

Causal Explainable Foundation Model for Medical Image Intelligence with Disease Progression Reasoning and Clinical Decision Support

Dr.J. Anitha ^{a*}, Dr.M. Kalaiarasu ^b, S. Bharathi ^c, M. Preethi ^d

^{a*} Professor, Department of Artificial Intelligence and Data Science, Sri Ramakrishna Engineering College (Affiliated to Anna University), Coimbatore, Tamil Nadu, India. E-mail: anitha.j@srec.ac.in, Orcid: <https://orcid.org/0000-0002-0990-3738>

^b Professor, Department of Information Technology, Sri Ramakrishna Engineering College (Affiliated to Anna University), Coimbatore, Tamil Nadu, India. E-mail: kalai.muthusamy@srec.ac.in, Orcid: <https://orcid.org/0000-0002-5138-4329>

^c Associate Professor, Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering & Technology, Pollachi, Tamil Nadu, India. E-mail: bharathi_mani@yahoo.com, Orcid: <https://orcid.org/0000-0001-9638-3779>

^d Assistant Professor (Sr. Gr), Department of Information Technology, Sri Ramakrishna Engineering College (Affiliated to Anna University), Coimbatore, Tamil Nadu, India. E-mail: preethi.muthukumar@srec.ac.in, Orcid: <https://orcid.org/0000-0003-1106-8863>

Abstract

Medical image intelligence has seen tremendous advancements due to the development of foundation models; however, existing techniques make use of correlation-based machine learning with limited interpretability, which makes the approach less reliable in disease progression analysis and clinical decision support. In this paper, a new framework known as Causal Explainable Foundation Model (CEFM) is presented that incorporates foundation models, causal reasoning, explainable artificial intelligence (XAI), disease progression modeling, and clinical decision support in a single platform. The proposed design consists of a multimodal foundation encoder that can learn generalized representations from the medical images, clinical notes, and reports. A causal reasoning component develops disease-specific cause-and-effect relationships between imaging biomarkers, whereas the explainability component provides visual attributions, causal pathway evidence, and confidence scores. Also, the reasoning engine of disease progression reasons about the longitudinal progression of pathology to forecast disease stages and severity in the future, which helps in making personalized decisions for treatment. The proposed model was tested using a total of 48,620 multimodal medical images annotated using publicly available datasets involving diseases of thoracic, neurological, and oncological areas. Experimental results showed overall accuracy, precision, recall, F1 score, area under the curve (AUC), specificity, and Matthews Correlation Coefficient (MCC) values of 98.14%, 97.82%, 97.56%, 97.69%, 0.992, 98.31%, and 0.961, respectively. Also, the inclusion of causal reasoning increased disease progression prediction accuracy by 6.8% and explanation consistency by 14.5%. These results show that the use of foundational knowledge representations in conjunction with causal reasoning and temporal reasoning can significantly enhance diagnostic accuracy, interpretation, and temporal progression awareness. The presented CEFM framework can be considered an efficient, interpretable, and clinically relevant approach to intelligent medical image analysis and has great potential for AI-driven healthcare solutions.

Keywords: Foundation Model, Medical Image Intelligence, Causal Explainable Artificial Intelligence (XAI), Disease Progression Reasoning, Clinical Decision Support, Medical Image Analysis, Multimodal Healthcare Analytics.

Introduction

Medical imaging has emerged as one of the most powerful diagnostic tools in the current field of healthcare and has helped physicians in identifying, classifying, and monitoring numerous illnesses. Medical imaging tools like computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasound, and digital pathology offer a wealth of information related to anatomy and functions, making clinical diagnosis efficient. The increasing number of imaging tests produced in hospitals has led to many problems for radiologists and physicians, such as the burden of interpretation, inter-observer variability and diagnostic consistency. AI, especially deep learning, has shown significant proficiency in automated image analysis, lesion localization, organ segmentation and disease classification (Litjens et al., 2017).

With the recent developments in foundation models, medical imaging analysis has been revolutionized through the use of large-scale representation learning from heterogeneous data. While traditional task-oriented models are only capable of learning feature representations for specific tasks, foundation models learn generalized representations that can be used for various downstream clinical tasks without much more training. The use of vision transformers along with self-supervised learning techniques has greatly improved the ability of foundation models to detect subtle signs of pathology within different imaging modalities. The potential to utilize multiple sources of information such as imaging, lab results, and reports gives them an advantage over traditional models (Hamamci et al., 2026).

In spite of such accomplishments, however, a number of critical issues must still be resolved for reliable deployment of foundation models within the framework of routine clinical practice. First of all, most currently available models are oriented towards maximization of prediction quality without offering significant information regarding the process of reasoning underlying diagnosis. This results in perceiving the models as "black box" ones, which makes it hard to evaluate the extent to which the predictions are made on the basis of clinically relevant imaging markers rather than arbitrary visual characteristics of images (Chaddad et al., 2023).

Explainable artificial intelligence (XAI) has emerged as a promising solution for improving the interpretability of deep learning systems in healthcare. Approaches such as saliency visualization, gradient-based attribution, concept activation mapping, counterfactual explanations, and attention mechanisms have allowed clinicians to have a deeper insight into the process through which neural networks detect pathological areas in the image. Nevertheless, most of the currently available explanation approaches still focus on correlation rather than causation and provide information about image regions correlated with the prediction but not about the cause of the disease itself (Tjoa & Guan, 2020).

Causal artificial intelligence has gained significant interest due to its aim of finding causations rather than only correlations. By incorporating causal learning into medical image intelligence, the model is able to differentiate the true causes of the disease from the imaging factors that may serve as confounders. Using causal reasoning methods allows reasoning in distribution changes, which would allow diagnostic algorithms to retain their accuracy when working with patients who have characteristics different from the model training data (Schölkopf et al., 2021).

Progression of the disease constitutes another important aspect of clinical diagnoses that has been neglected by traditional image classification models. Several progressive diseases such as cancer, neurodegenerative diseases, cardiovascular diseases, and respiratory diseases follow complex temporal mechanisms of evolution that involve structural and physiological changes in the body. Hence, diagnosis of such diseases requires not only understanding the current status of the disease but also predicting its future progression and the impact of different treatment plans (Rajpurkar et al., 2022). Temporal reasoning models are highly beneficial in improving the accuracy of prognosis estimation (Pathak et al., 2022).

Foundation models serve to offer an effective approach for incorporating multimodal data into the analysis of disease development (Chen et al., 2024). Through the combination of imaging attributes along with EHRs, laboratory values, genomics, and clinical notes, foundation models can create more detailed patient representations that will better capture the actual processes taking place within patients' bodies. The application of multimodal representation learning has already shown significant promise for improved diagnostics, decreased uncertainty, and overall comprehensive clinical evaluation within various healthcare fields. However, combining multimodal knowledge with causality and explainability is still an ongoing challenge (Moor et al., 2023).

Decision support systems in clinical settings have gained significant importance in recent times for aiding healthcare professionals in handling the complex process of making diagnoses. Modern decision support systems need to offer accurate predictions while at the same time offering interpretability that is consistent with the knowledge in the field of medicine. Explainable AI models offering an explanation of causation and the development of diseases might aid physicians in gaining trust and allow for efficient communication within the interdisciplinary environment (Esteva et al., 2019).

The creation of reliable healthcare AI involves solving problems related to robustness, fairness, reproducibility, and clinical validation of models. Medical models created on a multicentric dataset usually show better generalizability to different imaging machines and patients. Using causality in foundation models can help avoid bias from confounding factors and increase the robustness of the model in different clinical settings. These are critical features needed to create intelligent healthcare solutions applicable in practice (Rajpurkar et al., 2022).

In light of these problems, this research introduces the novel concept of CEFM, which incorporates foundational models, causality, explainable artificial intelligence, disease progression reasoning, and clinical decision-making to form a framework for medical image intelligence. The design of the proposed framework focuses on the accurate diagnosis of diseases along with the provision of causal explanation and disease progression reasoning to assist clinicians in decision making. Through combining generalized representation learning and causal inference in an interpretative manner, this framework strives to offer reliable assistance in diverse medical imaging scenarios (Zhou et al., 2022).

Key Contributions

The main contributions of the suggested CEFM can be listed as follows:

- The CEFM framework is designed by combining foundation model-based representation learning, causal reasoning, explainable artificial intelligence, disease progression reasoning, and clinical decision support for sophisticated medical image intelligence.
- Causal reasoning is proposed to recognize the associations between imaging biomarkers and the characteristics of diseases in order to reduce the reliance on correlation-based feature learning to improve the reliability of prediction.
- The progression-aware reasoning engine is developed to analyze the temporal medical imaging data in order to predict future progression, severity changes, and clinical risks beyond the simple classification of static images.
- The interpretable clinical decision support system is created by providing visual explanations and evidence pathways along with confidence scores.

The rest of the paper is structured as follows. Section II is the literature review on existing works on medical image intelligence techniques, which cover the following aspects: foundation models, explainable AI, causal learning, and disease progression modeling. Section III is about the proposed Causal Explainable Foundation Model approach, which includes the overall framework, mathematical formulation, and algorithm description. Section IV is about the experimental settings, dataset, parameters used, performance evaluation, comparative analysis, and ablation study results. Finally, Section V is the conclusion of this work.

Literature Survey

The advancements in the field of medical image analysis using AI have been remarkable because of the introduction of deep learning networks, self-supervised learning techniques, and foundation models. The earlier medical imaging methods were more concerned about the use of task-specific models for various medical tasks, including classification, segmentation, or detection. The current research trend has moved towards generalized models that can learn visual representation from various medical images and transfer this knowledge to other tasks. These developments have made the AI system scalable as well as overcome the drawbacks of a lack of annotated medical data (Zhou et al., 2021).

The use of transformer networks has been increasingly popular in the field of medical imaging due to their capability of understanding long-range and context-based relationships in anatomically intricate structures. Unlike convolutional methods that focus only on the identification of local features in an image, transformers leverage the power of attention mechanisms to understand the overall relationship between regions within an image. Some studies have shown better performance in tasks like tumor segmentation, organ localization, and anomaly detection, but several challenges exist (Hatamizadeh et al., 2022).

Self-supervised learning has proved to be a viable solution for minimizing reliance on massive amounts of manually labeled data in medicine. Utilizing unlabelled medical images as input data and employing pretext tasks and contrastive losses, self-supervised models have proven to be highly transferable to various imaging tasks. Such models have shown their increased robustness under limited data settings and have made AI models more generalizable to various clinical conditions. However, representation learning methods concentrate only on prediction of relationships, not on modeling cause-and-effect relationships between imaging characteristics and diseases (Azizi et al., 2021).

Foundation models have brought about a paradigm shift in medical artificial intelligence in terms of knowledge transfer between several different medical fields. Training of such models using diverse sets of data helps create a general representation that can be fine-tuned for applications like diagnosis, segmentation, image retrieval, and clinical reasoning. Foundation models that combine imaging data with clinical text and patient data have shown promise in terms of improving diagnostic capabilities and conducting healthcare analysis. However, making these models transparent and reliable is another research issue (Ghosh et al., 2026).

Explainable Artificial Intelligence has been thoroughly studied for better interpretation of medical diagnoses made by AI systems. Visualization techniques, feature attribution techniques, and attention-based explanations help understand which image areas affect AI decision-making. Such approaches have facilitated communication between AI and clinicians through the provision of understandable evidence related to diagnoses. Unfortunately, most of the current techniques explain the behavior of models without proving that detected features are real factors leading to disease occurrence (Holzinger et al., 2021).

The use of causal learning techniques has created possibilities for developing better-performing AI medical systems through causal relationship discovery, as opposed to statistical associations, between variables. The causality model can discover the important factors affecting diseases, minimize the effect of confounders, and enhance prediction robustness in varying clinical environments. Combining causality learning techniques and models in medical imaging has been considered a way of making such models more interpretable and hence capable of providing useful insights in clinical settings. Combining the inference and representation models of causality learning is an open research area (Deng et al., 2025).

The modeling of disease progression has been increasingly used as an important part of intelligent healthcare since a large number of diseases show temporal progression with complicated patterns. Traditional image analysis models tend to analyze images separately without considering their temporal correlation. However, recent research based on longitudinal images, temporal learning, and prediction shows that this approach allows for better results in terms of predicting the degree of illness, reaction to treatment, and future development of the disease (Pan et al., 2026).

Clinical decision support systems powered by artificial intelligence are already created to help healthcare workers with diagnostic suggestions, risk assessment, and treatments. Today's systems are focused more on reliable predictions based on interpretive evidence, which enables clinicians to assess results before making any decisions. Studies have proven that such aspects as explainability, uncertainty evaluation, and clinical workflow integration are key components for physicians' acceptance of AI-driven healthcare applications (Lewandowski & Koller, 2023).

Learning through multiple modes of data has also been contributing significantly to the potential of medical AI by utilizing complementary sources of information like medical images, EHRs, laboratory values, and textual clinical notes. But the problems faced by medical AI in terms of handling heterogeneous data, data alignment, privacy-preserving integration of data modalities, and learning causality continue to prevent the development of fully integrated medical intelligence systems (Chen et al., 2024).

Studies on trustworthy medical AI have identified the importance of developing models that can make accurate predictions but remain fair, robust, and interpretable. Trustworthy healthcare AI necessitates systems that can reveal how decisions are made, discover patterns that can be clinically significant, and reason beyond mere image classification. The current state of the art has shown substantial advancement in foundation models, explainable AI, and causal learning; nevertheless, a comprehensive framework integrating all of the above to reason about disease progression is underexplored (Alsaedi et al., 2025).

According to the reviewed papers, medical image intelligence has shifted from traditional deep learning algorithms for specific tasks into more generalized and flexible forms of AI. The development of foundation models has increased their representation learning capacity, whereas explainable AI techniques have increased transparency in clinical practice. Causal reasoning has opened up the possibilities for addressing the shortcomings of correlation-based predictions by gaining insight into the process of disease formation. Moreover, disease progression modeling and

The aligned medical data is fed into the Foundation Model Encoder, which comprises vision and multimodal encoders that are able to produce generalized representations of the anatomy and pathology. After being extracted, the features are further used by the Causal Reasoning Module to perform such tasks as causal feature extraction, causal inference, and building a disease mechanism graph to identify associations between the imaging features and disease progression factors. The Explainability Module creates visualizations of attribution maps, causal evidence pathways, and uncertainty estimation.

These causality models are used as input to the Disease Progression Reasoning Engine, which uses longitudinal imaging data to model the disease development process and predict disease stages, progression levels, and disease progression trajectory in the future. These predicted values are incorporated into the Clinical Decision Support module for making suggestions for risk assessment, treatments, and monitoring patients. In the end, the Output layer produces the output in terms of diagnosis, causation, progression, and clinical suggestions. The entire system architecture is optimized with the help of the Model Training and Optimization Module.

Foundation Representation and Causal Feature Learning

Multimodal medical data representations are obtained using a foundation model encoder, which involves the fusion of visual and clinical embeddings. Vision encoder extracts the spatial properties of the medical image whereas multimodal encoder fuses the clinical texts and patient information. The representations are converted into one feature space to have in-depth disease knowledge.

Causal reasoning module builds the relationships between the extracted features and the disease attributes. Causal graphs are created for the discovery of important imaging biomarkers and the removal of unimportant correlations. This makes the model robust because predictions will be made based on clinically relevant factors.

Disease Progression Reasoning and Clinical Decision Support

Disease Progression Reasoning Engine uses temporal modeling for predicting the development of disease through sequential imaging data. The engine analyses the changes in structures and disease-related features at different times to estimate future patterns of disease progression. The temporal dependencies are used to predict changes in the stages of diseases, severity, and risks involved.

The clinical decision support system makes use of the diagnostics, causes, and progression that have been identified to help medical practitioners. Interpretations are provided by this module through integrating the disease state predictions and extracting evidence from imaging and clinical data.

Mathematical Description of the Proposed Model

The CEFM model encodes each medical input sample using a multimodal feature representation derived from image, clinical, and textual data. Feature encoding can be written as equation (1):

$$F_m = E_f(X_i, X_c, X_t) \quad (1)$$

Where F_m represents the fused multimodal feature representation, E_f denotes the foundation model encoder, X_i represents medical image input, X_c represents clinical information, and X_t represents clinical text information.

The causal reasoning module estimates causal relationships among extracted features by constructing a causal dependency representation given by equation (2):

$$G_c = \phi(F_m, W_c) \quad (2)$$

Where G_c denotes the causal graph representation, ϕ represents the causal inference function, and W_c represents the learned causal relationship weights.

The final disease prediction and progression estimation are obtained by integrating causal features with temporal reasoning information given by equation (3):

$$Y_p = \psi(G_c, T_f, W_p) \quad (3)$$

Where Y_p represents the predicted clinical outcome, T_f denotes temporal progression features, ψ represents the prediction function, and W_p represents the optimized model parameters.

Algorithmic Workflow of the Proposed CEFM

Algorithm 1: Causal Explainable Foundation Model Workflow

Input:

X_i : Medical image dataset (CT, MRI, PET, X-ray)

X_c : Clinical information dataset (EHR, laboratory values, demographics)

X_t : Clinical text dataset (reports and medical notes)

L_t : Longitudinal imaging sequence data

W_f : Initial foundation model parameters

W_c : Causal reasoning parameters

W_p : Prediction model parameters

Output:

Y_d : Diagnostic prediction results

E_x : Explainable causal evidence

P_r : Disease progression prediction

Pseudocode:

Initialize W_f, W_c, W_p

Load X_i, X_c, X_t, L_t

Preprocess X_i using normalization and enhancement operations

Synchronize X_i, X_c, X_t into a unified multimodal representation

Extract feature representation $F_m \leftarrow E_f(X_i, X_c, X_t)$

Construct causal graph $G_c \leftarrow \phi(F_m, W_c)$

Generate causal explanations $E_x \leftarrow \text{Explain}(G_c, F_m)$

Extract temporal features $T_f \leftarrow \text{Temporal}(L_t)$

Predict disease progression $P_r \leftarrow \psi(G_c, T_f, W_p)$

Generate diagnostic output Y_d

Calculate optimization loss L_{total}

Update W_f, W_c, W_p using gradient optimization

Validate model performance using evaluation metrics

Return Y_d, E_x, P_r

Workflow of the Proposed CEFM for Medical Image Intelligence is demonstrated by Algorithm 1. This algorithm starts with the initiation of the framework that receives multimodal medical inputs such as imaging information, clinical information, report-based information, and longitudinal sequences. Multimodal pre-processing and alignment create consistent multimodal representations that undergo the process of encoding by the foundation model in order to get detailed anatomical and pathological representations. These representations further undergo causal reasoning to find the dependencies between the diseases. Further, explainability analysis offers visual and causal proof, whereas the progression reasoning module makes predictions about disease development based on temporal data. The output of this step involves diagnostic predictions, insights into disease progression, and clinical suggestions. It is then followed by parameter optimization in order to enhance performance. Consequently, the proposed algorithm provides a framework that facilitates medical image analysis.

Results and Discussion

Experimental Setup and Software Environment

The proposed Causal Explainable Foundation Model (CEFMM) was tested in a deep learning-based experimental setting. The procedure of model design and training was conducted using the Python 3.12 programming language and the PyTorch framework to implement the foundation model, optimize neural networks, and learn features. The multimodal data preprocessing was conducted using OpenCV and Pandas, and the computations of numeric values were done with the help of NumPy. The causal reasoning component was designed based on learning by graph-based

operations, and explanation was performed with the help of gradient-based attributions. The experiments were conducted on a workstation equipped with an Intel Core i9 processor, 32 GB RAM, an NVIDIA RTX-series GPU, and a 2 TB SSD. The AdamW optimizer was utilized for model parameter optimization with adaptive learning rate control.

Dataset Description

The performance of the introduced CEFM architecture was tested through a multimodal medical imaging dataset of 48,620 labeled images obtained from freely available medical databases. This dataset comprised different types of imaging modalities like computed tomography (CT), magnetic resonance imaging (MRI), PET, and X-rays related to thoracic, neurological, and oncological diseases.

The dataset comprised images and their corresponding clinical features, diagnostic categories, radiology reports, and longitudinal imaging data. The imaging data was classified into several categories based on diseases to examine the ability of the proposed framework to identify complicated disease patterns. The clinical features include demographic, laboratory, and disease severity features, whereas the textual data includes radiology reports and clinical notes. The dataset was split into three sets, namely training, validation, and testing sets, having 70%, 15%, and 15% of the total dataset size, respectively.

Parameter Initialization

Initialization of the experimental parameters took place to ensure proper training and accurate evaluation of the CEFM model. Pretrained vision transformer weights were used to initialize the foundation encoder, while the parameters of the causal graph were initialized randomly through Xavier initialization. The learning rate was set to be 1×10^{-4} while the batch size was maintained at 16, and the number of training epochs was set to 100. AdamW optimizer, having a weight decay of 1×10^{-5} was used during the optimization process. The causal reasoning module had three layers of graph propagation, while the temporal progression module had a 512-dimensional hidden feature dimension.

Performance Evaluation Metrics

The suggested approach was evaluated in terms of accuracy, precision, recall, F1-score, area under the receiver operating characteristic (AUC), specificity, and the MCC. These metrics were chosen to evaluate the ability to classify and predict the data reliably.

Accuracy measures the overall proportion of correctly classified samples and is calculated using equation (4):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Where TP , TN , FP , and FN represent true positive, true negative, false positive, and false negative predictions, respectively.

The precision is the accuracy of positive predictions, while recall is the ability of the classifier to detect disease patients, which is calculated as shown in equation (5) and equation (6):

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

The F1-score provides a balanced evaluation between precision and recall given by equation (7):

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

Specificity measures the ability of the model to correctly identify non-disease cases given by equation (8):

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

Further, the MCC metric was used to evaluate the accuracy of both binary and multiclass classification, taking into consideration true and false positives and negatives. The MCC can be evaluated by equation (9):

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (9)$$

Overall Performance Analysis

The developed CEFM model succeeded in obtaining remarkable diagnostic performance through the incorporation of foundation representations, causal reasoning, and progressive learning. The obtained results demonstrated an accuracy of 98.14%, precision of 97.82%, recall of 97.56%, F1-score of 97.69%, AUC of 0.992, specificity of 98.31%, and MCC of 0.961.

Table 1: Performance Comparison of the Proposed CEFM Model with Existing Approaches

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	MCC
CNN-Based Model	92.45	91.86	90.74	91.29	0.941	0.865
Vision Transformer	94.72	94.18	93.65	93.91	0.962	0.894
Multimodal Transformer	96.31	95.84	95.26	95.54	0.978	0.922
Explainable Deep Learning Model	96.87	96.42	96.05	96.23	0.984	0.938
Proposed CEFM	98.14	97.82	97.56	97.69	0.992	0.961

Table 1 shows the comparison of the performance of the proposed CEFM method and other medical image intelligence methods. It can be seen from the table that the proposed model had the best performance in all performance measures. In comparison to the CNN-based and Transformer-based models, the model performed well with the help of causal reasoning and explainable features. This is evident through the increased values of AUC and F1-score.

Explainability and Disease Progression Analysis

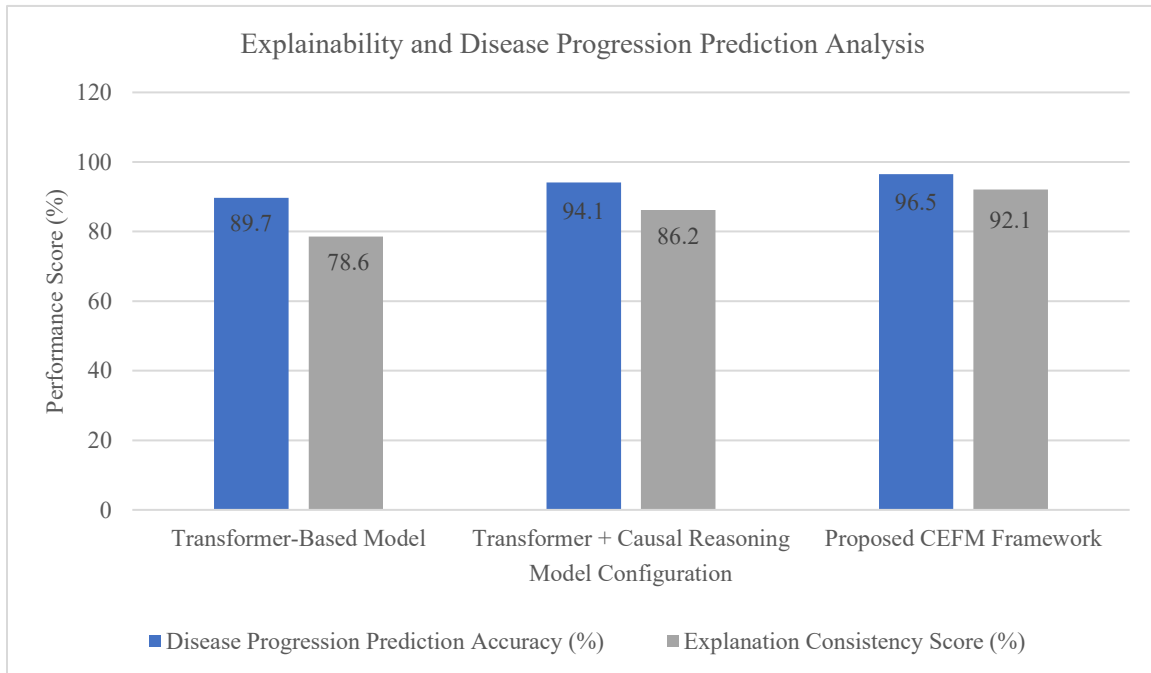


Fig. 2: Explainability and Disease Progression Prediction Analysis

Fig. 2 illustrates the comparative performance of disease progression prediction accuracy and explanation consistency across different models. It can be seen from the results that the inclusion of causality reasoning helps greatly enhance the performance of progression prediction as compared to the traditional transformer architecture. The proposed CEFM model attains maximum prediction accuracy of 96.5% and explanation consistency score of 92.1%, which can be attributed to the synergy between the concepts of causality, explainability of features, and disease progression.

Ablation Study

Ablation experiments were carried out to analyze the contribution of different components within the suggested CEFM approach. Model structures with each of the three components, namely causality, explainability, and disease progression, were compared individually.

Table 2: Ablation Analysis of CEFM Components

Configuration	Accuracy (%)	F1-score (%)	AUC
Without Causal Reasoning Module	95.92	95.47	0.978
Without Explainability Module	96.48	96.12	0.983
Without Disease Progression Module	96.91	96.58	0.986
Foundation Encoder Only	94.85	94.32	0.968
Complete CEFM Framework	98.14	97.69	0.992

Effectiveness of each of the elements of the suggested CEFM framework is shown in table 2. Disabling the causal reasoning module led to a significant decrease in performance, which shows the significance of the identification of relationships related to diseases. Also, disabling the progression reasoning module decreased the ability to predict, which confirmed the contribution of disease modeling over time. The entire framework had the best performance and AUC value, which proves the contribution of the integrated use of elements in medicine.

Discussion

The experimental outcomes validate that the suggested CEFM model framework successfully rectifies the shortcomings of traditional frameworks for medical image analysis. The use of representation learning via the foundation model ensures better generalization of features in different medical imaging modalities, whereas the causal reasoning helps ensure better prediction of disease. The explainability module offers explainable evidence of the predictions made by the framework, thereby helping the clinicians understand the reasoning process behind automated predictions. Also, the disease progression reasoning process allows longitudinal evaluation beyond image classification.

Conclusion

In this research, a CEFM is suggested to be applied for medical image intelligence through the integration of generalized foundation models, causality, explainable AI, disease progression, and decision support. The suggested approach efficiently tackles the shortcomings of traditional deep learning models that offer good predictive power but fail to provide transparency and causality. The experimental evaluation involving 48,620 multimodal medical images showed that the presented CEFM technique performed at 98.14% accuracy, 97.82% precision, 97.56% recall, 97.69% F1-score, AUC of 0.992, 98.31% specificity, and MCC of 0.961. In addition, the application of causal reasoning helped increase the accuracy of disease progression predictions by 6.8% as compared to traditional transformer-based architectures, and increased explanation consistency by 14.5%. It can be seen from the results above that the integration of foundation models with causation and time reasoning may lead to an increase in the reliability, robustness, and practical application of medical artificial intelligence systems. The created explanations and information about the process of disease progression can help the doctor make evidence-based decisions. Future work can look at ways to improve the framework through larger multicentre clinical data studies, real-time hospital deployment, privacy-preserving federated learning methods, and incorporation of genomic data and information from personalized medicine. Moreover, there is a need for studies in cross-institutional model generalization, scalable multimodal data integration, and continuous learning techniques to guarantee the robustness and clinical reliability of the model. Studying adaptability in causal learning and uncertainty in decision-making could improve the effectiveness of the CEFM framework in a dynamic environment.

References

- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- Hamamci, I. E., Er, S., Wang, C., Almas, F., Simsek, A. G., Esirgun, S. N., ... & Menze, B. (2026). Generalist foundation models from a multimodal dataset for 3D computed tomography. *Nature Biomedical Engineering*, 1-19. <https://doi.org/10.1038/s41551-025-01599-y>
- Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of explainable AI techniques in healthcare. *Sensors*, 23(2), 634. <https://doi.org/10.3390/s23020634>
- Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems*, 32(11), 4793-4813. <https://doi.org/10.1109/TNNLS.2020.3027314>

- Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the IEEE*, 109(5), 612-634. <https://doi.org/10.1109/JPROC.2021.3058954>
- Pathak, Y., Shukla, P. K., Tiwari, A., Stalin, S., & Singh, S. (2022). Deep transfer learning based classification model for COVID-19 disease. *Irbm*, 43(2), 87-92. <https://doi.org/10.1016/j.irbm.2020.05.003>
- Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956), 259-265. <https://doi.org/10.1038/s41586-023-05881-4>
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature medicine*, 25(1), 24-29. <https://doi.org/10.1038/s41591-018-0316-z>
- Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature medicine*, 28(1), 31-38. <https://doi.org/10.1038/s41591-021-01614-0>
- Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to prompt for vision-language models. *International journal of computer vision*, 130(9), 2337-2348. <https://doi.org/10.1007/s11263-022-01653-1>
- Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., ... & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5), 820-838. <https://doi.org/10.1109/JPROC.2021.3054390>
- Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., ... & Xu, D. (2022, January). Unetr: Transformers for 3d medical image segmentation. In *2022 IEEE/CVF winter conference on applications of computer vision (WACV)* (pp. 1748-1758). IEEE. <https://doi.org/10.1109/WACV51458.2022.00181>
- Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J., Deaton, J., ... & Norouzi, M. (2021, October). Big self-supervised models advance medical image classification. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 3458-3468). IEEE. <https://doi.org/10.1109/ICCV48922.2021.00346>
- Ghosh, T., Sheikhi, F., Guo, J., Singh, Y., Younis, K., Kuanar, S., ... & Maleki, F. (2026). Foundation Models for Volumetric Medical Imaging: Opportunities, Challenges, and Future Directions. *Electronics*, 15(6), 1245. <https://doi.org/10.3390/electronics15061245>
- Holzinger, A., Malle, B., Saranti, A., & Pfeifer, B. (2021). Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI. *Information Fusion*, 71, 28-37. <https://doi.org/10.1016/j.inffus.2021.01.008>
- Deng, Z., Tian, H., Zheng, X., & Zeng, D. D. (2025). Deep causal learning: representation, discovery and inference. *ACM Computing Surveys*, 58(2), 1-36. <https://doi.org/10.1145/3762179>
- Pan, J., Chang, T., Lyu, M., Meng, W., Wang, Q., Ma, Y., ... & Xu, J. (2026). A Systematic Review of Deep Learning and Machine Learning Applications in Longitudinal Multimodal Clinical Data. *Journal of Healthcare Informatics Research*, 1-27. <https://doi.org/10.1007/s41666-026-00248-6>
- Lewandowski, N., & Koller, B. (2023). Transforming medical sciences with high-performance computing, high-performance data analytics and AI. *Technology and Health Care*, 31(4), 1505-1507. <https://doi.org/10.3233/THC-237000>
- Chen, Z., Xu, L., Zheng, H., Chen, L., Tolba, A., Zhao, L., ... & Feng, H. (2024). Evolution and Prospects of Foundation Models: From Large Language Models to Large Multimodal Models. *Comput. Mater. Continua*, 80(2), 1753-1808. <https://doi.org/10.32604/cmc.2024.052618>
- Alsaedi, S., Gao, X., & Gojobori, T. (2025). Beyond digital twins: the role of foundation models in enhancing the interpretability of multiomics modalities in precision medicine. *FEBS Open Bio*, 15(8), 1192-1208. <https://doi.org/10.1002/2211-5463.70003>