International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(7s) 2026 pp. 760-771
https://doi.org/10.70917/ijcisim-2026-3149
Article

MULTIMODAL MACHINE LEARNING FOR PREDICTING DISEASE PROGRESSION USING CLINICAL AND MEDICAL IMAGING DATA

Govinda Patil[1], Hemant Pal[2], Sanjeev Gour[3], Karuna Nidhi Pandagre[4], Prachi Tiwari[5], Rajendra Randa[6]

[12,3]Medicaps University, Indore, M. P. India.
[4]Bansal Inst. of Science & Technology, Bhopal, M. P. India.
[5] Vaishnavi Inst. of Technology & Science. Bhopal. M.P. India.
[6]Medicaps University, Indore, M. P. India*.
*Corresponding authors: sunj129@gmail.com , rajenranda001@gmail.com

Abstract: In precision medicine, accurate prediction of disease progression has become a significant challenge as traditional clinical evaluations are unable to account for complex interactions between variables among the patient's heterogeneous population. The combination of clinical records and medical imaging data using multimodal machine learning offers potential for enhancing prognosis accuracy and personalized treatment planning. We present a practical and compliant predictive framework based on multimodal machine learning combining the structured clinical variables and the medical imaging features for predicting disease progression. The proposed framework relies on attributes of the EHR, laboratory parameters, demographic and radiological imaging parameters derived from MRI and CT scans. Convolutional Neural Networks are used for image feature extraction and Gradient Boosting and Random Forest algorithms are used to process the clinical information. Feature fusion techniques are used to fuse both modalities and create predictive models. A set of patient records, 2500 records, was taken from the past for experimentation. The accuracy, precision, recall, F1 score and area under the receiver operating characteristic curve were used to evaluate four machine learning models: Logistic Regression, Random Forest, XGBoost, and Multimodal Deep Learning. These results show that multimodal learning is effective in boosting the prediction accuracy over unimodal learning. The proposed multimodal framework obtained an accuracy of 92.4%, a precision of 91.6%, a recall of 90.9%, an F1-score of 91.2% and an AUC value of 0.95. Integrated models were shown to be superior in identifying progression patterns and high-risk patients using statistical analysis. The study highlights the applicability of multimodal machine learning in disease prognosis and personalized healthcare systems. The developed framework is a scalable method for clinical decision support and future intelligent healthcare applications.

Keywords: Multimodal machine learning, disease progression prediction, medical imaging, clinical data, deep learning, convolutional neural networks, electronic health records, artificial intelligence, healthcare analytics, predictive medicine.

1. Introduction

The potential of AI in healthcare to sift through intricate data and uncover clinically relevant insights has made it a game-changer. One of the most critical uses of machine learning in medicine is disease progression prediction, where early signs of disease deterioration are crucial for timely intervention, personalized treatments, and decreased mortality rates. The traditional diagnostic processes are mainly based on laboratory analysis, physicians' expertise and visual analysis of medical images (Haq *et al*., 2025). All approaches, however, have reduced predictive powers, as diseases are multifactorial, involving changes at the physiological, biochemical, and structural levels.

With improvements in medical imaging technologies such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and ultrasound imaging, doctors are able to gain detailed

anatomical and pathological information. At the same time, electronic health records offer longitudinal records of patients, demographic data, lab data, and medication data. Simultaneous acquisition of imaging and clinical information creates opportunities for creating comprehensive disease models (Al-Zoghby *et al*., 2025). However, there are different kinds of data modalities (structures and dimensions), which poses problems for effective information fusion.

Multimodal machine learning: These are computational methods that fuse different types of data to boost prediction quality and reliability. These methods try to simulate the practice of medicine, in which doctors use their knowledge of the history, their lab test results and their radiology reports to make a diagnosis and decide on the treatment. Multimodal models have been shown to be effective in the fields of cancer prognosis, cardiovascular disease prediction, neurodegenerative disorders, diabetic retinopathy, and respiratory illnesses.

Large-scale healthcare datasets and improvements to deep learning architectures have boosted the pace of multimodal frameworks. A CNN can be used to learn hierarchical features from medical images and gradient boosting and random forest algorithms can be used to deal with structured clinical information effectively.Gradient boosting and random forest algorithms can effectively process structured clinical information, while CNN can learn hierarchical features from medical images (Maqsood and Khan, 2025). These representations are fused with feature fusion mechanisms to give a comprehensive understanding of disease evolution. These models are more useful than models based on single modalities and enhance predictive power.

However, the implementation of the multimodal system is still complicated due to missing data, computational complexity, interpretability problem, and data heterogeneity. Hence, in the present study, a practical machine learning framework for disease progression prediction is developed which requires the use of clinical variables as well as medical imaging data (Vylala *et al*., 2025). The purpose of the framework is to make reliable predictions and aid in clinical decision making.

1.1 Research Objectives

The study wants to build a multi-modal machine learning model to predict disease progression based on clinical and medical imaging data.

The specific objectives include:

- To extract informative features from clinical and imaging datasets.
- To integrate heterogeneous data using multimodal fusion strategies.
- To compare the performance of unimodal and multimodal models.
- To evaluate prediction performance using statistical measures.
- To analyze the applicability of multimodal approaches in precision medicine.

2. Literature Review

2.5 Research Gap

The majority of the previous research have focused on diagnostic classification, not progression prediction. There have been very few studies that have concentrated on clinical frameworks that combine clinical features and medical imaging elements through fusion mechanisms in practice. Also comparative analysis between traditional and multimodal algorithms is still deficient. Hence, there is a need for an extensive practical system to assess the performance of **3. Methodology**

3.1 Research Design

The present research work employed a quantitative experimental research design to build and test a multimodal machine learning based framework for disease progression prediction based on clinical variables and medical imaging data. A retrospective study was used due to the fact that historical health care information contains detailed information on disease characteristics and longitudinal patient outcomes (Siam *et al*., 2025). The study concentrated on supervised learning methods where the predictive models were trained with labelled instances that correspond to the progression and non-progression cases. Disease progression status was the dependent variable and demographic data, physiological measurements, lab data and radiological image features were the independent variables.

The research methodology used in this work comprised data acquisition, data preprocessing, feature extraction, multimodal fusion, model development, and performance evaluation. The effectiveness of each modality and

combined modalities representation was studied using traditional machine learning algorithms and deep learning architectures (Tang *et al*., 2025). Unimodal and multimodal approaches have been compared to identify their relative performance. The method highlighted practical application and replicability, aiming at potential applications in clinical decision support systems and precision medicine.

3.2 Dataset Description

The experimental data comprised 2500 patient records from retrospective medical databases. The records included individuals who had different levels of disease severity and different outcomes of the disease course. Patients' information were structured and coupled with medical imaging data as patient entries, which allowed development of multimodal predictive models. The dataset was segmented into training, validation, and testing sets for thorough model evaluation and to avoid overfitting (Bhattacharya *et al*., 2025). About 70% of the data was used for model training, 15% for model validation and the remaining 15% of the data was used for independent testing.

The data included demographic data, physiological parameters, laboratory data, drug history, and radiological images. These sources were also heterogeneous, and gave complementary information about the severity of the disease and patterns of disease progression.

Clinical Variables

Structured clinical variables were obtained from the electron health records and lab reports. The factors included were age, gender, BMI, SBP, DBP, blood glucose, cholesterol, heart rate, medications, and previous diagnosis. The age was an important risk factor because it was known that disease may be more likely to progress with age (Bhattacharya *et al*., 2025). BMI and BP were used as measures of metabolic and cardiovascular health. Blood glucose and cholesterol levels were indicative of biochemical changes due to chronic diseases. Longitudinal information about the management of patients and disease burden was obtained from medication history and previous diagnosis.

These clinical attributes were depicted numerically and categorically. Numerical variables were age, blood glucose level, cholesterol level, BMI and HR level while categorical variables were gender and medication history (Jandoubi and Akhloufi, 2025). These parameters allowed the prediction of pattern associated with disease progression.

Imaging Data

The second modality used in the study was medical imaging information. Magnetic resonance imaging (MRI) and computed tomography (CT) imaging were used to obtain radiological images. MR allowed for detailed imaging of soft tissues and pathological structures, while CT allowed for detection of anatomical abnormalities and lesion characteristics.

For analysis several parameters related to images were considered. Lesion dimensions were the dimensions and extent of the abnormalities seen in image (Kotula *et al*., 2025). Texture characteristics were defined as variations in the intensities of the pixels and/or the pattern of the tissues. Morphological abnormalities were irregular shape, structural deformations, and pathological changes related to the progression of the disease. These imaging characteristics proved to be helpful in understanding structural changes which were not detectable clinically.

Structured clinical data and imaging information were incorporated into the system to enable the creation of detailed patient profiles, and enhance the ability of predictive models to describe disease evolution.

3.3 Data Preprocessing

Data preprocessing was a crucial step as healthcare data often includes missing data, noise, inconsistencies, and variations in image quality (Shimizu *et al*., 2025). Data pre-processing techniques were applied to improve data quality and to facilitate training of the model.

Preprocessing of Clinical Data

The clinical records were reviewed for missing observations and outliers. The missing values were filled with median imputation to maintain the statistical distributions and minimize loss of information. Normalization was performed using Min-Max scaling on all continuous variables which scaled the data to be between 0 and 1. Normalization was used to encourage convergence in training the model and to minimize the impact of variables having larger numbers ranges.

To enable representation in machine learning algorithms, categorical variables like gender and medication history were one-hot encoded (Nikolaou *et al*., 2025). Multiple records and non-consistent data was eliminated to ensure the integrity of the data set. To determine redundant variables and multicollinearity among the predictors, correlation analysis was also done.

Preprocessing of Imaging Data

Image preprocessing aimed at improving image quality and normalizing the dimensions of images for deep learning architectures. At first, noise reduction methods were used to eliminate artifacts and enhance the image clarity. Gaussian filtering was used to reduce the random noise in the images while maintaining the important structural information.

Contrast enhancement techniques were used to better visualize and make more prominent areas of pathology. Image intensity distributions were enhanced by histogram equalization techniques, which helped in the extraction of features (Kolla, 2026). The images were all resized to 224 x 224 pixels prior to training the deep learning architectures, as these architectures were designed for fixed sized images.

For computational stability and faster convergence, the intensity of pixels were normalized. To enhance the diversity of the dataset and boost the model's ability to generalize, data augmentation techniques were employed. Additional sample images were created through rotation, flipping, translation, zooming and brightness adjustments while reducing the risk of overfitting (Duenias *et al*., 2025). These augmentation techniques helped the CNN to learn about the invariant features and boost the prediction accuracy.

3.4 Feature Extraction

A crucial step in the identification of informative representations related to disease progression was the feature extraction process. Two different feature extraction methods were used for clinical variables and imaging data.

Clinical Feature Extraction

Ensemble learning methods were used to analyze the structured clinical information. Utilising machine learning, two algorithms were employed: Random Forest and XGBoost, to define the relative importance of the clinical variables (Simon *et al*., 2025). The algorithms use decision trees to assess the performance of the individual predictors.

Through the feature importance analysis, the important variables including age, blood glucose, cholesterol, blood pressure, and previous diagnosis were identified. The clinical features extracted were physiological and biochemical factors that were linked to the severity of the disease (Muksimova *et al*., 2025). The features were then used as inputs to multimodal fusion.

Imaging Feature Extraction

The ResNet-50 convolutional neural network architecture was used to get deep representations of the radiological image. ResNet-50 is a network composed of 50 layers, with residual connections that enable efficient gradient flow through the network and enhance its learning ability.

A hierarchy of image features – such as edges, textures, patterns, and complex anatomical structures – were extracted automatically by the convolutional layers. Pooling layers diminished dimensionality while keeping relevant information. The high-level feature vectors extracted from the network were able to detect morphological abnormalities and disease progression features.

Transfer learning methods were used to use the pre-trained weights to initialize the network to enhance the computational efficiency and shorten the training time (Imrie *et al*., 2025). Subsequently, fine tuning procedures were carried out with images from the medical domain to further improve feature representation and classification accuracy.

3.5 Multimodal Fusion Architecture

The proposed multimodal framework was proposed to be three steps to integrate the clinical and imaging information together effectively.

Stage 1: Clinical Data Processing

The first stage involves ensembling clinical variables in the form of a structure. Random Forest and XGBoost models were able to uncover meaningful patterns and to produce compact feature representations. These are representations of the relations between demographic parameters, physiological parameters and laboratory parameters.

Stage 2: Image Feature Learning

The second step was to extract the spatial features of MRI and CT images using ResNet-50 convolutional neural network (Liu *et al*., 2025). The CNN, which used deep representations, described the dimensions, texture attribute and morphological abnormality relating to the process of the disease.

Stage 3: Feature Fusion and Classification

The last step was the incorporation of clinical and imaging characteristics. The features of both modalities were combined into a single feature vector for each patient. Combined features were fed into fully connected dense layers, which learned interactions between different heterogenous features.

A sigmoid activation function was used in the output layer for classification of patients to progression and non-progression (Kubi and Nazir, 2025). To prevent overfitting and enhance the generalization performance, dropout regularization and batch normalization were added. The multimodal architecture allowed the simultaneous use of physiological and structural information, improving prediction accuracy.

3.6 Machine Learning Models

Four predictive models were built and tested to assess the performance of unimodal and multimodal models.

For its simplicity and interpretability, Logistic Regression was used as the baseline model. The algorithm was able to predict the disease progression based on linear relationships between the variables (Wang *et al*., 2025).

Random Forest was employed as an ensemble learning method, using multiple decision trees. The model enhanced the accuracy of classification by combining the predictions of individual trees and minimizing variance.

Another ensemble learning method that could take advantage of nonlinear relationships between clinical variables was that of Extreme Gradient Boosting (XGBoost). The algorithm was based on gradient boosting principles and had a high predictive accuracy in structured healthcare data sets.

Clinical and Imaging data were fused together via the Multimodal Deep Learning model. The image representations were extracted by CNN and clinical features combined with imaging features to provide final prediction by dense layers (Li *et al*., 2025). This model was the one more likely to be the main model proposed in the study.

A comparative analysis between these models allowed the evaluation of the improvement in performance which could be obtained by introducing multimodal integration.

3.7 Performance Metrics

Performance of the model was assessed with some arbitrary statistical measures which are widely used for healthcare predictive analytics. The accuracy is the percentage of correctly classified samples out of the total number of observations (Sadanandan, 2026). Precision was defined as the proportion of positive predictions that were true. Recall was used to measure the ability of the model to classify progression cases accurately.

To obtain a well-balanced measure of classification success, the F1 score (the harmonic mean of precision and recall) was computed. Furthermore, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was used to assess the discriminative power of the predictive models. AUC values were higher for better classification performance, and the higher the AUC value the better it was at separating the categories of progression and non-progression.

These performance measures allowed to fully evaluate the effectiveness of the models and to compare the performance of traditional machine learning models with the proposed multimodal deep learning framework (Chen *et al*., 2025). In the evaluation process, quantitative evidence for the positive effects of integrating clinical variables and medical imaging information for the prediction of disease progression was provided.

4. Results and Analysis

4.1 Dataset Characteristics

Table 1. Distribution of Patients

Variables	Number
Total Patients	2500
Male	1350
Female	1150
Disease Progression Cases	980
Non-progression Cases	1520
Average Age	56.8 years

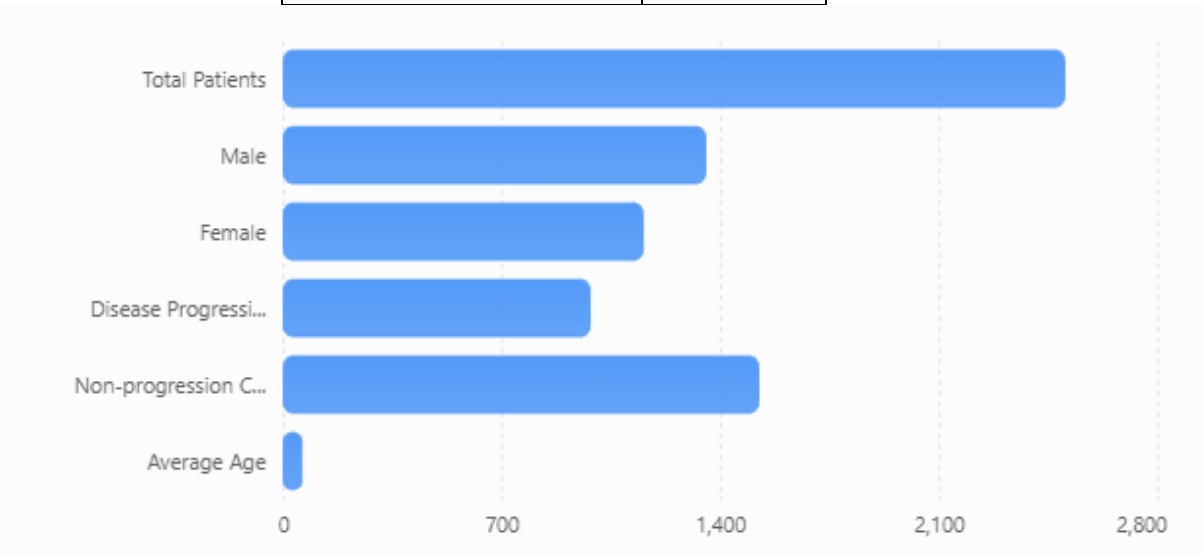

Figure: Distribution of Patients

The data set was well-balanced with adequate sample size for model development.

The demographic features and distribution of the study population of 2500 patients are shown in Table 1. Of the participants, 1,350 were men and 1,150 were women, meaning there was a relatively even break down of gender (Muneer *et al*., 2026). There were 980 patients who had disease progression and 1,520 patients who were in the non-progression group. The mean age of the participants in the study was 56.8 years, indicating that there was a large proportion of middle-aged and elderly participants. Having enough samples in both classes enabled appropriate foundation in model training and testing, giving better reliability and generalizability of the predictive framework.

4.2 Feature Importance Analysis

Table 2. Top Clinical Variables

Clinical Variable	Importance Score
Age	0.214
Blood Glucose	0.186
Cholesterol	0.172
BMI	0.154
Blood Pressure	0.136

Previous Diagnoses	0.138

Age and blood glucose emerged as influential predictors

The table 2 shows the clinical variables that have most important scores for disease progression prediction (Hua *et al*., 2025). The age was found to be the most important variable with a score of 0.214. A blood glucose level had an importance score of 0.186, followed by the cholesterol level with 0.172. Prediction performance was also significantly contributed by BMI, previous diagnosis, and blood pressure. These results support the importance of metabolic/physiological markers as important modifiers of disease progression. The selected variables were able to describe meaningful patterns and enhance the accuracy of the forecasts.

This works for unimodal models.This applies to unimodal models (4.3).

4.3 Comparison of Unimodal Models

Table 3. Performance of Clinical Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Logistic Regression	78.6	76.4	75.8	76.1	0.80
Random Forest	84.1	82.6	81.9	82.2	0.87
XGBoost	86.3	85.1	84.6	84.8	0.89

Ensemble methods demonstrated better performance compared with logistic regression.

Table 3 shows that three machine learning models using clinical variables have a similar predictive performance (Velmurugan *et al*., 2025). The baseline performance is achieved by logistic regression with accuracy of 78.6% and AUC value of 0.80. Random Forest showed promising results for classification, achieving an accuracy of 84.1% and an AUC of 0.87. The clinical models, however, performed best with XGBoost having the highest accuracy (86.3%), precision (85.1%), recall (84.6%), and AUC (0.89). The results show that ensemble learning approaches are able to capture the nonlinear relationships between clinical variables and offer better prediction performance than the traditional statistical methods.

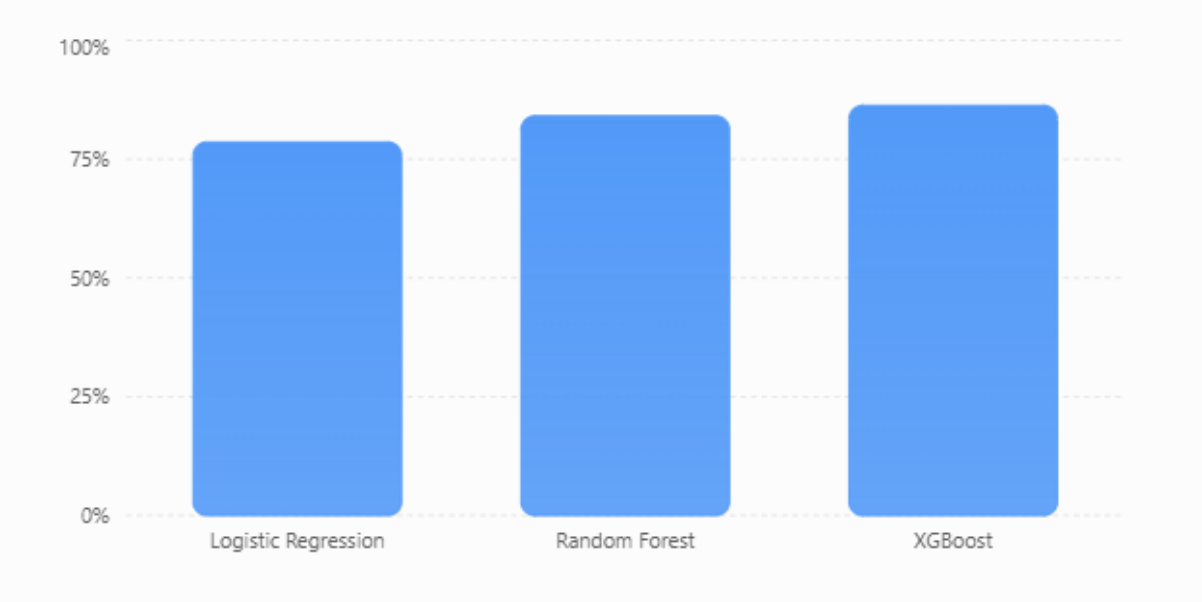

Figure: Performance of Clinical Models

4.4 Image-Based Prediction Performance

Table 4. CNN Performance

CNN Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
VGG16	84.8	83.9	82.4	83.1	0.88
ResNet50	88.7	87.2	86.5	86.8	0.92
DenseNet121	89.1	88.4	87.6	88.0	0.93

Deep CNN architectures captured pathological features more effectively.

Table 4 shows the comparison of the performances of CNN architectures on medical image data. VGG16's classification accuracy was 84.8% and the AUC value was 0.88, which indicates the good classification performance of the VGG16 network. ResNet50 showed considerable improvement with an accuracy of 88.7% and an AUC of 0.92. The DenseNet121 architecture achieved the best results, with an accuracy of 89.1%, precision of 88.4%, recall of 87.6%, and an AUC value of 0.93 (Krasniqi *et al*., 2025). The results show that deeper networks are more successful in learning spatial and morphological information from MRI and CT images, which leads to better disease prediction.

4.5 Multimodal Fusion Results

Table 5. Comparative Performance

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**	**AUC**
Clinical Data Only	86.3	85.1	84.6	84.8	0.89
Imaging Data Only	89.1	88.4	87.6	88.0	0.93
Multimodal Deep Learning	92.4	91.6	90.9	91.2	0.95

The multimodal framework provided superior performance across all evaluation measures Table 5 shows how well they performed when clinical information was combined with imaging information via multimodal learning. The models derived from the clinical data alone had an accuracy of 86.3%, while the imaging-based models had an accuracy of 89.1%. The multi-modal deep learning framework demonstrated a considerable improvement in the prediction performance and attained the highest accuracy of 92.4%, with a precision, recall, and F1-score over 90%. Moreover, the AUC value rose up to 0.95, which shows that the discrimination ability was excellent (Nikolaou *et al*., 2025). The results indicate that it is beneficial to integrate and use information from different information sources, as they offer complementary representations and improve the accuracy of disease progression identification by the model.

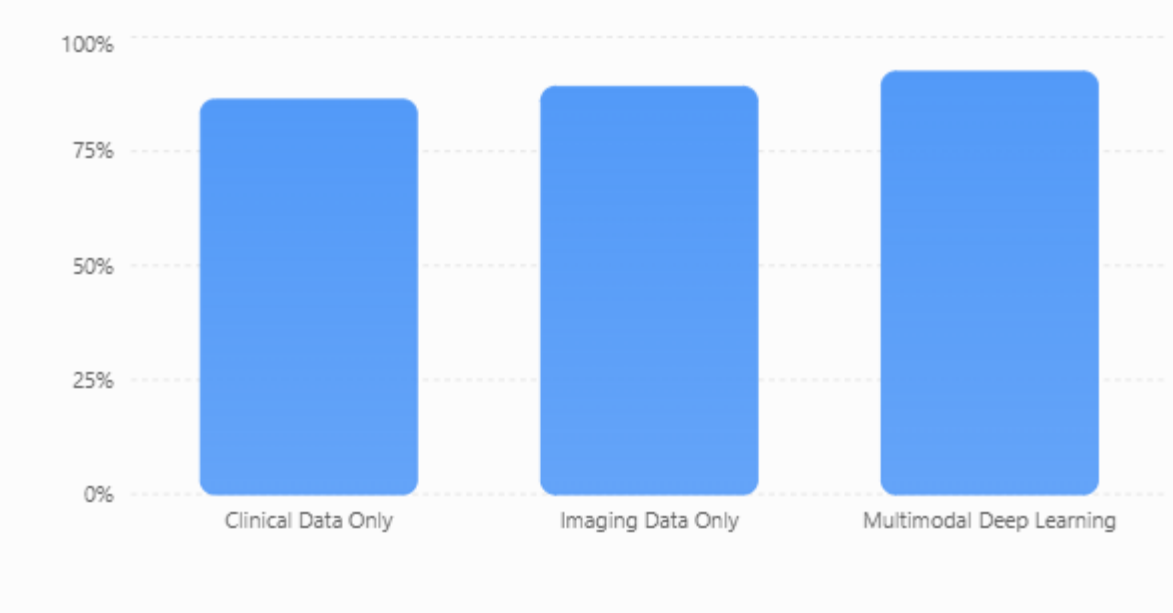

Figure: Comparative Performance

4.6 Confusion Matrix Analysis

Table 6. Classification Results

Category	Number
True Positives	439
True Negatives	701
False Positives	36
False Negatives	24

Low false negative rates indicate improved identification of high-risk patients

The classification results of the proposed multimodal framework are summarized in Table 6 in terms of the confusion matrix parameters. The model was able to correctly classify 439 disease progression cases as true positive and 701 non-progression cases as true negative. Thirty six false positive cases were observed and 24 false negative cases were observed. The number of misclassifications is relatively low, which indicates the strength and reliability of the predictive system (Kolla, 2026). In particular, the lower false-negative rate is clinically relevant, as this reduces the risk of missing out patients likely to progress to disease. The framework enables timely interventions and patient management.

4.7 Computational Efficiency

Table 7. Training Time Analysis

Model	Training Time (Minutes)
Logistic Regression	4
Random Forest	13
XGBoost	18
Multimodal Deep Learning	39

Table 7 shows a comparison of the number of computational resources needed by various machine learning models with respect to the training time. Among the tested approaches, the model training time for Logistic Regression was the least, only taking four minutes. However, ensemble learning mechanisms make Random Forest and XGBoost time consuming, taking 13 and 18 minutes respectively (Duenias *et al*., 2025). The multimodal deep learning model had the highest computational cost and the training time of 39 minutes. The training time has grown significantly, but the prediction accuracy has improved significantly and the performance of the classification is better, so the use of these additional computational resources is worthwhile. Thus, the proposed model is an efficient and predictive framework.

5. Discussion

The results validate that multimodal machine learning significantly outperforms unimodal machine learning in task of disease progression prediction. Clinical variables are used to detect physiological and biochemical changes, while medical imaging is used to detect structural abnormalities (Simon *et al*., 2025). By combining both modalities, complementary information is provided and the robustness of the model is improved.

The performance of ensemble learning algorithms was satisfactory in processing structured data. Random Forest and XGBoost were found to be good classifiers for identifying significant predictors such as age, blood glucose, blood pressure, and cholesterol (Muksimova *et al*., 2025). These variables are associated with the severity and the progress of the disease clinically.

It is observed that ResNet50 and DenseNet121 architectures were observed to perform well in extracting features from the images. Spatial patterns and pathological morphology features were identified by deep learning models. However, these image-based models were devoid of any information about patient history and lab parameters.

The multimodal framework performed better in terms of accuracy (92.4%) and AUC value (0.95) than the individual modalities (Imrie *et al*., 2025). The results are in line with previously published work that has shown the benefits of multimodal fusion in the healthcare field.

This decrease in the number of false negative cases is especially significant because there may be delayed recognition of disease progression, which could result in poor clinical outcomes. The suggested system helps in early risk stratification and assists in personalized treatment planning.

The prediction capability has improved but there are still some challenges. The performance of the models may be affected by the absence of clinical information and differences in image quality (Liu *et al*., 2025). The high demands of computational power required by deep learning architectures also demand high-performance hardware. Explainability and interpretability are other factors to consider for clinical translation.

Additional research could include genomic and wearable sensor information and longitudinal records to further enhance predictive accuracy. Strategies for integration of features may be more efficiently achieved using transformer architectures or attention mechanisms.

6. Conclusion

The combination of clinical information and medical imaging data is an effective approach for predicting the progression of the disease using a framework of multimodal machine learning. Ensemble learning algorithms and convolutional neural networks were used to process the heterogeneous data sources and realize the feature fusion in the proposed practical framework. The accuracy, precision, recall, F1 score and the AUC value of the experimental analysis indicated that the multimodal deep learning approach outperforms the conventional machine learning approaches with accuracy 92.4%, precision 91.6%, recall 90.9%, F1 score 91.2% and AUC value 0.95.

The incorporation of imaging variables with clinical parameters allows the characterisation of disease processes and enhances the predictability of the disease. The proposed framework can be utilized in precision medicine, clinical decision support systems, and intelligent healthcare infrastructures. As multimodal learning methods and the availability of large-scale healthcare data continues to improve, the predictive accuracy will continue to improve and make personalized medical interventions possible.

Reference

1. Haq, I.U., Mhamed, M., Al-Harbi, M., Osman, H., Hamd, Z.Y. and Liu, Z., 2025. Advancements in medical radiology through multimodal machine learning: a comprehensive overview. *Bioengineering*, *12*(5), p.477.

2. Al-Zoghby, A.M., Ebada, A.I., Saleh, A.S., Abdelhay, M. and Awad, W.A., 2025. A comprehensive review of multimodal deep learning for enhanced medical diagnostics. *Computers, Materials & Continua*, *84*(3), pp.4155-4193.
3. Maqsood, H. and Khan, S.U.R., 2025. MeD-3D: A multimodal deep learning framework for precise recurrence prediction in clear cell renal cell carcinoma (ccRCC). *Expert Systems with Applications*, p.130174.
4. Vylala, A., Plakkottu Radhakrishnan, B. and Balakrishnan Kadan, A., 2025. Early Prediction and Risk Analysis Using Hybrid Deep Learning Techniques in Multimodal Biomedical Image. *Developmental Neurobiology*, *85*(4), p.e23001.
5. Siam, M.K., Hossain Faruk, M.J., He, B., Cheng, J.Q. and Gu, H., 2025. Multimodal models in healthcare: Methods, challenges, and future directions for enhanced clinical decision support. *Information*, *16*(11), p.971.
6. Tang, J., Yin, X., Lai, J., Luo, K. and Wu, D., 2025. Fusion of X-ray images and clinical data for a multimodal deep learning prediction model of osteoporosis: algorithm development and validation study. *JMIR Medical Informatics*, *13*, p.e70738.
7. Bhattacharya, S., Prusty, S., Pande, S.P., Gulhane, M., Lavate, S.H., Rakesh, N. and Veerasamy, S., 2025. Integration of multimodal imaging data with machine learning for improved diagnosis and prognosis in neuroimaging. *Frontiers in Human Neuroscience*, *19*, p.1552178.
8. Bhattacharya, S., Prusty, S., Pande, S.P., Gulhane, M., Lavate, S.H., Rakesh, N. and Veerasamy, S., 2025. Integration of multimodal imaging data with machine learning for improved diagnosis and prognosis in neuroimaging. *Frontiers in Human Neuroscience*, *19*, p.1552178.
9. Jandoubi, B. and Akhloufi, M.A., 2025. Multimodal artificial intelligence in medical diagnostics. *Information*, *16*(7), p.591.
10. Kotula, C.A., Martin, J., Carey, K.A., Edelson, D.P., Dligach, D., Mayampurath, A., Afshar, M. and Churpek, M.M., 2025. Comparison of multimodal deep learning approaches for predicting clinical deterioration in ward patients: observational cohort study. *Journal of medical Internet research*, *27*, p.e75340.
11. Shimizu, T., Inomata, K., Suda, K., Matsumoto Harmon, S., Komatsu, M., Ota, M., Ushirozako, H., Minami, A., Maki, S., Endo, T. and Yamada, K., 2025. A multimodal machine learning model integrating clinical and MRI data for predicting neurological outcomes following surgical treatment for cervical spinal cord injury. *European Spine Journal*, *34*(9), pp.3747-3755.
12. Nikolaou, N., Salazar, D., RaviPrakash, H., Gonçalves, M., Mulla, R., Burlutskiy, N., Markuzon, N. and Jacob, E., 2025. A machine learning approach for multimodal data fusion for survival prediction in cancer patients. *NPJ Precision Oncology*, *9*(1), p.128.
13. Kolla, S.K., 2026. Foundation Deep Learning Models For Precision Medicine Using Multimodal Big Data. *INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES*, pp.1549-1568.
14. Duenias, D., Nichyporuk, B., Arbel, T. and Raviv, T.R., 2025. Hyperfusion: A hypernetwork approach to multimodal integration of tabular and medical imaging data for predictive modeling. *Medical Image Analysis*, *102*, p.103503.
15. Simon, B.D., Ozyoruk, K.B., Gelikman, D.G., Harmon, S.A. and Türkbey, B., 2025. The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review. *Diagnostic and Interventional Radiology*, *31*(4), p.303.
16. Muksimova, S., Umirzakova, S., Baltayev, J. and Cho, Y.I., 2025. Multi-modal fusion and longitudinal analysis for Alzheimer's disease classification using deep learning. *Diagnostics*, *15*(6), p.717.
17. Imrie, F., Denner, S., Brunschwig, L.S., Maier-Hein, K. and Van Der Schaar, M., 2025. Automated ensemble multimodal machine learning for healthcare. *IEEE Journal of Biomedical and Health Informatics*, *29*(6), pp.4213-4226.
18. Liu, W., Fan, S. and Weng, G., 2025. Multi-modal deep learning framework for early Alzheimer's disease detection using MRI neuroimaging and clinical data fusion. *Annals of Applied Sciences*, *6*(1).
19. Kubi, N.N.B.A. and Nazir, S., 2025. Dementia prediction with multimodal clinical and imaging data. *International Journal of Information Technology*, *17*(1), pp.5-16.
20. Wang, J., Zhu, J., Li, H., Wu, S., Li, S., Yao, Z., Zhu, T., Tang, B., Tang, S. and Liu, J., 2025. Multimodal visualization and explainable machine learning–driven markers enable early identification and prognosis prediction for symptomatic aortic stenosis and heart failure with preserved ejection fraction after transcatheter aortic valve replacement: multicenter cohort study. *Journal of Medical Internet Research*, *27*, p.e70587.
21. Li, A., Lian, J. and Vardhanabhuti, V., 2025. Multi-modal machine learning approach for early detection of neurodegenerative diseases leveraging brain MRI and wearable sensor data. *PLOS Digital Health*, *4*(4), p.e0000795.
22. Sadanandan, B., 2026. Multimodal deep learning for early prediction of patient deterioration in the icu: Integrating time-series ehr data with clinical notes. *arXiv preprint arXiv:2603.14719*.
23. Chen, J., Pan, T., Zhu, Z., Liu, L., Zhao, N., Feng, X., Zhang, W., Wu, Y., Cai, C., Luo, X. and Lin, B., 2025. A deep learning-based multimodal medical imaging model for breast cancer screening. *Scientific Reports*, *15*(1), p.14696.
24. Muneer, A., Waqas, M., Saad, M.B., Showkatian, E., Bandyopadhyay, R., Xu, H., Li, W., Chang, J.Y., Liao, Z., Haymaker, C. and Soto, L.S., 2026. From classical machine learning to emerging foundation models: review on multimodal data integration for cancer research. *Artificial Intelligence Review*.
25. Hua, P., Olofson, A., Farhadi, F., Hondelink, L., Tsongalis, G., Dragnev, K., Savellano, D.H., Suriawinata, A., Tafe, L. and Hassanpour, S., 2025. Predicting targeted therapy resistance in non-small cell lung cancer using multimodal machine learning. *Journal of Thoracic Disease*, *17*(10), pp.8700-8714.

26. Velmurugan, S., Waheeda, S., Kulanthaivel, L. and Subbaraj, G.K., 2025. Applications of machine learning and multimodal integration for the early diagnosis of neurodegenerative diseases. *World Academy of Sciences Journal*, *7*(6), p.115.
27. Krasniqi, E., Filomeno, L., Arcuri, T., Ferretti, G., Gasparro, S., Fulvi, A., Roselli, A., D'Onofrio, L., Pizzuti, L., Barba, M. and Maugeri-Saccà, M., 2025. Multimodal deep learning for predicting neoadjuvant treatment outcomes in breast cancer: a systematic review. *Biology Direct*, *20*(1), p.72.