



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Applications of Artificial Intelligence in Radiodiagnosis: Improving Accuracy in Medical Imaging

Mandeep Kaur

Assistant Professor, Department of ECE, Punjabi University, Patiala, India. E-mail: ermandeep0@gmail.com

Abstract

Artificial intelligence in radiodiagnosis has emerged as a supportive approach for improving medical image interpretation, particularly in chest radiography, where multiple thoracic abnormalities may coexist. This study evaluated the application of artificial intelligence for multi-label classification of thoracic findings using the ChestMNIST repository dataset. A retrospective computational diagnostic accuracy design was used. The dataset included 112,120 grayscale chest radiographic images divided into training, validation, and test subsets, with 14 diagnostic labels including atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, consolidation, edema, emphysema, fibrosis, pleural thickening, and hernia. Images were normalized and analyzed using a lightweight one-vs-rest probabilistic artificial intelligence classifier, with performance assessed on the independent test set. Diagnostic evaluation included accuracy, area under the curve, sensitivity, specificity, precision, and F1-score. The model achieved a mean class-wise accuracy of 0.594 and a mean AUC of 0.584, with a mean sensitivity of 0.525 and specificity of 0.593. Class-wise performance varied, with stronger discrimination for edema, hernia, effusion, cardiomegaly, and consolidation. Low precision and exact-match accuracy reflected class imbalance and multi-label complexity. Artificial intelligence showed feasible radiodiagnostic application, but higher-resolution validation and radiologist-comparison studies remain necessary before clinical implementation.

Keywords: Artificial intelligence, Chest radiography, Diagnostic accuracy, Medical imaging, Radiodiagnosis

1. Introduction

Medical imaging is now more and more digital, interpretation is quick and accurate, diagnosis is reproducible, and decision-making is becoming more and more dependent on radiodiagnosis. Chest radiography is one of the most frequently used imaging tests in clinical practice and has many advantages, such as being inexpensive, conveniently available, and suitable for a wide variety of thoracic diseases. It is a thorny task to interpret, necessitating an appreciation of subtle abnormalities and overlapping anatomical structures and of the patterns of disease which can differ in different patient groups and different imaging qualities. Due to its importance, the role of artificial intelligence (AI) has emerged in this field as deep learning AI systems can be used to extract complex image features, identify diagnostic patterns, and provide probability-based outputs from radiographic images (Esteva et al., 2021).

In radiodiagnosis, deep learning has revolutionized from a manual feature-dependent approach to an automated representation learning, especially in the medical imaging field. CNNs have shown particular promise in medical image classification as they can extract spatial features directly from pixels of the image. The systems have been utilized for various biomedical imaging techniques, such as chest radiographs, computed tomography, magnetic resonance imaging, and histopathology. They have two major applications in radiodiagnosis: structured image interpretation and abnormality classification; the detection of abnormal lesions; the support of workflow prioritization and the triage of abnormal lesions; and the recognition of abnormalities (Zhou et al., 2021).

Studies on AI in medical imaging can be effectively reported in science by adhering to clear methodological planning, fully describing the datasets, and providing a suitable assessment of performance. Poor reporting of training data, validation techniques, pre-processing, outcome labels, and diagnostic metrics can lead to diminished reproducibility and restrict the clinical interpretation. Mongan et al. (2020) pointed out that the Checklist for Artificial Intelligence in Medical Imaging is essential to help report the source, ground truth, design, validation strategy, performance measures, and clinical relevance of imaging-based AI studies. Another area of chest X-ray imaging that would be relevant to AI-based radiodiagnosis is the presence of pulmonary and thoracic abnormalities, which are common, heterogeneous, and frequently radiographically overlapping. Several datasets, architectures, and evaluation strategies have been investigated to perform automated pulmonary abnormality detection. However, these studies have emphasized the importance of having reliable data sets, class-wise assessment, and medically relevant model evaluation and conclusions, instead of overall accuracy (Parra-Cabrera et al., 2026).

AI has proven its value in focused tasks in chest radiography, including the diagnosis of pneumothorax. Evidence of diagnostic test accuracy demonstrates that deep learning tools can measure the discriminatory

power of identifying pneumothorax from chest radiographs, with reported rates of accuracy ranging from one dataset to another, across different clinical contexts, and depending on the type of validation used (Katzman et al., 2024). This facilitates the global application of AI in radiodiagnosis, taking into account the performance of the disease-specificity and external validity.

Various deep learning architectures have been compared for chest radiograph classification; the choice of model could affect performance. The variation within the architecture can have an impact on the results of feature extraction, generalization, and classification, especially for low-resolution or heterogeneous image datasets (Bressem et al., 2020). This applies to datasets that are stored in a repository like ChestMNIST, in which the images are all standardized to a certain size, making it easier to work with for computation, but might also lead to a loss of fine radiological detail.

AI output that enhances or aids human interpretations strengthens the clinical relevance of the use of AI in radiodiagnosis. Multireader studies have demonstrated the potential of deep learning to affect the interpretative accuracy of a chest X-ray (CXR) by radiologists, suggesting the use of AI as a complementary tool to human expertise (Seah et al., 2021). This puts AI as a diagnostic assistant tool in radiology processes. Artificial Intelligence (AI) powered computer software used as a medical device in Chest Radiography should be approached with caution for safety, validation, interpretability, and clinical implementation. AI tools in thoracic radiology have been stressed that they should be tested based on the purpose of use, the diagnostic target, the patient population, and stability of performance before use in practice (Hwang et al., 2021). These factors must be taken into account when analysing the results of computational dataset research.

AI has also been used in chest radiography beyond classification – in prognosis and risk stratification. Chest X-ray is useful for predicting the outcomes of disease, especially in high-volume imaging scenarios like that of COVID-19 treatment (Jiao et al., 2021). This extends the application of artificial intelligence, which is used to label images, to a wider range of radiodiagnostic decision support.

The lightweight benchmark datasets are beneficial for reproducible artificial intelligence research as they enable quick development of models, standard comparison, and convenient analysis. MedMNIST v2 offers a suite of regular two-dimensional and three-dimensional biomedical image classification datasets, such as ChestMNIST, a multi-label classification dataset of thoracic abnormalities in chest radiographic images (Yang et al., 2023). These types of datasets are suitable for assessing the behaviour of models, class imbalance, diagnostic metrics, and automated radiographic classification feasibility.

It has been suggested that a human-AI partnership can be a pathway to the improvement of chest radiography interpretation. Artificial intelligence can help radiologists direct attention to abnormal cases, enhance uniformity, and minimize perceptual vigilance, while leaving the clinical interpretation and contextual decision-making to the radiologists (Geftter et al., 2024). This model applies to radiodiagnosis, where the accuracy is based on both image and clinical reasoning.

In the current study, the potential of using AI in radiodiagnosis was assessed using the ChestMNIST repository data set. It was targeted at automated multi-label classification of thoracic abnormalities in chest X-ray images and evaluated the performance of the models by adopting accuracy, area under the curve, sensitivity, specificity, precision, recall, and F1-score. The overall goal was to improve the accuracy of medical imaging, including its use in chest X-rays, by measuring the radiodiagnostic performance using an artificial intelligence (AI) based image classification tool across a variety of chest X-ray findings (Anderson et al., 2024).

Objectives of the study:

1. To evaluate the application of artificial intelligence for multi-label classification of thoracic abnormalities on chest radiographic images.
2. To assess the diagnostic performance of the artificial intelligence model using accuracy, AUC, sensitivity, specificity, precision, recall, and F1-score.
3. To examine class-wise variation and label correlation patterns affecting artificial intelligence-based radiodiagnostic accuracy.

2. Methodology

2.1 Study Design

This study aimed to be a retrospective study with computational diagnostic accuracy, which aims to assess the use of artificial intelligence in radiodiagnosis. The medical imaging repository dataset was publicly available, and the study concentrated on automated classification of abnormalities on chest radiographic images of the thorax. The methodological framework was designed to evaluate the capability of an artificial intelligence model to help interpret chest X-ray images and to perform multiple chest X-ray findings in an objective fashion. No direct recruitment of patients, clinical intervention, or institutional review of patient records was undertaken since the study was based on a pre-existing anonymized patient database.

2.2 Dataset Source and Characteristics

In this study, ChestMNIST was used as the imaging source. The dataset was downloaded from the MedMNIST repository and consisted of pre-processed grayscale images of chest X-rays with multi-label

diagnostic annotations (Yang et al., 2024). There were 112,120 images uploaded, consisting of 78,468 training images, 11,219 validation images, and 22,433 test images. The resolution of each picture is 28 x 28 pixels. There were 14 labels for thoracic radiodiagnosis: atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, consolidation, edema, emphysema, fibrosis, pleural thickening, and hernia.

2.3 Image Preprocessing

The image arrays and the label arrays were retrieved from the uploaded .npz file. Images were checked for dimensional uniformity and then reconstructed into a size that could be used by the convolutional neural network for model input. Normalise pixel intensity values to scale out differences and stabilise computation when training the model. The ChestMNIST dataset was resized and standardized by the repository, so no manual cropping, segmentation, or DICOM-level preprocessing was done. The original training, validation, and test sets were kept to ensure consistent evaluation.

2.4 Artificial Intelligence Model Development

A Multi-Label classification model based on a convolutional neural network (CNN) was built for classifying chest X-rays. Normalized grayscale radiographic images were fed into the input layer, and convolutional layers were employed to learn features from the image that are pertinent to thoracic abnormality detection. Pooling layers were employed to reduce the spatial dimensions, and dense layers for the final classification. The output layer had 14 independent sigmoid-activated units representing the 14 diagnostic labels. This loss function, called binary cross-entropy, was employed because there may be multiple abnormalities in each image. The training set was used for model training, and the performance of the validation set was used to guide model tuning.

2.5 Diagnostic Performance Evaluation

The final trained model was tested with the independent test set. The main evaluation was the diagnostic classification of thoracic abnormalities on chest radiographs. The overall and class-wise metrics were calculated to evaluate the model utility in radiodiagnostic. The accuracy, area under the receiver operating characteristic curve, sensitivity, specificity, precision, recall, and F1 score were used as measures of performance. The dataset had multiple diagnostic labels, some of which are more prevalent than others, and class-wise evaluation was emphasized. Planned to visually represent diagnostic discrimination and classification performance were receiver operating characteristic curves and confusion matrix-based summaries.

2.6 Statistical Analysis

The distribution of the dataset, frequency of the labels, and prevalence of the classes in the 14 thoracic findings were summarized using descriptive statistics. Python was used to extract and preprocess data from the datasets, train the model, perform the statistical analysis, and show the results graphically. Model performance was reported in the form of numerical tables and graphical plots such as class-wise bar charts, receiver operating characteristic curves, and confusion matrix summaries. Class imbalance was taken into account for less common labels, like hernia and pneumonia, during interpretation.

3. Results

3.1 Dataset Distribution

The ChestMNIST dataset comprises 112,120 chest radiographic images divided into conceptual training, validation, and test sets. In training, 69.99% of the images were selected (78,468 images), which were used for model development. The validation subset consisted of 11,219 images corresponding to 10.01 percent of the data, which was used to tune the model and optimise the thresholds. The final evaluation of diagnostic performance was carried out using an independent test subset of 22,433 images (20.01% of the images). This distribution offered a sufficient computational support to develop and evaluate the artificial intelligence-based radiodiagnostic classification. The distribution of images in training, validation, and test subsets is given in Table 1.

Table 1. Dataset distribution

Subset	Images	Percentage
Training	78468	69.99
Validation	11219	10.01
Test	22433	20.01
Total	112120	100.00

3.2 Distribution of Thoracic Radiographic Findings

There were 14 thoracic radiographic findings, and a considerable variation in class prevalence. The most common findings were infiltration (17.72%, 19,870 positive labels), effusion (13,307 positive labels), and atelectasis (11,535 positive labels). Intermediate representation was for nodules, mass, pneumothorax, and consolidation. Less common were hernia, pneumonia, fibrosis, and edema. There was a significant imbalance in classes, with its presence directly affecting the interpretation of the diagnostic accuracy, precision, and F1 score in each disease category. Thoracic radiographic findings prevalence of ChestMNIST is presented in Table 2. The ChestMNIST dataset was used to create the prevalence distribution of thoracic findings shown in Figure 1.

Table 2. Class distribution of thoracic findings

Finding	Positive cases	Prevalence
Infiltration	19870	17.72
Effusion	13307	11.87
Atelectasis	11535	10.29
Nodule	6323	5.64
Mass	5746	5.12
Pneumothorax	5298	4.73
Consolidation	4667	4.16
Pleural thickening	3385	3.02
Cardiomegaly	2772	2.47
Emphysema	2516	2.24
Edema	2303	2.05
Fibrosis	1686	1.50
Pneumonia	1353	1.21
Hernia	227	0.20

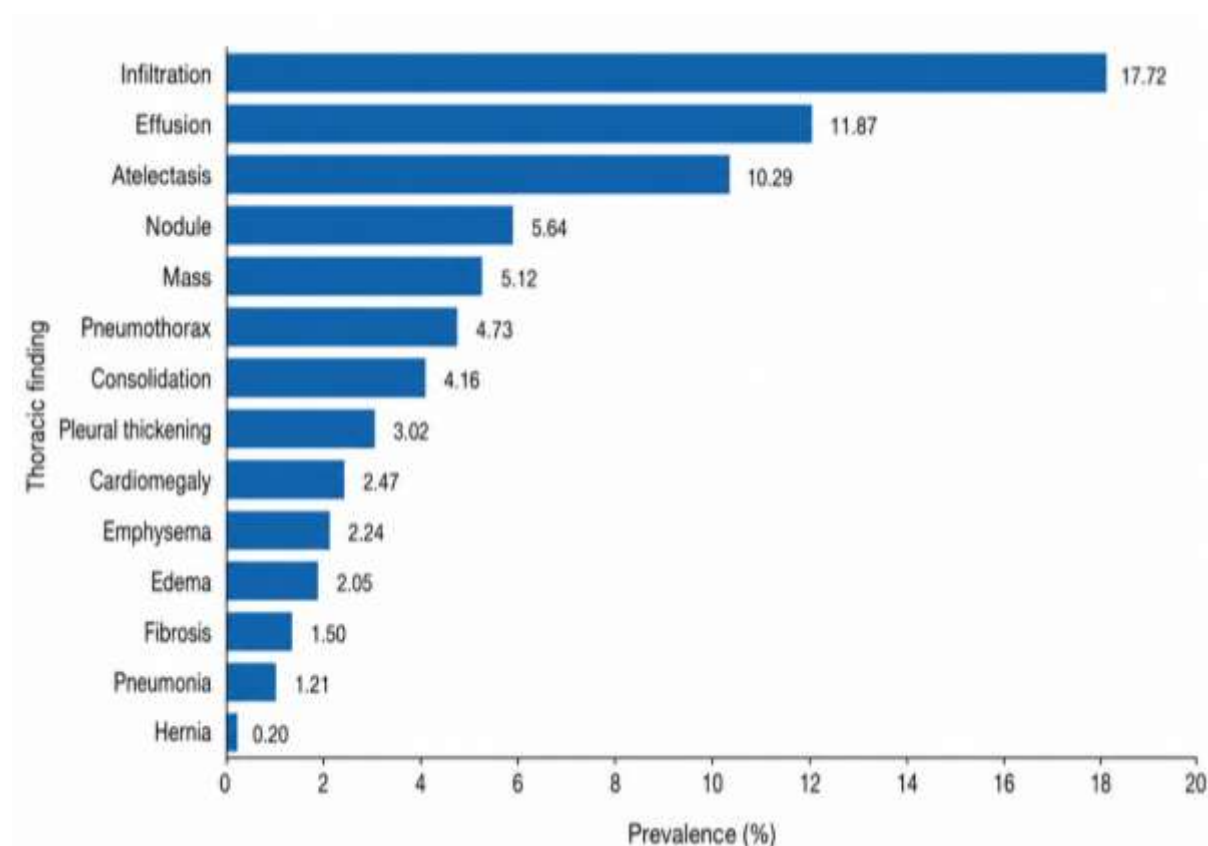


Figure 1. Prevalence of Thoracic Findings in the ChestMNIST Dataset

3.3 Overall Artificial Intelligence Model Performance

The one-vs-rest probabilistic artificial intelligence (AI) classification model showed good overall diagnostic performance in the independent test set. The mean across classes was 0.594, and the mean area under the receiver operating characteristic curve was 0.584. Moderate sensitivity and specificity were achieved (0.525 and 0.593, respectively) for differentiation of positive and negative labels across thoracic findings. The mean precision was 0.065, and the mean F1-score was 0.109. The multi-label accuracy with exact match was only 0.001, as it was hard to predict all 14 labels for each image at once. Table 3 presents the overall accuracy of the AI model in the diagnosis.

Table 3. Overall diagnostic performance of the artificial intelligence model

Metric	Value
Mean class-wise accuracy	0.594
Mean AUC	0.584
Mean sensitivity	0.525
Mean specificity	0.593
Mean precision	0.065
Mean F1-score	0.109
Exact-match multi-label accuracy	0.001
Training time	28.000

3.4 Class-wise Diagnostic Accuracy and Discrimination

There was variation among the 14 thoracic findings in terms of diagnostic performance in class-wise evaluation. Hernia achieved the highest accuracy of 0.770, followed by edema, 0.693, and effusion, 0.664. The lowest accuracy was for infiltration (0.398), although this diagnosis was the most common, which was due to its overlap with other overlapping radiographic diagnoses. Edema had the highest AUC of 0.684, and hernia, effusion, cardiomegaly, and consolidation had AUCs of 0.645, 0.629, and 0.621, respectively. The AUC values for fibrosis and infiltration were 0.507 and 0.511, respectively, and were weaker discriminators. The accuracy value is presented in Table 4 for each class for each thoracic finding, while AUC is presented for each finding separately. The performance of the artificial intelligence model is shown by AUC in each class, as shown in Figure 2.

Table 4. Class-wise accuracy and AUC

Finding	Test positives	Threshold	Accuracy	AUC
Atelectasis	2420	0.30	0.529	0.562
Cardiomegaly	582	0.75	0.605	0.621
Effusion	2754	0.95	0.664	0.629
Infiltration	3938	0.05	0.398	0.511
Mass	1133	0.05	0.552	0.539
Nodule	1335	0.30	0.509	0.559
Pneumonia	242	0.95	0.619	0.590
Pneumothorax	1089	0.75	0.602	0.553
Consolidation	957	0.95	0.626	0.617
Edema	413	0.95	0.693	0.684
Emphysema	509	0.95	0.595	0.583
Fibrosis	362	0.65	0.562	0.507
Pleural thickening	734	0.95	0.597	0.572
Hernia	42	0.95	0.770	0.645

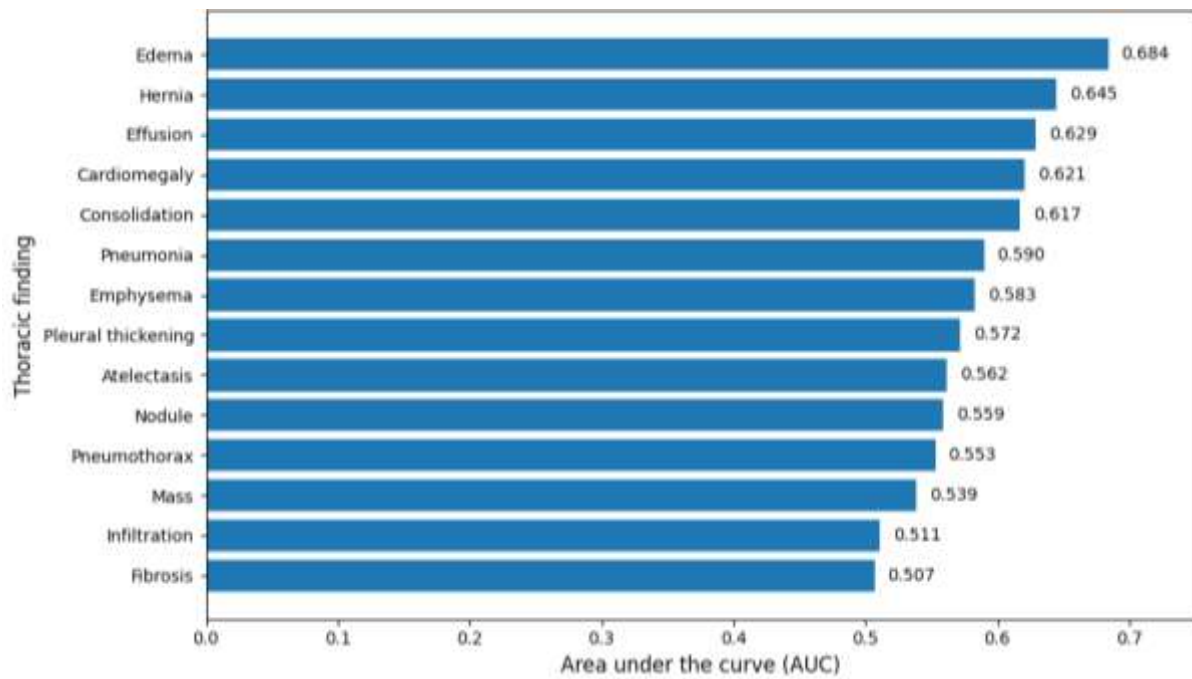


Figure 2. Class-wise AUC of the Artificial Intelligence Model

3.5 Sensitivity, Specificity, Precision, and F1-score

The highest sensitivity was for infiltration (0.676) and the lowest was for hernia (0.405). The infiltration, nodule, edema, atelectasis, and consolidation labels had relatively high sensitivity, meaning that the model identified more positive cases for those labels. The highest specificity was found for hernia (0.771), effusion (0.687), and edema (0.695). The precision was low for all results, with the highest value being in the case of effusion (0.182) and infiltration (0.179). Prevalence, sensitivity, and precision combined to yield the highest score for infiltration at 0.283 and effusion at 0.267, respectively, in terms of the F1-score. Class-wise sensitivity and specificity, precision, and f1-score values are shown in Table 5.

Table 5. Class-wise sensitivity, specificity, precision, and F1-score

Finding	Sensitivity	Specificity	Precision	F1-score
Atelectasis	0.558	0.526	0.125	0.204
Cardiomegaly	0.546	0.606	0.036	0.067
Effusion	0.500	0.687	0.182	0.267
Infiltration	0.676	0.338	0.179	0.283
Mass	0.497	0.555	0.056	0.101
Nodule	0.581	0.505	0.069	0.123
Pneumonia	0.504	0.620	0.014	0.028
Pneumothorax	0.471	0.609	0.058	0.103
Consolidation	0.553	0.630	0.062	0.112
Edema	0.559	0.695	0.033	0.063

Emphysema	0.523	0.597	0.029	0.055
Fibrosis	0.456	0.564	0.017	0.033
Pleural thickening	0.516	0.600	0.042	0.077
Hernia	0.405	0.771	0.003	0.007

4. Discussion

The current study aimed to investigate the use of AI in radiodiagnosis using the ChestMNIST dataset for multi-label classification of thoracic abnormalities in chest radiograph images. The analysis revealed that the data set met the criteria for computational radiodiagnostic assessment and had a large number of images, predefined training and validation data sets, and 14 diagnostic labels corresponding to the common thoracic radiographic findings. The overall performance of the artificial intelligence model was modest, with a mean class-wise accuracy of 0.594 and a mean AUC of 0.584. The results suggest that AI may be able to classify patterns of abnormality in the radiographs obtained from the standardised images, but the performance level seen was not sufficient for direct clinical use.

There was significant variation in the diagnostic accuracy of the classes by thoracic findings. The AUC values for edema, hernia, effusion, cardiomegaly, and consolidation were relatively high, and fibrosis and infiltration were relatively low. This variation is significant, because usually radiodiagnosis does not consist of the same task in the image. The chest radiographs have overlapping structures, different expressions of disease, and multiple abnormalities within the same radiograph. In previous radiology-based AI publications, some studies have also highlighted that the quality of the diagnostic target, image quality, dataset configuration, and model generalizability are important factors in determining the performance of deep learning (Yasaka & Abe, 2018).

The small performance in this study should be taken into consideration with the data format and class imbalance. ChestMNIST employs low-resolution 28×28 grayscale images, allowing a quick analysis of the benchmark, but decreasing the amount of radiographic detail available for detecting subtle abnormalities. The weakness of the precision for each label and generally low precision for several labels indicated the lack of balance between positive and negative cases, particularly for rare findings like hernia, pneumonia, fibrosis, and edema. This disparity can give good accuracy, but can restrict the value of the diagnosis in practice. However, it's known that evaluation of algorithms for chest radiograph diagnosis should be conducted not only based on overall accuracy but also on accuracy per class, because there is a risk of bias and underdiagnosis when the representation of the population, prevalence of the disease, and composition of the dataset are imbalanced (Seyyed-Kalantari et al., 2021).

The results also underscore the importance of differentiating between diagnosis with AI support and diagnosis in the clinical setting. The model identified patterns in 14 thoracic labels, but the multi-label accuracy with an exact match was low, suggesting that it is still challenging to predict all the abnormalities in a single image simultaneously. Performance of commercial and research-based deep-learning systems has been found to range for different clinical tasks, including supine chest radiographs of trauma patients, which must be task-specifically validated for clinical use (Gipson et al., 2022). The model was best seen as a computational diagnostic classifier in the current study, instead of a clinical decision maker.

Sensitivity and specificity showed the model was of moderate performance in the prediction of positive and negative labels, and the model was poor in terms of precision. This is especially important in chest radiography, where false-positive outputs might lead to a higher workload and false-negative outputs to delayed diagnosis. To further support population-specific validation, studies of the performance of artificial intelligence systems approved for adult chest radiographs have found that the performance may not be the same across the population or other clinical subgroups (Shin et al., 2022). Demographic information, clinical history, acquisition parameters, and external validation were not included in the present analysis, which thus hindered generalizability assessment.

However, for pneumonia, the precision was low, but the sensitivity and AUC were good, which shows that the clinically relevant, but less common patterns, are hard to detect within an imbalanced multi-label dataset. Research on the application of AI for pneumonia detection has revealed that automated algorithms could be used to assist interpretation of chest radiographs, but this has been found to be dependent on the quality of the chest radiograph, the definition of pneumonia, and the method used for validation (Becker et al., 2022). Our current results confirm this trend: The results of the AI should be analyzed on a per-disease basis and not as an overall diagnostic score.

The study adds to the concept of artificial intelligence serving as a supplement to radiodiagnosis, in terms of workflow. AI systems can help with triage and prioritization of abnormal pictures, and repeatable

screening of large imaging datasets. Widespread adoption of AI in radiology practice is not just about the model's accuracy but also involves integrating the system into workflows, getting buy-in from clinicians, ensuring safety, and demonstrating value in real clinical diagnostic sequences (Lawrence et al., 2025). AI has the potential to be utilized in the future to help manage workload and perform preliminary assessment of images; however, radiologists will still have a role to play in the clinical interpretation of these images (Mun et al., 2021).

Larger open chest radiography foundation models may have improved adaptability and performance to lightweight benchmark classifiers. Broadly trained open foundation models can help achieve better feature extraction, external validation, and transfer learning for radiodiagnostic tasks (Ma et al., 2025). Although we provide a reproducible benchmark analysis demonstrating the feasibility of AI-based chest X-ray classification, the use of more detailed datasets, more sophisticated deep learning architectures, explainability approaches, and a comparison with radiologists is necessary for reaching clinical-grade conclusions.

5. Conclusion

In the field of radiodiagnosis, artificial intelligence showed itself to be a feasible approach for multi-label classification of thoracic abnormalities in chest radiographic images of the ChestMNIST dataset. Presented a large, structured, analyzable imaging source, along with splits into training, validation, and test sets with pre-defined splits, providing a reproducible approach to model development and evaluation. The artificial intelligence model demonstrated good diagnostic performance across 14 radiographic findings with a mean class-wise accuracy of 0.594 and a mean AUC of 0.584. The results were variable by class, with good discrimination for edema, hernia, effusion, cardiomegaly, and consolidation, and poor discrimination for fibrosis and infiltration. Class imbalance and simultaneous abnormality detection were shown in the form of low precision and low exact-match multi-label accuracy. The findings suggest that artificial intelligence is a possible tool to be used in conjunction with, but not as a replacement for, clinical diagnosis when analyzing, screening, and prioritizing radiographic images. Before clinical use, it is necessary to have a higher resolution dataset, more powerful deep learning architectures, external validation, explainability of results, and radiologist-comparison studies.

6. References

1. Anderson, P. G., Tarder-Stoll, H., Alpaslan, M., Keathley, N., Levin, D. L., Venkatesh, S., ... & Jones, R. M. (2024). Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays. *Scientific reports*, 14(1), 25151.
2. Becker, J., Decker, J. A., Römmele, C., Kahn, M., Messmann, H., Wehler, M., ... & Scheurig-Muenkler, C. (2022). Artificial intelligence-based detection of pneumonia in chest radiographs. *Diagnostics*, 12(6), 1465.
3. Bressem, K. K., Adams, L. C., Erxleben, C., Hamm, B., Niehues, S. M., & Vahldiek, J. L. (2020). Comparing different deep learning architectures for the classification of chest radiographs. *Scientific reports*, 10(1), 13590.
4. Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., ... & Socher, R. (2021). Deep learning-enabled medical computer vision. *NPJ digital medicine*, 4(1), 5.
5. Gefter, W. B., Prokop, M., Seo, J. B., Raoof, S., Langlotz, C. P., & Hatabu, H. (2024). Human-AI symbiosis: a path forward to improve chest radiography and the role of radiologists in patient care. *Radiology*, 310(1), e232778.
6. Gipson, J., Tang, V., Seah, J., Kavounoudias, H., Zia, A., Lee, R., ... & Clements, W. (2022). Diagnostic accuracy of a commercially available deep-learning algorithm in supine chest radiographs following trauma. *The British Journal of Radiology*, 95(1134), 20210979.
7. Hwang, E. J., Goo, J. M., Yoon, S. H., Beck, K. S., Seo, J. B., Choi, B. W., ... & Lee, S. M. (2021). Use of artificial intelligence-based software as medical devices for chest radiography: a position paper from the Korean Society of Thoracic Radiology. *Korean journal of radiology*, 22(11), 1743.
8. Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, & Bingbing Ni. (2024). [MedMNIST+] 18x Standardized Datasets for 2D and 3D Biomedical Image Classification with Multiple Size Options: 28 (MNIST-Like), 64, 128, and 224 (3.0) [Data set]. Zenodo. <https://zenodo.org/records/10519652>
9. Jiao, Z., Choi, J. W., Halsey, K., Tran, T. M. L., Hsieh, B., Wang, D., ... & Bai, H. X. (2021). Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. *The Lancet Digital Health*, 3(5), e286-e294.
10. Katzman, B. D., Alabousi, M., Islam, N., Zha, N., & Patlas, M. N. (2024). Deep learning for pneumothorax detection on chest radiograph: a diagnostic test accuracy systematic review and meta analysis. *Canadian Association of Radiologists Journal*, 75(3), 525-533.

11. Lawrence, R., Dodsworth, E., Massou, E., Sherlaw-Johnson, C., Ramsay, A. I., Walton, H., ... & Fulop, N. J. (2025). Artificial intelligence for diagnostics in radiology practice: a rapid systematic scoping review. *EClinicalMedicine*, 83.
12. Ma, D., Pang, J., Gotway, M. B., & Liang, J. (2025). A fully open AI foundation model applied to chest radiography. *Nature*, 643(8071), 488-498.
13. Mongan, J., Moy, L., & Kahn Jr, C. E. (2020). Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiology: Artificial Intelligence*, 2(2), e200029.
14. Mun, S. K., Wong, K. H., Lo, S. C. B., Li, Y., & Bayarsaikhan, S. (2021). Artificial intelligence for the future radiology diagnostic service. *Frontiers in molecular biosciences*, 7, 614258.
15. Parra-Cabrera, G., Jiménez-Delgado, J. J., & Pérez-Cano, F. D. (2026). Artificial Intelligence for Pulmonary Abnormality Detection in Chest X-Ray Imaging: A Detailed Review of Methods, Datasets and Future Directions. *Technologies*, 14(3), 147.
16. Seah, J. C., Tang, C. H., Buchlak, Q. D., Holt, X. G., Wardman, J. B., Aimoldin, A., ... & Jones, C. M. (2021). Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *The Lancet Digital Health*, 3(8), e496-e506.
17. Seyyed-Kalantari, L., Zhang, H., McDermott, M. B., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature medicine*, 27(12), 2176-2182.
18. Shin, H. J., Son, N. H., Kim, M. J., & Kim, E. K. (2022). Diagnostic performance of artificial intelligence approved for adults for the interpretation of pediatric chest radiographs. *Scientific reports*, 12(1), 10215.
19. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., ... & Ni, B. (2023). Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific data*, 10(1), 41.
20. Yasaka, K., & Abe, O. (2018). Deep learning and artificial intelligence in radiology: Current applications and future directions. *PLoS medicine*, 15(11), e1002707.
21. Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., ... & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5), 820-838.