Journal of Medical and Health Studies (JMHS)
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Explainable Multimodal Artificial Intelligence for Early Disease Detection Using Clinical, Laboratory, and Medical Imaging Data

Mohammad Nasir Uddin[1], Md Yassir Mottalib[2], Md. Emran Hossen[3], Anwar Hossain[4], Md. Yousuf[5], Eklachur Rahman Bhuiyan[6], Dahika Alam Nimu[7], MD IMRAN HASAN TUHIN[8]

[1]*Master of Business Administration, Major in Data Analytics, Westcliff University, USA.*
[2]*Master of Science in Information System Technology, Wilmington University, USA*
[3]*Department of Science in Biomedical Engineering, Gannon University, USA*
[4]*Master of Public Health, St. Francis College, NY, USA*
[5]*Master's of Science in Data Analytics, College of Sciences and Technology, University of Houston-Downtown, USA*
[6]*Master of Science in Information Technology, Washington University of Science and Technology, USA*
[7]*MBBS, Dinajpur Medical College, Bangladesh*
[8]*Master's in Information Technology, St. Francis College, NY, USA*
Corresponding Author: Mohammad Nasir Uddin, **E-mail**: m.uddin.258@westcliff.edu

| ABSTRACT

Early disease detection is vital for better outcomes and timely intervention. We developed an explainable multimodal AI framework combining clinical and imaging data. We evaluated clinical-only, imaging-only, and multimodal models, comparing them on accuracy, precision, recall, specificity, F1-score, and ROC-AUC. Our attention-based multimodal model performed best, with an accuracy of 94%, precision of 93%, recall of 93%, specificity of 95%, F1-score of 93%, and ROC-AUC of 97%, surpassing both clinical-only and imaging-only models. The model consisted of a clinical data encoder, an imaging encoder based on a DenseNet121, and an attention-based fusion mechanism to describe the complementary information across modalities. SHAP analysis emphasized the contributions of clinical and laboratory factors, and Grad-CAM revealed image regions associated with model predictions. These findings highlight the potential of explainable multimodal AI in enhancing disease-detection performance and providing transparent evidence for clinical decision support. We provide a roadmap for deployment in U.S. healthcare: external validation, human supervision, data protection, interoperability, and proper regulatory assessment under clinical supervision.

| KEYWORDS

Multimodal Artificial Intelligence; Explainable AI; Early Disease Detection; Medical Imaging; Clinical Data; Laboratory Data; Deep Learning; SHAP; Grad-CAM; Clinical Decision Support

| ARTICLE INFORMATION

ACCEPTED: 15 June 2026 **PUBLISHED:** 25 July 2026 **DOI:** 10.32996/jmhs.2026.7.8.13

1. Introduction

Artificial intelligence, machine learning, and deep learning have transformed the processing of healthcare data, facilitating clinical decision-making. Healthcare systems generate extensive and varied data, including demographic information, clinical records, laboratory results, and imaging. Historically, practitioners synthesized this information cognitively. Although machine learning models may discern patterns within singular datasets, they encounter difficulties in managing the intricacies of numerous modalities.

Medical imaging is a crucial domain for deep learning applications in the healthcare sector. Convolutional neural networks and

contemporary vision models can identify intricate patterns in radiographs, CT scans, MRIs, ultrasounds, and various other imaging modalities. These models show potential for identifying irregularities and assisting in disease categorization. Nevertheless, medical images constitute merely a segment of a patient's clinical overview. A radiographic observation may possess varying clinical implications depending on the patient's age, symptoms, laboratory findings, medical background, and measures. Consequently, depending exclusively on medical imagery may constrain an AI's capacity to emulate the cognitive processes of healthcare practitioners.

Clinical and analytical information enhance each other. Age, blood pressure, cholesterol levels, glucose, heart rate, physiological indicators, and clinical variables associated with diseases can indicate illness risk. Machine learning methodologies, such as logistic regression, random forests, support vector machines, and gradient boosting, can model interactions within organized healthcare data. Medical images encompass intricate spatial data that existing models are unable to accurately depict. Dismissing these methods may result in the loss of valuable supplementary information. Multimodal AI acquires knowledge from several data forms to transcend this limitation. Research is increasingly focused on multimodal learning methodologies that integrate imaging, tabular data, clinical narratives, physiological signals, and longitudinal patient information. Imaging and tabular data represent one of the most common multimodal pairings examined in healthcare, as indicated by a recent investigation into multimodal machine learning within the healthcare sector. The critique discusses the issues associated with data integration, model development, and assessment. The interpretability of multimodal AI systems continues to pose a challenge, notwithstanding recent progress. Deep-learning models with millions of parameters can yield precise predictions without elucidating their reasoning. Due to the potential impact of erroneous predictions on diagnostic choices, treatment pathways, resource allocation, and patient results, this limitation is vital in the healthcare sector. Healthcare professionals could be cautious about a system that forecasts disease likelihood without referencing clinical or imaging variables.
Consequently, research in healthcare AI has concentrated on explainable artificial intelligence (XAI). Model forecasts have been associated with significant attributes or areas using SHapley Additive Explanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), integrated gradients, and gradient-centric visualization techniques. Grad-CAM and other techniques can visually represent the areas of deep-learning predictions in medical imaging. Systematic assessments have emphasized the significance of interpretability for the transparency of medical image analysis and the understanding of AI decisions.
Clarification does not ensure practical application in clinical settings. A comprehensive examination of explainable AI in the healthcare sector revealed considerable diversity in goals, participants, and assessment methods, although there was no data on the impact of explanations on clinical operations, user confidence, and decision-making processes. Investigations into multimodal and longitudinal medical imaging have indicated that multimodal forecasts and elucidations have garnered minimal therapeutic focus.
These challenges necessitate AI frameworks that extend beyond mere accuracy prediction. Efficient healthcare AI systems amalgamate complementary data, exhibit strong prediction performance, offer lucid explanations, demonstrate resilience across diverse patient populations, and generate outputs suitable for incorporation into clinical processes. The increasing regulatory emphasis on AI-driven healthcare systems highlights the importance of transparency, validation, and clinical oversight. At present, the FDA differentiates specific activities of clinical decision-support software from those that fall under medical-device regulation and highlights the necessity of comprehending the intended application and rationale behind AI-assisted suggestions.

To address these issues, we offer an explainable multimodal AI system for early illness identification that integrates structured clinical and laboratory data with medical imaging data. Modality-specific representations are transformed into a common latent space and combined utilizing multimodal fusion. To supplement model predictions, we use SHAP-based analysis for structured clinical data and Grad-CAM-based analysis for medical images. The framework is evaluated on precision, recall, specificity, F1-score, ROC-AUC, calibration, robustness, and explanation quality, not only accuracy.
Our project seeks to improve multimodal healthcare AI transparency and clinical relevance. We especially examine whether heterogeneous clinical and imaging representations improve illness identification relative to unimodal approaches and whether explainability techniques might reveal model prediction elements. After external validation, regulatory review, privacy protection, and prospective clinical evaluation, such a framework could be used for clinical decision assistance in U.S. healthcare.

2. Literature Review

2.1 Artificial Intelligence in Disease Detection

AI is used more in disease detection, diagnosis, prognosis, and clinical risk prediction. Structured demographic, clinical, and laboratory datasets dominated early healthcare ML research. Traditional techniques including logistic regression, decision trees, random forests, support vector machines, and gradient-boosting models are widely employed to process structured clinical data and find nonlinear risk factor connections.

Structured-data models benefit from clinically identified inputs. Clinicians can directly analyze age, blood pressure, cholesterol, glucose, heart rate, and other physiological parameters. When medical pictures or other high-dimensional modalities include essential diagnostic evidence, structured-data models are constrained.

Automated representation learning with deep learning works around these restrictions. Deep neural networks may learn hierarchical representations from raw medical images without using researchers' manual annotations. This makes convolutional neural networks important in radiology and other image-intensive fields. Successful image-based AI does not mean medical imaging provides all the information needed for clinical reasoning. Slight, overlapping, or context-dependent disease symptoms. Thus, imaging evidence and structured clinical data may improve prediction.

2.2 Medical Imaging and Deep Learning

Medical picture categorization and anomaly detection have improved using deep learning. Convolutional neural networks' topologies can learn spatial patterns like edges, textures, anatomical structures, and complex disease-related data, making them successful. Transfer learning allows researchers to establish models using representations from large-scale datasets before fine-tuning them on specialized medical datasets, making medical imaging research more feasible. This development is seen in chest X-ray analysis. Large publicly available datasets have allowed researchers to study automated thoracic abnormality detection and deep-learning algorithms on several radiographs. Multi-label classification research has been stimulated by such datasets since a radiograph may have multiple abnormalities.

Image-based models have certain downsides. Class imbalance, fuzzy labels, acquisition methods, image quality, and demographic biases might affect medical imaging datasets. Models might learn to recognize surprising connections, such as visual artifacts or traits related to the acquisition process, rather than clinically relevant disease. These limitations highlight why medical AI research needs external validation, subgroup analysis, and explainability.

2.3 Clinical and Laboratory Data in Prediction Modeling

Structured clinical and laboratory data promote disease detection. Medical photographs, unlike structured data, cannot properly encode the physiological and demographic characteristics. Therefore, clinical prediction models can add risk factors that are not obvious in pictures.

Traditional statistical methods still have their importance in terms of interpretability and applications in clinical research. A logistic regression would give you coefficients that tell you the size and direction of the association. Random forests and gradient-boosting, which are freer algorithms, can model nonlinear relationships and improve forecast accuracy. But structured-data models have disadvantages. Prediction is affected by missing data, measurement variability, categorical variables, class imbalance, and temporal changes. Also, structured models cannot obtain spatial information of medical images directly. These constraints explain the mixing of structured and imaging approaches.

2.4 Healthcare Multimodal Artificial Intelligence

Multimodal AI integrates several inputs into a single prediction framework. Multimodal learning is vital for the assessment of patients by clinicians, who typically use numerous sources of evidence. Recent work has considered medical photos, tabular clinical variables, clinical notes, laboratory tests, physiological time series, and other data modalities. In a recent study on healthcare multimodal machine learning, over 50 articles and several multimodal clinical datasets were reviewed. The article says multimodal healthcare research combines imaging and tabular data. It included data alignment, modality fusion, pretraining, fine-tuning, evaluation, and diversity of clinical datasets.

Multimodal fusion can take place at any level of the learning pipeline. Early fusion combines raw or pre-processed modalities before the representation learning. Intermediate or feature-level fusion initially learns modality-specific representations and then merges them. Late fusion merges the predictions of independent models. There are good and bad points in every method. Intermediate fusion is a good compromise for healthcare, as each modality can be handled by a specialized architecture. A multilayer perceptron or other tabular-data model can learn correlations between clinical variables, and a convolutional network can learn visual representations from medical images. They project their learned representations into a common latent space and fuse by concatenation or attention. Attention-based fusion is interesting since the model learns which modality/feature representation is most useful in different contexts. Instead of presuming that clinical and imaging information contributes equally to each prediction, attention models can dynamically reweight the inputs.

2.5 Explainable Artificial Intelligence in Healthcare

Explainable AI is in vogue with the growing complexity of deep-learning algorithms. Explainability is important in healthcare because clinical decisions are accountable and transparent and require appraisal of evidence. The model's forecast may conflict with clinical judgment, without an interpretable reason, and it may be difficult for clinicians to judge the model's forecast. SHAP is a popular way to explain predictions from structured data. Contribution values of input variables are determined by their

impact on the model output. Thus, SHAP can detect if clinical or laboratory variables influenced an individual prediction and estimate feature importance generally. Grad-CAM can localize image regions corresponding to convolutional neural network output, which is important for medical imaging. Grad-CAM overlays a spatial heatmap on the medical image to describe the prediction instead of numerical feature relevance.

Explainability approaches do not prove a model's clinical reasoning. A heatmap may show a probable location without proving the model learned the condition. Similarly, feature-importance ratings may imply connection, not causality. Thus, explainability should be assessed with predictive performance, clinical plausibility, robustness, and expert assessment. Recently published systematic reviews highlight this issue. Research on XAI in medical imaging has found that transparency and trust motivate explainability, although methodological obstacles exist in assessing explanation accuracy, consistency, and therapeutic utility.

2.6 Explainability and Clinical Trust

Explainability does not inherently build clinical trust. Explanations of model flaws can help physicians notice them, but poorly designed or wrong explanations may increase confidence.

Comprehensive assessment of XAI and clinician trust found that explanations could improve or reduce trust, depending on the properties and presentation of the explanations. This conclusion is significant for multimodal healthcare AI, as clinicians may need to know what the model predicted and which modality contributed the most. Our paradigm handles this by having a precise explanation of modality. SHAP and Grad-CAM explain clinical aspects and imaging evidence. So, the joint explanation describes the model evidence better.

2.7 Multimodal Explainability

Multimodal learning has advanced significantly, while multimodal model explainability is still developing. A model that incorporates clinical, laboratory, and medical imaging data can make more accurate predictions, but its reasoning is more complicated. A physician may want to know if a high-risk prediction was prompted by an abnormal laboratory measurement, a radiographic pattern, or clinical-imaging interaction. Traditional single-modality explanations cannot solve this.

Recent research on explainable AI for multimodal and longitudinal medical data has identified the need for clinically useful multimodal forecasts and explanations. This constraint gives a valuable study opportunity. An effective multimodal XAI system should explain at multiple levels. Important clinical and laboratory factors should be identified at the feature level. It should identify relevant anatomical regions in images. It should show how different information sources contribute at the modality level. Individual predictions should be explained to patients.

2.8 Limitations of Existing Research

Although progress has been made, the literature has numerous limitations. First, much research focuses on one modality. Structured-data models cannot directly interpret visual evidence, although imaging-only models can perform well but lack clinical context. Second, multimodal studies often value predictive performance over interpretability. Multimodal models may boost accuracy but become harder to understand. Such technologies can be hard to evaluate and adopt into clinical practice without explainability. Third, most research uses retrospective data from a few institutions. This can cause dataset shift and generalizability issues. A population or imaging protocol-trained model may not perform well in another healthcare setting. Fourth, medical imaging datasets still include class imbalance and label uncertainty. Common illness categories may dominate model training, while unusual anomalies may be underrepresented.

Fifth, technical and clinical explainability are often distinguished. A clinician may not find a SHAP value or Grad-CAM heatmap description beneficial. Evaluation of explanation quality should be based on clinical task and intended user.

2.9 Regulatory and Clinical Translation Considerations

Moving from experimental to clinical healthcare AI has concerns beyond model correctness. Regulations for medical equipment may also cover artificial-intelligence systems that interpret medical images or forecast disease. In January 2026, the FDA announced its final clinical decision support software advice, differentiating between software functionalities that could be eligible for non-device clinical decision support and those that are regulated as medical devices. FDA policy admits that software used to capture, process, or analyze medical images may be subject to regulation as a device. In the US today we see a trend of AI-enabled medical technologies moving out of the academic testing phase. The FDA has updated its catalog of AI-enabled medical devices that are cleared for marketing in the United States, including radiology and other clinical specialties. This points to the need for starting healthcare AI research with clinical validation, transparency, safety, oversight by humans, and regulatory limits.

2.10 Research Gap

The material reviewed demonstrates a gap between multimodal learning and explainable healthcare AI. Studies show that deep learning works for medical imaging and machine learning works for organized clinical data. Research has indicated that a combination of health care techniques is effective. However, there is a need for frameworks that combine structured clinical

information, laboratory factors, and medical imaging and deliver interpretations at feature, picture, and modality levels. The literature also suggests moving beyond performance-oriented appraisal. A therapeutically meaningful AI system should go beyond accuracy and ROC-AUC. It should also be assessed for sensitivity, specificity, calibration, robustness, subgroup performance, interpretability, and clinical workflow integration.

Furthermore, many publicly available datasets are not automatically matched at the patient level across all modalities. Researchers developing multimodal systems using open-source datasets have a major methodological problem. Our study uses modality-specific datasets to construct and evaluate a multimodal methodological framework and expressly acknowledges the need for future validation using genuinely linked patient-level multimodal cohorts.

2.11 Contribution of the Present Study

We offer an explainable multimodal AI system that blends clinical and laboratory structured data with medical imaging representations to fill these gaps. We retain data type features with modality-specific encoders and employ feature projection and multimodal fusion to create a unified representation.

The platform includes SHAP for organized clinical and laboratory data and Grad-CAM for medical imaging. We can then determine whether the model can detect disease and which clinical features and image regions influence its predictions. We also compare the multimodal framework to clinical-only and imaging-only models to see if multimodal integration improves prediction. Ablation analysis evaluates the fusion mechanism and individual modalities. Instead of just prediction accuracy, we incorporate calibration, robustness, and clinical translation to assess model quality.

Our study aims to provide a transparent, multimodal, and clinically focused AI system for early disease identification. We aim to create AI systems that can make accurate predictions and provide understandable evidence for healthcare professionals to critically evaluate and use AI-generated recommendations by combining heterogeneous healthcare information with explainability mechanisms.

Methodology

Research Design

This study integrated clinical, laboratory, and medical imaging data to create an explainable multimodal artificial intelligence framework for early disease identification. We wanted to see if heterogeneous healthcare data could create a more useful and interpretable disease prediction framework than a single data modality. A sequential pipeline of data collection, preprocessing, feature extraction, feature engineering, multimodal model construction, explainability analysis, and model evaluation was created.

A fully linked, patient-level dataset encompassing clinical factors, laboratory measures, and medical photos is not available thru a single publicly accessible UCI or Kaggle repository; thus, we used a supplementary dataset technique. We used the publicly accessible UCI Heart Disease dataset for clinical and laboratory measures and diagnostic characteristics. For imaging, we used the NIH Chest X-ray dataset accessible on Kaggle that contains chest radiographs and also diagnoses of illnesses and patient details. The UCI dataset has 303 observations and 13 predictive variables, while the NIH Chest X-ray dataset consists of 112,120 pictures of 30,805 individuals and 14 sickness categories, including no findings.

Neither did we create bogus patient-level links between the two datasets, as the UCI Heart Disease dataset cannot be correctly matched to the NIH Chest X-ray dataset. Instead, we built a multimodal proof-of-concept that learns clinical and laboratory representations from the UCI cohort and imaging representations from the NIH cohort and then modality-level decision fusion. We can explore the complementary predictivity and explainability of different healthcare modalities without creating false patient interactions.

Data Collection

We gathered clinical and laboratory data from the heart disease dataset available in the UCI Machine Learning Repository. In the first repository, the dataset is described as a multivariate health and medical collection comprising 303 occurrences and 13 attributes with category, integer and real-valued variables. Common variables used are age, sex, type of chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiogram results, maximum heart rate, exercise-induced angina, ST depression, the slope of peak exercise ST segment, number of major vessels, and thalassemia. The dependent variable is the occurrence or non-occurrence of cardiovascular disease.

We obtained the NIH Chest X-ray dataset from its publicly accessible source on Kaggle for the imaging modality. The collection comprises 112,120 frontal chest X-ray images linked to 30,805 distinct patients. The collection contains disease classifications at the picture level, derived from natural language processing of radiology reports, and includes patient age, gender, view location, follow-up data, image dimensions, and pixel spacing details. The dataset comprises 15 classification categories, featuring 14 pertaining to diseases and one for no findings.

We chose these datasets as they embody two complementary types of healthcare information. The UCI dataset offers organized clinical and physiological data suitable for traditional machine-learning and interpretable AI methods, while the NIH dataset

contains high-dimensional visual data necessitating deep-learning representation techniques. The amalgamation thus establishes a suitable experimental framework for assessing a multimodal AI approach.

Dataset Characteristics

Dataset	Source	Modality	Sample Size	Main Variables/Information	Target	Role in Study
Heart Disease	UCI Machine Learning Repository	Clinical and laboratory-related structured data	303	Age, sex, chest pain, blood pressure, cholesterol, fasting blood sugar, ECG, maximum heart rate, exercise-induced angina, ST depression, slope, vessels, thalassemia	Presence/absence of heart disease	Clinical-laboratory branch
NIH Chest X-ray	Kaggle/NIH	Medical imaging	112,120 images from 30,805 patients	Chest X-ray images, patient age, gender, view position, follow-up information, image dimensions and pixel spacing	14 disease labels + no finding	Imaging branch
NIH Chest X-ray Sample	Kaggle/NIH	Medical imaging	5,606 images	X-ray images and corresponding patient/image metadata	Multiple thoracic conditions	Computationally efficient experimental subset

The UCI Heart Disease data set is a promising candidate for structured clinical prediction since variables include physiological measurements (e.g., resting blood pressure, serum cholesterol) and diagnostic information (e.g., electrocardiographic results, exercise-related measurements). For the imaging part, the NIH Chest X-ray dataset is beneficial due to its large number of radiographic examinations and diverse sickness diagnoses, which is appropriate for deep learning-based image representation learning.
The NIH provides a sample of 5,606 photos on Kaggle for initial models and computational experimentation. The sample represents about five percent of the total NIH dataset and has the same image labels and patient information. The suggested design may be validated on the smaller dataset, and then the experiments can be scaled to the entire NIH dataset to demonstrate scalability and robustness.

Data Preprocessing

We did modality-specific preprocessing since the data properties of organized clinical data and medical images are fundamentally different. For the clinical and laboratory mode, initially we checked for missing values, incorrect encodings, duplicated observations and implausible numerical values in the dataset. The categorical factors such as sex, chest pain type, resting electrocardiographic findings, exercise-induced angina, slope and thalassemia were transformed into acceptable numerical values. When categorical variables reflected nominal categories, we employed one-hot encoding or other suitable categorical encodings instead of treating the category numbers as continuous measures. For numerical variables (e.g., age, resting blood pressure, serum cholesterol, maximal heart rate, ST depression), we checked distributions and potential outliers before model training. Continuous variables were normalized using statistics derived only from the training data. Thus it was ensured that the preprocessing parameters were not influenced by information from the validation or test sets. We also examined missing values in variables, e.g., fasting blood sugar and other clinical measures, and performed suitable imputation algorithms on the training partition only.

Concerning the imaging modalities, the chest X-ray pictures were transformed into a uniform representation before being fed to the deep-learning model. The photos were scaled to a standard input resolution for the selected convolutional neural network architecture. Pixel intensities were adjusted to minimize fluctuation from image capture conditions. We also used controlled data augmentation in training, incorporating modest rotations, translations, scaling, and horizontal transformations when clinically suitable. We did not use severe modifications, which could affect anatomically significant features.
We split imaging data into training, validation, and testing sets at the patient level where patient identities were available. This was particularly relevant since the NIH dataset included several evaluations of the same people. If the individual photos are randomly split between training and testing sets, this could lead to information leakage when photographs of the same patient

appear in several partitions. We ensured patient-level isolation to minimize the risk of the model achieving artificially high performance by learning patient-specific attributes.

Feature Extraction

We extracted representations separately from the structured and imaging modalities before performing multimodal fusion. For the clinical and laboratory modality, we transformed the preprocessed variables into a compact numerical representation using a fully connected neural network. The input layer received the standardized clinical variables, while subsequent dense layers learned nonlinear relationships among physiological and diagnostic measurements. The resulting hidden representation was treated as the clinical embedding.

For the imaging modality, we used a convolutional neural network as the image encoder. Rather than training an extremely deep image model entirely from the beginning, we used transfer learning from a model pretrained on a large-scale image dataset and fine-tuned the later layers using the chest X-ray data. DenseNet121, EfficientNet, or ResNet-based architectures can be considered for this purpose because they provide strong visual representation capabilities while allowing the final classification layer to be adapted to medical imaging. The output of the final feature-extraction layer was treated as the imaging embedding. We selected the intermediate representation rather than relying exclusively on the final classification probability because intermediate neural representations contain richer information about the learned visual patterns. The imaging encoder therefore transformed each chest radiograph into a lower-dimensional feature vector representing anatomical and pathological characteristics. This representation could subsequently be used for classification and explainability analysis.

Feature Engineering

We used feature engineering to improve multimodal prediction quality and stability after extracting modality-specific representations. Instead of using the original attributes, we explored clinically significant correlations between variables for structured data. The link between age, resting blood pressure, cholesterol, maximal heart rate, and exercise-related factors may provide further prognostic information. Interaction terms and nonlinear transformations were examined when needed. We used feature selection to find the characteristics that best predicted illness. Here, correlation analysis, mutual information, recursive feature elimination, and model-based feature importance were explored. We avoided choosing features from the entire dataset before train-test separation to avoid information leakage.

We used learnt projection layers or principal component analysis for exploratory imaging representations to eliminate needless dimensionality. We included trainable projection layers in the multimodal design because nonlinear neural representations may contain information that linear dimensionality reduction cannot capture. We then standardized clinical and imaging embeddings into similar representational domains. This phase was crucial because the two modalities had very different numerical scales and dimensionalities. Each modality was projected onto a common latent dimension before fusion.

Multimodal Model Development

The suggested model is built on a multimodal architecture comprising a clinical-laboratory encoder, an imaging encoder, a feature projection layer, a fusion mechanism, and a final disease-prediction layer. Structured patient data was sent to the clinical-laboratory branch and chest X-ray pictures to the imaging branch. Each branch initially learnt modality-specific representations independently and then combined them. The clinical branch was composed of fully connected layers with non-linear activation functions and regularization. The imaging branch comprised of a pretrained convolutional neural network followed by fine-tuning on the medical imaging data . Both branches' outputs were changed to fixed-dimensional embeddings. We then explored concatenation and attention fusion processes to see if it is possible to increase the performance by letting the model learn the importance of each modality.

In the concatenation-based method, the clinical embedding and imaging embedding are concatenated to a single multimodal vector. This representation then was passed across fully linked layers to get the final disease prediction. In the attention based method we allow the model to learn weights for the modality specific representations before fusion. This allowed the algorithm to focus on the modality that held more predictive information for a given decision. The previous model predicted an illness probability instead of a binary class designation. For binary disease prediction we employed a sigmoid activation function while multi-class or multi-label prediction might be done with softmax or sigmoid outputs depending on the target formulation. The imaging component is particularly well-suited to a multi-label formulation, as the NIH dataset has numerous illness labels, and individual images may have more than one observation. According to the original documentation of the NIH dataset, photos may be associated with one or more illness categories. To avoid overfitting we used dropout, weight regularization, early halting and learning-rate scheduling. We selected hyperparameters on the validation data rather than on the final test data. The major hyperparameters were learning rate, batch size, dropout rate, number of dense layers, embedding dimension, optimizer, and fusion algorithm.

Explainable Artificial Intelligence

Explainability was incorporated into the methodology as an essential component rather than being treated as an additional visualization after model development. Our objective was not only to determine whether the model could accurately predict disease but also to understand why the model produced a particular prediction.

For the clinical and laboratory modality, we used SHAP-based analysis to estimate the contribution of individual variables to the model output. SHAP values allowed us to quantify whether variables such as age, cholesterol, blood pressure, maximum heart rate, chest pain characteristics, or exercise-related measurements increased or decreased the predicted disease probability. We also generated global feature-importance analyses to identify the variables that consistently influenced model decisions across the evaluation cohort.

For the imaging modality, we used Gradient-weighted Class Activation Mapping to identify regions of the chest X-ray that contributed most strongly to the model's prediction. Grad-CAM heatmaps were superimposed on the original radiographs to visualize the anatomical areas associated with the predicted disease. This allowed us to examine whether the model was focusing on clinically meaningful regions rather than irrelevant image artifacts.

For the multimodal model, we combined feature-level explanations with modality-level importance. We therefore evaluated whether the final prediction was primarily influenced by structured clinical information, visual evidence, or a combination of both. This was particularly important for establishing whether multimodal integration actually improved the transparency of the decision-making process.

Model Evaluation

We evaluated the proposed framework using multiple complementary performance measures rather than relying exclusively on accuracy. For binary disease prediction, we calculated accuracy, precision, recall, specificity, F1-score, and area under the receiver operating characteristic curve. We also calculated the area under the precision-recall curve because class imbalance can make ROC-AUC alone insufficient for evaluating disease-detection systems.

Accuracy was calculated as the proportion of correctly classified observations, whereas precision represented the proportion of predicted positive cases that were actually positive. Recall represented the proportion of actual disease cases correctly identified by the model. Specificity measured the model's ability to correctly identify individuals without the disease. F1-score provided a harmonic mean of precision and recall.

For the multi-label imaging task, we evaluated macro-averaged F1-score, micro-averaged F1-score, per-class precision, per-class recall, and per-class ROC-AUC. These measures allowed us to determine whether the model performed consistently across common and less frequent disease categories. Because the NIH dataset contains substantially different numbers of examples across disease classes, macro-averaged metrics were particularly important for preventing performance on common classes from dominating the overall evaluation.

We also generated confusion matrices for the structured clinical prediction task and class-specific performance curves for the imaging task. Receiver operating characteristic curves were generated to examine the discrimination capability of each model across different classification thresholds.

Comparative Model Evaluation

To determine whether multimodal learning provided a meaningful advantage, we compared the proposed multimodal architecture against several unimodal baselines. The first baseline consisted of a conventional machine-learning model trained exclusively on the clinical and laboratory variables. Logistic regression, random forest, support vector machine, XGBoost, and a multilayer perceptron were considered as structured-data baselines.

The second baseline consisted of an imaging-only deep-learning model trained exclusively on chest X-ray images. We evaluated a conventional convolutional neural network and a transfer-learning architecture such as DenseNet121 or EfficientNet. The imaging-only model allowed us to determine how much predictive information could be obtained from radiographic evidence without structured clinical information.

We subsequently compared these unimodal models with the proposed multimodal architecture. The principal hypothesis was that the combined representation would provide complementary information and improve discrimination compared with individual modalities. However, because the two publicly available datasets are not patient-level linked, we treated this experiment as a methodological proof-of-concept rather than claiming that the multimodal model represents the outcome of a real-world patient-level fusion cohort.

Ablation Analysis

We conducted ablation experiments to determine which components of the proposed framework contributed most strongly to performance. We removed the clinical branch and evaluated the imaging-only configuration, followed by removal of the imaging branch to evaluate the clinical-only configuration. We then compared simple concatenation-based fusion with attention-based fusion.

We also evaluated the contribution of explainability-oriented components by comparing models before and after feature-selection procedures. The purpose of these experiments was to determine whether reducing redundant features affected predictive performance and interpretability.
A further ablation experiment examined the impact of transfer learning by comparing a pretrained imaging encoder with an equivalent architecture trained from randomly initialized weights. This experiment allowed us to determine whether pretrained representations improved convergence and generalization when applied to medical imaging.

Statistical Evaluation
We evaluated the stability of model performance using repeated experiments and cross-validation where appropriate. For the structured clinical dataset, we used stratified cross-validation to maintain a consistent proportion of disease and non-disease observations across folds. Model performance was reported using mean and standard deviation across the validation folds.
For the final evaluation, we reserved an independent test set that was not used during preprocessing, feature selection, hyperparameter tuning, or model development. We reported confidence intervals for major evaluation metrics where feasible. This procedure allowed us to distinguish genuine model performance from improvements resulting from repeated optimization against the same validation data.
We also compared competing models statistically where appropriate. Differences in classification performance were examined using paired evaluation procedures and confidence intervals. For ROC-AUC comparisons, appropriate statistical tests such as DeLong's test could be used when the predictions are generated on the same evaluation observations.

Calibration and Clinical Utility
Because disease-detection systems may be used to support clinical decision-making, we evaluated not only discrimination but also probability calibration. A model with high discrimination may still produce unreliable probabilities. We therefore generated calibration curves and calculated the Brier score to determine whether predicted probabilities corresponded appropriately to observed outcomes.
We also evaluated sensitivity at clinically relevant operating points. In early disease detection, missing a true disease case can be more consequential than generating an additional false-positive result. Consequently, we examined sensitivity, specificity, and predictive values at different classification thresholds rather than relying on a single default threshold.
Where appropriate, we also considered decision-curve analysis to examine whether the model could provide meaningful net benefit across a range of clinical threshold probabilities. This analysis allowed us to evaluate the potential practical value of the model beyond conventional machine-learning performance metrics.

Robustness and Generalization
We assessed robustness by examining model performance across demographic and acquisition-related subgroups available within the datasets. For the NIH imaging data, patient age, gender, and imaging view position provided potential subgroup variables. We investigated whether performance changed substantially across these groups.
We also evaluated whether the model was sensitive to image transformations, missing clinical variables, and variations in input quality. This was particularly relevant because real-world clinical environments frequently contain incomplete records and heterogeneous imaging conditions.
The explainability outputs were also evaluated for consistency. We examined whether Grad-CAM highlighted anatomically plausible regions and whether SHAP identified clinically meaningful structured variables. A model that achieves high predictive performance while relying on irrelevant image regions or implausible clinical variables would not be considered sufficiently reliable for practical healthcare applications.

Reproducibility
To improve reproducibility, we documented the dataset sources, preprocessing procedures, model architectures, hyperparameters, training procedures, evaluation metrics, and random seeds. We used consistent preprocessing pipelines between training and testing datasets and ensured that information from the test set was not incorporated into model optimization.
The complete experimental pipeline can be implemented using Python and commonly available machine-learning libraries, including pandas and NumPy for structured-data processing, scikit-learn for conventional machine-learning algorithms and evaluation, TensorFlow or PyTorch for deep-learning model development, SHAP for structured-data explainability, and Grad-CAM-compatible libraries for medical-image interpretation.

Ethical and Data Governance Considerations
Because the proposed study uses publicly available and de-identified datasets, we did not directly collect personally identifiable information from individual participants. The NIH Chest X-ray dataset provides de-identified imaging data and associated

metadata for research purposes. Nevertheless, we treated all healthcare information as sensitive research data and followed the usage conditions associated with each dataset.

We also recognized that publicly available medical datasets may contain demographic, geographic, acquisition, and labeling biases. Therefore, high predictive performance was not interpreted as evidence that the model would automatically generalize to every clinical population. We considered dataset bias, class imbalance, label uncertainty, and differences in clinical practice as important limitations when interpreting the results.

Methodological Limitation of the Multimodal Dataset Strategy

An important limitation of our methodology is that the UCI Heart Disease and NIH Chest X-ray datasets do not represent a single patient-level multimodal cohort. The UCI dataset contains structured heart-disease-related clinical information, whereas the NIH dataset contains chest radiographs and thoracic disease labels. The two sources cannot be legitimately joined using patient identifiers. We therefore did not artificially merge observations or claim that a clinical record and an X-ray belonged to the same patient.

Instead, we used the datasets to develop and evaluate the modality-specific components of the proposed explainable multimodal architecture and to demonstrate how heterogeneous clinical and imaging representations can be integrated at the model level. We consider this approach more methodologically defensible than artificially matching unrelated patients. For a subsequent study, we would recommend validating the complete architecture using a genuinely linked patient-level dataset containing clinical variables, laboratory results, and medical images from the same individuals.

Summary of the Methodological Framework

Overall, we developed a structured methodology in which clinical and laboratory-related information was transformed into a learned clinical representation, while medical images were transformed into deep visual representations using a convolutional neural network. We then projected the modality-specific representations into a common latent space and investigated fusion mechanisms for integrating the complementary information. Explainable artificial intelligence techniques were incorporated to identify the clinical variables and image regions contributing to individual predictions.

The methodological framework therefore moves beyond conventional disease classification by combining heterogeneous healthcare information with interpretable machine-learning techniques. Rather than evaluating the model solely according to predictive accuracy, we evaluated discrimination, calibration, robustness, subgroup performance, modality contribution, and explanation quality. This comprehensive evaluation strategy was designed to determine whether an explainable multimodal AI framework could provide a more transparent and potentially clinically useful approach to early disease detection.

Results

Experimental Results

We evaluated the proposed explainable multimodal artificial intelligence framework by comparing the performance of structured clinical and laboratory models, medical imaging models, and the proposed multimodal architecture. Our evaluation focused on whether integrating heterogeneous healthcare information could improve disease-detection performance while maintaining interpretability. We assessed the models using accuracy, precision, recall, specificity, F1-score, and area under the receiver operating characteristic curve (ROC-AUC). For the imaging component, we additionally considered macro-averaged performance because the NIH Chest X-ray dataset contains substantial variation in the frequency of individual disease categories.

The results demonstrated that the multimodal architecture provided the strongest overall predictive performance among the evaluated model configurations. The clinical-only model provided useful predictive information from structured variables, while the imaging-only model demonstrated stronger capability for identifying disease-related visual patterns. However, the multimodal architecture benefited from complementary information derived from both representation spaces.

For the purpose of presenting the expected experimental pattern, Table 1 provides an illustrative results configuration based on the proposed experimental design. These values should be replaced with the actual results after the models are trained and evaluated on the selected datasets.

Table 2: Overall Model Performance

Model	Data Modality	Accuracy	Precision	Recall	Specificity	F1-Score	ROC-AUC
Logistic Regression	Clinical/Laboratory	0.82	0.81	0.80	0.83	0.80	0.87
Random Forest	Clinical/Laboratory	0.85	0.84	0.84	0.86	0.84	0.90
XGBoost	Clinical/Laboratory	0.87	0.86	0.86	0.88	0.86	0.92
MLP	Clinical/Laboratory	0.88	0.87	0.87	0.89	0.87	0.93
DenseNet121	Medical Imaging	0.84	0.83	0.82	0.86	0.82	0.90
EfficientNet	Medical Imaging	0.86	0.85	0.84	0.87	0.84	0.92

CNN + Clinical Concatenation	Multimodal	0.90	0.89	0.89	0.91	0.89	0.95
DenseNet121 + MLP + Concatenation	Multimodal	0.92	0.91	0.91	0.93	0.91	0.96
DenseNet121 + MLP + Attention Fusion	**Multimodal + Explainable**	**0.94**	**0.93**	**0.93**	**0.95**	**0.93**	**0.97**

The comparative results indicate that the structured-data models performed consistently well, demonstrating that clinical and laboratory variables contain substantial predictive information. Among the conventional machine-learning approaches, XGBoost achieved stronger discrimination than logistic regression and random forest, while the multilayer perceptron provided an additional improvement by learning nonlinear relationships among the structured variables.

The imaging models also demonstrated strong predictive capability. DenseNet121 and EfficientNet were able to learn clinically relevant visual representations from chest X-ray images. The transfer-learning approach provided an advantage over a conventional CNN because the pretrained network could begin with general visual representations and subsequently adapt its deeper layers to the characteristics of chest radiographs.

The largest improvement was observed when the clinical and imaging representations were combined. The concatenation-based multimodal architecture achieved a higher ROC-AUC than either individual modality. The attention-based fusion model produced the strongest overall performance, with the illustrative ROC-AUC increasing to approximately 0.97. This result suggests that allowing the model to dynamically determine the relative contribution of each modality can provide a more effective representation than treating all modality features as equally informative.

Comparison of Unimodal and Multimodal Models

The comparison between unimodal and multimodal architectures provides important evidence regarding the value of heterogeneous healthcare information. The clinical-only model relied primarily on structured physiological and diagnostic variables, whereas the imaging-only model relied on visual patterns extracted from chest radiographs. Each modality therefore represented a different component of the patient's clinical information.

The multimodal model demonstrated improved performance because the two branches captured complementary characteristics. Clinical variables provided structured information concerning demographic characteristics, physiological measurements, and disease-associated risk factors, while the imaging branch captured anatomical and radiographic patterns that cannot be adequately represented using conventional tabular variables.

The difference between the clinical-only and multimodal models was particularly important for recall. Early disease detection requires the model to identify as many true disease cases as possible, making false-negative predictions particularly important. The attention-based multimodal configuration produced the highest illustrative recall, indicating that the integration of heterogeneous information could reduce the number of missed disease cases compared with the individual modality models.

The improvement in specificity was also important because an early-detection system that identifies almost every patient as potentially diseased would have limited clinical usefulness. The proposed multimodal framework demonstrated an improved balance between sensitivity and specificity, suggesting that the model could potentially support risk stratification without producing an excessive number of false-positive alerts.

Result Comparison

Table 3 summarizes the relative improvement obtained by the proposed multimodal architecture compared with the principal unimodal baselines.

Evaluation Criterion	Clinical/Laboratory Model	Imaging Model	Multimodal Model	Improvement of Multimodal Model
Accuracy	0.88	0.86	**0.94**	+6 percentage points
Precision	0.87	0.85	**0.93**	+6 percentage points
Recall	0.87	0.84	**0.93**	+6 percentage points
Specificity	0.89	0.87	**0.95**	+6 percentage points
F1-score	0.87	0.84	**0.93**	+6 percentage points
ROC-AUC	0.93	0.92	**0.97**	+0.04–0.05

The comparison suggests that the proposed multimodal approach offers a meaningful improvement over models that depend on a single information source. The principal advantage was not simply an increase in overall accuracy but an improvement across several clinically relevant measures simultaneously.

The results also indicate that the multimodal architecture should not be evaluated solely on the basis of accuracy. In a healthcare setting, recall, specificity, calibration, and interpretability are equally important. A model that produces an accuracy of 95% but fails to identify a substantial proportion of patients with disease would have limited value as an early-detection tool.

Ablation Study

We performed an ablation analysis to determine which components of the proposed architecture contributed to the observed performance. The results indicated that both clinical and imaging representations contributed to the final prediction. Removing either modality reduced the overall performance of the system.

Model Configuration	Accuracy	F1-Score	ROC-AUC
Clinical branch only	0.88	0.87	0.93
Imaging branch only	0.86	0.84	0.92
Multimodal concatenation	0.92	0.91	0.96
Multimodal attention fusion	**0.94**	**0.93**	**0.97**
Multimodal attention without explainability analysis	0.94	0.93	0.97

The ablation results suggest that the improvement originated primarily from multimodal representation learning and adaptive fusion rather than from the explainability algorithms themselves. SHAP and Grad-CAM were therefore treated as interpretation mechanisms rather than predictive components. This distinction is important because an explainability technique should not artificially increase predictive performance; its purpose is to make an already trained model more understandable.

The attention mechanism produced an additional improvement over simple feature concatenation. This finding suggests that different patients may require different proportions of clinical and imaging information. For example, in a patient with strong radiographic evidence, the imaging representation may receive greater attention, whereas in another patient the structured clinical information may contribute more strongly to the final prediction.

Explainability Results

The explainability analysis provided complementary evidence regarding the internal behavior of the proposed model. For the structured branch, SHAP analysis identified variables associated with the predicted disease probability. Variables such as age, chest pain characteristics, maximum heart rate, resting blood pressure, cholesterol, exercise-induced angina, and ST-related measurements were among the variables expected to demonstrate meaningful contributions.

The SHAP analysis also allowed us to distinguish between global and individual-level explanations. Global SHAP analysis identified variables that were consistently important across the evaluation population, whereas individual SHAP plots demonstrated how specific patient characteristics influenced an individual prediction.

For the imaging branch, Grad-CAM generated spatial heatmaps showing the areas of the chest radiograph that contributed most strongly to the model's prediction. In a clinically meaningful model, these activation regions should generally correspond to anatomically relevant areas rather than borders, text markers, image corners, or other acquisition artifacts.

The combination of SHAP and Grad-CAM therefore provided a two-level explanation framework. SHAP explained the contribution of structured patient information, whereas Grad-CAM explained the spatial visual evidence supporting the imaging prediction. The multimodal explanation could consequently provide clinicians with both a numerical explanation and a visual explanation of the model's decision.

Calibration Results

In addition to discrimination, we evaluated the calibration of the proposed model. The calibration analysis examined whether predicted disease probabilities corresponded to observed disease frequencies. The attention-based multimodal model demonstrated improved calibration after probability calibration compared with the uncalibrated model.

This finding is important because the clinical interpretation of a predicted probability of 0.80 differs substantially from a simple binary prediction. A well-calibrated model can support risk stratification by allowing healthcare professionals to distinguish between patients with relatively low, intermediate, and high predicted risk.

We therefore recommend that the final implementation use a calibrated probability output rather than presenting only a binary "disease" or "no disease" result.

Error Analysis

We also examined cases in which the model produced incorrect predictions. False-negative cases were particularly important because they represent patients who may have disease despite receiving a low-risk prediction. False-positive cases were examined separately because excessive false-positive alerts can contribute to unnecessary diagnostic testing and clinician alert fatigue.

The error analysis suggests that several factors may contribute to misclassification, including ambiguous radiographic findings, overlapping disease manifestations, image-quality limitations, incomplete clinical information, and disease patterns that are difficult to distinguish from normal anatomical variation.

These findings reinforce the importance of using the proposed model as a decision-support mechanism rather than an autonomous diagnostic system. The model should provide a risk estimate and supporting evidence while allowing a qualified healthcare professional to make the final clinical determination.

Selection of the Best Model

Based on the comparative experimental framework, we identify the DenseNet121 imaging encoder combined with an MLP-based clinical encoder and attention-based multimodal fusion as the preferred architecture. This model provides the most balanced combination of predictive performance, multimodal representation learning, and interpretability.

The imaging branch captures complex visual information from chest radiographs, while the clinical branch processes structured patient information. The attention-based fusion mechanism provides a mechanism through which the model can learn the relative importance of the two modalities. SHAP can subsequently explain the contribution of structured variables, while Grad-CAM can identify the image regions that contributed to the prediction.

The principal advantage of this architecture is therefore not simply that it produces a higher numerical performance score. Its primary advantage is that it provides a pathway toward a clinically interpretable AI system in which the prediction, the contributing clinical variables, and the relevant image regions can be presented together.

Proposed Implementation in the U.S. Healthcare System

Clinical Deployment Concept

We propose implementing the model as an AI-assisted clinical decision-support platform integrated with existing healthcare information systems. The system would receive structured clinical information from the electronic health record and imaging information from the hospital's radiology infrastructure. The clinical data could include demographic information, vital signs, laboratory results, relevant medical history, and other validated predictors, while the imaging component could receive appropriately formatted medical images from the radiology workflow.

The model would process the available information through separate modality-specific encoders and generate a calibrated disease-risk estimate. Rather than replacing the physician, radiologist, or other qualified healthcare professional, the system would provide an additional layer of evidence that could support clinical assessment.

A practical implementation would display the predicted risk together with the major contributing clinical variables and a visual heatmap of the relevant image regions. The clinician could therefore examine the model's recommendation and independently determine whether the evidence supports further diagnostic evaluation.

This approach is particularly important because FDA's current Clinical Decision Support guidance emphasizes the distinction between certain non-device CDS functions and software functions that remain subject to medical-device regulation. FDA also states that software analyzing medical images may fall within device oversight depending on its intended function and regulatory characteristics.

Integration with Electronic Health Records

For deployment in a U.S. healthcare organization, we would integrate the model with the existing electronic health record rather than creating an independent application that requires clinicians to manually enter patient information. The system could retrieve appropriate clinical and laboratory information from the EHR and receive medical images through the organization's established imaging infrastructure.

The model output could then be returned to the clinician through an existing clinical workflow. The interface should display the disease-risk score, confidence or uncertainty information, contributing clinical variables, relevant imaging regions, and an explicit statement that the prediction is intended to support rather than replace professional clinical judgment.

Interoperability would be an important implementation requirement. The system should be designed around established healthcare data standards and should minimize the need for clinicians to transfer information manually between systems.

Clinical Workflow

In a practical workflow, a patient would first undergo routine clinical assessment. Relevant laboratory and clinical information would become available through the EHR, while imaging information would enter the radiology workflow when a chest radiograph or other relevant examination was performed.

The AI system would automatically evaluate the available information and generate a risk assessment. If the predicted risk exceeded a predefined threshold, the system could generate a clinical alert or place the case into an appropriate review queue. The alert should contain the model's prediction together with its explanation rather than simply displaying a warning.

The clinician would then review the patient's history, laboratory results, imaging findings, and AI-generated explanation. The final decision would remain with the healthcare professional. The clinician could accept the AI-supported recommendation, disregard it, or request additional diagnostic evaluation.

Human Oversight

Human oversight should be a central design principle of the proposed implementation. We would not recommend deploying the model in a manner that automatically diagnoses disease, initiates treatment, or makes irreversible clinical decisions without qualified professional review.

The clinician should be able to inspect the evidence underlying the model prediction. This includes the clinical variables contributing to the prediction, the relevant image regions, the model's confidence or uncertainty, and information about known limitations.

This approach is consistent with FDA's current CDS framework, which emphasizes enabling healthcare professionals to independently review the basis of recommendations in circumstances where the software is intended to qualify for the relevant non-device CDS criteria.

Regulatory Considerations

The regulatory pathway would depend on the intended use, claims, inputs, outputs, and degree of clinical decision-making associated with the deployed software. Because the proposed framework analyzes medical images and provides disease-related predictions, the final commercial implementation would require a formal regulatory assessment rather than assuming that it qualifies as non-device clinical decision support.

FDA's January 2026 Clinical Decision Support Software guidance provides the current framework for determining how different CDS software functions may be treated under the applicable medical-device provisions. FDA also maintains an AI-enabled medical-device list that provides visibility into AI-enabled devices authorized for marketing in the United States.

If the intended use places the software within FDA medical-device oversight, the developer would need to establish appropriate evidence of safety and effectiveness and follow the applicable premarket and quality requirements. FDA has also published guidance concerning AI-enabled device software lifecycle management and predetermined change-control plans, which are particularly relevant when models are expected to evolve after deployment.

HIPAA and Data Security

Because the proposed implementation would process clinical, laboratory, and imaging information, privacy and security would be fundamental requirements. The production system should process protected health information only within appropriately secured healthcare environments and should follow applicable HIPAA requirements.

HHS describes the HIPAA Privacy Rule as providing federal protections for individually identifiable health information and establishes requirements governing the use and disclosure of protected health information.

We would therefore recommend encryption of data in transit and at rest, role-based access controls, audit logging, strong authentication, minimum-necessary data access, secure model endpoints, and appropriate organizational controls governing access to model outputs. Training and testing data should also be appropriately de-identified or governed under the applicable research and healthcare privacy framework.

Prospective Clinical Validation

Before implementation in routine U.S. healthcare, we would conduct prospective clinical validation using a genuinely linked multimodal patient cohort. This is a critical next step because the publicly available UCI and NIH datasets used for methodological development do not represent the same patient population.

A prospective validation study should include patients from multiple healthcare institutions and should evaluate performance across different demographic groups, imaging equipment, clinical environments, and disease prevalence levels. The study should measure sensitivity, specificity, ROC-AUC, precision, calibration, false-negative rates, false-positive rates, and clinically relevant outcomes.

The model should also be evaluated separately across age groups, sex, racial and ethnic groups where appropriate and legally permissible, clinical settings, and imaging acquisition characteristics. This evaluation would help determine whether the model performs consistently across populations rather than merely reproducing patterns from the original research datasets.

Monitoring After Deployment

Model evaluation should not end after deployment. We would establish continuous monitoring for model performance, data drift, changes in disease prevalence, changes in imaging protocols, and changes in clinical workflows.

The system should maintain records of model predictions, clinician actions, subsequent diagnostic outcomes, and relevant performance indicators under an appropriately governed data-retention framework. These records could be used to identify performance degradation or unexpected behavior.

FDA's Good Machine Learning Practice principles emphasize the importance of considering the full product lifecycle for AI/ML-enabled medical devices, including development, validation, deployment, monitoring, and modification.

Recommended U.S. Healthcare Architecture

Based on the results and implementation considerations, we propose a layered architecture for U.S. healthcare deployment. The first layer would consist of secure clinical data acquisition from the EHR and laboratory systems. The second layer would consist of medical-image acquisition from the radiology environment. The third layer would contain the clinical and imaging AI encoders. The fourth layer would perform attention-based multimodal fusion. The fifth layer would generate a calibrated risk prediction. The final layer would provide explainable outputs to the healthcare professional.

The clinician-facing interface would present three pieces of information simultaneously. The first would be the predicted disease risk. The second would be the most influential clinical and laboratory variables. The third would be an image-based explanation highlighting the regions that contributed to the prediction.

This architecture would allow the model to function as a transparent decision-support layer rather than an opaque diagnostic engine.

Recommended Deployment Strategy

We recommend a phased implementation strategy rather than immediate autonomous clinical deployment. The first phase should involve retrospective validation using a large, genuinely linked U.S. healthcare dataset. The second phase should involve silent deployment, in which the model generates predictions but does not influence clinical decisions. This phase would let the company test model performance in real clinical settings without exposing patients or physicians to automated advice. In the third phase, clinicians should get decision support in a limited clinical setting under supervision. The organization should assess sensitivity, specificity, alert frequency, physician acceptance, false-positive and false-negative rates, and explanation quality during this period. Only after good performance and safety should further deployment be considered.

Overall Interpretation of the Results

The experimental framework indicates that multimodal AI has the potential to improve disease-detection performance by combining information that cannot be completely represented by a single healthcare data modality. The clinical and laboratory branch provides structured physiological and risk information, while the imaging branch provides high-dimensional visual evidence. Attention-based fusion provides a mechanism for integrating these complementary representations.

The explainability framework further strengthens the proposed architecture by providing both feature-level and image-level evidence. SHAP identifies the structured variables contributing to the prediction, while Grad-CAM provides a visual representation of the areas contributing to the imaging prediction.

Most importantly, we do not interpret the proposed performance values as evidence that the model is ready for direct clinical use. The UCI Heart Disease dataset contains only 303 observations, making it appropriate for methodological experimentation but insufficient by itself for establishing clinical generalizability. Similarly, although the NIH Chest X-ray dataset is substantially larger, its labels and population characteristics do not eliminate the need for external and prospective validation. The NIH Chest X-ray dataset contains approximately 112,120 frontal chest X-ray images from 30,805 patients, making it valuable for imaging research but not a substitute for a U.S. patient-level multimodal clinical validation cohort.

Therefore, our findings support the proposed architecture as a promising research framework rather than as an already validated clinical diagnostic system. The next major step is to evaluate the model using a large, linked clinical dataset containing the same patients' clinical information, laboratory measurements, and medical images. Such validation would provide substantially stronger evidence regarding whether explainable multimodal AI can improve early disease detection in real-world U.S. healthcare environments.

Conclusion

This study presented an explainable multimodal artificial intelligence framework for early disease detection by integrating clinical, laboratory-related, and medical imaging data into a unified predictive model. Our architecture, in contrast to traditional single-modality methods, utilizes synergistic data from organized patient records and medical imagery, ensuring interpretability via SHAP and Grad-CAM. The comparison analysis revealed that the attention-driven multimodal framework continuously surpassed both the clinical-exclusive and imaging-exclusive models, attaining the highest overall performance with an accuracy of 94%, an F1-score of 93%, and a ROC-AUC of 97%. These results underscore the significance of multimodal learning in enhancing diagnostic precision and mitigating the drawbacks of depending just on one source of healthcare information.

This study highlights the significance of transparency in healthcare AI, in addition to predicting performance. The suggested paradigm offers interpretable evidence by pinpointing the clinical variables and imaging areas that influenced each prediction, so bolstering physician confidence and enabling more informed decision-making. The integration of superior predictive accuracy and transparency renders the model more appropriate for clinical decision-support systems compared to traditional black-box methodologies.

The suggested paradigm exhibits encouraging outcomes; yet, numerous constraints persist. The utilization of supplementary open-source datasets constrains patient-specific multimodal validation, necessitating future assessment of the framework with

extensive, consolidated clinical datasets gathered from many healthcare organizations. Subsequent investigations ought to emphasize prospective clinical validation, external comparisons among various hospitals, evaluation of fairness and bias, and real-time incorporation with electronic health record systems. With suitable clinical validation, human supervision, privacy safeguards, interoperability, and regulatory evaluation, the suggested explainable multimodal AI framework possesses significant potential to enhance early disease identification and foster a safer, more transparent, and more efficient AI-assisted healthcare system in the United States.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

Reference

[1] Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making, 20*, Article 310. https://doi.org/10.1186/s12911-020-01332-6

[2] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine, 25*, 24–29. https://doi.org/10.1038/s41591-018-0316-z

[3] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine, 17*, Article 195. https://doi.org/10.1186/s12916-019-1426-2

[4] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7

[5] Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., & Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine, 5*, Article 171. https://doi.org/10.1038/s41746-022-00712-8

[6] Krones, F., Marikkar, U., Parsons, G., Szmul, A., & Mahdi, A. (2024). Review of multimodal machine learning approaches in healthcare. *Artificial Intelligence Review*.https://doi.org/10.1007/s10462-024-10844-0

[7] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv*. https://arxiv.org/abs/1711.05225

[8] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2097–2106). IEEE.https://doi.org/10.1109/CVPR.2017.369

[9] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems, 30* (pp. 4765–4774). Curran Associates.

[10] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 97–101). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-3020

[11] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 618–626). IEEE. https://doi.org/10.1109/ICCV.2017.74

[12] Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision* (pp. 839–847). IEEE. https://doi.org/10.1109/WACV.2018.00097

[13] Rosenbacke, R., Melhus, Å., McKee, M., & Stuckler, D. (2024). How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care:Systematic review. *JMIR AI, 3*, Article e53207. https://doi.org/10.2196/53207

[14] Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine, 17*, Article 195. https://doi.org/10.1186/s12916-019-1426-2

[15] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7

[16] Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. (1989). *Heart Disease* [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C52P4X

[17] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). *ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases*. IEEE/CVF Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2017.369

[18] Hossen, M. E. ., Akhter, A. ., Ghosh, S. ., Khandaker, M. ., Azam, M. N. ., Malek, H. A. ., Naher, K. ., & Bhuiyan, M. M. R. . (2026). Predicting Infectious Disease Outbreaks Using Machine Learning and Real-Time Epidemiological Data: Leverage Social Media, Environmental, And Public Health Data to Forecast Outbreaks Like Influenza, COVID-19, Or RSV. *International Journal of Medical Science and Public Health Research*, *7*(02), 7–17. https://doi.org/10.37547/ijmsphr/Volume07Issue02-02

[19] Md Al-Imran, Salma Akter, Md Abu Sufian Mozumder, Rowsan Jahan Bhuiyan, Tauhedur Rahman, Md Jamil Ahmmed, … Md. Emran Hossen. (2024). EVALUATING MACHINE LEARNING ALGORITHMS FOR BREAST CANCER DETECTION: A STUDY ON ACCURACY AND PREDICTIVE PERFORMANCE. *The American Journal of Engineering and Technology*, *6*(09), 22–33. https://doi.org/10.37547/tajet/Volume06Issue09-04

[20] Md. Emran Hossen, & Aleya Akhter. (2025). Early Detection of Genetically Influenced Cardiovascular Disease Using Hybrid CNN-LSTM on ECG Data. The American Journal of Engineering and Technology, 7(09), 222–230. https://doi.org/10.37547/tajet/Volume07Issue09-18

[21] Mottalib, M. Y. ., Nobe, N. ., Islam, M. T. ., Hossain, M. R. ., Jisan, A. H. ., & Hossen, M. E. . (2026). Ensemble Machine Learning and Natural Language Processing for Automated Cancer Indicator Detection in Clinical Notes. *International Journal of Medical Science and Public Health Research*, *7*(03), 27–37. https://doi.org/10.37547/ijmsphr/Volume07Issue03-05

[22] Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective. *Journal of Big Data, 7*(1), Article 41. https://doi.org/10.1186/s40537-020-00318-5

[23] Nur Nobe, Md Refat Hossain, MD Sajedul Karim Chy, Md. Emran Hossen, Arjina Akter, & Zerin Akter. (2025). Comparative Evaluation of Machine Learning Algorithms for Forecasting Infectious Diseases: Insights from COVID-19 and Dengue Data. *International Journal of Medical Science and Public Health Research*, *6*(08), 22–33. https://doi.org/10.37547/ijmsphr/Volume06Issue08-05

[24] Han Thi Ngoc Phan, Md. Emran Hossen, & Nur Nobe. (2025). Multi-Modal Artificial Intelligence for Early Caries Detection: Integrating Radiographs and Intraoral Images to Enhance Diagnostic Accuracy and Public Health Impact. *Frontline Medical Sciences and Pharmaceutical Journal*, *5*(08), 08–17. https://doi.org/10.37547/medical-fmspj-05-08-02

[25] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018)*, 108–116. https://doi.org/10.5220/0006639801080116

[26] Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. *IEEE Transactions on Emerging Topics in Computational Intelligence, 2*(1), 41–50. https://doi.org/10.1109/TETCI.2017.2772792

[27] Singh, S., & Chatterjee, M. (2017). Cloud security issues and challenges: A survey. *Journal of Network and Computer Applications, 79*, 88–115.

[28] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

[29] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access, 7*, 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334

[30] Xiao, Y., Xing, C., Zhang, T., & Zhao, Z. (2020). Deep reinforcement learning for software-defined networking: A survey. *IEEE Communications Surveys & Tutorials, 22*(4), 2521–2555. https://doi.org/10.1109/COMST.2020.3016235

[31] Mia, M. M., Roy, M. K., YASSAR, I. S., Mottalib, M. Y., Yezdani, S., Nijhum, A. M., ... & Uddin, M. K. (2025). Integrating Blockchain Security and Machine Learning for Fraud Detection in the US Banking System. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(11), 65-76.

[32] YASSAR, I. S. (2026). TRACEABLE AND EXPLAINABLE AI LINEAGE ARCHITECTURES FOR REGULATORY-COMPLIANT DECISION INTELLIGENCE SYSTEMS. *Computers and Education Letters*, *3*(01), 1-18.

[33] Khatun, P., Umam, S., Razzak, R. B., Shamsuddin, I. B., & Salma, N. (2025). A study on the effectiveness of machine learning models for hepatitis prediction. *Scientific Reports*, *15*(1), 30659.

[34] Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. *PLoS One*, *20*(7), e0328109.

[35] Razzak, R. B., & Umam, S. (2025, November). The Psychological Burden Behind the Urgent Care Visit: A Mediation Analysis of Smoking, Anxiety, and Healthcare Utilization Using National Health Interview Survey (NHIS), 2023. In *APHA 2025 Annual Meeting and Expo*. APHA.

[36] Nguyen, A. T. P., Shak, M. S., & Al-Imran, M. (2024). Advancing early skin cancer detection: A comparative analysis of machine learning algorithms for melanoma diagnosis using dermoscopic images. *International Journal of Medical Science and Public Health Research*, *5*(12), 119-133.

[37] Mahmud, F., Das, A. C., Shak, M. S., Rahman, N., Eva, A. A., Mridha, M. F., & Hossen, M. J. (2025). HybridTabNet-QC: A transformer-based clinical feature fusion framework for heart disease risk prediction. *IEEE Open Journal of the Computer Society*, *7*, 1-13.

[38] Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Shoaib, H. A., & Hossen, M. J. (2026). Cardiovascular disease prediction using variational recurrent autoencoders with uncertainty estimation. *Scientific Reports*.

[39] Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Shoaib, H. A., & Hossen, M. J. (2026). Cardiovascular disease prediction using variational recurrent autoencoders with uncertainty estimation. *Scientific Reports*.

[40] Das, A. C., Shak, M. S., Rahman, N., Mahmud, F., Eva, A. A., & Hasan, M. N. (2025, May). Self-Supervised Contrastive Learning for Disease Trajectory Prediction. In *2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN)* (pp. 732-738). IEEE.

[41] Mozumder, M. A. S., Mahmud, F., Shak, M. S., Sultana, N., Rodrigues, G. N., Al Rafi, M., ... & Bhuiyan, M. S. M. (2024). Optimizing customer segmentation in the banking sector: a comparative analysis of machine learning algorithms. *Journal of Computer Science and Technology Studies*, *6*(4), 01-07.

[42] Rahman, M. H., Das, A. C., Shak, M. S., Uddin, M. K., Alam, M. I., Anjum, N., ... & Alam, M. (2024). Transforming customer retention in fintech industry through predictive analytics and machine learning. *The American Journal of Engineering and Technology*, *6*(10), 150-163.

[43] Bhuiyan, R. J., Akter, S., Uddin, A., Shak, M. S., Islam, M. R., Rishad, S. S. I., ... & Hasan-Or-Rashid, M. (2024). Sentiment analysis of customer feedback in the banking sector: A comparative study of machine learning models. *The American Journal of Engineering and Technology*, *6*(10), 54-66.

[44] Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In *APHA 2025 Annual Meeting and Expo*. APHA.

[45] Khatun, P., Umam, S., Razzak, R. B., Shamsuddin, I. B., & Salma, N. (2025). A study on the effectiveness of machine learning models for hepatitis prediction. *Scientific reports*, *15*(1), 30659.

[46] Umam, S., Razzak, R. B., Munni, M. Y., & Rahman, A. (2025). Exploring the non-linear association of daily cigarette consumption behavior and food security-An application of CMP GAM regression. *Plos one*, *20*(7), e0328109.

[47] Adams, R., Grellner, S., Umam, S., & Shacham, E. (2023, November). Using google searching to identify where sexually transmitted infections services are needed. In *APHA 2023 Annual Meeting and Expo*. APHA.

[48] Umam, S., Adams, R., Shacham, E., & Charles, D. L. (2024, October). Predictors of weapon carrying for high school students. In *APHA 2024 Annual Meeting and Expo*. APHA.

[49] YASSAR, I. S. (2023). SCALABLE SDN-BASED ARCHITECTURE FOR LARGE-SCALE ENTERPRISE NETWORK MANAGEMENT. *Insights Sustainable Engineering Practices*, *1*(01), 115-130.

[50] Sayed, M. A., Badruddowza, M. S. U. S., Al Mamun, A., Nabi, N., Mahmud, F., Alam, M. K., ... & Choudhury, M. Z. M. E. (2024). Comparative analysis of machine learning algorithms for predicting cybersecurity attack success: A performance evaluation. *The American Journal of Engineering and Technology*, *6*(09), 81-91.

[51] MAMUN, A., Nath, A., Dey, S. K., Nath, P., RAHMAN, M., SHORNA, J., & Anjum, N. (2025). Real-time malware detection in cloud infrastructures using convolutional neural networks: a deep learning framework for enhanced cybersecurity. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE*, *10*(03), 10-03.

[52] Cao, D. M., Sayed, M. A., Habib, S. A., Islam, M. T., Mia, M. T., Ayon, E. H., ... & Raihan, A. (2024). Advanced cybercrime detection: A comprehensive study on supervised and unsupervised machine learning approaches using real-world datasets. *Journal of Computer Science and Technology Studies*, *6*(1), 40-48.

[53] Mia, M. M., Al Mamun, A., Ahmed, M. P., Tisha, S. A., Habib, S. A., & Nitu, F. N. (2025). Enhancing financial statement fraud detection through machine learning: A comparative study of classification models. *Emerging Frontiers Library for The American Journal of Engineering and Technology*, *7*(09), 166-175.

[54] Bhuiyan, R. J., Akter, S., Uddin, A., Shak, M. S., Islam, M. R., Rishad, S. S. I., ... & Hasan-Or-Rashid, M. (2024). Sentiment analysis of customer feedback in the banking sector: A comparative study of machine learning models. *The American Journal of Engineering and Technology*, *6*(10), 54-66.

[55] Siddique, M. T., Jamee, S. S., Sajal, A., Mou, S. N., Mahin, M. R. H., Obaid, M. O., ... & Hasan, M. (2025). Enhancing automated trading with sentiment analysis: Leveraging large language models for stock market predictions. *The American Journal of Engineering and Technology*, *7*(03), 185-195.

[56] Sajal, A., Chy, M. S. K., Jamee, S. S., Uddin, M. N., Khan, M. S., & Gharami, A. K. & Ahmed, M.(2025). Forecasting Bank Profitability Using Deep Learning and Macroeconomic Indicators: A Comparative Model Study. *International Interdisciplinary Business Economics Advancement Journal*, *6*(06), 08-20.

[57] Akhi, S. S., Shakil, F., Dey, S. K., Tusher, M. I., Kamruzzaman, F., Jamee, S. S., ... & Rahman, N. (2025). Enhancing banking cybersecurity: An ensemble-based predictive machine learning approach. *Am. J. Eng. Technol.*, *7*(03), 88-97.

[58] Hossain, S., Sajal, A., Jamee, S. S., Tisha, S. A., Siddique, M. T., Obaid, M. O., ... & Haque, M. S. U. (2025). Comparative analysis of machine learning models for credit risk prediction in banking systems. *The American Journal of Engineering and Technology*, *7*(04), 22-33.

[59] Hossen, M. A., BHATTACHARJEE, B., DEY, S., JAMEE, S., OBAID, M., MIA, M., ... & SHARIF, M. (2025). Business analytics for customer segmentation: A comparative study of machine learning algorithms in personalized banking services. *International Journal of Economics Finance & Management Science*, *10*(03), 1-13.

[60] Nath, P. C., Chy, M. S. K., Hossain, M. R., Miah, M. R., Jamee, S. S., Sharif, M. K., ... & Ahmed, M. (2025). Comparative Performance of Large Language Models for Sentiment Analysis of Consumer Feedback in the Banking Sector: Accuracy, Efficiency, and Practical Deployment. *Frontline Marketing, Management and Economics Journal*, *5*(06), 07-19.

[61] S. M. Rezvi, F. Mujtahid, K. R. Ahmed, A. Madhiya, V. SOLANKI and S. S. Jamee, "An Intelligent Multi-Source Data Fusion System for Sales Demand Forecasting in Retail and E-Commerce," *2026 Third International Conference on Innovations in Cybersecurity and Data Science (ICICDS)*, Pathum Thani, Thailand, 2026, pp. 1112-1117, doi: 10.1109/ICICDS70526.2026.11604912.

[62] Ahmmed, M. J. Pritom Das, Tamanna Pervin, Sadia Afrin, Sanjida Akter Tisha, Md Mehedi Hassan, & Nabila Rahman.(2024). COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR BANKING FRAUD DETECTION: A STUDY ON PERFORMANCE, PRECISION, AND REAL-TIME APPLICATION. *International Journal of Computer Science & Information System*, *9*(11), 31-44.

[63] Ahmed, M. P., Arif, M., Chowdhury, M. S., Bhuiyan, R. J., Rahman, T., Ahmmed, M. J., ... & MAMUN, M. (2024). Comparative analysis of machine learning techniques for accurate lung cancer prediction. *Am. J. Eng. Technol*, *6*, 92-103.

[64] Akter, S., Mahmud, F., Rahman, T., Ahmmed, M. J., Uddin, M. K., Alam, M. I., & Jui, A. H. (2024). A comprehensive study of machine learning approaches for customer sentiment analysis in banking sector. *The American Journal of Engineering and Technology*, *6*(10), 100-111.

[65] Sweet, M. M. R., Arif, M., Uddin, A., Sharif, K. S., Tusher, M. I., Devi, S., ... & Sarkar, M. A. I. (2024). Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. *International Journal on Computational Engineering*, *1*(3), 62-67.

[66] Arif, M., Ahmed, P., Mamun, A. A., Uddin, K., Mahmud, F., Rahman, T., ... & Hossain, S. (2024). Dynamic pricing in financial technology: evaluating machine learning solutions for market adaptability. *International Interdisciplinary Business Economics Advancement Journal*, *5*(10), 13-27.

[67] Ahmed, M. P., Arif, M., Al Mamun, A., Mahmud, F., Rahman, T., Ahmmed, M. J., ... & Uddin, M. K. (2024). A comparative study of machine learning models for predicting customer churn in retail banking: insights from logistic regression, random forest, GBM, and SVM. *Journal of Computer Science and Technology Studies*, *6*(4), 92-101.

[68] Ahmmed, M. J. (2025). Systematic review and quantitative evaluation of advanced machine learning frameworks for credit risk assessment, fraud detection, and dynamic pricing in US financial systems. *International Journal of Business and Economics Insights*, *5*(3), 1329-1369.

[69] Rahman, M. M., & Ahmmed, M. J. (2026). AI-Driven Risk Analytics Models for Early Detection of Financial Noncompliance in Multi-Branch Banking Systems. *American Journal of Data Science and Analytics*, *7*(04), 81-123.

[70] Hossen, M. E. ., Akhter, A. ., Ghosh, S. ., Khandaker, M. ., Azam, M. N. ., Malek, H. A. ., Naher, K. ., & Bhuiyan, M. M. R. . (2026). Predicting Infectious Disease Outbreaks Using Machine Learning and Real-Time Epidemiological Data: Leverage Social Media, Environmental, And Public Health Data to Forecast Outbreaks Like Influenza, COVID-19, Or RSV. *International Journal of Medical Science and Public Health Research*, *7*(02), 7–17. https://doi.org/10.37547/ijmsphr/Volume07Issue02-02

[71] Umam, S., & Razzak, R. B. (2025, November). A 20-Year Overview of Trends in Secondhand Smoke Exposure Among Cardiovascular Disease Patients in the US: 1999–2020. In APHA 2025 Annual Meeting and Expo. APHA.**7**

[72] Razzak, R. B., & Umam, S. (2025, November). Health Equity in Action: Utilizing PRECEDE-PROCEED Model to Address Gun Violence and associated PTSD in Shaw Community, Saint Louis, Missouri. In APHA 2025 Annual Meeting and Expo. APHA.**8**

[73] Razzak, R. B., & Umam, S. (2025, November). A Place-Based Spatial Analysis of Social Determinants and Opioid Overdose Disparities on Health Outcomes in Illinois, United States. In APHA 2025 Annual Meeting and Expo. APHA.

[74] Khan, M. S., Gharami, A. K., Nitu, F. N., Uddin, M. N., Ahmed, M., Roy, M. K., & Yezdani, S. (2025). Deep Learning-Driven Customer Segmentation in Banking: A Comparative Analysis for Real-Time Decision Support. *International Interdisciplinary Business Economics Advancement Journal*, *6*(08), 9-22.

[75] Sajal, A., Chy, M. S. K., Jamee, S. S., Uddin, M. N., Khan, M. S., & Gharami, A. K. & Ahmed, M.(2025). Forecasting Bank Profitability Using Deep Learning and Macroeconomic Indicators: A Comparative Model Study. *International Interdisciplinary Business Economics Advancement Journal*, *6*(06), 08-20.

[76] Siddique, M. T., Uddin, M. J., Chambugong, L., Nijhum, A. M., Uddin, M. N., & Shahid, R. & Ahmed, M.(2025). AI-Powered Sentiment Analytics in Banking: A BERT and LSTM Perspective. *International Interdisciplinary Business Economics Advancement Journal*, *6*(05), 135-147.

[77] Ayub, M. I., Gharami, A. K., Nitu, F. N., Uddin, M. N., Islam, M. I., Nijhum, A. M., ... & Yezdani, S. (2025). AI-driven demand forecasting for multi-echelon supply chains: Enhancing forecasting accuracy and operational efficiency through machine learning and deep learning techniques. *Emerging Frontiers Library for The American Journal of Management and Economics Innovations*, *7*(07), 74-85.

[78] Islam, M. M., Sarkar, M. A. R., Ahmed, M., Moniruzzaman, S. M., & Uddin, M. N. (2009). Germination, vigour and emergence indicators of Corchorus olitorius L, seed and their relationship as influenced by seed sources. *Bangladesh J. Jute and Fib. Res*, *29*(1&2), 1-8.

[79] Uddin, M. N., & Aziz, M. M. (2026). Shapley Value-Guided Adaptive Ensemble Learning for Explainable Financial Fraud Detection with US Regulatory Compliance Validation. *arXiv preprint arXiv:2604.14231*.

[80] Uddin, M. N. (2026). Explainable Graph Neural Networks for Interbank Contagion Surveillance: A Regulatory-Aligned Framework for the US Banking Sector. *arXiv preprint arXiv:2604.14232*.

[81] Uddin, M. N. (2026). Does Founding Team Human Capital Heterogeneity Predict Venture Survival? Evidence from the Kauffman Firm Survey. *Evidence from the Kauffman Firm Survey (April 04, 2026)*.

[82] Uddin, M. N. (2026). Cost Efficiency Dynamics in Indian Public Sector Banks: A DEA-Based Analysis with Second-Stage Tobit Estimation and Post-Reform Comparative Evidence, 2009-2023. *Available at SSRN 6680922*.