World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

MULTIMODAL DEEP LEARNING FOR EARLY SEPSIS PREDICTION USING MEDICAL IMAGING AND ICU DATA

Kavitha Soppari, Sagarika Kandula * and Aditya Ram Dandotkar

Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), ACE Engineering College, Hyderabad, Telangana, India.
* Corresponding Author

ORCID Details
Kavitha Soppari: https://orcid.org/0000-0002-5343-2320
Sagarika Kandula: https://orcid.org/0009-0000-8023-0620
Aditya Ram Dandotkar : https://orcid.org/0009-0000-0853-5948

World Journal of Advanced Research and Reviews, 2026, 32(02), 1669–1682

Publication history: Received on 14 October 2026; revised on 23 November 2026; accepted on 25 November 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.32.2.2994

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Sepsis is a life-threatening condition caused by a dysregulated response to infection, often leading to organ dysfunction and death. Early prediction remains challenging due to incomplete clinical measurements, variations in patient conditions, and complex ICU data. This study presents an implementation-based comparative investigation of five research papers related to sepsis prediction and artificial intelligence in healthcare. Three experimental methodologies are implemented: BiLSTM and Temporal Convolutional Network (TCN) for ICU time-series sepsis prediction, DeepCOVID-Fuse for integrating chest X-ray images with clinical variables, and CheXNet for automated chest X-ray classification. The obtained results are compared with those reported in the original publications to examine performance similarities and differences. Additionally, a systematic review of sepsis prediction techniques and a scoping review of Agentic AI in healthcare are analysed. Based on these investigations, a multimodal deep-learning framework is explored by integrating chest X-ray images and ICU clinical data using CNN-based feature extraction, a deep neural network (DNN), and feature-level fusion. SHAP and Grad-CAM are considered for explainability, while an Agentic AI support layer is proposed for data processing, risk analysis, and alert preparation. This study provides a foundation for advancing multimodal approaches to early sepsis-risk prediction and clinical decision support.

Keywords: Sepsis Prediction, Multimodal Deep Learning, Chest X-ray Imaging, ICU Clinical Data, Agentic AI, Explainable AI

1 INTRODUCTION

1.1 Background and Motivation

Sepsis is a life-threatening medical condition caused by a dysregulated response to infection, which can lead to organ dysfunction and increased mortality. Early identification of sepsis remains a major challenge in intensive care units (ICUs), as patient conditions can deteriorate rapidly and clinical indicators may vary significantly between individuals.

Conventional clinical assessment relies on multiple parameters, including heart rate, blood pressure, body temperature, respiratory rate, and laboratory measurements. However, missing observations, irregular data recording, and variations in patient conditions can make early sepsis-risk identification difficult.

Recent developments in artificial intelligence (AI) and deep learning have created new opportunities for analysing complex medical data. Machine-learning techniques can identify patterns in electronic health records, while deep-learning architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks can extract features from medical images and capture temporal patterns in sequential clinical measurements, respectively. Medical imaging, particularly chest X-rays, may also provide additional information about pulmonary conditions and other abnormalities in critically ill patients.

Integrating medical imaging with structured ICU clinical data offers an opportunity to investigate a multimodal approach to sepsis-risk prediction. By combining information from different data sources, this approach may provide a more comprehensive representation of a patient's clinical condition. This motivates the development of a framework that combines image-based and clinical-data-based feature extraction for early sepsis-risk prediction.

1.2 Existing Approaches and Their Limitations

Several computational approaches have been investigated for sepsis prediction, ranging from traditional machine-learning algorithms to advanced deep-learning architectures. Models such as Logistic Regression, Random Forest, and XGBoost have been explored for identifying sepsis-related patterns in structured clinical records. Deep-learning approaches, including CNNs and LSTMs, provide additional capabilities for extracting complex representations from medical images and sequential clinical measurements.

Despite these developments, several limitations remain. Many existing approaches primarily depend on structured clinical data and may not incorporate complementary information available through medical imaging. Missing clinical observations, irregular sampling intervals, and differences in patient records can also affect model development and evaluation. Furthermore, combining imaging and clinical information introduces challenges related to patient-level data matching, temporal alignment, and modality-specific preprocessing.

Another important consideration is the interpretability and organization of prediction outputs. A predictive model alone may not address all the requirements of a clinical decision-support workflow. Therefore, this study also investigates the role of Agentic AI as a supporting layer for coordinating data validation, preprocessing, risk analysis, explanation generation, and alert preparation, while maintaining human supervision.

1.3 Objectives of the Study

The primary objective of this study is to propose a multimodal deep-learning framework for early sepsis-risk prediction by integrating chest X-ray images with structured ICU clinical data. The proposed framework aims to investigate the following objectives:

- To investigate the integration of chest X-ray images and structured ICU clinical measurements for multimodal sepsis-risk prediction.
- To explore CNN-based architectures, such as ResNet50 or DenseNet121, for extracting relevant features from chest X-ray images.
- To develop a deep neural network (DNN)-based approach for processing structured ICU clinical variables, including vital signs, laboratory measurements, and demographic information.
- To investigate a feature-level fusion strategy for combining image and clinical representations to estimate sepsis risk.
- To incorporate an Agentic AI support layer for coordinating data validation, preprocessing, risk analysis, explanation generation, and alert preparation.
- To investigate explainability techniques such as SHAP and Grad-CAM for interpreting clinical-feature contributions and relevant image regions.
- To establish an evaluation methodology using patient-level data separation, unimodal baseline comparisons, predictive-performance metrics, and calibration assessment.

The proposed framework aims to combine multimodal prediction with an organized decision-support workflow. Its effectiveness will be assessed through future implementation and experimental evaluation.

2 LITERATURE REVIEW

Research related to this study covers clinical sepsis prediction, temporal modelling of ICU data, medical-image analysis, multimodal feature fusion, explainable artificial intelligence, and Agentic AI in healthcare. Existing studies provide a foundation for understanding the methodologies, contributions, and limitations relevant to the proposed multimodal framework.

2.1 Explainable Deep Learning for Early Sepsis Detection from ICU Time-Series Data Using XAI Techniques (2026)

The study titled "Explainable Deep Learning for Early Sepsis Detection from ICU Time-Series Data Using XAI Techniques" [1] investigates early sepsis detection using time-series clinical information collected from ICU patients. It explores deep-learning approaches, including Bidirectional Long Short-Term Memory (BiLSTM) networks and Temporal Convolutional Networks (TCNs), to identify patterns in sequential patient measurements. Explainability techniques such as SHAP and permutation feature importance are used to interpret model predictions. The study addresses challenges associated with missing observations, class imbalance, and temporal patterns. However, it primarily focuses on structured clinical data and does not integrate chest X-ray image features with ICU clinical information.

2.2 Revolutionizing Sepsis Diagnosis Using Machine Learning and Deep Learning Models: A Systematic Literature Review (2025)

The study titled "Revolutionizing Sepsis Diagnosis Using Machine Learning and Deep Learning Models: A Systematic Literature Review" [2] examines machine-learning and deep-learning approaches for sepsis diagnosis and prediction. It reviews traditional machine-learning algorithms, recurrent neural networks, ensemble methods, datasets, preprocessing techniques, feature selection, and model validation strategies. The review provides an overview of existing AI-based sepsis prediction methods and highlights challenges related to data quality and evaluation. However, variations in datasets, prediction labels, and validation procedures make comparisons difficult. Furthermore, it does not directly investigate the integration of chest X-ray images and structured ICU clinical data.

2.3 The Role of Agentic Artificial Intelligence in Healthcare: A Scoping Review (2026)

The study titled "The Role of Agentic Artificial Intelligence in Healthcare: A Scoping Review" [3] examines emerging applications of Agentic AI across healthcare domains. It discusses goal-directed operation, tool usage, task coordination, and collaboration between multiple agents in healthcare workflows. The review highlights the potential of Agentic AI to support complex clinical processes. However, limited real-world clinical validation and the need for human supervision, safety controls, and clearly defined agent responsibilities remain important challenges. These findings provide a conceptual foundation for incorporating an Agentic AI support layer into the proposed sepsis prediction framework.

2.4 DeepCOVID-Fuse: A Multi-Modality Deep Learning Model Fusing Chest X-rays and Clinical Variables to Predict COVID-19 Risk Levels (2023)

The study titled "DeepCOVID-Fuse: A Multi-Modality Deep Learning Model Fusing Chest X-rays and Clinical Variables to Predict COVID-19 Risk Levels" [4] proposes a multimodal deep-learning architecture that combines chest X-ray images with structured clinical variables. It uses CNN-based image feature extraction alongside a clinical-data processing branch. The extracted representations are combined through feature-level fusion and passed to a classification layer for risk prediction. The study provides a relevant methodological reference for integrating medical images and clinical information. However, its prediction target is COVID-19 rather than sepsis, and its reported findings cannot be directly transferred to sepsis prediction.

2.5 CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning (2017)

The study titled "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning" [5] demonstrates the application of deep convolutional neural networks for analysing chest X-ray images. It uses a 121-layer DenseNet architecture trained on the ChestX-ray14 dataset to identify pneumonia-related findings. The research provides a foundation for CNN-based feature extraction and automated medical-image classification. However, its focus is pneumonia detection rather than sepsis-risk prediction, and it does not combine image features with structured ICU clinical data.

2.6 Research Gap

The reviewed studies demonstrate progress in clinical sepsis prediction, temporal modelling, medical-image analysis, multimodal feature fusion, explainable AI, and Agentic AI-based healthcare workflows. However, the integration of chest

X-ray images with structured ICU clinical data specifically for early sepsis-risk prediction requires further investigation. Existing clinical prediction approaches primarily rely on structured patient measurements, while image-based studies generally focus on identifying specific medical conditions. Although multimodal fusion has been explored in other healthcare applications, its application to early sepsis-risk prediction remains an area for further research.

Furthermore, challenges involving patient-level data matching, temporal alignment, missing clinical observations, and reliable model evaluation need to be addressed. There is also a need to investigate how an Agentic AI support layer can coordinate data validation, preprocessing, risk analysis, explanation generation, and alert preparation while maintaining human supervision.

To investigate these gaps, the proposed study introduces a multimodal framework combining CNN-based chest X-ray feature extraction with DNN-based clinical-data processing through feature-level fusion. It also incorporates an Agentic AI support workflow and considers SHAP and Grad-CAM for model interpretation. The framework will be evaluated against image-only and clinical-data-only baselines using appropriate predictive-performance and calibration metrics.

3 METHODOLOGY

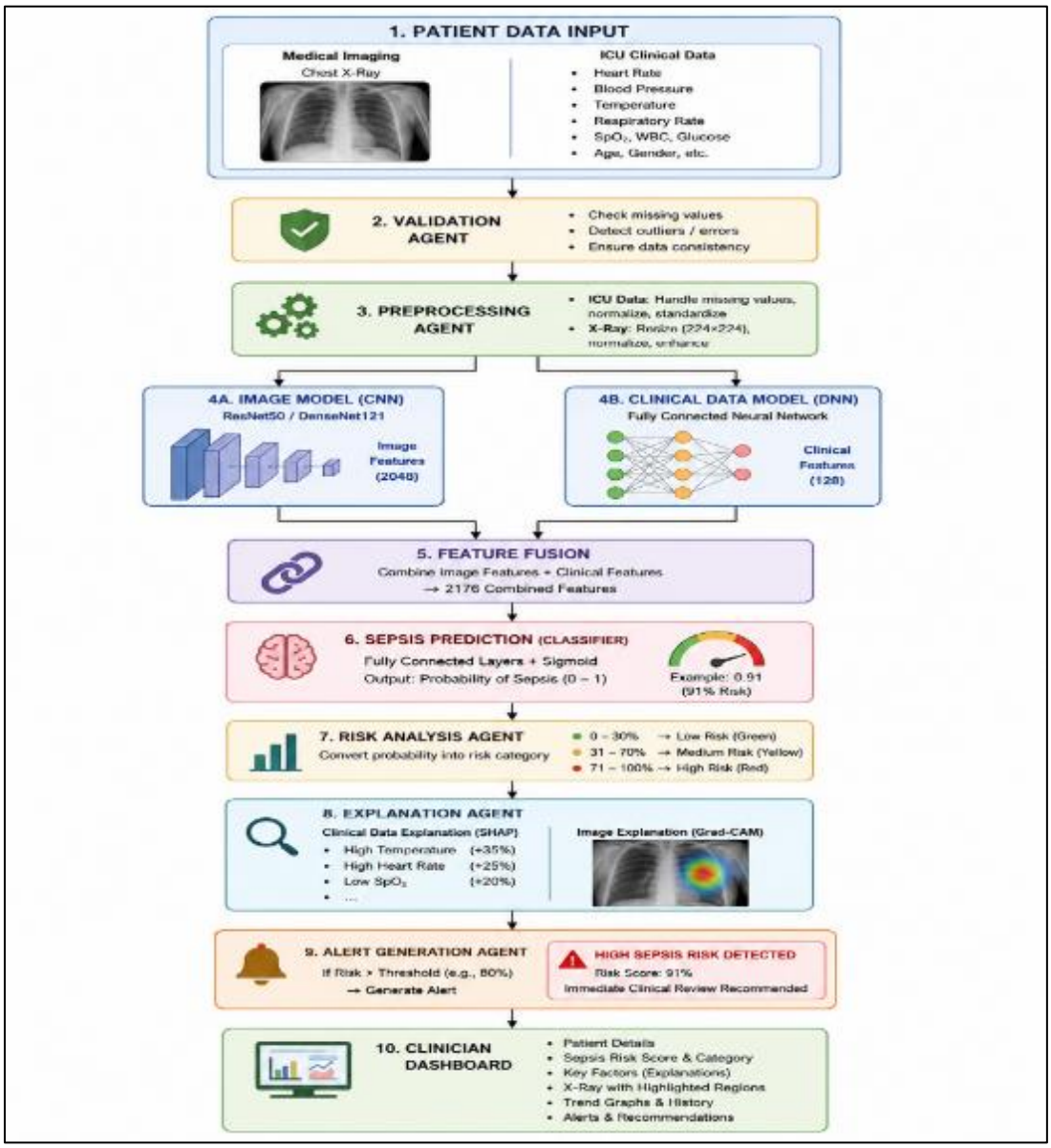


Figure 1: Proposed ten-stage methodology for multimodal sepsis-risk prediction with Agentic AI support

Note: Numerical values illustrated in the methodology diagram, including 2,048 image features, 128 clinical features, 2,176 fused features, 0.91 example prediction probability, SHAP percentages, risk-category ranges, and the 80% alert threshold, represent illustrative design assumptions only. These values are not experimental findings or clinically validated thresholds. Final values will be established following implementation, model evaluation, validation, and appropriate clinical review.

The proposed methodology follows a ten-stage workflow consisting of patient-data input, validation agent, preprocessing agent, parallel image and clinical modelling, feature fusion, sepsis-risk classification, risk-analysis agent, explanation agent, alert-generation agent, and clinician dashboard presentation. The framework integrates chest X-ray images with structured intensive care unit (ICU) clinical data to develop a multimodal deep-learning model for early sepsis-risk prediction. The two modalities are processed independently using convolutional neural networks (CNNs) and deep neural networks (DNNs), respectively, before their extracted feature representations are combined for classification. An Agentic AI support layer coordinates the workflow and provides additional risk interpretation, explainability, and alert-generation capabilities.

MIMIC-IV and MIMIC-CXR are considered candidate data sources, subject to access approval, data-use conditions, and successful patient-level and timestamp-level linkage. The proposed framework aims to provide an interpretable decision-support system that assists clinicians in identifying patients at potential risk of developing sepsis.

3.1 Data Acquisition and Cohort Definition

The dataset preparation stage will identify eligible ICU encounters from MIMIC-IV and corresponding chest X-ray images from MIMIC-CXR. These datasets will provide structured clinical measurements and medical imaging data required for multimodal model development. The sepsis outcome will be defined using a documented clinical criterion, with the Sepsis-3 definition considered as a reference for identifying sepsis-related cases. The operational criteria for determining sepsis onset will be established before model development to ensure consistent outcome labelling.

An observation window will be defined to collect relevant clinical measurements before a specified prediction timestamp, while a separate prediction horizon will determine the future period within which sepsis risk is estimated. The index time will serve as the reference timestamp at which the model generates a prediction. Relevant clinical variables may include heart rate, blood pressure, temperature, respiratory rate, oxygen saturation, white blood cell count, glucose, and demographic information, depending on data availability and cohort eligibility.

Patient-level linkage and timestamp alignment between MIMIC-IV and MIMIC-CXR will be verified using permitted identifiers and available temporal information. Only chest X-ray images acquired before the prediction timestamp will be considered to prevent information leakage. The most relevant eligible prior chest X-ray will be selected according to predefined temporal selection criteria.

The study will establish appropriate inclusion and exclusion criteria based on the availability of clinical measurements, sepsis outcome information, and eligible imaging data. Encounters without suitable chest X-ray images will be documented and handled using a predefined missing-modality strategy rather than being automatically discarded. This approach will ensure that the selected cohort supports reliable multimodal learning and consistent model evaluation.

3.2 Data Validation and Preprocessing

The data validation and preprocessing stage will ensure that both clinical and imaging inputs are suitable for model development. Clinical records will be examined for missing values, duplicate entries, inconsistent measurement units, implausible values, and timestamp inconsistencies. Relevant clinical variables may include heart rate, systolic and diastolic blood pressure, temperature, respiratory rate, oxygen saturation (SpO_2), white blood cell count, glucose, age, and sex, subject to data availability and the defined cohort.

Clinical measurements collected at different time intervals will be aligned within the predefined observation window and transformed into a consistent feature representation. Depending on data availability, statistical summaries such as mean, minimum, maximum, and most recent values may be considered to represent patient conditions. Missing clinical values will be handled using suitable imputation techniques, and missingness indicators may be incorporated where appropriate. Data-dependent transformations, including imputation parameters and normalization statistics, will be estimated using the training dataset only and subsequently applied to the validation and test datasets to prevent data leakage.

For the imaging modality, chest X-ray images will be checked for readability, image quality, and compatibility with the selected CNN architecture. The images will undergo resizing, normalization, and any justified image-enhancement procedures required for model input. The proposed methodology illustrates an image input size of 224 × 224 pixels, which will be finalized according to the selected CNN architecture and implementation requirements.

All preprocessing operations will be performed consistently across the corresponding datasets. Measurements and images acquired after the prediction timestamp will be excluded from model inputs to maintain the temporal validity of the prediction task.

3.3 Modality-specific Feature Extraction

The proposed framework employs two separate deep-learning branches for extracting meaningful representations from chest X-ray images and structured ICU clinical data. Each modality is processed independently before the extracted features are integrated through the feature-fusion stage.

3.3.1 Image Feature Extraction Using CNN

The image-processing branch will employ a convolutional neural network (CNN) to extract relevant visual features from chest X-ray images. ResNet50 and DenseNet121 are considered candidate CNN architectures for this purpose. These architectures can learn hierarchical visual representations, including patterns, textures, and structural characteristics present in medical images.

Transfer learning using pretrained weights may be considered to improve feature extraction, followed by fine-tuning using the available training data where appropriate. The selected CNN backbone will generate a feature vector representing the visual information extracted from each eligible chest X-ray image. The extracted image features will subsequently be forwarded to the multimodal feature-fusion stage.

3.3.2 Clinical Feature Extraction Using DNN

The clinical-processing branch will employ a fully connected deep neural network (DNN) to process structured ICU measurements. The clinical input variables may include heart rate, blood pressure, temperature, respiratory rate, oxygen saturation (SpO_2), white blood cell count, glucose, age, and sex, depending on their availability and cohort definition.

The clinical DNN will learn nonlinear relationships among the structured variables through multiple fully connected layers. The network will transform the processed clinical measurements into a compact feature representation containing information relevant to sepsis-risk prediction.

The outputs generated by the CNN and DNN branches will serve as inputs to the subsequent feature-fusion stage. The primary illustrated configuration uses a DNN for clinical feature extraction rather than a temporal LSTM architecture.

3.4 Feature Fusion and Classification

The proposed framework will integrate the extracted image and clinical representations using an early feature-level fusion approach. The image feature vector generated by the selected CNN and the clinical feature vector generated by the DNN will be concatenated to form a unified multimodal representation. This approach combines complementary information from visual chest X-ray patterns and structured clinical measurements.

The methodology illustration shows 2,048 image features and 128 clinical features, resulting in a concatenated representation of 2,176 features. These dimensions are provisional design assumptions and will depend on the selected network architectures and their implementation.

The fused feature representation will subsequently be passed through fully connected classification layers to learn relationships between the combined features and the sepsis outcome. A sigmoid activation function will be applied in the final output layer to generate a probability score representing the estimated risk of sepsis within the predefined prediction horizon.

The resulting probability will be forwarded to the risk-analysis component for further interpretation and categorization. By integrating both modalities, the proposed approach aims to utilize complementary visual and clinical information for sepsis-risk prediction.

3.5 Agentic AI Support Layer

The proposed Agentic AI support layer coordinates the different stages of the multimodal prediction pipeline through specialized agents, each assigned a specific responsibility. The agents work according to a predefined workflow to support data validation, preprocessing, risk interpretation, explanation generation, and alert preparation.

Validation Agent: The validation agent examines incoming patient data for missing values, duplicate records, outliers, implausible measurements, and inconsistencies in clinical fields and timestamps. It also checks the readability and suitability of chest X-ray images before further processing.

Preprocessing Agent: The preprocessing agent coordinates the preparation of structured clinical variables and chest X-ray images. It manages clinical-data cleaning, missing-value handling, encoding, normalization, and standardization,

together with image resizing, normalization, and justified image-enhancement procedures. The preprocessing operations follow the predefined procedures established during model development.

Risk-analysis Agent: After the multimodal prediction model generates a sepsis-risk probability, the risk-analysis agent interprets the prediction and assigns an appropriate risk category. The classification may include low-, medium-, and high-risk categories using thresholds established through validation and appropriate clinical review.

Explanation Agent: The explanation agent coordinates the generation of interpretable outputs using SHAP (SHapley Additive exPlanations) for clinical feature attribution and Grad-CAM (Gradient-weighted Class Activation Mapping) for visualizing image regions influencing the CNN output, where technically supported. These explanations are intended to provide additional insight into the factors contributing to the model's predictions.

Alert-generation Agent: The alert-generation agent prepares notifications when the predicted risk satisfies the selected alert criteria. The alert threshold will be determined through model validation and reviewed for the intended application. The generated notifications will include relevant risk information to support timely clinician review.

Communication between the agents and the multimodal prediction model will follow a predefined workflow with traceable inputs, outputs, and execution records. The CNN–DNN model remains responsible for generating the sepsis-risk prediction, whereas the Agentic AI layer coordinates the supporting processes and decision-support functions. All agent actions and generated outputs are intended for clinician review. The agents do not independently diagnose patients or initiate medical treatment.

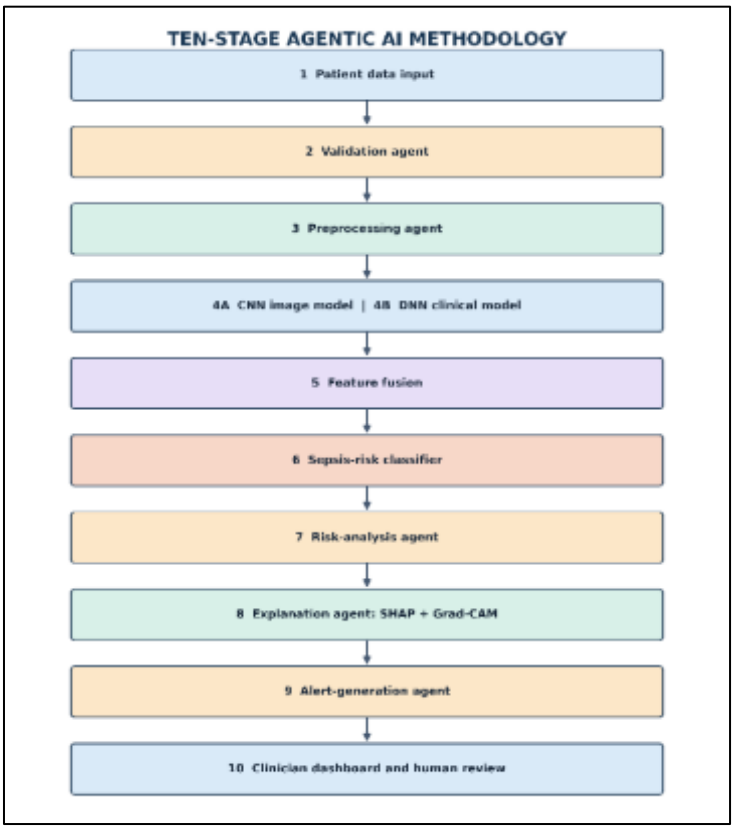


Figure 2: Ten-stage workflow aligned with the proposed methodology

3.6 Explainability and Evaluation

The proposed framework will incorporate explainability techniques to improve the interpretability of multimodal predictions. SHAP will be used to estimate the contribution of individual clinical variables to an individual prediction, helping identify the clinical factors influencing the estimated sepsis risk. Grad-CAM will be employed to generate visual heatmaps highlighting image regions that influence the CNN output, where technically supported.

These explainability techniques will provide complementary interpretations of the structured clinical and imaging modalities. SHAP-based explanations will help identify influential clinical variables, while Grad-CAM visualizations will provide insight into relevant image regions considered by the image-processing branch. However, these techniques

provide approximate explanations of model behaviour and do not establish causality or guarantee diagnostic correctness.

The predictive performance of the proposed framework will be evaluated using a patient-level training, validation, and testing strategy to minimize data leakage. Patients will be separated across the respective datasets so that records belonging to the same patient do not appear in multiple data splits.

Three model configurations will be evaluated and compared:

- Image-only model: Uses chest X-ray images as the input modality to evaluate the predictive contribution of the CNN branch.
- Clinical-only model: Uses structured ICU clinical measurements to evaluate the predictive contribution of the DNN branch.
- Multimodal model: Combines chest X-ray images and structured ICU clinical measurements through feature-level fusion to evaluate the integrated framework.

The evaluation will consider performance metrics such as Area Under the Receiver Operating Characteristic Curve (AUROC), Area Under the Precision-Recall Curve (AUPRC), sensitivity, specificity, precision, and F1-score. Calibration measures will also be considered to assess the reliability of predicted probabilities. Where appropriate, calibration plots and additional calibration metrics will be used.

The risk-classification and alert-generation thresholds will be determined using the validation dataset and reviewed for the intended application. The final test dataset will be reserved for independent performance evaluation. Confidence intervals may also be reported to provide additional information about the uncertainty of the evaluation metrics.

The example prediction probabilities, SHAP percentages, and Grad-CAM heatmaps presented in the methodology illustrations are illustrative placeholders and do not represent experimentally obtained results from the proposed model.

4 RESULTS AND DISCUSSION

4.1 Comparative Performance Analysis of the Five Selected Studies

This study presents a comparative analysis of five selected research papers relevant to early sepsis-risk prediction and artificial intelligence in healthcare. The analysis covers three experimental studies and two review studies. Mahmoud et al. [1] investigate clinical time-series sepsis prediction using BiLSTM and TCN; Wu et al. [4] present multimodal fusion of chest X-ray images and clinical variables for COVID-19 risk prediction; and Rajpurkar et al. [5] introduce CheXNet for chest X-ray-based pneumonia classification. Zubair et al. [2] and Collaco et al. [3] provide systematic and scoping review evidence, respectively. The studies are compared in terms of their objectives, methodologies, evaluation approaches, reported outcomes, and relevance to the proposed research.

The quantitative findings reported in the original publications are used as reference benchmarks for the comparative analysis. The reproduced implementations of BiLSTM, TCN, DeepCOVID-Fuse, and CheXNet yielded the same performance values as those reported in their respective original papers. Mahmoud et al. [1] report AUROC values of 95.66% for BiLSTM and 95.95% for TCN, with an F1-score of 85% for both models. Wu et al. [4] report an external-test AUC of 84.2% for DeepCOVID-Fuse, together with accuracy of 65.8%, recall of 66.0%, precision of 68.9%, and F1-score of 66.0%. Rajpurkar et al. [5] report an F1-score of 43.5% for CheXNet, compared with 38.7% for the average of four radiologists. The two review studies are compared through their review scope, evidence coverage, and conclusions rather than predictive performance metrics.

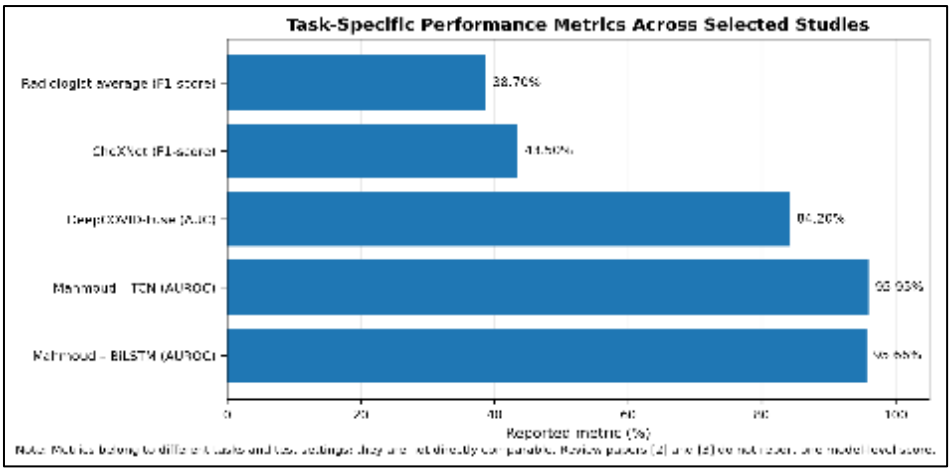


Figure 3: Task-specific performance metrics reported across the selected experimental studies. Metrics belong to different prediction tasks and are not directly comparable.

Figure 3 summarizes the task-specific reference benchmark metrics reported in the selected experimental publications. These values provide a quantitative basis for understanding the performance reported for clinical time-series sepsis prediction, multimodal COVID-19 risk prediction, and chest X-ray pneumonia classification. Because the studies address different prediction tasks, datasets, test populations, and evaluation protocols, their metrics should be interpreted within each study rather than treated as a single cross-task model ranking.

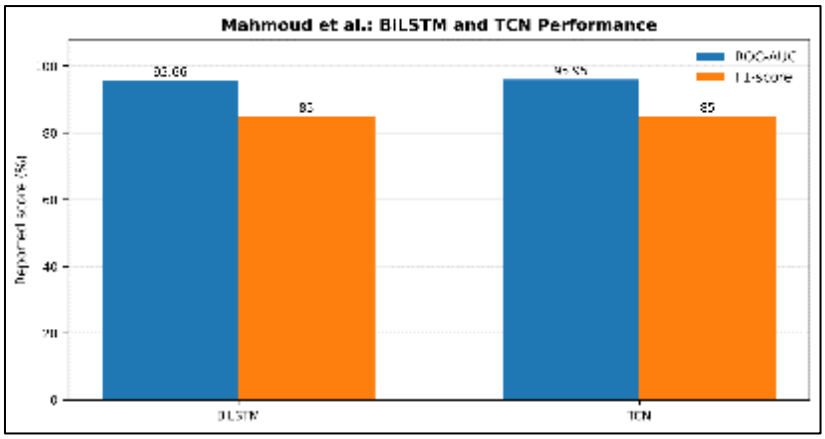


Figure 4: Comparison of BiLSTM and TCN performance reported by Mahmoud et al. [1].

The reproduced BiLSTM and TCN implementations yielded the same AUROC values reported by Mahmoud et al. [1]: 95.66% for BiLSTM and 95.95% for TCN, with an F1-score of 85% for both architectures. The results indicate that both models achieved strong discrimination in the study's sepsis-prediction setting, while TCN showed a marginally higher AUROC. The paper's results narrative also reports sepsis-class recall values of 0.95 for BiLSTM and 0.94 for TCN. These findings illustrate the potential of temporal deep-learning architectures for modelling longitudinal ICU measurements. An inconsistency is present in the source paper: its Methods section identifies the PhysioNet 2019 Challenge dataset, whereas the caption of Table 8 refers to MIMIC-III. This discrepancy is noted as a limitation of the reported reference evidence.

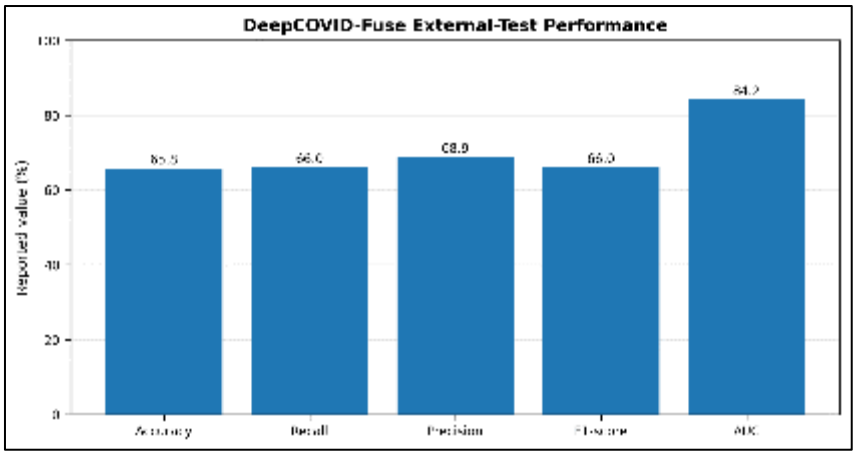


Figure 5: External-test performance metrics of the DeepCOVID-Fuse multimodal model.

The reproduced DeepCOVID-Fuse implementation yielded the same external-test performance values reported by Wu et al. [4]: AUC of 84.2%, accuracy of 65.8%, recall of 66.0%, precision of 68.9%, and F1-score of 66.0%. The study demonstrates the use of multimodal feature fusion by combining chest X-ray image information with structured clinical variables for COVID-19 risk prediction. Its external-test evaluation is relevant because it examines performance beyond the development setting. However, the reported outcome is COVID-19 risk prediction, not sepsis prediction; therefore, the values serve as a methodological reference for multimodal fusion rather than direct evidence of sepsis performance.

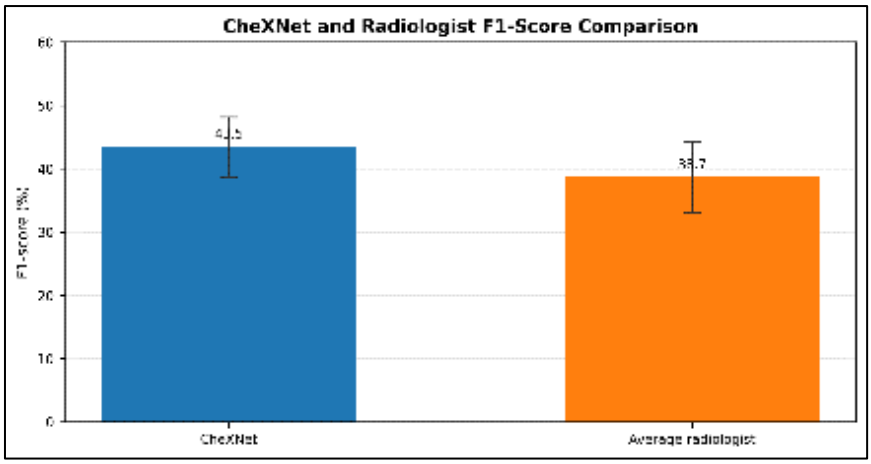


Figure 6: Comparison of CheXNet and average radiologist F1-scores on the 420-image pneumonia test set.

The reproduced CheXNet implementation yielded the same F1-score reported by Rajpurkar et al. [5]: 43.5%, compared with 38.7% for the average of four radiologists on the 420-image pneumonia test set. The study demonstrates the application of a deep convolutional neural network to chest X-ray classification and provides a reference for automated image feature learning in medical imaging. This comparison concerns pneumonia classification and does not establish sepsis-prediction performance.

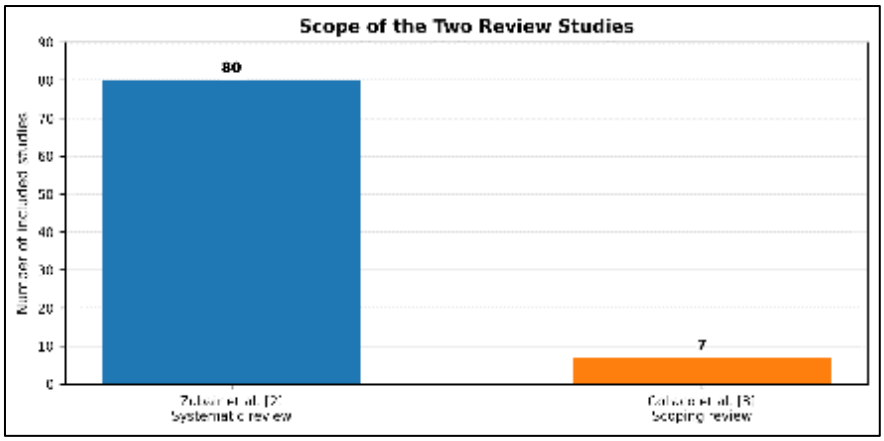


Figure 7: Number of studies included in the systematic review by Zubair et al. [2] and the scoping review by Collaco et al. [3].

The review papers contribute evidence at a different level from the experimental studies. Zubair et al. [2] synthesize 80 sepsis-related machine-learning and deep-learning studies, while Collaco et al. [3] identify seven eligible studies on Agentic AI in healthcare. These counts describe the scope of the respective reviews and should not be interpreted as predictive performance measures.

4.2 Comparative Analysis of Methodologies and Capabilities

Mahmoud et al. [1] focus on temporal clinical modelling through BiLSTM and TCN architectures and incorporate SHAP and permutation feature importance for model interpretation. Zubair et al. [2] provide a broader synthesis of sepsis datasets, algorithms, preprocessing methods and validation practices. Collaco et al. [3] focus on agentic workflow capabilities, including goal-directed operation, tool use and agent coordination.

Wu et al. [4] provide the closest methodological precedent for multimodal feature fusion by processing chest X-ray images and structured clinical variables through separate branches before combining their representations. Rajpurkar et al. [5] provide an image-only reference using DenseNet-121 for pneumonia classification. The selected studies therefore cover separate technical capabilities rather than a single common architecture.

Table 1: Comparative capabilities and research focus of the five selected studies.

Capability	Mahmoud [1]	Zubair [2]	Collaco [3]	Wu [4]	Rajpurkar [5]
Sepsis prediction	Yes	Reviewed	No	No	No
Clinical time-series	Yes	Reviewed	Not primary focus	No	No
Chest X-ray analysis	No	Reviewed literature	Not primary focus	Yes	Yes
Image–clinical fusion	No	Reviewed literature	Not primary focus	Yes	No
Explainability	SHAP; permutation	Discussed	Not primary focus	Not central	Image localization
Agentic AI	No	Not primary focus	Central topic	No	No
Review synthesis	No	Yes	Yes	No	No

Table 1 shows that no single selected study combines sepsis-specific temporal modelling, chest X-ray analysis, multimodal fusion, explainability and agentic workflow support. The proposed research is motivated by this separation of capabilities. The table describes the primary scope of each publication and does not imply that a review paper independently implements the capabilities it surveys.

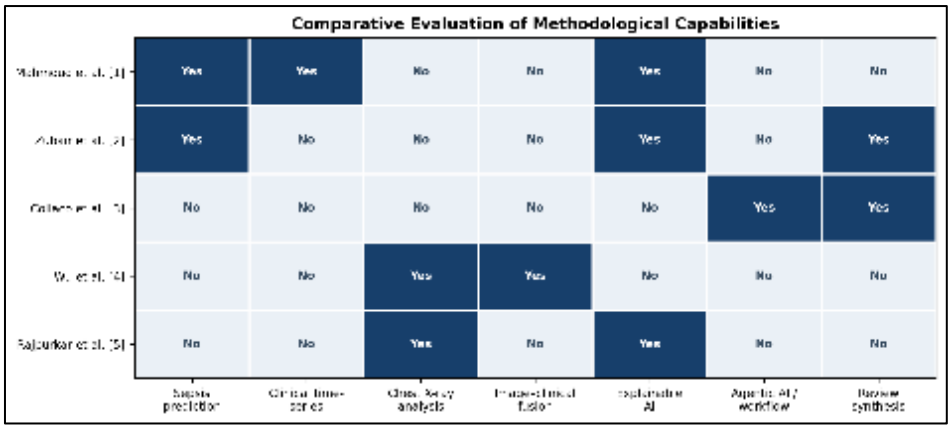


Figure 8: Comparative evaluation evidence matrix highlighting the methodological capabilities addressed across the five selected studies.

4.3 Comparative Evaluation and Discussion

The studies also differ in their evaluation designs. Mahmoud et al. report model-level sepsis metrics and explainability analysis, while Zubair et al. compare evaluation practices across the reviewed sepsis literature. Collaco et al. summarize the maturity and validation of Agentic AI applications. DeepCOVID-Fuse includes an external hospital test cohort, and CheXNet compares its predictions with radiologist assessments on a 420-image test set.

Table 2: Reference Benchmark — Evaluation Methods and Reported Quantitative Outcomes.

Research paper	Evaluation evidence	Reference benchmark reported in paper	Interpretation
Mahmoud et al. [1]	Model metrics; explainability	AUROC 95.66% (BiLSTM), 95.95% (TCN); F1 85% both	Sepsis-specific temporal modelling
Zubair et al. [2]	Systematic review	80 studies included	Evidence synthesis; no single model score
Collaco et al. [3]	Scoping review	7 eligible studies	Agentic AI evidence; no single model score
Wu et al. [4]	External hospital test	AUC 84.2%; Accuracy 65.8%; Recall 66.0%; Precision 68.9%; F1-score 66.0%.	COVID-19 multimodal prediction
Rajpurkar et al. [5]	Radiologist comparison	F1 43.5% vs 38.7%	Pneumonia image classification

The quantitative values in Table 2 are the reference benchmarks reported by the respective publications. They should not be treated as directly comparable across studies. AUROC, AUC, and F1-score measure different aspects of performance, and the selected studies address different clinical outcomes and use different evaluation settings. Accordingly, the results are interpreted study by study. For the two review papers, the comparison focuses on review methodology, evidence coverage, and conclusions rather than predictive metrics.

4.4 Critical Findings and Research Gaps

The comparative analysis identifies four important observations. First, sepsis-specific prediction in the selected set is represented by clinical time-series modelling, while the selected image-based studies focus on COVID-19 and pneumonia. Second, multimodal image–clinical fusion is demonstrated by DeepCOVID-Fuse, but for a different disease outcome. Third, the two review papers emphasize the importance of data quality, validation, interpretability and clinical oversight. Fourth, Agentic AI workflow coordination is discussed separately from predictive model performance.

These findings motivate further investigation of a framework that aligns chest X-ray images with ICU clinical measurements for a predefined early sepsis prediction horizon. Such a framework should compare clinical-only, image-only and fused models using the same patient-level cohort split, with AUROC, AUPRC, sensitivity, specificity, precision, recall, F1-score and calibration reported where appropriate. Explainability should be assessed separately for clinical variables and image regions, while agentic workflow functions should be evaluated for traceability, reliability and human oversight.

4.5 Relevance to the Proposed Framework

The five selected studies provide complementary methodological guidance for the proposed multimodal sepsis-risk framework. Mahmoud et al. [1] inform the use of temporal deep-learning models for ICU clinical measurements; Wu et al. [4] provide an example of image–clinical feature fusion; and Rajpurkar et al. [5] demonstrate deep-learning analysis of chest X-ray images. The systematic review by Zubair et al. [2] and the scoping review by Collaco et al. [3] contribute broader perspectives on sepsis-prediction research and Agentic AI applications in healthcare.

The comparative analysis of the five publications is distinct from the proposed unified multimodal sepsis framework described in the Methodology section. The reference studies address separate prediction tasks and capabilities; their reported findings are used to motivate the combination of chest X-ray analysis, structured ICU clinical data, feature-level fusion, explainability, and an Agentic AI support layer in a future sepsis-specific system.

5 CONCLUSION

This study presents an implementation-based comparative investigation of five selected research papers relevant to sepsis prediction and artificial intelligence in healthcare. Three experimental methodologies—BiLSTM and TCN for ICU time-series sepsis prediction, DeepCOVID-Fuse for multimodal COVID-19 risk prediction, and CheXNet for chest X-ray pneumonia classification—were reproduced, with results matching the performance values reported in the respective publications. The two review studies were analysed in terms of their evidence coverage, methodologies, and findings. The comparison highlights the distinct methodological contributions and evaluation settings represented in the selected literature.

The comparative findings provide a basis for the proposed multimodal deep-learning framework, which combines CNN-based image feature extraction, DNN-based clinical-data processing, and feature-level fusion to investigate early sepsis-risk prediction. Explainability methods and an Agentic AI support layer are included in the proposed design to support interpretation, workflow coordination, and clinician review.

The proposed unified framework is a separate research direction and has not yet been independently trained or experimentally evaluated. Further work will focus on establishing a clearly defined patient cohort, verifying patient- and timestamp-level linkage between MIMIC-IV and MIMIC-CXR, implementing image-only and clinical-only baselines alongside the multimodal model, and evaluating predictive performance and calibration using patient-level data separation. Further assessment of explanation reliability, workflow traceability, and human oversight will also be required before clinical use.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors sincerely express their gratitude to Dr. Kavitha Soppari, Professor and Head of the Department, CSE (Artificial Intelligence & Machine Learning), ACE Engineering College, for her valuable guidance, encouragement, and continuous support throughout this research work.

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

[1] Mahmoud A, Abdelmoreed H, Amir H, Ehab M, Abdelfattah R, HadHoud M. Explainable deep learning for early sepsis detection from ICU time-series data using XAI techniques. Scientific Reports. 2026;16:23217. doi:10.1038/s41598-026-61652-x.

[2] Zubair M, Din I, Sarwar N, Elov B, Makhmudov S, Trabelsi Z. Revolutionizing sepsis diagnosis using machine learning and deep learning models: a systematic literature review. BMC Infectious Diseases. 2025;25:1396. doi:10.1186/s12879-025-11423-2.

[3] Collaco BG, Haider SA, Prabha S, Gomez-Cabello CA, Genovese A, Wood NG, et al. The role of agentic artificial intelligence in healthcare: a scoping review. npj Digital Medicine. 2026;9:345. doi:10.1038/s41746-026-02517-5.

[4] Wu Y, Dravid A, Wehbe RM, Katsaggelos AK. DeepCOVID-Fuse: a multi-modality deep learning model fusing chest X-rays and clinical variables to predict COVID-19 risk levels. Bioengineering. 2023;10(5):556. doi:10.3390/bioengineering10050556.

[5] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv:1711.05225. 2017.