Analysis and Metaphysics
ISSN: 1584-8574, e-ISSN: 2471-0849
Vol 24 (1), 2025
pp. 153-165

Artificial Intelligence in Nuclear Medicine: From Image Reconstruction to Clinical Decision Support

Abdullah S. Alharbi[1], Basel A. Qenam[2], and Anas M. Alsawat[3]

[1]Email: Aalbalaji@moi.med.sa

[1]Department of Radiology and Nuclear Medicine, Security Forces Hospital, Riyadh, Saudi Arabia

[2]Email: bqenam@ksu.edu.sa

[2]Department of Radiological Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

[3]Email: aalsawat@moi.med.sa

[3]Department of Radiology and Nuclear Medicine, Security Forces Hospital, Riyadh, Saudi Arabia

Abstract

This review aimed to answer the PICO question: In patients undergoing nuclear medicine imaging, does the use of artificial intelligence for image reconstruction and clinical decision support, compared to traditional methods, improve diagnostic accuracy and clinical workflow efficiency? Google Scholar was searched to identify the relevant papers, and the PRISMA flow process was used to screen and select 15 papers, based on certain inclusion and exclusion criteria, for this review. The review showed that AI tools help in image reconstruction, improve diagnostic accuracy, clinical workflow efficiency and clinical decision support. AI also helps to provide an inclusive and equitable personalised care in different populations. Some AI tools may be better than others; therefore, the choice of the most appropriate AI tool for a given context is important. AI tools facilitate decision-making by multi-disciplinary teams and in preventive therapies, especially when considering pandemics like COVID-19. Some limitations of this review and the scope for future research have been indicated.

Keywords: Nuclear medicine, AI tools, image reconstruction, clinical decision support, diagnostic accuracy.

Introduction

The integration of Artificial Intelligence (AI) into nuclear medicine marks a transformative shift in both diagnostic precision and clinical workflow efficiency. Traditionally reliant on complex imaging modalities and expert interpretation, nuclear medicine now stands at the cusp of a paradigm change driven by advances in machine learning, deep learning, and data-driven analytics. From enhancing image reconstruction fidelity to enabling real-time clinical decision support, AI technologies are redefining the boundaries of what is possible in molecular imaging and personalised care.

This review provides a comprehensive overview of AI applications across the nuclear medicine continuum, beginning with algorithmic innovations in image acquisition and reconstruction, progressing through automated segmentation and quantification, and culminating in predictive modelling and decision support systems. It critically examines the technical foundations, clinical validations, and translational challenges associated with AI deployment, while highlighting emerging trends such as federated learning, explainable AI, and multimodal data fusion.

By synthesising current literature and identifying gaps in implementation, this article aims to guide researchers, clinicians, and policymakers in harnessing AI's full potential to improve diagnostic accuracy, streamline workflows, and ultimately enhance patient outcomes in nuclear medicine.

The next section outlines the methodology followed for this review. This is followed by the Results section, describing each selected paper. The subsequent section discusses and synthesises the findings from the reviewed papers along with some numerical trends. The limitations of this review are mentioned in the next section. After the Conclusion section, scope for further work is presented.

Methodology

Google Scholar was used for the identification of papers related to the topic. The review topic itself was used as the search terms, as other search terms like"AI for medical nuclear imaging and image reconstruction", "automated nuclear image segmentation and quantification", "AI innovations in predictive modelling of nuclear images" and "decision support systems for AI-based nuclear imaging" did not identify the required papers correctly.

The identified papers were screened and selected repeatedly using some inclusion and exclusion criteria (Table 1) and the PRISMA flow process to finally select 15 papers. The PRISMA flow process used for this review is appended.

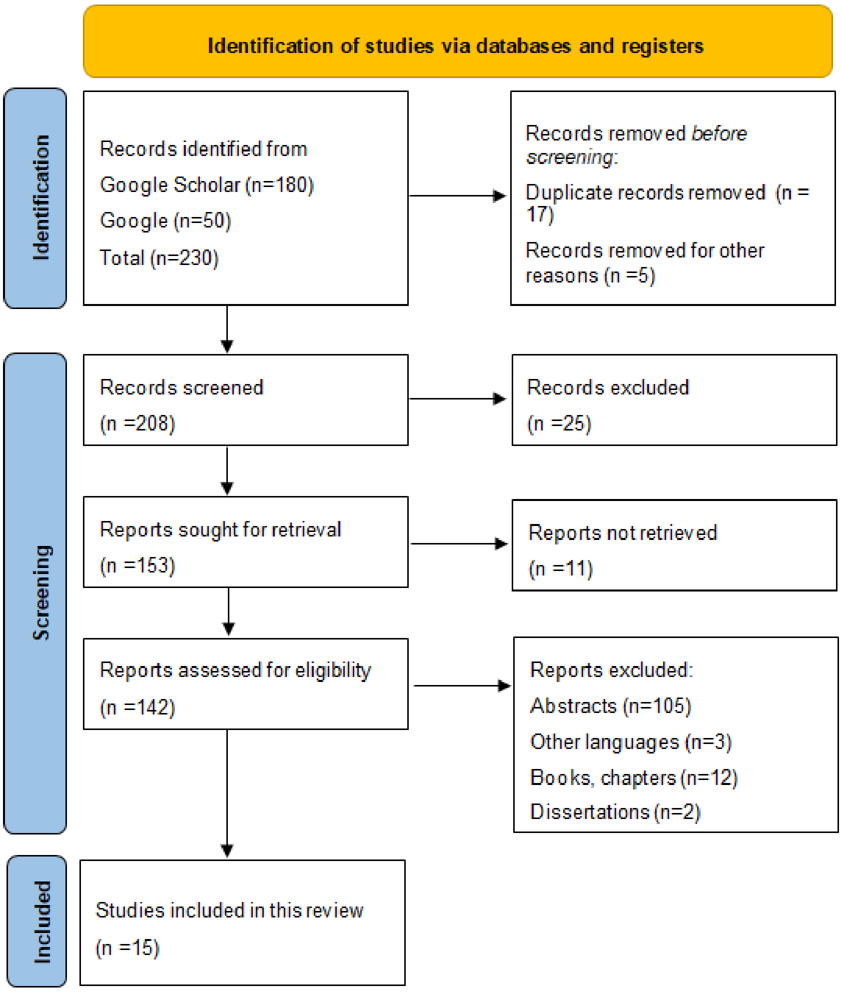
Table 1

Inclusion and exclusion criteria

Inclusion criteria	**Exclusion criteria**	**Comments**
Full-text research papers and reports.	Abstracts.	Abstracts may not contain all required information.
Papers in English	Papers in other languages.	Even the best translation can distort the information in papers published in other languages.
Published: 2024-2025.	Published earlier.	To reflect the latest trends.
	Dissertations.	Guided research.
	Books and chapters.	Full texts are not usually available in Google Scholar,
	Editorials, comments, etc.	Not research papers.
	Inadequate reference details.	Cannot be cited.

Figure 1

PRISMA



A PICO question was framed as the review question:

In patients undergoing nuclear medicine imaging, does the use of artificial intelligence for image reconstruction and clinical decision support, compared to traditional methods, improve diagnostic accuracy and clinical workflow efficiency?

(P (Population/Problem): Patients undergoing diagnostic imaging in nuclear medicine **I** (Intervention): Integration of Artificial Intelligence for image reconstruction and clinical decision support **C** (Comparison): Conventional nuclear medicine imaging and decision-making workflows without AI assistance **O** (Outcome): Improved diagnostic accuracy, workflow efficiency, predictive model efficiency, and clinical decision-making quality).

Assessment of the quality of papers

To assess the quality, an Excel file was constructed, in which the reference details, country of the study, the aim, method, findings, implications, recommendations, limitations, and risk of bias were entered for each paper. In the risk of bias, all types of biases are included. The quality scoring was as follows:

For an appropriate description of each entry, one point was awarded. Less than one point was awarded depending on the extent to which the entry deviates from appropriateness. Thus, the range of scores is 0 to 7 (for entries on aim to risk of bias).

This method of quality assessment was adopted for simplicity and convenience. An Excel file has been constructed with all the above points for each paper.

Results

In nuclear medicine, AI can be used for image generation and shorten the time for image acquisition. Provided an overview of the application of AI in image generation for single-photon emission computed tomography (SPECT) and positron emission tomography(PET), either without or with anatomical information [CT or magnetic resonance imaging (MRI)].

A review by Currie, Hawk, and Rohren (2025) pointed out that generative AI in nuclear medicine-spanning text and image generation-poses risks of gender and ethnic bias, misrepresentation, and output inaccuracies that may compromise safety. Its adoption across education, diagnostics, patient engagement, and marketing necessitates critical awareness of potential bias amplification and error propagation. To ensure ethical and effective use, integration must align with defined quality standards and professional criteria. While generative AI offers transformative potential in clinical, research, and educational domains, its assimilation demands both new competencies and a cautious, evaluative approach.

Hanna, Wakene, Lehmann, and Medford (2025) tasked four large language models (LLMs) with generating discharge instructions for HIV patients using identical encounter data, varied only by race/ethnicity (African American, Asian, Hispanic White, Non-Hispanic White). Prompts excluded explicit mention of race. Outputs were evaluated for sentiment, subjectivity, readability, and word usage. While most linguistic measures showed no significant variation across racial groups, GPT-4 revealed a statistically significant difference in entity count. However, post hoc analysis found no significant pairwise differences after Bonferroni correction. The urgent need to establish global standards for measuring bias in LLM-generated outputs was stressed by the authors. The limitations are the small sample size of 100 patients, focusing only on encounters with HIV patients, and only specific racial/ethnic groups and linguistic metrics may not indicate racial/ethnic bias.

Iglesias, Talavera, Troya, Díaz-Álvarez, and García-Remesal (2024) presented an AI-driven comparative diagnostic system for brain tumour evaluation, enabling the retrieval of similar cases from institutional datasets to support new patient assessments. Unlike prior models, it generated enriched image descriptors from binary data, bypassing the need for costly tumour segmentation. Achieving a Dice coefficient of 0.474 across both healthy and tumoral regions, the architecture outperformed existing methods. By effectively extracting anatomical and pathological features from brain MRIs using minimal label information, the model reduces training costs while enhancing diagnostic precision and efficiency. Using only segmentation labels to evaluate the model's performance, and the need for labels to train are some limitations to address in future research.

Myocardial perfusion imaging (MPI) is a critical tool for assessing cardiac function and diagnosing coronary artery disease (CAD). To enhance diagnostic accuracy and patient outcomes, artificial intelligence (AI) techniques have been integrated into MPI workflows. Erdagli, Ozsahin, and Uzun (2024) tested a novel framework to evaluate how diagnostic criteria and performance metrics influence the selection of MPI modalities and AI models for CAD detection. The analysis encompasses leading convolutional neural networks (CNNs)-InceptionV3, VGG16, ResNet50, DenseNet121, and machine learning (ML) algorithms such as random forests (RF), K-nearest neighbour (KNN), support vector machine (SVM), and Naïve Bayes (NB). It also compares nuclear MPI methods (PET and SPECT) with the non-nuclear approach of cardiovascular magnetic resonance imaging (CMR). AI models were assessed using metrics like F1-score, recall, specificity, precision, accuracy, and AUC-ROC, while MPI techniques were evaluated based on sensitivity, specificity, radiation exposure, cost, and scan duration. Using the PROMETHEE multi-criteria decision-making method, RF emerged as the topperforming AI model for SPECT-based CAD diagnosis (Φnet = 0.3778), and CMR was identified as the most effective MPI technique (Φnet = 0.3666). SPECT ranked lowest in overall diagnostic utility.

de Bourbon, Ahamed, Blanc-Durand, Rahmim, and Klein (2025) evaluated the detection limits of AI algorithms for identifying synthetic lesions in PET/CT imaging. A library of 565 synthetic lesions was generated using the Lesion Synthesis Toolbox (LST) and embedded into raw PET data and reconstructed CT images from 56 disease-free patients. Two external teams tested their AI models on this dataset, with lesion detection outcomes classified as hits, misses, or false positives. Psychophysical response models were applied to assess AI performance, revealing improved detection with larger lesion size and higher contrast. One AI consistently outperformed the other in sensitivity, precision, and lesion count. For instance, 10 mm lesions were detected with 90% sensitivity at 8:1 and 16:1 contrast ratios, respectively. Similarly, 3:1 contrast lesions required diameters of ~16 mm and ~31 mm for reliable detection. AI-segmented lesions showed reduced contrast and size, reflecting partial volume effects inherent to PET imaging. One limitation in this study is the method of characterising detected and missed lesions. Using the proximity (7 mm) of centres of mass as a criterion for lesion detection may have mis-scored large, overlaying lesions as a miss if they were off-centred. Use of only spherical lesions is another limitation. More obvious lesions would have been desirable for inclusion in the lapse rate parameter and modelling for lesion detection precision.

Hajianfar, et al. (2025) aimed to develop deep learning (DL) models to diagnose coronary artery disease (CAD) using supervised and semi-supervised approaches. They leveraged techniques like transfer learning and data augmentation, with SPECT-MPI and invasive coronary angiography (ICA) as reference standards. The dataset consisted of 940 patients who underwent SPECT-MPI. A subset of 281 patients was used in the study. Quantitative Perfusion SPECT (QPS) was used for image processing. Two tasks were defined. Task 1 consisted of CAD diagnosis using an expert reader (ER) assessment of SPECT-MPI as reference. Task 2 consisted of CAD diagnosis using ICA reports as a reference. DL models with 13 architectures, four variations of inputs, six different training strategies (especially for Task 2), with or without augmentation (WAug and WoAug) were used for model development. Totally, 728 models were trained. For evaluation, a subset of 100 ICA patients was used per vessel and per-patient analysis. The evaluation metrics were AUC, accuracy, sensitivity, specificity, precision, balanced accuracy, and statistical comparison (DeLong and Wilcoxon rank sum tests with 1000 bootstraps. In the case of Task 1, the best per-vessel model was DenseNet201 Late Fusion (AUC = 0.89), and the best per-patient model was ResNet152V2 Late Fusion (AUC = 0.83). In the case of Task 2, the best per-vessel model was Strategy 3 + WoAug InceptionResNetV2 Early Fusion (AUC = 0.71), and the best per-patient model was Strategy 5 (semi-supervised) + WoAug ResNet152V2 Early Fusion (AUC = 0.77). Saliency maps helped to highlight relevant regions in images, aiding interpretability. The limitations of this study are the small sample size with ICA reference, the use of acquired data within six months for SPECT imaging, with a threshold of 70%, not using scatter- and attenuation-corrected SPECT-MPI images in model development, and the use of only stress and rest polar maps as inputs to DL models.

This retrospective comparative study by Zabaleta, et al. (2025) evaluated whether OpenAI's GPT-3.5 Turbo (NLP-based AI) could assist a thoracic multi-disciplinary tumour board (MTB) in decision-making for patients with suspected or newly diagnosed non-small-cell lung cancer (NSCLC) between January and June 2023. The intervention consisted of GPT-3.5 with clinical case summaries and SEPAR lung cancer guidelines. For the task, one of five treatment options- follow-up, surgery, chemotherapy, radiotherapy, or chemoradiotherapy- was recommended.

The sample was 52 patients who were presented to the MTB. There was 76% agreement (moderate level) of AI-supported decisions with MTB decisions. Surgical recommendation replicability was 92.3% across four repetitions. Thus, AI demonstrated promising potential as a consistent and supportive tool for clinical decision-making in thoracic MTBs.

Urinary tract infection (UTI) remains one of the most prevalent infectious conditions in paediatric populations. However, the clinical utility of prophylactic antibiotics for children newly diagnosed with febrile pyelonephritis is debatable. To address this issue, Chen, et al. (2024) conducted a study leveraging ^{99m}Tc-DMSA renal static imaging to assess the necessity of preventive antibiotic therapy in children under two years of age at initial diagnosis. The self-curated dataset included 64 children who did not receive prophylactic treatment and 112 who did. By applying four established deep learning architectures ((a) AlexNet, (b) VGG16, (c) ResNet50, and (d) DenseNet201), the authors demonstrated the feasibility of stratifying patients based on their need for preventive antibacterial intervention. Notably, the AlexNet model achieved an accuracy of 84.05%, sensitivity of 81.71%, and specificity of 86.70%. These findings suggest that deep learning approaches hold promise for enabling computer-assisted decision support in the diagnostic evaluation of febrile pyelonephritis. This study has some limitations. Although AlexNet has shown promise in classifying images, the ideal accuracy is yet to be reached for direct use in the clinical scenario. AlexNet may not show such promising results in all cases of pyelonephritis. Radiation damage and image abnormalities may not be distinguishable. Only four models were tested. The sample size was small for generalisation.

Cell nuclei in pathology images are critical for diagnostic insights, yet segmentation remains challenging due to blurring, noise, and indistinct boundaries. To address these limitations, Li, He, Zhu, Gou, and Wu (2024) proposed DRPVit-a segmentation framework integrating deblurring and region proxies. The approach employs IE-IFANet to enhance nucleus clarity and RegProxy to segment regions via semantic agents, transformer encoding, and classification. A combined boundary–Tversky loss optimises edge and region accuracy, while post-processing removes artefacts and enables morphological analysis. Experimental results show DRPVit surpasses baseline models, improving IoU, Dice, and Recall by 3.3%, 2.3%, and 2.1%, respectively. The study is limited to a single modality.

Non-small cell lung cancer (NSCLC) often presents as Solitary Pulmonary Nodules (SPNs), making early and accurate diagnosis essential. While CNNs effectively detect SPNs from CT and PET scans, they lack interpretability. To address this, Papandrianos, et al. (2024) integrated DeepFCM CNN outputs with Fuzzy Cognitive Maps (FCMs), combining clinical data (e.g., age, gender, BMI, glucose) and SPN features (e.g., diameter, SUVmax, location) to predict malignancy. RGB-CNN predictions from PET images served as inputs to DeepFCM, whose fuzzy logic structure, initially expert-defined, is optimised via Particle Swarm Optimisation and Genetic Algorithms. Embedded in a Medical Decision Support System (MDSS), DeepFCM enhanced transparency using Grad-CAM for image region relevance and Natural Language Generation (NLG) to explain feature contributions in human-readable terms. Compared to the reports in the literature and the RGB-CNN model, the proposed model recorded higher accuracy, sensitivity, specificity, and precision and lower losses. The authors used 456 patients, consisting of 222 benign and 234 malignant cases.

The INCISIVE project, funded by the EU, aims to transform cancer imaging through an AI-powered toolbox. To support its implementation, a mixed-methods study by Hesso, et al. (2024) explored healthcare professionals' (HCPs) needs, expectations, and barriers. The study consisted of a Phase 1 using UX design workshops with 16 HCPs from six European countries, and a Phase 2 consisting of a Delphi survey with 12 HCPs ranking key features and challenges.

There was a strong consensus on the toolbox's potential to improve diagnosis accuracy, lesion detection, treatment planning, prognosis, and care integration. The major barriers identified were limited resources, lack of institutional support, inconsistent data entry and a high demand for AI explainability, including feature relevance and visual explanations. The limitations were the focus only on feature prioritisation, not broader success criteria, no comparative analysis with other AI tools, a small, geographically limited sample and a cross-sectional design lacking longitudinal insights. By integrating end-user perspectives, INCISIVE aims to build a clinically relevant, adoptable AI solution for cancer imaging.

Dabas, Kapp, and Gefen (2025) developed a generalisable algorithm for automated wound image analysis in clinical practice, focusing on accurate surface area measurement of pressure ulcers. Using image processing and the Hue-Saturation-Value (HSV) colour space to detect red tones, the algorithm segmented wound regions from nurse-captured images and calculated real-world wound size. Validation against annotated images showed high accuracy, with up to 0.85 Intersection over Union and full overlap with expert annotations. The method demonstrates strong potential for supporting clinical decision-making by tracking wound progression. Limitations include a small, variable-quality dataset, affecting generalizability and introducing bias. Future research directions are outlined to enhance robustness and scalability.

Transthyretin amyloid cardiomyopathy (ATTR-CM) is commonly underdiagnosed in patients with severe aortic stenosis (AS). In a prospective, single-centre study, Shiri, et al. (2025) evaluated AI-based models using preprocedural and routine data to detect ATTR-CM in AS patients scheduled for transcatheter aortic valve implantation (TAVI). In this study, 263 AS patients (mean age 83, 57% male) were screened for ATTR-CM using [99mTc]-DPD scintigraphy. Data sources included clinical, lab, ECG, echocardiography, invasive haemodynamic, 4D cardiac CT strain, and CT radiomics. Machine learning models were trained (70%) and tested (30%) across 100 random seeds. ATTR-CM was confirmed in 27 patients (10.3%). The lowest performance was for ECG and invasive data (AUC < 0.68). Moderate performance was for clinical, lab, imaging, and radiomics (AUC 0.70–0.76). Good performance was for Echocardiography (AUC 0.79). The best performance was for 4D-CT strain (AUC 0.85, sensitivity 0.90). The AUC for the multi-modality model was 0.84, but not better than 4D-CT strain alone. The model also predicted all-cause mortality risk over 13 months. Thus, AI models using pre-TAVI data can effectively detect ATTR-CM in severe AS patients, potentially replacing more invasive diagnostics like scintigraphy or biopsy.

Ferrández, et al. (2024) aimed to validate a previously developed deep learning model for predicting 2-year progression in diffuse large B-cell lymphoma (DLBCL) across five independent clinical trials, comparing its performance with the International Prognostic Index (IPI) and two radiomic PET/CT-based models. The HOVON dataset was used. A total of 1,132 patients were included, with 296 used for training and 836 for external validation. The deep learning model was trained on maximum-intensity projections from PET/CT scans, while the clinical PET model incorporated metabolic tumour volume, maximum lesion distance, SUV peak, age, and performance status, and the PET model included only the radiomic features. Model performance was evaluated using the area under the curve (AUC) and Kaplan–Meier survival analysis. The IPI yielded an AUC of 0.60, whereas the deep learning model achieved a significantly higher AUC of 0.66 ($P < 0.01$), consistently outperforming IPI across all trials. Both radiomic models demonstrated superior prognostic accuracy, with the clinical PET and deep learning models showing comparable performance (AUC 0.69), and the PET model achieving the highest AUC of 0.71 ($P < 0.05$). The deep learning model effectively predicted outcomes and separated survival curves across trials without requiring tumour delineation, although its prognostic performance was slightly lower than that of radiomics-based approaches. The limitations include assessing only 2-y TTP as the outcome variable, and the lack of activation maps to explain the predictions due to the design of the CNN with two branches.

Boukhennoufa, et al. (2024) proposed an automated approach for detecting parathyroid glands (PGs) by integrating statistical normalisation with artificial intelligence to differentiate between 99mTc-MIBI and iodine-123 images. This technique was tested on PINHOLE scans from 88 retrospective cases at a single centre, yielding a high average correlation of 0.95 when compared to expert physician assessments.

Salimi, et al. (2025) developed deep learning models to automatically detect and score transthyretin amyloid cardiomyopathy (ATTR-CM) from total body scintigraphy images across multi-tracer, multi-scanner, and multi-centre datasets. Six datasets were used for different tasks and purposes. Dataset 1 (93 patients, 99mTc-MDP) was used to develop the 2D planar segmentation and localisation models. Dataset 2 (216 patients,99mTc-DPD) was used for the detection (grade 0 vs. grades 1, 2, and 3) and scoring (0 and 1 vs. grades 2 and 3) of ATTR-CM. Datasets 3 (41 patients, 99mTc-HDP), 4 (53 patients, 99mTc-PYP), and 5 (129 patients, 99mTc-DPD) were utilised as external centres. Two physicians at each centre performed ATTR-CM detection and scoring. Dataset 6, comprising 3215 patients without labels, was used for retrospective evaluation of the model's performance. Different regions of interest were cropped and fed into the classification model for the detection and scoring of ATTR-CM. Ensembling was performed on the outputs of various models to enhance their performance. Model performance was assessed by classification accuracy, sensitivity, specificity, and AUC. Grad-CAM and saliency maps were generated to explain the model's decision-making process. In internal evaluations, all models demonstrated robust performance, achieving AUC values exceeding 0.95 and F1 scores above 0.90 for both detection and scoring tasks. External validation across datasets 3, 4, and 5 yielded detection AUCs of 0.93, 0.95, and 1.0, respectively, while scoring AUCs were 0.95, 0.83, and 0.96. In the retrospective analysis of dataset 6, the model identified ten potential ATTR-CM cases, with four subsequently validated by nuclear medicine specialists as plausible. Visualisation techniques such as Grad-CAM and saliency mapping confirmed that the models consistently attended to diagnostically relevant cardiac regions. The limitations were: the scoring models were divided into only 0 and 1 vs. 2 and 3 grades due to the limited number of cases in different classes, and a highly unbalanced training dataset for four-class classifications; not involving human experts for evaluations, use of only one large dataset for model performance evaluation; not evaluating the impact of the diagnostic model on patient outcomes; and all ground truth for the ATTR-CM was defined visually by nuclear medicine physicians. The proposed framework presents multiple clinically impactful applications that strengthen diagnostic workflows across diverse environments. It serves as an effective screening mechanism for identifying ATTR-CM across various imaging modalities. By minimising both inter- and intra-observer discrepancies, the algorithm promotes greater diagnostic consistency and reliability. Additionally, its scalability supports extensive retrospective and prospective screening, facilitating the identification of ATTR-CM within large repositories of bone scintigraphy data.

Discussion

Some quantitative trends are discussed first.

Country of study

The frequency distribution of the 15 papers among different countries is presented in Table 2.

Table 2

Country of study

Country	No.
USA	1
Spain	2
Turkey & UAE	1
Canada	1
International	3
China	2
Greece	1
Israel	1
Switzerland	1
Netherlands	1
France	1

Studies involving three or more countries dominated with three papers. There were two papers, one from Spain and one from China. Only one paper was found with a combination of two countries, and from the other seven countries. This distribution of countries is not generalisable as the trend is based on only 15 papers.

Aim of the study

The frequency distribution of papers according to the aim of the study is presented in Table 3.

Table 3

Aim of the study

Aim	No.
New model. Approach, framework,	6
Enhance	2
Evaluate	4
Support	1
Others	2

There were six papers describing the development of a new model, approach, or framework. This was followed by four evaluation studies. Two papers studied the enhancement effects of AI-based applications. Two others were undifferentiated. This trend is limited to this review as it is based on 15 papers only.

Method of study

The frequency distribution of papers according to their methods of study is presented in Table 4.

Table 4

Method of study

Method	No.
Cross-sectional	1
Evaluation	8
Comparison	3
Integration	1
Mixed	1
Case-study	1

Evaluation of models or frameworks was done in eight papers. Three papers compared models or datasets. There was one cross-sectional, one integration, one mixed approach and one case study. These trends may not reflect the general trend since it is based on only 15 papers.

Mention of implications

Out of 15, 14 papers mentioned implications, and one did not.

Recommendations

While 11 papers provided recommendations from their findings, four papers did not.

Study limitations

Out of 15 papers, 12 mentioned limitations, and the remaining three did not.

Risk of bias

Table 5 provides the different levels of risk of biases identified from their full texts.

Table 5

Risk of bias

RoB	No.
Low	5
Medium	8
High	2

Overall, medium levels of risk of bias could be detected from the reviewed papers. There were five, eight and two papers containing low, medium and high levels of risks, respectively.

Quality of papers

The overall quality scores of the reviewed papers are presented in Table 6.

Table 6

Quality of papers

Quality	No.
< 4	2
4 to 5.5	8
>5.5	5

The total score possible was seven. Suppose we consider four or less as poor, 4-5.5 as medium and over 5.5 as high quality, medium quality dominated with eight papers. There were five high-quality papers. Only two papers were of low quality. Thus, the overall trend was medium to high quality.

Integration of findings

Overall, there is strong evidence for a positive role of AI applications in medical image reconstruction and decision support.

AI-based prompts can help to exclude reference to race, ethnicity or gender in clinical records, which can lead to biased care (Hanna et al., 2025). AI can support clinical assessments of new patients by enhancing images through their reconstruction (Iglesias et al., 2024). Some AI tools are more efficient than others in the diagnosis of many chronic ailments (Erdagli et al., 2024; de Bourbon et al., 2025; Papandrianos et al., 2024; Hesso et al., 2024; Shiri et al., 2025; Ferrández et al., 2024; Boukhennoufa et al., 2024; Salimi et al., 2025). However, some task-wise differences may be observed (Hajianfar et al., 2025). Image enhancement by AI tools was noted by Li et al. (2024) and by Dabas et al. (2025) to support decision-making systems.

Some AI tools can function as decision-support systems. Zabaleta et al. (2025) noted AI support for multi-disciplinary teams in the case of thoracic tumour. Similarly, Chen et al. (2024) observed AI support for preventive antibiotic interventions in children with febrile pyelonephritis.

Thus, research on AI tools as decision support systems seems to be comparatively rare.

Limitations of this review

Using only Google Scholar to identify papers might have missed many papers available in databases. Even after scanning both 12 pages of Google Scholar and five pages of Google, only 15 papers could be selected which met all the criteria. Limiting the search to the papers published from 2024 to 2025 might have missed some good papers published in earlier years. Limiting to the papers published in English might have missed many good papers in other languages.

Although a meta-analysis was preferable, the inconsistencies of variables and their units of expression across the selected papers prevented the scope for this.

Since this review used only 15 papers, any generalisation of the findings needs to be done carefully.

Conclusions

This review aimed to answer the PICO question-

In patients undergoing nuclear medicine imaging, does the use of artificial intelligence for image reconstruction and clinical decision support, compared to traditional methods, improve diagnostic accuracy and clinical workflow efficiency?

The review showed that AI tools help in image reconstruction, improve diagnostic accuracy, clinical workflow efficiency and clinical decision support. AI also helps to provide an inclusive and equitable personalised care in different populations. Some AI tools may be better than others; therefore, the choice of the most appropriate AI tool for a given context is important.

AI tools facilitate decision-making by multi-disciplinary teams and in preventive therapies, especially when considering pandemics like COVID-19.

Scope for future research

This review indicated some aspects requiring more research. Research on this topic appears to be less from certain countries. This situation can be addressed by undertaking multi-country studies, which can include these countries. Research on medical image enhancement and decision support seems to be much less. More research in these aspects will be useful. Among methods, integration, case studies and use of mixed approaches were used by very few papers. Many papers did not address the risks of various biases appropriately. Future papers need to pay more attention to this aspect. Attention to the quality of the paper can improve the current trend of medium to high, to high, to very high range.

References

Boukhennoufa, O., Comas, L., Nicod, J.-M., Ungureanu, C., Zerhouni, N., & Boulahdour, H. (2024). Automatic detection of parathyroid glands in nuclear medicine. *2024 IEEE International Symposium on Biomedical Imaging (ISBI), 27-30 May 2024, Athens, Greece* (pp. 1-4). IEEE. doi:https://doi.org/10.1109/ISBI56570.2024.10635117

Chen, Z., Li, N., Chen, Z., Zhou, L., Xiao, L., & Zhang, Y. (2024). Computer-assisted decision support for the usage of preventive antibacterial therapy in children with febrile pyelonephritis: A preliminary study. *Heliyon, 10*(10), e31255. doi:https://doi.org/10.1016/j.heliyon.2024.e31255

Currie, G. M., Hawk, K. E., & Rohren, E. M. (2025). Generative artificial intelligence biases, limitations and risks in nuclear medicine: an argument for appropriate use framework and recommendations. *Seminars in Nuclear Medicine, 55*(3), 423-436. doi:https://doi.org/10.1053/j.semnuclmed.2024.05.005

Dabas, M., Kapp, S., & Gefen, A. (2025). Utilising image processing techniques for wound management and evaluation in clinical practice: Establishing the feasibility of implementing artificial intelligence in routine wound care. *Advances in Skin & Wound Care, 38*(1), 31-39. doi:https://doi.org/10.1097/ASW.0000000000000246

de Bourbon, Q., Ahamed, S., Blanc-Durand, P., Rahmim, A., & Klein, R. (2025). Characterisation of artificial intelligence performance for lesion detection using synthetic lesions in PET imaging. *Medical Physics, 52*(6), 3994–4007. doi:https://doi.org/10.1002/mp.17694

Erdagli, H., Ozsahin, D. U., & Uzun, B. (2024). Evaluation of myocardial perfusion imaging techniques and artificial intelligence (AI) tools in coronary artery disease (CAD) diagnosis through multi-criteria decision-making method. *Cardiovascular diagnosis and therapy, 14*(6), 1134. doi:https://doi.org/10.21037/cdt-24-237

Ferrández, M. C., Golla, S. S., Eertink, J. J., Wiegers, S. E., Zwezerijnen, G. J., Heymans, M. W., & Lugtenburg, P. J. (2024). Validation of an artificial intelligence–based prediction model using 5 external PET/CT datasets of diffuse large B-cell lymphoma. *Journal of Nuclear Medicine, 65*(11), 1802-1807. doi:https://doi.org/10.2967/jnumed.124.268191

Hajianfar, G., Gharibi, O., Sabouri, M., Mohebi, M., Amini, M., Yasemi, M. J., & Chehreghani, M. e. (2025). Artificial intelligence-powered coronary artery disease diagnosis from SPECT myocardial perfusion imaging: a comprehensive deep learning study. *European Journal of Nuclear Medicine and Molecular Imaging, 52*, 3019–3035. doi:https://doi.org/10.1007/s00259-025-07145-x

Hanna, J. J., Wakene, A. D., Lehmann, C. U., & Medford, R. J. (2025). Assessing Racial and Ethnic Bias in Text Generation by Large Language Models for Health Care–Related Tasks: Cross-Sectional Study. *Journal of Medical Internet Research, 27*, e57257. doi:https://doi.org/10.2196/57257

Hesso, I., Zacharias, L., Kayyali, R., Charalambous, A., Lavdaniti, M., Stalika, E., . . . Nabhani-Gebara, S. (2024). Artificial Intelligence for Optimising Cancer Imaging: User Experience Study. *JMIR cancer, 10*, e52639. doi:https://doi.org/10.2196/52639

Iglesias, G., Talavera, E., Troya, J., Díaz-Álvarez, A., & García-Remesal, M. (2024). Artificial intelligence model for tumoral clinical decision support systems. *Computer Methods and Programs in Biomedicine, 253*, 108228. doi:https://doi.org/10.1016/j.cmpb.2024.108228

Li, L., He, K., Zhu, X., Gou, F., & Wu, J. (2024). A pathology image segmentation framework based on deblurring and region proxy in medical decision-making system. *Biomedical Signal Processing and Control, 95*, 106439. doi:https://doi.org/10.1016/j.bspc.2024.106439

Papandrianos, N. I., Feleki, A. A., Apostolopoulos, I. D., Papageorgiou, E. I., Papathanasiou, N. D., Dimitrios, J., . . . Andreu-Perez, J. (2024). A Medical Decision Support System for Explainable Multimodal Detection of Non-Small Cell Lung Cancer Using Clinical and PET Data. *EXPLAINABILITY 2024 : The First International Conference on Systems Explainability* (pp. 27-33). IARIA. doi:https://doi.org/10.5281/zenodo.14631853

Salimi, Y., Shiri, I., Mansouri, Z., Sanaat, A., Hajianfar, G., Hervier, E., & Bitarafan, A. e. (2025). Artificial intelligence-based cardiac transthyretin amyloidosis detection and scoring in scintigraphy imaging: multi-tracer, multi-scanner, and multi-center development and evaluation study. *European Journal of Nuclear Medicine and Molecular Imaging, 52*(7), 2513-2528. doi:https://doi.org/10.1007/s00259-025-07117-1

Shiri, I., Balzer, S., Baj, G., Bernhard, B., Hundertmark, M., Bakula, A., & Nakase, M. e. (2025). Multi-modality artificial intelligence-based transthyretin amyloid cardiomyopathy detection in patients with severe aortic stenosis. *European Journal of Nuclear Medicine and Molecular Imaging, 52*(2), 485-500. doi:https://doi.org/10.1007/s00259-024-06922-4

Zabaleta, J., Aguinagalde, B., Lopez, I., Fernandez-Monge, A., Lizarbe, J. A., Mainer, M., . . . De Assas, M. (2025). Utility of artificial intelligence for decision making in thoracic multi-disciplinary tumor boards. *Journal of Clinical Medicine, 14*(2), 399. doi:https://doi.org/10.3390/jcm14020399