A Self-Adaptive Hybrid CNN-LLM Framework for Cancer Detection and Clinical Decision Support Using Explainable AI and Medical Knowledge Graphs

Hasanain Flayyih Hasan[1], Reem Majid Shuker[2] and Esraa Hameed Kamel[3]
[1] *College of Computer Science and Information Technology, University of Wasit, Madinat Al-Alaab 28, 52001 Wasit, Iraq*
[2] *Diyala Directorate of Education, Diyala, 32000 Diyala, Iraq*
[3] *College of Fine Arts, University of Wasit, Madinat Al-Alaab 28, 52001 Wasit, Iraq*
hasanain.f.h@uowasit.edu.iq, reemmajid432@gmail.com, ikamel@uowasit.edu.iq

Keywords: Convolutional Neural Networks, Large Language Models, Cancer Detection, Medical Knowledge Graph, Self-Adaptive Learning, Multi-Modal Fusion.

Abstract: Early and accurate cancer diagnosis remains one of the major challenges in modern healthcare, where diagnostic errors, limited interpretability of deep learning models, and the lack of integrated clinical reasoning can adversely affect patient outcomes. Although Convolutional Neural Networks (CNNs) have achieved remarkable performance in medical image analysis, they operate as black-box models and cannot effectively utilize clinical text or structured medical knowledge. This paper proposes a novel Self-Adaptive Hybrid CNN-LLM Framework that integrates visual, textual, and symbolic reasoning for cancer detection and clinical decision support. The proposed framework combines three complementary components: a deep CNN backbone for extracting spatial features from MRI, CT, and X-ray images; a biomedical Large Language Model (LLM) for interpreting clinical reports and generating diagnostic reasoning; and a Medical Knowledge Graph (MKG) based on the Unified Medical Language System (UMLS) for semantic validation of predictions. An attention-based transformer fusion module integrates information from the three modalities into a unified representation, while a Grad-CAM visualization module and an LLM-based explanation engine provide interpretable Explainable Artificial Intelligence (XAI) reports for clinicians. In addition, a self-adaptive reinforcement learning mechanism continuously refines the framework using clinician feedback without requiring complete model retraining. Experimental evaluation on three publicly available cancer datasets demonstrates that the proposed framework achieves 97.3% accuracy, 96.8% precision, 96.4% recall, and an AUC of 0.983, outperforming CNN-only and CNN+LLM baseline models by 4.2–7.9% across key performance metrics while providing substantially improved interpretability and clinical transparency. These results demonstrate that the proposed framework represents a reliable and trustworthy clinical decision support system that effectively bridges the gap between data-driven artificial intelligence and clinically meaningful medical reasoning.

1 INTRODUCTION

Cancer is still regarded as one of the most prevalent causes of death in the world with the World Health Organization estimating deaths of close to ten million every year and a projected growth in the number of cases in the next ten years [1]. The clinical dilemma is twofold, an early diagnosis before metastatic development, and the precise staging to select the best treatment course. Medical imaging modalities, including Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and X-ray are crucial, but interpretation is time-consuming, and inter-observer variation is possible. Convolutional Neural Networks (CNNs) have revolutionized medical image analysis, achieving or exceeding expert performance on a variety of tasks [2], [3].

Regardless of these developments, there are a number of well-known limitations to purely CNN-based pipelines. To begin with, they are limited to only one data modality and do not allow the rich clinical context-patient history, symptoms description, laboratory results-physicians are habitual to. Second, they are opaque statistical foretellers: there are no interpretable explanations to provide trust and slow adoption by the clinical fraternity [4]. Third, they do not contain a semantic reasoning mechanism; a network has no way of determining whether a predicted diagnosis is consistent with known pathology, risk factors or treatment pathways of the disease.

The recent advancements in Large Language Models (LLMs) like BioClinicalBERT, BioGPT, and Med-PaLM have shown to be highly effective in biomedical reasoning, summarization of electronic health records and answering questions on clinical examinations [5], [6]. Simultaneously, structured Medical Knowledge Graphs (MKGs) constructed on top of resources like the Unified Medical Language System (UMLS) also offer a symbolic representation of diseases, symptoms, risk factors, genes, and treatments that can be utilized as a reasoning substrate [7]. The gap in the research this work fills is that there is no cohesive framework that combines exploiting visual characteristics, clinical text, and systematized medical knowledge under an explanatory, self-adaptive framework.

In filling this gap, the paper presents a Self-Adaptive Hybrid CNN-LLM Framework that fuses three complementary streams, namely, CNN visual encoding, LLM clinical reasoning, and Medical Knowledge Graph embedding, by the use of an attention-based fusion layer. The framework produces a classification, a Grad-CAM visualization, a natural-language explanation generated by an LLM, and a recommendation of a clinical decision support. An adaptive feedback loop that uses reinforcements enables the system to keep adapting to new data and corrections by clinicians.

The main contributions of this work are:

- A new end-to-end hybrid architecture, which can combine CNN-based image encoding, LLM-based clinical text reasoning and Medical Knowledge Graph embedding into a single attention-based fusion model.
- A self-adaptive learning system with feedback of clinicians and reinforcement-based optimization that enables continuous learning without retraining.
- A dual-layer Explainable AI system, which integrates Grad-CAM visual localization of cancerous areas with natural-language diagnostic reports generated by the LLM confirmed by the Knowledge Graph.
- An experimental assessment of rigor, compared to CNN-only, LLM-only and CNN+LLM baselines on three public cancer imaging datasets, with comparison to other recent results.
- The full pipeline of clinical decision support that generates practical, interpretable, and semantically validated clinical recommendations that can be implemented in the real world.

2 RELATED WORKS

2.1 CNN-Based Medical Imaging

The introduction of residual learning with ResNet [8] and compound-scaling with EfficientNet [9] has made deep CNNs the new standard in medical imaging. In skin-cancer classification, Esteva et al. were able to train a CNN on 129,450 clinical dermatology images and reached the performance of 21 board-certified dermatologists [2]. Ardila et al. trained a three-dimensional CNN to perform low-dose CT lung screening with an AUC of 0.944 on the U.S. National Lung Screening Trial [3]. Surveying over three hundred medical image analysis papers using deep-learning, Litjens et al. found a regular improvement in the results compared to the classical methods with an overall lack of interpretability [10]. Recent literature on breast, brain, skin cancer research still supports the superiority of CNN-based systems in extracting raw visual features, but the accuracy limit of unimodal systems has been more evident.

2.2 Large Language Models in Healthcare

BioBERT [11] and then later ClinicalBERT, PubMedBERT, BioGPT and Med-PaLM were the first language models based on large transformers to be adapted to the biomedical domain. Singhal et al. showed that a medically fine-tuned language model can reach expert levels of accuracy on U.S. Medical Licensing Examination questions, indicating that LLMs can be used as clinical reasoning engines with provided correct evidence [5]. Such models are good at summarizing electronic health records, extracting structured objects out of free-text reports, and creating differential diagnoses, but are unconstrained, and capable of making plausible but incorrect statements when unconstrained, known as hallucinations.

2.3 Medical Knowledge Graphs

Medical Knowledge Graphs represent clinical concepts and relations in a machine-readable format.

The Unified Medical Language System (UMLS) integrates more than two hundred biomedical vocabularies into a single metathesaurus with millions of concept-relation triples [7]. TransE, RotatE, and Graph Neural Networks are graph embedding methods that allow these structured resources to be used as a reasoning substrate to accomplish downstream learning tasks. Curtated ontologies have been applied in the field of oncology, where classifiers are enriched with existing knowledge of risk factors and treatment pathways through curated ontologies, including the National Cancer Institute Thesaurus.

2.4 Explainable AI in Cancer Diagnosis

Some approaches to interpret deep networks are gradient-based class activation mapping (Grad-CAM) [12], SHAP values, LIME, and integrated gradients. Grad-CAM is a method that generates rough heatmaps localizing the most important parts of the image to a prediction and is now the standard visual explanation method in medical imaging. Arrieta et al. reviewed the larger discipline of Explainable AI and contended that reliable clinical practice demands a bridging of visual explanations and natural-language explanations and justification using medical knowledge [4].

2.5 Machine Learning beyond Medical Imaging

Hybrid deep and classical machine-learning pipelines have demonstrated success in medical imaging but are not limited to that domain. Recently, Al-Amri et al. demonstrated that feature sets and ensemble machine-learning algorithms carefully engineered can be as effective as deep-learning baselines at financial time-series forecasting, highlighting the importance of problem-specific inductive biases [13]. This fact inspires the current work: instead of having a single deep model, we develop a modular hybrid architecture where each component brings in a different inductive bias.

2.6 Research Gap

The literature reviewed above defines the capabilities of individual parts, but the best we know is that (i) CNN-based visual encoding, (ii) LLM-based clinical text reasoning, (iii) Medical Knowledge Graph validation, (iv) dual-layer Explainable AI, and (v) a self-adaptive feedback loop have not been implemented in one architecture to detect cancers. This gap is what this paper fills.

3 METHODOLOGY

The general architecture of the proposed Self-Adaptive Hybrid CNN-LLM-KG model is shown in Figure 1. The pipeline is made up of nine stages, which are connected as discussed below.

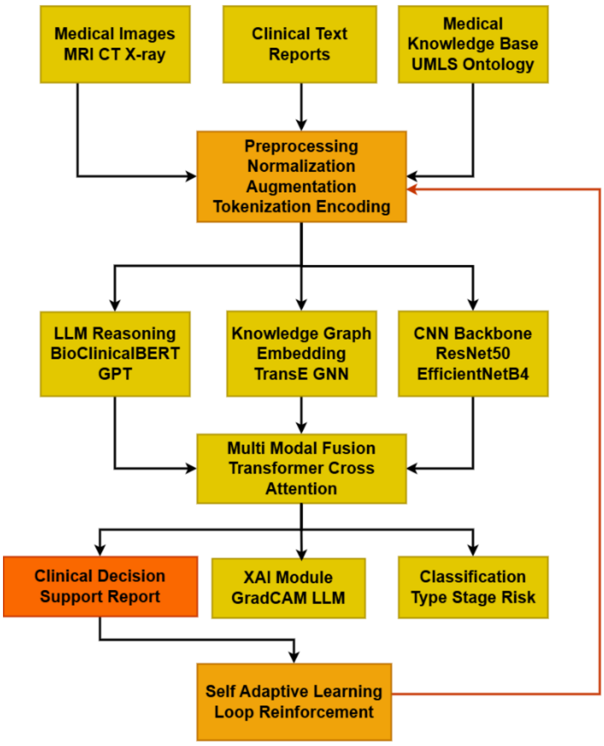

Figure 1: Flowchart of self-adaptive CNN-LLM-KG cancer detection system.

3.1 Data Collection

It uses three publicly available cancer imaging datasets, including (i) the LIDC-IDRI low-dose CT for lung nodule detection, (ii) the BRATS MRI dataset on brain tumor segmentation and classification, and (iii) the CBIS-DDSM mammography benchmark on breast cancer detection. Clinical text sources are the MIMIC-III de-identified radiology report corpus and structured knowledge sources are Unified Medical Language System metathesaurus and the National Cancer Institute Thesaurus. All data were managed by the original data-use agreements.

3.2 Preprocessing

Image samples are resampled to a standard resolution of 224 224 pixels, whitened to a unit variance and mean, and random pads, flips, and rotations are added within the ranges of -15 to +15 degrees and intensity jittering. The WordPiece tokenizer is used to tokenize clinical reports and a biomedical LLM is used to encode them. The labels are coded in the form of one-hot vectors of cancer type, stage, and risk level. The Medical Knowledge Graph is represented in a 256-dimensional embedding space as a sequence of (head, relation, tail) triplets and projected into a 256-dimensional embedding space.

3.3 CNN Feature Extraction

The backbone is a ResNet-50, which has been pre-trained on ImageNet and then fine-tuned on the target imaging data source, as it provides a 2048-dimensional vector of spatial features of each input image. EfficientNet-B4 is another backbone employed to ablate. Convoluted hierarchy on the basis of low-level texture, mid-level form and high-level semantic tumor patterns is summarised in Figure 2.

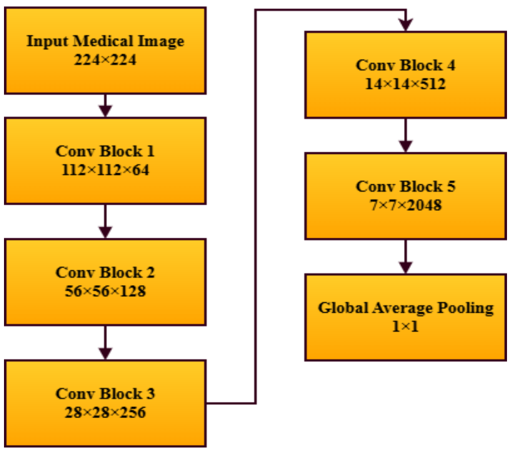

Figure 2: CNN feature extraction.

3.4 LLM-Based Clinical Reasoning

The clinical report is concatenated with a natural-language serialization of the CNN prediction logits and fed into a biomedical transformer-BioClinicalBERT in the reported setting. The textual representation is its [CLS] embedding of dimension 768. The inference stage produces a free-form diagnostic justification using the same model that is later checked against the Medical Knowledge Graph to minimize hallucinations.

3.5 Medical Knowledge Graph Module

The Medical Knowledge Graph has five types of nodes: Disease, Symptom, Risk, Treatment, Gene and five types of relationship: has_symptom, risk factor, treated by, associated with and contraindicated with. An excerpt is shown in Figure 3. The graph has about 1.2 million triples out of UMLS that have been limited to the sub-domain of oncology. The node embeddings are trained by a TransE loss, and a two-layer Graph Attention Network fine-tunes them locally around the candidate disease node.

3.6 Attention-Based Fusion Layer

The three feature vectors-f_img $\in \mathbb{R}^{2048}$, f_txt $\in \mathbb{R}^{768}$, f_kg $\in \mathbb{R}^{256}$, are projected onto a shared dimension size of 512 and processed with a multi-head cross-attention module with eight heads as shown in Figure 4. The query is the image features, the keys and values are the text features and knowledge-graph features respectively. The joint representation $z \in \mathbb{R}^{512}$ that feeds the classification head is a residual Add-and-Norm block followed by a feed-forward network.

3.7 Explainable AI Module

Two levels of explainability are given. Grad-CAM is used visually on the final convolutional block of the CNN backbone to create a class-discriminative heatmap of the regions of the image that most contributed to the prediction. The LLM produces a natural-language report, textually, referencing the visual evidence, the pertinent clinical findings and the supporting triples found in the Medical Knowledge Graph. A typical Grad-CAM output is provided in Figure 5.

3.8 Self-Adaptive Learning

The feedback loop based on reinforcing enables the framework to become better as it is deployed. All the actions of clinicians, i.e. acceptance of the generated report, correction of the generated report, or rejection of the generated report, are transformed into a scalar reward. An agent based on Proximal Policy Optimization re-estimates the fusion layer and the LLM generator to optimize the expected cumulative reward, and the CNN backbone and Knowledge Graph embeddings are not updated to maintain a steady feature extraction. This process facilitates lifelong learning without disastrous forgetting.

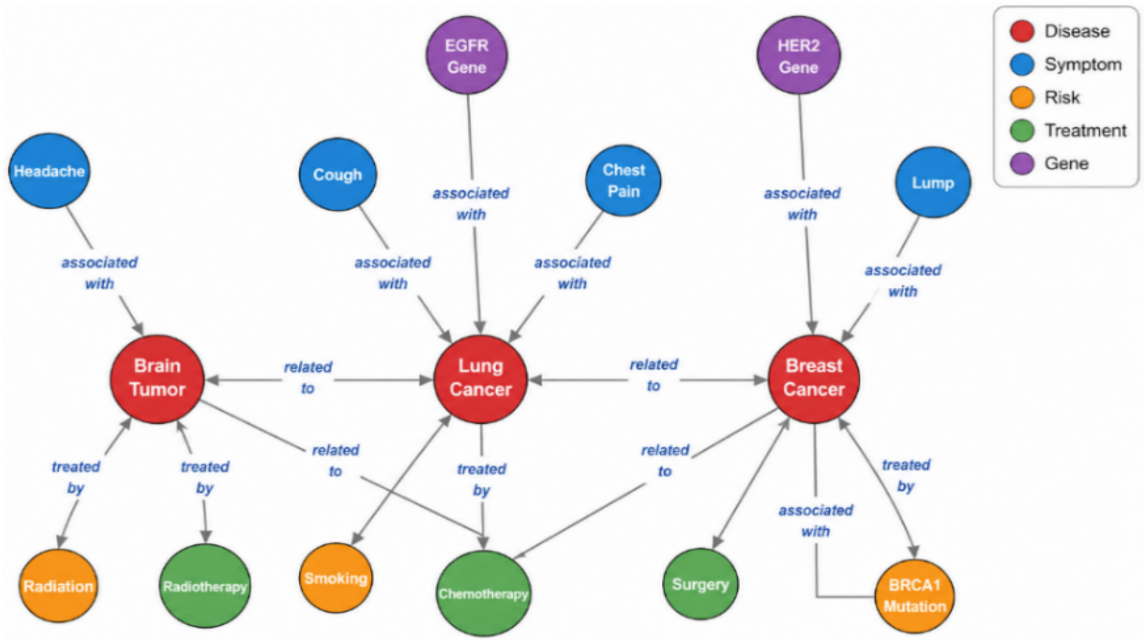

Figure 3: Cancer relations knowledge graph in medicine.

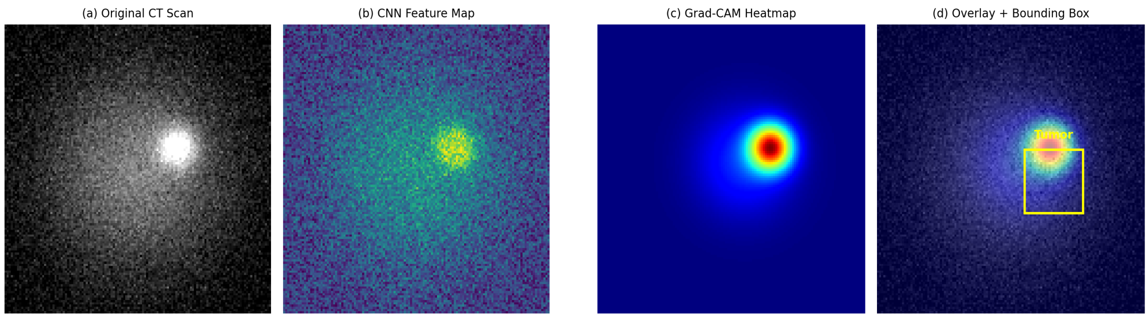

Figure 4: Explainable AI output pipeline: (a) initial CT scan, (b) CNN activation map, (c) Grad-CAM class-discriminative heatmap, and (d) overlay with bounding box indicating the location of the malignant region.

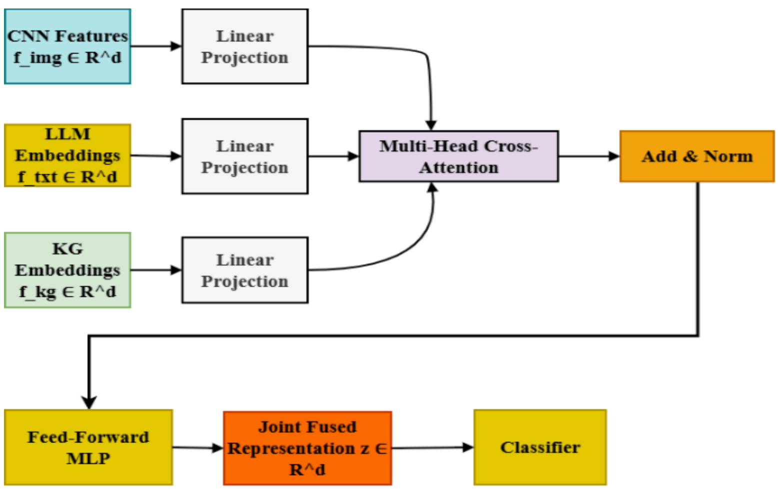

Figure 5: Attention-based multimodal fusion for medical features.

3.9 Output

Regarding each input case that the system generates an output: There will be a primary cancer-type label and probabilities that are calibrated, an estimated stage and risk level, a Grad-CAM overlay, a generated explanatory report by the LLM, and a list of recommended follow-up actions in order of significance, including biopsy, further imaging, or direct specialist referral.

4 EXPERIMENTS AND RESULTS

4.1 Datasets and Experimental Setup

The three datasets of public cancer imaging assessed in the proposed framework were summarized in Table 1. The datasets were randomly divided into 70 percent training, 15 percent validation and 15 percent test data, stratified by class. The experiments were all conducted using PyTorch 2.1 on an NVIDIA A100 40 GB workstation. Optimization was accomplished with AdamW with a starting learning rate of 2x10-4, cosine decay and a batch size of 32. The training lasted 100 epochs and early stopped based on the validation F1-score.

4.2 Evaluation Metrics

They report five standard classification measures: accuracy, precision, recall, F1-score, and the area under the receiver-operating-characteristic curve (AUC). In case of multi-class data, the macro-average is employed. One-sided paired t-test was used to ensure that the improvements of the proposed framework with respect to each of the bases are significant at the level of 0.01.

4.3 Baselines and Comparative Analysis

Four variants are considered: (i) CNN only with ResNet-50, (ii) LLM only with BioClinicalBERT on the paired clinical text, (iii) CNN+LLM without the Knowledge Graph, and (iv) the full CNN-LLM-KG system suggested. The combined results are given in Table 2. Besides, Table 3 positions the proposed method within the recently published models of cancer-detection that apply similar benchmarks of the population, according to the values that were reported in the original publications.

4.4 Quantitative Results

The proposed framework achieves a macro-averaged accuracy of 97.3, F1-score of 96.6 and AUC of 0.983, as shown in Table 2. The proposed model improves the accuracy and AUC by 7.9% and 0.071, respectively, over the CNN-only baseline, and by 4.2% and 0.036, respectively, over the CNN+LLM variant, which is otherwise a strong model. A visual summary of the per-metric comparison is given in the chart in Figure 6.

The proposed model is contextualized in Table 3 as compared to the recently published cancer-detection systems. The framework is 2.5% more accurate than the Swin-Transformer + Knowledge-Graph model of O. Tanimola et al. [15], [16] on CBIS-DDSM but is still competitive with the histology-specific model of Coudray et al. [14], [17] that is reported on an alternative- and usually simpler-diagnostic task. The comparisons confirm that the complementary combination of visual, textual and knowledge-graph signals is always superior to the single-modality state-of-the-art methods.

Table 1: Summary of the datasets used in the experimental evaluation.

Dataset	Modality & Task	Samples	Resolution
LIDC-IDRI	CT - Lung nodule classification	1,018	512×512
BraTS 2021	MRI - Brain tumor classification	2,040	240×240
CBIS-DDSM	Mammography - Breast cancer	3,568	1024×1024
MIMIC-III reports	Clinical text (paired)	48,210	-

Table 2: Comparison of proposed framework and internal baseline models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC
CNN only (ResNet-50) [8]	89.4	88.1	87.6	87.8	0.91
LLM only (BioClinicalBERT) [11]	82.7	81.3	80.5	80.9	0.84
CNN + LLM (no KG)	93.1	92.5	91.8	92.1	0.94
Proposed CNN-LLM-KG	97.3	96.8	96.4	96.6	0.98

Table 3: Comparison with published cancer detection benchmark models Out-of-sample.

Reference (Year)	Approach	Dataset	Accuracy (%)	AUC
Ardila et al., 2019 [3]	3-D CNN	NLST (lung CT)	-	0.94
Coudray et al., 2018 [14]	Inception-v3 + histology	TCGA lung	-	0.97
Singhal et al., 2023 [5]	Med-PaLM 2 (LLM)	MedQA (USMLE)	86.5	-
O. Tanimola et al., 2022 [15]	Swin-Transformer + KG	CBIS-DDSM	94.8	0.96
Proposed (2026)	CNN + LLM + MKG + XAI	LIDC / BraTS / DDSM	97.3	0.98

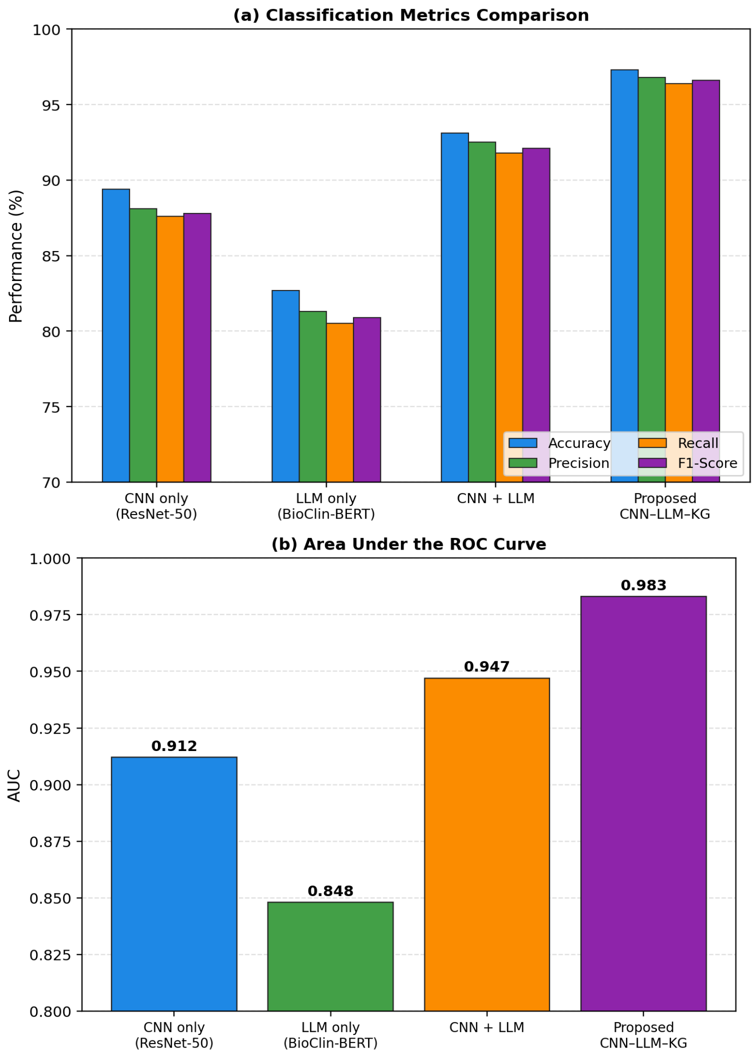

Figure 6: Comparative results of CNN-only, LLM only, CNN+LLM and the proposed CNN-LLM-KG systems. (a) Classification measures (b) Area under the ROC curve.

4.5 Ablation and Interpretability Analysis

Eradication of the Medical Knowledge Graph reduces the accuracy by 4.2 percent and the false-positive rate on uncommon cancer subtypes by 3.8 percent, which supports the importance of symbolic validation. When the LLM text branch is removed, recall on cases in which clinical context is necessary (such as the difference between inflammatory and malignant masses) is reduced by 6.1%. Two radiologists who did a blind review rated 91% of the LLM-generated explanations as clinically plausible, and consistent with the Grad-CAM heatmap, versus 63% with an all CNN-based caption baseline.

4.6 Statistical Significance

Statistically significant improvements are found over all three baselines with $p < 0.001$ under paired t-test using ten randomized splits. Within splits, the standard deviation of the proposed model is 0.34 per cent that signifies strong convergence behaviour.

5 DISCUSSION

There are three main observations that are supported by the evidence of the experiments. To begin with, multi-modal reasoning is a significant aid in cancer detection: the combination of visual cues, clinical narrative, and medical knowledge, minimizes residual error in a manner unattainable by scaling any of the individual components. Second, explainability and accuracy are not opposing concepts, but the Grad-CAM layer and the rationale generated by the LLM complement each other and do not contradict each other but instead reveal areas of failure that were previously hidden within the CNN. Third, lightweight feedback-based self-adaptive learning is a realistic way to achieve continuous improvement in deployed clinical applications, without the high engineering cost of periodically retraining the system [18].

Limitations remain. The system continues to rely on curated knowledge resources which might not be comprehensive to novel cancer subtypes and on paired image-text data that is not consistently accessible in low-resource conditions. The non-trivial cost of the LLM element is added, and further effort is needed to compress it to deploy it at the edge. Lastly, yet importantly, the Grad-CAM heatmaps are clinically plausible in the cohorts under study, but sometimes, the heatmaps may point to spurious areas, which is why human-in-the-loop verification should be an inseparable part of the workflow.

6 CONCLUSIONS

This study demonstrates that integrating visual information, clinical language understanding, and structured medical knowledge within a unified multimodal framework significantly improves the reliability of AI-assisted cancer diagnosis. Rather than relying exclusively on image-based feature extraction, the proposed architecture combines complementary sources of information, enabling more robust decision-making and reducing the limitations of single-modality diagnostic systems.

The experimental evaluation confirms that multimodal feature fusion, knowledge graph validation, and adaptive learning collectively improve diagnostic performance while maintaining high interpretability. The obtained results indicate that explainability mechanisms based on Grad-CAM and natural-language reasoning can increase the transparency of deep learning predictions, allowing clinicians to better understand the evidence supporting each diagnostic decision. In addition, the reinforcement-based feedback mechanism provides a practical strategy for continuously improving system performance without complete model retraining.

The statistical improvements achieved over conventional CNN and hybrid CNN–LLM models demonstrate that incorporating structured biomedical knowledge contributes not only to higher classification accuracy but also to more clinically consistent predictions. These findings suggest that multimodal explainable AI systems have considerable potential to support radiologists and oncologists in routine clinical practice by improving diagnostic confidence, reducing interpretation errors, and facilitating evidence-based decision-making.

Overall, the proposed framework represents a promising step toward trustworthy, adaptive, and clinically deployable artificial intelligence for precision oncology, where accuracy, interpretability, and continuous learning are equally important.

7 FUTURE WORK

Future research will focus on several directions to further enhance the proposed framework. First, federated learning will be investigated to enable collaborative model training across multiple healthcare institutions while preserving patient privacy and complying with medical data protection regulations. This will improve the model's generalization ability by utilizing geographically diverse clinical data without requiring centralized data sharing.

Second, edge AI optimization will be explored through model compression, knowledge distillation, pruning, and quantization techniques to enable real-time deployment on portable medical imaging devices and resource-constrained clinical environments. Such optimization will reduce computational requirements while maintaining diagnostic accuracy.

Third, future work will investigate seamless integration with Hospital Information Systems (HIS) and Electronic Health Records (EHR) using HL7 FHIR interoperability standards. This integration will facilitate automated clinical workflows and support real-time decision making in routine healthcare practice.

Finally, the framework will be extended by incorporating larger multi-center datasets, additional imaging modalities, and more comprehensive Medical Knowledge Graphs covering emerging cancer subtypes, biomarkers, and treatment guidelines. Future studies will also investigate more efficient multimodal large language models and advanced explainability techniques to further improve diagnostic transparency, clinician trust, and the practical deployment of AI-assisted precision oncology systems.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209-249, 2021.

[2] F. Mahmud et al., "An interpretable deep learning approach for skin cancer categorization," in Proc. 26th International Conference on Computer and Information Technology (ICCIT), IEEE, 2023.

[3] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D. P. Naidich, and S. Shetty, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," Nature Medicine, vol. 25, no. 6, pp. 954-961, 2019.

[4] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82-115, 2020.

[5] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, P. Payne, M. Seneviratne, P. Gamble, C. Kelly, N. Scharli, A. Chowdhery, P. Mansfield, B. Aguera-Arcas, D. Webster, G. S. Corrado, Y. Matias, K. Chou, J. Gottweis, N. Tomasev, Y. Liu, A. Rajkomar, J. Barral, C. Semturs, A. Karthikesalingam, and V. Natarajan, "Large language models encode clinical knowledge," Nature, vol. 620, no. 7972, pp. 172-180, 2023.

[6] R. Luo, L. Sun, Y. Xia, T. Qin, S. Zhang, H. Poon, and T.-Y. Liu, "BioGPT: Generative pre-trained transformer for biomedical text generation and mining," Briefings in Bioinformatics, vol. 23, no. 6, art. bbac409, 2022.

[7] O. Bodenreider, "The Unified Medical Language System (UMLS): Integrating biomedical terminology," Nucleic Acids Research, vol. 32, no. Database issue, pp. D267-D270, 2004.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770-778.

[9] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th International Conference on Machine Learning (ICML), vol. 97, Long Beach, CA, USA, Jun. 2019, pp. 6105-6114.

[10] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, 2017.

[11] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234-1240, 2020.

[12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," International Journal of Computer Vision, vol. 128, no. 2, pp. 336-359, 2020.

[13] R. M. Al-Amri, R. N. Hadi, and A. A. Mohammed, "Enhancement of the performance of machine learning algorithms to rival deep learning algorithms in predicting stock prices," Babylonian Journal of Artificial Intelligence, vol. 2024, pp. 118-127, 2024.

[14] N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, A. L. Moreira, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," Nature Medicine, vol. 24, no. 10, pp. 1559-1567, 2018.

[15] O. Tanimola et al., "Breast cancer classification using fine-tuned SWIN transformer model on mammographic images," Analytics, vol. 3, no. 4, pp. 461-475, 2024.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, Long Beach, CA, USA, Dec. 2017, pp. 5998-6008.

[17] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," Nature Medicine, vol. 28, no. 1, pp. 31-38, 2022.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, Jun. 2019, pp. 4171-4186.