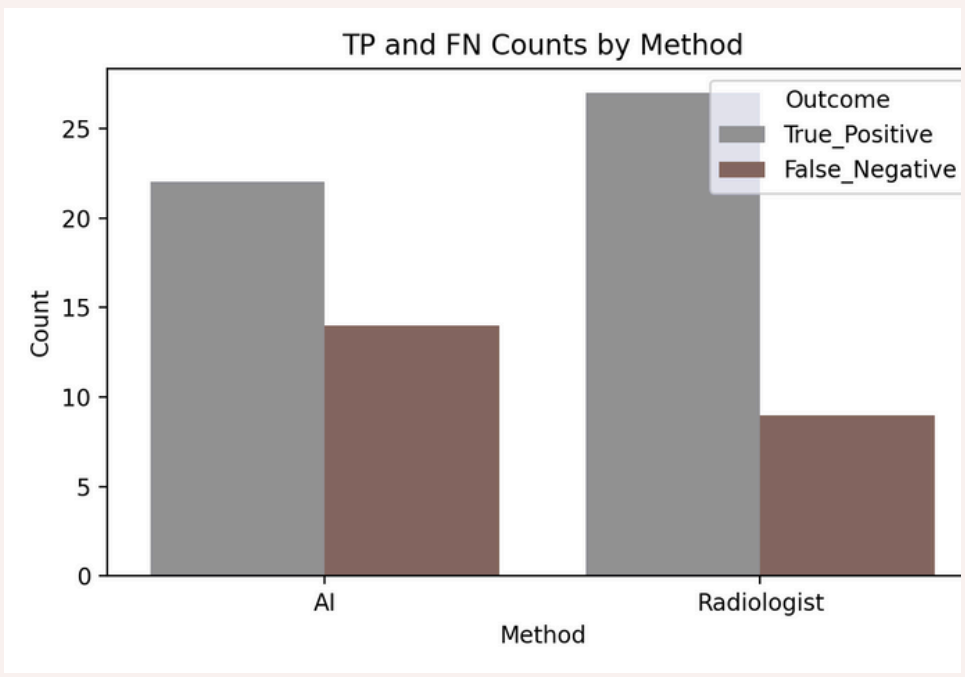
Diagnosing acute surgical abdomen (like obstruction or perforation) in children is challenging. While abdominal X-rays are common, interpreting them is difficult due to pediatric anatomical differences. Artificial intelligence (AI) has successfully aided in diagnosing other pediatric conditions, like pneumonia and fractures, but its use for acute abdominal pathologies is underexplored.

## METHODOLOGY

A total of 36 pediatric abdominal X-rays were interpreted using ChatGPT. Interpretations were generated using only clinical summaries and radiographic images. These were compared against intraoperative findings as the gold standard for diagnostic correctness. Additionally, in cases where radiologist reports were available, ChatGPT's interpretations were compared for agreement using Cohen's kappa statistic.

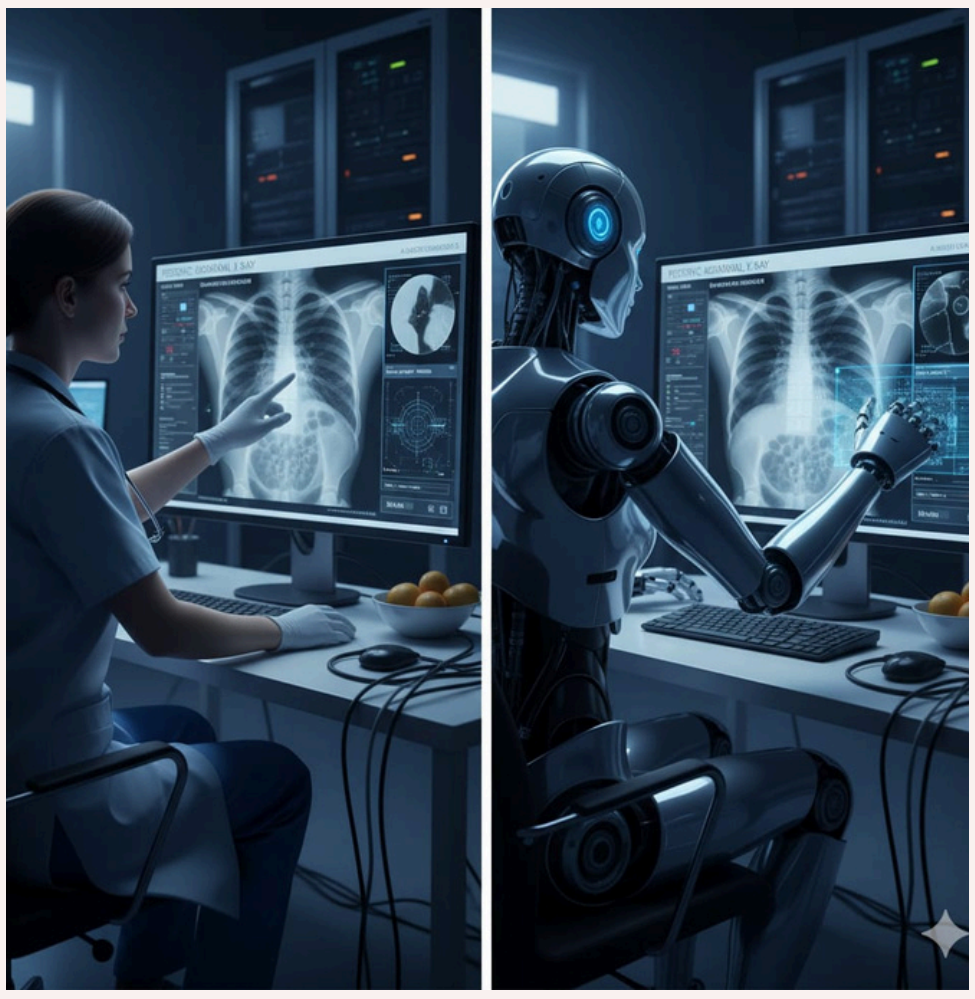
## RESULTS

ChatGPT accurately interpreted 22 out of 36 cases (61.1%) based on intraoperative findings. It agreed with radiologist assessments in 18 out of 22 cases (81.8%), yielding a Cohen's kappa (k) of 0.19, indicating slight to fair agreement. Notably, ChatGPT missed pneumoperitoneum in both instances where it was present - a critical oversight. The tool performed reasonably well in identifying obstructive and Hirschsprung's disease, but showed limitations in detecting subtle or life-threatening findings.




## OBJECTIVE

To evaluate the diagnostic performance of ChatGPT in interpreting pediatric abdominal radiographs by comparing its outputs against intraoperative findings (gold standard) and radiologist reports, and to quantify agreement using Cohen's kappa.



## CONCLUSION

ChatGPT demonstrated moderate diagnostic accuracy in pediatric abdominal X-ray interpretation, with slight to fair inter-rater agreement compared to radiologists. While it shows potential as an educational or triage support tool, its current limitations - particularly in missing emergent findings like pneumoperitoneum - underscore the need for cautious and supervised clinical integration.

## REFERENCES

1. Kim S, Yoon H, Lee MJ, Kim MJ, Han K, Yoon JK, et al. Performance of deep learning-based algorithm for detection of ileocolic intussusception on abdominal radiographs of young children. Scientific Reports. 2019 Dec;9
2. Davendralingam N, Sebire NJ, Arthurs OJ, Shelmerdine SC. Artificial intelligence in paediatric radiology: Future opportunities. The British Journal of Radiology [Internet]. 2021 Jan 1;94(1117):20200975.