

Original Research Article

Digital image display manipulation and observer variability in pediatric chest radiograph interpretation

Chris Neil James¹, Dinesh Kumar Narayanasamy^{2*}, Vasanthan Tanigasalam²,
Jeyaraj Munusamy², Harish Vijayakandeean Kumaresan³

¹JIPMER, Karaikal Puducherry, India

²Department of Pediatrics, JIPMER Karaikal, Puducherry, India

³Department of Respiratory Medicine, Indira Gandhi Medical College & Research Institute, Puducherry, India

Received: 25 September 2026

Accepted: 21 October 2026

*Correspondence:

Dr. Dinesh Kumar Narayanasamy,
E-mail: dinesh.paed@gmail.com

Copyright: © the author(s), publisher and licensee Medip Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Interpretation of pediatric chest radiographs is subject to observer variability, and digital image processing may further influence visual assessment. We evaluated interobserver agreement and the effect of controlled contrast manipulation on the interpretation of pediatric chest radiographs.

Methods: Sixty anonymised pediatric chest radiographs, comprising 30 classified as normal and 30 as abnormal, were presented in three display conditions: original (OI), contrast-reduced (CRR), and contrast-enhanced (CER). Five pediatricians independently assessed all 180 image presentations for image quality, radiographic features, and final interpretation as normal or abnormal. Interobserver agreement was assessed using Fleiss' kappa and intraobserver consistency using Cohen's kappa. Display-related changes in abnormal and lung-opacity assessments were compared at the underlying radiograph level using paired Wilcoxon signed-rank tests.

Results: Overall interobserver agreement for final interpretation was moderate ($\kappa=0.528$), with complete agreement in 48.3% of presentations. Agreement was $\kappa=0.496$, 0.594, and 0.479 for OI, CRR, and CER, respectively. Among radiographs classified as normal, lung-opacity identification increased from 8.7% with OI to 32.7% with CRR ($p=0.012$) and 27.3% with CER ($p=0.018$). Final normal/abnormal classification did not differ significantly between display conditions.

Conclusions: Controlled contrast manipulation influenced recognition of lung opacity despite unchanged underlying radiographic content. Standardised display conditions and awareness of perceptual variability may be important in pediatric chest radiograph interpretation.

Keywords: Paediatric, Chest radiography, Observer variability, Agreement, Digital radiography, Lung opacity

INTRODUCTION

Chest radiography is frequently used in children to evaluate respiratory disease, cardiopulmonary abnormalities, and other thoracic conditions. Accurate interpretation requires systematic assessment of image quality and positioning, followed by evaluation of the airways, lungs, pleura, heart, mediastinum, and other thoracic structures.¹ However, interpretation is not completely consistent between observers, with studies

involving pediatricians, other clinicians, and radiologists demonstrating variability in intra- and interobserver agreement, particularly for pulmonary abnormalities.²⁻⁵ Such variability may affect clinical decision-making and can also influence outcome classification and comparability in epidemiological and research settings.

In addition to observer-related factors, the characteristics of the displayed radiographic image may influence interpretation. Digital radiography permits manipulation

of image-processing parameters, and experimental work has shown that changes in image processing can alter the perceived visibility of anatomical structures and pulmonary lesions.⁶ Thus, the displayed image is not necessarily a fixed visual representation of the acquired radiograph, and its presentation may influence perception.

Observer variability may arise not only from differences in expertise but also from differences in the perception of subtle findings under different viewing conditions.⁷ Recent pediatric data continue to show moderate agreement between clinicians interpreting chest radiographs in real-world practice, indicating that interpretation variability remains relevant despite advances in imaging technology.⁸

Most previous studies of pediatric chest radiograph variability have focused on differences between observers, professional groups, or diagnostic criteria. Less attention has been given to whether changing the display of the same underlying radiograph alters interpretation. This distinction is important because display-related effects may represent an additional source of variability independent of the underlying radiographic content.

Therefore, this study evaluated interobserver agreement among pediatricians interpreting chest radiographs and examined whether controlled manipulation of image contrast influenced recognition of radiographic abnormalities, particularly lung opacity, and the overall final interpretation.

METHODS

Study design

This was an image-based observer-variability study designed to assess interobserver agreement and the influence of controlled digital image-display manipulation on pediatric chest radiograph interpretation. The study was approved by the Institutional Ethics Committee of JIPMER (JIP/IEC-OS/2026/107). The requirement for informed consent was waived because the study used anonymised archival radiographs. The study used an anonymised archival set of pediatric digital chest radiographs from the district hospital Picture Archiving and Communication System (PACS) during September to December 2025. The radiographs were obtained from children aged 28 days to <12 years. Each underlying radiograph was evaluated in three display conditions, allowing display-related effects to be examined while keeping the radiographic content constant. Observers interpreted the images independently and were blinded to clinical information and the underlying classification.

Image dataset and display manipulation

The dataset comprised 60 pediatric chest radiographs, of which 30 were classified as normal and 30 as abnormal during dataset preparation by the study pulmonologist. This classification was used to stratify the analysis and was not considered an independent reference standard for diagnostic accuracy. Radiographs with identity markers or potentially confounding visible objects (including talismans, foreign bodies, or electrocardiographic leads) and multiple radiographs from the same child were excluded. The final dataset, therefore, contained one radiograph per child.

Each radiograph was prepared in three display conditions: original (OI), contrast-reduced (CRR), and contrast-enhanced (CER). Display manipulation was performed using Adobe Lightroom. The original image was used as the baseline; contrast alone was adjusted from the baseline setting to -70 for CRR and $+70$ for CER. All other image-processing parameters were kept unchanged. The resulting images were exported as separate CRR and CER files. Thus, the three display conditions represented different presentations of the same underlying radiograph rather than different acquisitions.

Observers and reading procedure

Five pediatricians independently interpreted all 180 image presentations. The observers had a median of 7 years of pediatric clinical experience, had received structured training in pediatric chest radiograph interpretation using the Indian Academy of Paediatrics (IAP) approach, based on the IAP educational video, and routinely interpreted approximately 30 chest radiographs per day. Observers were not provided with clinical information, previous interpretations, or the underlying classification of the radiographs.

To determine the presentation sequence, the three versions of each underlying radiograph (OI, CRR, and CER) were first listed. Computer-generated random numbers were assigned to the three versions, and the versions were then sorted in ascending order of the random number. This procedure determined the presentation order. The 180 image presentations were distributed as 60 images per day over 3 days, with each day containing 30 presentations from radiographs classified as normal and 30 from radiographs classified as abnormal. Each observer therefore completed all 180 presentations independently.

Image interpretation

A standardized electronic assessment form was used for image interpretation. All observers assessed the images at the same time each day using an iPad with automatic brightness adjustment disabled and recorded their responses independently through a Google Form on their mobile devices. Observers were instructed to assess each

image independently and submit their response before proceeding to the next image. The assessment included overall image quality, rotation, degree of inspiration, anatomical coverage, cardiothoracic ratio, mediastinal width, presence and pattern of lung opacity, laterality, zone involved, hyperinflation, pleural effusion, pneumothorax, and final interpretation. The final interpretation was recorded as normal or abnormal. Detailed radiographic features were retained for secondary analysis of observer variability across individual features.

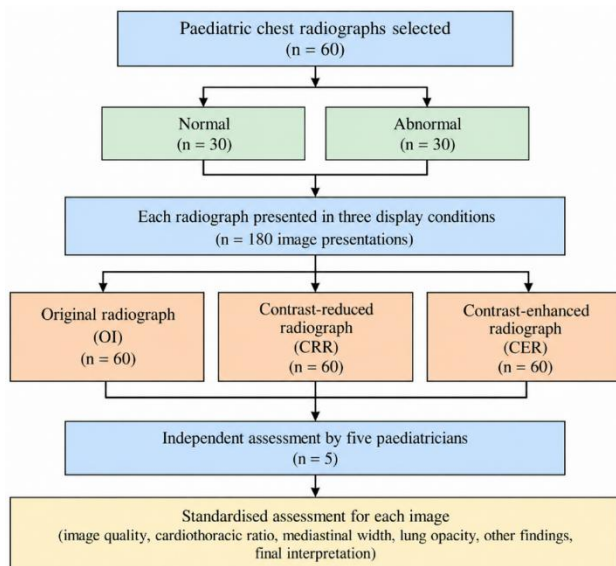


Figure 1: Study flow diagram showing selection of 60 paediatric chest radiographs, classification into normal and abnormal groups, presentation under three display conditions, and independent assessment by five paediatricians.

Interobserver agreement

The primary analysis assessed agreement among all five observers for the final interpretation. Because more than two observers independently rated the same categorical outcome, Fleiss' kappa was used to quantify interobserver agreement.⁹ Observed percentage agreement was reported alongside kappa because kappa is influenced by the distribution of ratings and may not fully represent raw agreement. Interobserver agreement was calculated for all 180 image presentations, each display condition (OI, CRR, and CER), and the normal and abnormal subsets within each condition. Feature-level interobserver agreement was assessed separately as a secondary analysis.

Intraobserver agreement

For each observer, intraobserver consistency of final interpretation was assessed by comparing interpretation of the same underlying radiographs between OI and CRR and between OI and CER. Cohen's kappa was calculated

separately for radiographs classified as normal and abnormal in the underlying study dataset.

Effect of display manipulation

The principal analysis of the effect of display manipulation was performed at the underlying-radiograph level rather than treating the 900 observer assessments as independent observations. For each underlying radiograph and display condition, the number of observers who classified the image as abnormal was calculated, yielding a score from 0 to 5. The same approach was used for lung-opacity recognition. Paired comparisons between display conditions were performed using the Wilcoxon signed-rank test, separately for radiographs classified as normal and abnormal. This preserved the paired nature of the observations because each display condition represented the same underlying radiograph. Each display condition contributed 150 observer assessments within each underlying classification (30 radiographs $\times$ 5 observers).

Statistical considerations

Kappa statistics were interpreted alongside observed agreement rather than relying on rigid categorical labels alone. Interpretation of agreement statistics considered the study design, outcome prevalence, and scale of measurement.¹⁰ Feature-level analyses were considered secondary and exploratory. The primary inference focused on the final interpretation and lung opacity assessment. A two-sided p value <0.05 was considered statistically significant for paired comparisons.

RESULTS

Study dataset and observer assessments

The final dataset comprised 60 paediatric chest radiographs, of which 30 were classified as normal and 30 as abnormal. Each was presented in three display conditions, yielding 180 image presentations. All five observers completed the assessments, resulting in 900 observer-image assessments.

Interobserver agreement

Overall interobserver agreement for the final interpretation was moderate (Fleiss' $\kappa=0.528$), with complete agreement among all five observers in 48.3% of image presentations. Across display conditions, κ was 0.496 for OI, 0.594 for CRR, and 0.479 for CER (Table 1). Feature-level agreement varied across radiographic domains.

Effect of display manipulation on final interpretation

Among radiographs classified as normal, abnormal interpretations occurred in 50.0% of observer assessments with OI, 48.7% with CRR, and 62.7% with

CER. None of the paired comparisons was statistically significant. Among radiographs classified as abnormal, abnormal interpretations occurred in 72.0%, 68.7%, and

73.3% of assessments with OI, CRR, and CER, respectively, with no significant paired differences (Table 2).

Table 1: Interobserver agreement for final interpretation across display conditions.

Display condition	Image presentations, N	Fleiss' κ	Complete agreement, N (%)
All image presentations	180	0.528	87 (48.3)
OI	60	0.496	27 (45.0)
CRR	60	0.594	32 (53.3)
CER	60	0.479	28 (46.7)
OI – normal	30	0.427	11 (36.7)
OI – abnormal	30	0.521	16 (53.3)
CRR – normal	30	0.626	16 (53.3)
CRR – abnormal	30	0.520	16 (53.3)
CER – normal	30	0.501	13 (43.3)
CER – abnormal	30	0.438	15 (50.0)

OI, original radiograph; CRR, contrast-reduced radiograph; CER, contrast-enhanced radiograph; κ , Fleiss' kappa.

Table 2: Effect of display manipulation on final interpretation according to underlying radiograph classification.

Underlying classification	Display condition	Abnormal interpretation, n/150 (%)	Paired comparison	P value
Normal	OI	75/150 (50.0)	OI vs CRR	0.852
	CRR	73/150 (48.7)	OI vs CER	0.153
	CER	94/150 (62.7)	CRR vs CER	0.310
Abnormal	OI	108/150 (72.0)	OI vs CRR	0.750
	CRR	103/150 (68.7)	OI vs CER	0.874
	CER	110/150 (73.3)	CRR vs CER	0.634

P values are from paired Wilcoxon signed-rank tests performed at the underlying radiograph level using the number of observers (0–5) assigning an abnormal interpretation.

Table 3: Effect of display manipulation on recognition of lung opacity according to underlying radiograph classification.

Underlying classification	Display condition	Opacity identified, n/150 (%)	Paired comparison	P value
Normal	OI	13/150 (8.7)	OI vs CRR	0.012*
	CRR	49/150 (32.7)	OI vs CER	0.018*
	CER	41/150 (27.3)	CRR vs CER	0.549
Abnormal	OI	62/150 (41.3)	OI vs CRR	0.159
	CRR	37/150 (24.7)	OI vs CER	0.607
	CER	68/150 (45.3)	CRR vs CER	0.040*

*Unadjusted $P < 0.05$ from paired Wilcoxon signed-rank tests performed at the underlying-radiograph level using the number of observers (0–5) identifying opacity. Pairwise analyses were exploratory and were not adjusted for multiple comparisons.

Table 4: Intraobserver agreement for final interpretation across display conditions.

Observer	Normal radiographs		Abnormal radiographs	
	OI vs CRR κ^*	OI vs CER κ^*	OI vs CRR κ^*	OI vs CER κ^*
P1	–0.009	0.148	–0.054	–0.304
P2	–0.009	0.211	–0.023	–0.154
P3	0.154	–0.154	–0.164	0.321
P4	–0.143	0.211	0.024	–0.118
P5	–0.086	0.156	0.153	–0.078

*Analyses were performed separately for radiographs classified as normal and abnormal in the underlying study dataset. κ , Cohen's kappa

Effect of display manipulation on lung-opacity recognition

Among radiographs classified as normal, lung opacity was identified in 8.7% of assessments with OI, compared with 32.7% with CRR and 27.3% with CER. The paired comparisons showed higher recognition with CRR than OI ($p=0.012$) and with CER than OI ($p=0.018$). Among radiographs classified as abnormal, opacity identification was 41.3%, 24.7%, and 45.3% with OI, CRR, and CER, respectively; the CRR-versus-CER comparison was statistically significant ($p=0.040$), whereas comparisons involving OI were not (Table 3). These pairwise analyses were exploratory and were not adjusted for multiple comparisons.

Intraobserver consistency

Intraobserver agreement for final interpretation was low across display conditions. For normal radiographs, κ ranged from -0.143 to 0.211 for OI versus CRR and from -0.154 to 0.211 for OI versus CER. For abnormal radiographs, κ ranged from -0.054 to 0.153 for OI versus CRR and from -0.304 to 0.321 for OI versus CER (Table 4).

DISCUSSION

The principal findings of this study were threefold. First, interobserver agreement for the final interpretation of pediatric chest radiographs was moderate. Second, changing image contrast did not significantly alter the overall binary classification of the radiographs as normal or abnormal when assessed at the paired underlying-radiograph level. Third, display manipulation substantially changed recognition of lung opacity among radiographs classified as normal, indicating that display characteristics may influence an intermediate perceptual step even when the final binary interpretation does not change.

The moderate interobserver agreement observed in the present study is consistent with previous reports of variability in pediatric chest radiograph interpretation. Johnson et al reported fair-to-moderate interobserver agreement among clinicians, while studies of pediatric pneumonia have demonstrated disagreement in the recognition and characterization of pulmonary abnormalities.²⁻⁵ More recent real-world data also showed moderate agreement between pediatric emergency clinicians and radiologists, with a Cohen's kappa of 0.505 across more than 1300 pediatric radiographs.⁸ The present findings therefore remain consistent with the broader observation that interpretation variability persists across clinical settings and observer groups.

The contrast-manipulation findings offer a different perspective from studies that primarily examine disagreement among observers. In the present study, the underlying radiograph remained unchanged, while

contrast alone was modified. Among radiographs classified as normal, lung-opacity identification increased from 8.7% with the original display to 32.7% with reduced contrast and 27.3% with enhanced contrast. This indicates that the visual threshold for perceiving or labelling pulmonary opacity may vary depending on how the image is presented. Importantly, these findings do not establish diagnostic misclassification because no independent reference standard was available to determine whether an opacity identified by an observer was truly abnormal.

The finding is also consistent with experimental evidence that digital image-processing parameters can alter the visibility of thoracic anatomy and lesions.⁶ Kheddache et al demonstrated that different processing levels affected ratings of the visibility of chest anatomy, rounded opacities, and vessels. The present study extends this concept by examining how controlled changes in contrast may influence clinicians' interpretation of pediatric chest radiographs. The different responses to reduced and enhanced contrast, particularly among radiographs classified as abnormal, further suggest that display manipulation does not produce a simple linear change in abnormality detection.

The low intraobserver agreement across display conditions provides additional evidence that variability was not confined to differences between observers. The same observers showed inconsistent final classifications when viewing different display versions of the same underlying radiograph. This finding supports the concept that visual presentation can contribute to within-observer variability, although kappa estimates should be interpreted in the context of outcome prevalence and marginal distributions.

The study has practical implications for clinical practice and medical education. The findings support attention to consistent image-display conditions during radiographic interpretation and highlight the importance of systematic assessment rather than reliance on a single visual impression. For learners, awareness that image presentation can influence perception may be useful when teaching interpretation of subtle pulmonary findings. However, the present study evaluated contrast manipulation alone and should not be extrapolated directly to other display factors such as monitor calibration, ambient lighting, or windowing.

The principal strength of the study was its paired design: the same underlying radiograph was presented in three display conditions, allowing display-related effects to be examined without changing the radiographic content. Randomized presentation order was used to reduce systematic order effects, and all five observers assessed the complete image set. The study also used an underlying-radiograph-level analysis for the effect of display manipulation, avoiding the assumption that the 900 observer assessments were independent.

Several limitations should be considered. The number of underlying radiographs was small, limiting precision and generalizability. The observers were pediatricians rather than radiologists and were drawn from a teaching setting, which may limit applicability to other observer groups and settings. No independent reference standard was used; therefore, the study evaluates agreement and perceptual change rather than diagnostic accuracy. Only contrast was manipulated, while other display characteristics were held constant and were not evaluated. Finally, interpretation without clinical information differs from routine clinical practice, where imaging findings are integrated with the clinical context.

Future studies should include larger and more diverse image sets, observers from different professional backgrounds, calibrated displays, and an independent reference standard. Studies incorporating additional display parameters and clinical information could determine whether display-related perceptual effects translate into clinically meaningful differences in diagnosis or management.

CONCLUSION

This study demonstrated moderate interobserver agreement in pediatric chest radiograph interpretation and showed that controlled contrast manipulation altered recognition of lung opacity, particularly in radiographs classified as normal. These findings suggest that display characteristics may contribute to perceptual variability even when the underlying radiograph is unchanged. Standardized display conditions and awareness of this potential source of variability may be relevant to clinical interpretation and radiographic education.

ACKNOWLEDGEMENTS

Authors would like to thank Dr Devi J, NFSG, Department of Pediatrics, GGH Karaikal, and Dr Abarna L, NFSG, Department of Pediatrics, for coordinating the X-ray examination and reporting.

Funding: This study has been submitted for funding under the GJ Strauss Research Grant at JIPMER

Conflict of interest: None declared

Ethical approval: The study was approved by the Institutional Ethics Committee

REFERENCES

1. Arthur R. Interpretation of the paediatric chest X-ray. *Paediatr Respir Rev.* 2000;1(1):41-50.
2. Johnson J, Kline JA. Intraobserver and interobserver agreement of the interpretation of pediatric chest radiographs. *Emerg Radiol.* 2010;17(4):285-90.
3. Elemraid MA, Muller M, Spencer DA, Rushton SP, Gorton R, Thomas MF, et al. Accuracy of the interpretation of chest radiographs for the diagnosis of paediatric pneumonia. *PLoS One.* 2014;9(8):e106051.
4. Voigt GM, Thiele D, Wetzke M, Weidemann J, Parpatt PM, Welte T, et al. Interobserver agreement in interpretation of chest radiographs for pediatric community acquired pneumonia: findings of the pedCAPNETZ-cohort. *Pediatr Pulmonol.* 2021;56(8):2676-85.
5. Cherian T, Mulholland EK, Carlin JB, Ostensen H, Amin R, de Campo M, et al. Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bull World Health Organ.* 2005;83(5):353-9.
6. Kheddache S, Denbratt L, Angelhed JE. Digital chest radiography-optimizing image processing parameters for the visibility of chest lesions and anatomy. *Eur J Radiol.* 1996;22(3):241-5.
7. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging.* 2017;8(1):171-82.
8. Rotem-Grunbaum B, Scheuerman O, Tamary O, Lakovsky Y, Shkalim Zemer V, Goldberg L, et al. Pediatric chest radiograph interpretation in a real-life setting. *Eur J Pediatr.* 2024;183(10):4435-44.
9. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* 1971;76(5):378-82.
10. Geijer M, Båth M, Wessman C. Some common statistical methods for assessing rater agreement in radiological studies. *Acta Radiol.* 2025;66(6):675-83.

Cite this article as: James CN, Narayanasamy DK, Tanigasalam V, Munusamy J, Harish VK. Digital image display manipulation and observer variability in pediatric chest radiograph interpretation. *Int J Contemp Pediatr* 2026;13:2301-6.