International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct-2026, PP 01-07

Artificial Intelligence For Medical Image And Report Detection: A Review Of Deep Learning, Ocr And Clinical Nlp For Healthcare Information Analysis

Kharat Yash Amar; Prof. M. C. Kolse; Prof. S. K. Said
Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India

Abstract—The growth of digital healthcare has produced large volumes of medical images and clinical reports, and reviewing them manually demands considerable time from radiologists, physicians and records staff. Artificial Intelligence (AI) can support these professionals by analysing images to highlight possible abnormalities and by extracting meaningful information from scanned and typed reports. This paper reviews an integrated AI approach to medical image and report detection. On the image side, it surveys deep convolutional networks for classification, segmentation networks for outlining structures, and visual explanation techniques that show where a model is looking. On the report side, it reviews Optical Character Recognition (OCR) for converting printed and scanned reports into text, layout-aware document models for extracting fields from forms, and biomedical natural language processing for recognizing findings, diagnoses, medications and negations. The link between the two, in which report text is used to label images and image findings are summarized in text, is also examined. A layered architecture covering input acquisition, preprocessing, parallel image and text pipelines, fusion and a clinician-review interface is proposed, together with research gaps, objectives and a comparison of techniques. The review concludes that combining image analysis with report understanding can reduce repetitive work and organize patient information, provided that systems are validated and qualified professionals verify every output.

Index Terms—Medical image analysis, deep learning, optical character recognition, clinical natural language processing, medical reports, healthcare informatics, decision support

I. Introduction

A patient's medical history is spread across images, such as X-rays, CT and MRI scans, and documents, such as laboratory reports, discharge summaries and prescriptions, many of which are still printed or scanned. When a doctor needs a quick picture of the patient, someone must find, read and interpret these records. Manual review is slow, inconsistent across reviewers and difficult to scale as hospitals digitize.

AI offers support on both fronts. Deep learning has transformed medical image analysis [1], while advances in OCR and natural language processing (NLP) make it possible to turn unstructured reports into searchable, structured data. Combining both capabilities allows a system to flag an image that may show an abnormality and, at the same time, extract the key findings from the accompanying report.

This paper reviews AI-based image and report detection, proposes an integrated architecture and discusses the conditions under which such systems can be safely used. The emphasis throughout is on assisting, not replacing, clinical judgement.

II. Related Work

Image analysis. Litjens *et al.* [1] surveyed deep learning across medical imaging and concluded that CNNs had become the dominant approach for classification, detection and segmentation. Rajpurkar *et al.* [2] trained a 121-layer DenseNet on chest X-rays and reported pneumonia detection comparable to radiologists. Ronneberger *et al.* [3] introduced U-Net, whose encoder–decoder structure with skip connections segments biomedical images accurately from relatively few training examples. McKinney *et al.* [4] demonstrated that an AI system could reduce both false positives and false negatives in breast-cancer screening.

Linking images and reports. Wang *et al.* [5] used NLP to mine disease labels from radiology reports, creating a large labelled chest X-ray dataset without manual annotation. Irvin *et al.* [6] improved this approach with a rule-based labeller that recognizes uncertainty and negation in reports, such as "no evidence of effusion," which is critical for accurate labels.

Document understanding. Smith [7] described the Tesseract OCR engine, widely used for converting scanned documents to text. Xu *et al.* [8] proposed LayoutLM, which jointly models text and its position on the page, significantly improving information extraction from forms, an important capability for structured laboratory reports.

Clinical NLP. BioBERT [9], pre-trained on biomedical literature, outperformed general-purpose BERT [10] in recognizing diseases, chemicals and genes. Clinical databases such as MIMIC-III [11] supply notes and reports for training and evaluation. Table I summarizes the surveyed literature.

Ref.	Author(s) / Year	Technique / Approach	Key Contribution	Limitation / Gap
[1]	Litjens et al., 2017	Survey of deep learning in medical imaging	Established CNN dominance across tasks	Few prospective studies
[2]	Rajpurkar et al., 2017	DenseNet-121 on chest X-rays	Radiologist-level pneumonia detection	Noisy report-derived labels
[3]	Ronneberger et al., 2015	U-Net encoder–decoder	Accurate segmentation from few images	2-D design in original form
[4]	McKinney et al., 2020	AI for mammography screening	Fewer false positives and negatives	Retrospective evaluation
[5]	Wang et al., 2017	NLP label mining from radiology reports	Large labelled X-ray dataset	Label errors from NLP
[6]	Irvin et al., 2019	Labeller with uncertainty and negation	More accurate report-derived labels	Rule-based; English only
[7]	Smith, 2007	Tesseract OCR engine	Open, widely used text recognition	Weak on poor scans and handwriting
[8]	Xu et al., 2020	LayoutLM text + layout model	Improved form field extraction	Requires fine-tuning per format
[9]	Lee et al., 2020	BioBERT biomedical language model	Better biomedical entity recognition	Trained on literature, not notes

TABLE I Summary of Literature Survey

III. Preliminaries

3.1 Medical Image Analysis Tasks

Image analysis tasks include classification (does this image show a condition?), detection (where is the lesion?) and segmentation (which pixels belong to an organ or tumour?). Residual networks [12] and EfficientNet [13] are common for classification, and U-Net [3] for segmentation.

3.2 Report Processing Tasks

Report processing begins with OCR for scanned documents, followed by layout analysis to identify fields such as patient name, test name and result, and NLP to recognize medical entities and their relationships. Fig. 1 classifies the AI functions involved.

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct-2026, PP 01-07

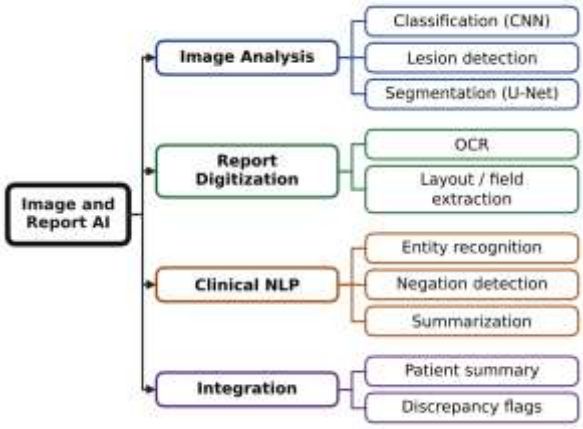


Fig. 1. AI functions for medical image and report detection.

IV. Problem Statement and Objectives

4.1 Identified Research Gaps

The following gaps are identified:

G1: Image and report pipelines are usually built separately, with little integration into a single patient view.

G2: OCR accuracy drops on low-quality scans, handwritten notes and the varied report formats used by Indian laboratories.

G3: Clinical NLP models are trained mainly on English text from foreign hospitals and handle local abbreviations poorly.

G4: Explanations for image predictions are not consistently provided.

G5: Validation in real hospital workflows, including measurement of time saved, is limited.

4.2 Objectives of the Review

The objectives of this review are:

- To review deep learning methods for medical image classification, detection and segmentation.
- To review OCR, layout analysis and clinical NLP for medical report processing.
- To propose an integrated architecture for image and report detection with clinician review.
- To compare techniques in terms of accuracy, data needs and explainability.
- To identify challenges related to data quality, privacy and validation.

4.3 Proposed Framework

Fig. 2 shows the proposed architecture, which processes images and reports in parallel and fuses the results into a structured patient summary.

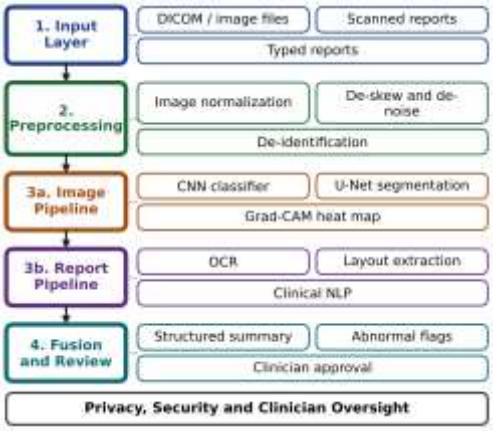


Fig. 2. Proposed integrated architecture for medical image and report detection.

A. Input Layer

Images are received in DICOM or standard image formats, and reports arrive as scanned PDFs, photographs or typed text from hospital systems.

4.5 Preprocessing Layer

Images are resized, normalized and checked for quality. Scanned reports are de-skewed, de-noised and binarized to improve OCR accuracy. Personal identifiers are protected according to privacy rules.

4.6 Image Analysis Pipeline

A CNN classifies the image for selected conditions and a segmentation network outlines relevant regions. Grad-CAM [14] produces a heat map showing the areas that influenced the prediction.

4.7 Report Analysis Pipeline

OCR converts the report to text, a layout model extracts key–value pairs such as test names and values, and a clinical NLP model identifies findings, diagnoses, medications and negations.

4.8 Fusion and Review Layer

Image predictions and extracted report findings are combined into a structured summary that highlights abnormal values and possible discrepancies. A clinician reviews, edits and approves the summary, which is then stored in the electronic record.

V. Working Methodology

The workflow is given in Fig. 3. Segmentation quality is commonly measured by the Dice coefficient between the predicted region A and the expert annotation B:

$$Dice = 2|A \cap B| / (|A| + |B|) \qquad (1)$$

OCR quality is measured by the character error rate

$$CER = (S + D + I) / N \qquad (2)$$

where S, D and I are the numbers of substituted, deleted and inserted characters, and N is the number of characters in the reference text. Entity extraction is evaluated with precision, recall and F1-score.

International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct-2026, PP 01-07

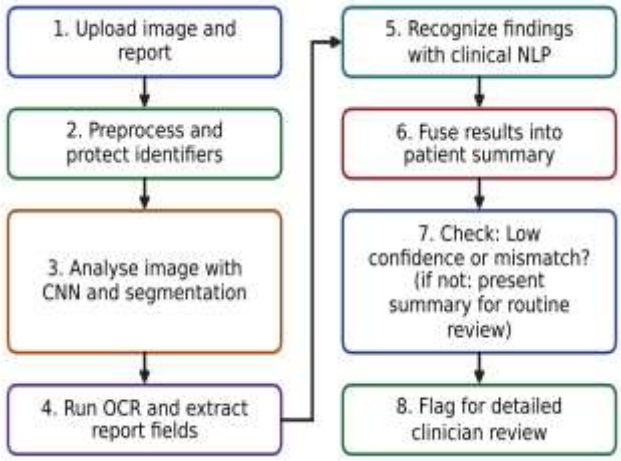


Fig. 3. Workflow of the image and report detection system.

Every output includes a confidence value; low-confidence predictions are explicitly marked so that clinicians give them closer attention.

VI. Discussion

6.1 Comparison of Approaches

Table II compares techniques for each pipeline. Deep CNNs and U-Net are the established choices for images; for reports, rule-based extraction is transparent but brittle, while layout-aware and biomedical language models are more robust to format variation.

Technique	Task	Strength	Limitation
CNN (ResNet / DenseNet)	Image classification	High accuracy	Needs labelled data
U-Net	Segmentation	Precise regions	Annotation effort
Tesseract OCR	Report digitization	Free, fast	Poor on bad scans
LayoutLM	Field extraction	Layout-aware	Compute cost
BioBERT / clinical NLP	Finding extraction	Domain knowledge	Local terms missing

TABLE II Comparison of Techniques

6.2 Open Challenges and Future Work

Table III lists key challenges. Future work includes multimodal models that read an image and its report together, automatic draft report generation that a radiologist edits, support for Indian languages in reports, and integration with national digital health records under strong consent and data protection rules [15].

Challenge	Mitigation
Poor image or scan quality	Quality checks; request re-capture
Varied report formats	Layout models; template learning
Prediction errors	Confidence flags; mandatory clinician review
Patient privacy	De-identification; encryption; access control
Limited local data	Transfer learning; local validation sets

TABLE III Challenges and Mitigation

VII. Conclusion and Future Work

AI can make the review of medical images and reports faster and more consistent. The reviewed literature shows that deep learning detects and segments abnormalities with high accuracy, that OCR and layout-aware models can digitize reports, and that biomedical NLP can extract findings from text. Integrating these capabilities produces organized, searchable patient information and reduces repetitive manual work. Because errors can affect patient care, such systems must be validated on local data, protect privacy and keep qualified professionals responsible for every final interpretation.

7.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

References

1. G. Litjens et al., “A survey on deep learning in medical image analysis,” Med. Image Anal., vol. 42, pp. 60–88, 2017.
2. P. Rajpurkar et al., “CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning,” arXiv:1711.05225, 2017.
3. O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Proc. MICCAI, 2015, pp. 234–241.
4. S. M. McKinney et al., “International evaluation of an AI system for breast cancer screening,” Nature, vol. 577, no. 7788, pp. 89–94, 2020.
5. X. Wang et al., “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in Proc. IEEE CVPR, 2017, pp. 2097–2106.
6. J. Irvin et al., “CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison,” in Proc. AAAI Conf. Artif. Intell., vol. 33, 2019, pp. 590–597.
7. R. Smith, “An overview of the Tesseract OCR engine,” in Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR), 2007, pp. 629–633.
8. Y. Xu et al., “LayoutLM: Pre-training of text and layout for document image understanding,” in Proc. 26th ACM SIGKDD, 2020, pp. 1192–1200.
9. J. Lee et al., “BioBERT: A pre-trained biomedical language representation model for biomedical text mining,” Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.
10. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.
11. A. E. W. Johnson et al., “MIMIC-III, a freely accessible critical care database,” Sci. Data, vol. 3, Art. no. 160035, 2016.
12. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.
13. M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.

14. R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in Proc. IEEE ICCV, 2017, pp. 618–626.

15. Government of India, The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023). New Delhi: Ministry of Law and Justice, 2023.