Journal of Multimedia Information System VOL. 13, NO. 1, March 2026 (pp. 29-42): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2026.13.1.29

Multimodal Prostate Cancer Diagnosis via Knowledge-Enhanced Deep Learning Networks

Xiaodan Zhang[1], Chao Wang[2*]

Abstract

Prostate cancer remains one of the most prevalent malignancies among men worldwide, where early and accurate diagnosis significantly impacts patient outcomes. Current deep learning approaches for prostate cancer diagnosis face two fundamental limitations: heavy reliance on large-scale annotated multimodal data and insufficient incorporation of clinical expertise embedded in medical knowledge structures. These limitations often result in models that struggle to capture the nuanced relationships between imaging biomarkers, clinical indicators, and pathological states. To address these concerns, a multimodal prostate cancer diagnosis framework integrating medical knowledge graphs with deep learning is proposed, termed Multimodal Prostate Cancer Diagnosis using Knowledge-Enhanced Networks (MP-KDNet). The core architecture employs a knowledge-driven convolutional neural network that fuses heterogeneous data sources, including MRI sequences, clinical laboratory results, and patient history documentation. Through entity disambiguation and knowledge graph embedding techniques, structured clinical knowledge regarding prostate pathology is extracted and transformed into continuous vector representations. These knowledge entity vectors, alongside multimodal feature representations, serve as multi-channel inputs to the convolutional architecture. Multi-scale convolutional kernels capture diagnostic patterns across both fine-grained clinical observations and broader symptom constellations encoded in medical knowledge. Experimental validation on the MIMIC-IV dataset containing 12,847 prostate-related cases demonstrates that MP-KDNet achieves 82.7% diagnostic accuracy, outperforming conventional multimodal fusion, transformer-based imaging analysis, and knowledge graph reasoning baselines. Results confirm that integrating structured clinical expertise with patient-specific multimodal data yields more accurate discrimination among benign prostatic hyperplasia, prostatitis, and adenocarcinoma subtypes than either data-driven or knowledge-driven approaches alone.

Key Words: Multimodal Learning, Knowledge Graph, Deep Learning, Prostate Cancer Diagnosis.

I. INTRODUCTION

Prostate cancer diagnosis represents a critical challenge in modern urological oncology, where accurate early detection directly correlates with treatment efficacy and patient survival rates [1]. Traditional diagnostic pathways rely heavily on a combination of serum prostate-specific antigen (PSA) measurements, digital rectal examination findings, and transrectal ultrasound-guided biopsy results [2]. However, these conventional methods exhibit notable limitations in sensitivity and specificity, frequently leading to overdiagnosis of indolent tumors or delayed detection of aggressive malignancies [3]. The integration of multimodal medical data—encompassing magnetic resonance imaging sequences, histopathological patterns, laboratory biomarkers, and clinical narratives—offers promising avenues for enhancing diagnostic precision. However, the heterogeneous nature of these data sources, combined with their complex interrelationships, poses substantial computational challenges for automated analysis systems.

Recent advances in deep learning have catalyzed significant progress in medical image analysis and clinical decision support [4-5]. Convolutional neural networks demonstrate remarkable capability in extracting hierarchical features from radiological images, while recurrent architectures excel at processing sequential clinical records [6]. Nevertheless, these data-driven approaches exhibit fundamental weaknesses when confronting medical diagnosis tasks [7]. Deep neural networks typically require extensive annotated datasets for adequate training, which remain scarce in specialized domains like prostate pathology [8]. Furthermore, purely data-driven models lack mechanisms for incorporating decades of accumulated clinical expertise regarding disease manifestations, risk factors, and diagnostic

Manuscript received January 27, 2026; Revised February 21, 2026; Accepted March 03, 2026. (ID No. JMIS-26M-01-005)
Corresponding Author (*): Chao Wang, +86-15940669973, wangc@jzmu.edu.cn
[1]The First Clinical Medical College, Jinzhou Medical University, Jinzhou, China, 15640497496@163.com
[2]Department of Urology, First Affiliated Hospital of Jinzhou Medical University, Jinzhou, China, wangc@jzmu.edu.cn

criteria [9]. This absence of structured medical knowledge frequently results in models that capture superficial statistical patterns while missing crucial clinical relationships that experienced urologists intuitively recognize [10].

Medical knowledge graphs represent a paradigm shift in how clinical expertise can be formalized and computationally leveraged [11-12]. These structured knowledge bases encode entities such as diseases, symptoms, biomarkers, and anatomical structures, along with their semantic relationships, including causal links, diagnostic associations, and hierarchical taxonomies [13]. For prostate cancer specifically, knowledge graphs can capture intricate relationships between Gleason scoring patterns, PSA kinetics, imaging characteristics on T2-weighted and diffusion-weighted MRI sequences, and ultimate pathological diagnoses [14]. Several large-scale medical knowledge graphs have emerged, including specialized oncology knowledge bases that document evidence-based relationships between clinical findings and cancer subtypes [15]. Unlike traditional knowledge representation formats, graph-structured knowledge facilitates efficient reasoning and enables machine learning models to traverse relationship paths during inference [16].

The intersection of knowledge graphs with deep learning architectures presents compelling opportunities for advancing prostate cancer diagnosis [17-18]. By embedding structured clinical knowledge into continuous vector spaces, these entities can be integrated directly into neural network training pipelines [19]. Consider the diagnostic scenario where a patient presents with moderately elevated PSA levels, an enlarged prostate on imaging, and nocturia symptoms [20]. A purely data-driven model might focus predominantly on the PSA elevation, potentially leading to unnecessary biopsies [21]. However, an approach augmented with knowledge graph representations would recognize that these symptoms collectively align more strongly with benign prostatic hyperplasia rather than malignancy, particularly when certain imaging features are absent [22]. The knowledge graph captures that while PSA elevation occurs in both conditions, specific patterns of elevation combined with particular MRI characteristics differentiate the two entities.

Building upon these observations, the MP-KDNet framework is introduced for multimodal prostate cancer diagnosis that synergistically combines knowledge graph embeddings with convolutional neural architectures. The methodology extracts structured pathological knowledge from medical knowledge graphs through entity alignment and graph embedding techniques, transforming this knowledge into dense vector representations. These knowledge vectors are then fused with multimodal clinical feature representations—derived from MRI imaging data, laboratory measurements, and clinical documentation—to form multi-channel inputs for a specialized convolutional network. Through this knowledge-enhanced learning process, the model simultaneously learns from objective multimodal patient data and subjective clinical expertise encoded in knowledge structures.

The contribution of this paper can be summarized as follows.

(1) We propose MP-KDNet, a novel framework that synergistically integrates medical knowledge graphs with multimodal deep learning for prostate cancer diagnosis, addressing the limitations of purely data-driven approaches in capturing clinical expertise.

(2) We construct a comprehensive prostate cancer knowledge graph containing 25,264 entities and 70,049 triples, then employ TransD embedding to transform structured clinical knowledge into continuous representations compatible with neural processing.

(3) We design a knowledge-enhanced multi-channel convolutional architecture that simultaneously processes multimodal clinical features, aligned knowledge entity embeddings, and contextual relationship vectors through parallel information streams with multi-scale pattern detection.

(4) We develop entity alignment and contextual knowledge extraction mechanisms that automatically link clinical observations from MRI reports, laboratory measurements, and patient documentation to relevant knowledge graph entities and their semantic neighborhoods.

The rest of the paper is organized as follows: Section II establishes the foundation for knowledge integration by describing how the prostate cancer knowledge graph is constructed and represented. Section III presents the complete MP-KDNet framework, walking through each component of the diagnostic pipeline. Section IV provides experimental validation of the proposed approach. Section V concludes the paper.

II. PROSTATE CANCER KNOWLEDGE GRAPH CONSTRUCTION

2.1. Knowledge Graph Schema Definition

A prostate cancer knowledge graph $\mathcal{G}_{\mathrm{PCa}}$ constitutes a structured representation of clinical entities and their semantic relationships within the prostate cancer domain. The graph encodes medical knowledge through a collection of triples, where each triple formalizes a specific clinical relationship. Formally, $\mathcal{G}_{\mathrm{PCa}}$ is defined as:

$$\mathcal{G}_{\mathrm{PCa}} = \langle \xi_h, \rho, \xi_t \rangle, \tag{1}$$

where ξ_h denotes the head entity, ξ_t represents the tail entity, and $\rho \in \{\rho_1, \rho_2, \ldots, \rho_{|\mathcal{R}|}\}$ specifies the relationship

type connecting them. Both ξ_h and ξ_t belong to the entity set $\mathcal{E}$ of $\mathcal{G}_{\text{PCa}}$, while ρ belongs to the relationship set $\mathcal{R}$ containing $|\mathcal{R}|$ distinct relationship types.

The entity set $\mathcal{E}$ encompasses diverse medical concepts relevant to prostate cancer diagnosis. Primary entities include pathological conditions (adenocarcinoma, prostatic intraepithelial neoplasia, benign prostatic hyperplasia), imaging biomarkers (PI-RADS scores, apparent diffusion coefficient values, T2-weighted signal characteristics), laboratory measurements (PSA levels, PSA density, free-to-total PSA ratio), anatomical structures (peripheral zone, transition zone, anterior fibromuscular stroma), clinical manifestations (lower urinary tract symptoms, hematuria, bone pain), and therapeutic interventions (active surveillance, radical prostatectomy, radiation therapy).

The relationship set $\mathcal{R}$ captures various semantic associations between entities. Key relationship types include diagnostic_indicator (linking imaging findings to pathological states), risk_factor (connecting predisposing conditions to cancer development), prognostic_marker (relating biomarkers to disease outcomes), anatomical_location (specifying tumor spatial distribution), treatment_response (associating therapies with clinical outcomes), and disease_progression (tracking temporal evolution of pathological states).

For instance, the clinical knowledge that "peripheral zone adenocarcinoma typically demonstrates restricted diffusion on DWI sequences with ADC values below 1.0×10^{-3} mm²/s translates into the triple representation. The comprehensive knowledge graph aggregates thousands of such triples extracted from clinical guidelines, peer-reviewed literature, and electronic health records to construct a holistic representation of prostate cancer domain knowledge.

2.2. Knowledge Graph Embedding Framework

Knowledge graph embedding transforms discrete symbolic entities and relationships from $\mathcal{G}_{\text{PCa}}$ into continuous low-dimensional vector spaces while preserving the graph's semantic structure [23]. This vectorization enables downstream integration with neural network architectures [24]. Given the complexity of prostate cancer relationships—including one-to-many, many-to-one, and many-to-many associations—the TransD embedding model is adopted for its superior handling of heterogeneous relationship patterns.

For a triple $\langle \xi_h, \rho, \xi_t \rangle$, TransD posits that entities connected through relationship ρ occupy distinct semantic spaces. The model employs projection matrices $\mathbf{M}_h$ and $\mathbf{M}_t$ to map head and tail entities into the relationship-specific space. These projection matrices are decomposed as:

$$\mathbf{M}_h = \rho_p \xi_{hp}^{\top} + \mathbf{I}_{\delta \times \kappa}, \quad \mathbf{M}_h = \rho_p \xi_{tp}^{\top} + \mathbf{I}_{\delta \times \kappa'}, \tag{2}$$

where $\rho_p \in \mathbb{R}^{\kappa}$ represents the relationship projection vector, $\xi_{hp}, \xi_{tp} \in \mathbb{R}^{\delta}$ denote entity-specific projection vectors, and $\mathbf{I}_{\delta \times \kappa}$ is the identity matrix of appropriate dimensions. The scoring function measuring triple plausibility becomes:

$$\phi_\rho(\xi_h, \xi_t) = \|\mathbf{M}_h \xi_h + \rho - \mathbf{M}_t \xi_t\|_{L_1/L_2}. \tag{3}$$

This formulation allows each entity-relationship pair to define unique projection dynamics, accommodating the diverse semantic characteristics present in medical knowledge. Training optimizes embeddings by minimizing ϕ_ρ for valid triples while maximizing scores for corrupted negatives, yielding entity vectors $\xi_h, \xi_t \in \mathbb{R}^{\kappa}$ and relationship vectors $\rho \in \mathbb{R}^{\kappa}$ that encode structured clinical knowledge in continuous space.

III. MULTIMODAL PROSTATE CANCER DIAGNOSIS VIA KNOWLEDGE-ENHANCED NETWORKS

3.1. Framework Overview and Architecture

The MP-KDNet framework addresses prostate cancer diagnosis through three synergistic components: multimodal feature extraction, structured knowledge integration, and knowledge-enhanced convolutional classification. Fig. 1 illustrates the complete architecture, demonstrating information flow from raw multimodal inputs through knowledge-augmented processing to diagnostic outputs.

The input layer accepts heterogeneous clinical data sources: (1) MRI sequences including T2-weighted, diffusion-weighted, and dynamic contrast-enhanced imaging; (2) laboratory measurements encompassing PSA kinetics, complete blood counts, and metabolic panels; (3) clinical documentation comprising patient histories, symptom descriptions, and physical examination findings. These diverse modalities undergo specialized preprocessing pipelines tailored to their respective data types. Missing data were addressed through domain-informed imputation. For imaging data, missing sequences were imputed using zero-padding with a missing indicator channel. Laboratory values were imputed using cohort-specific median values stratified by age and diagnostic category. Text data missing specific sections was processed with available portions only. We evaluated model performance under systematic missing data scenarios, where randomly excluding one modality reduced accuracy by an average of 3.2 percentage points, demonstrating reasonable robustness to incomplete records.

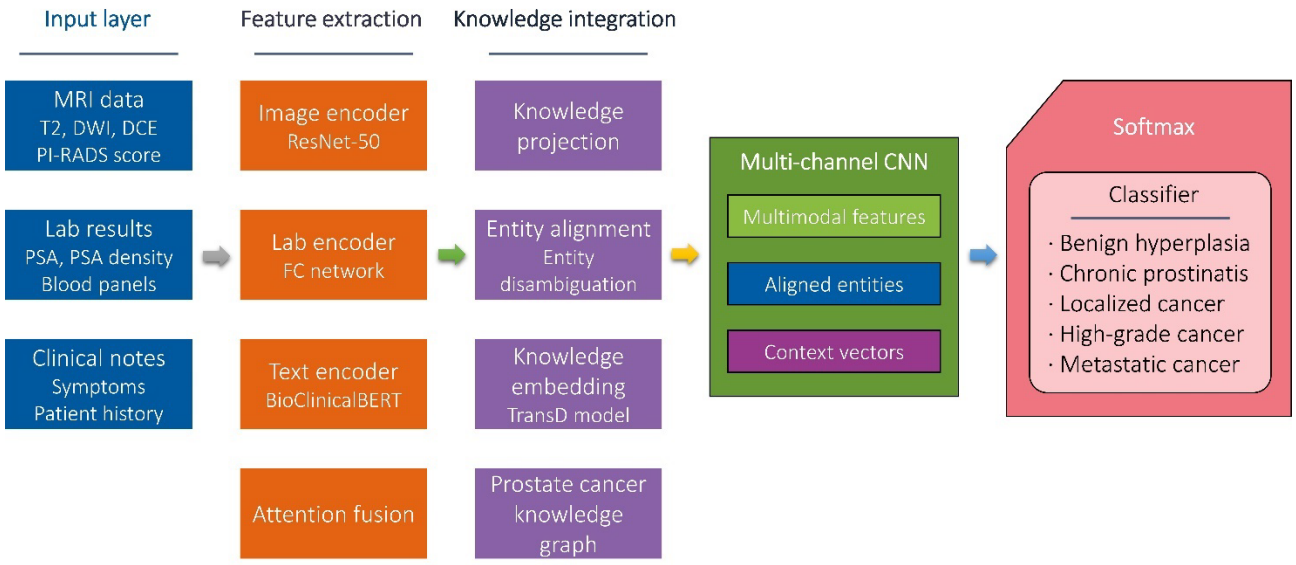

Fig. 1. Overall architecture of MP-KDNet framework.

The knowledge integration module establishes connections between extracted clinical features and entities within $\mathcal{G}_{\text{PCa}}$. Through entity alignment mechanisms, clinical descriptors from multimodal inputs are mapped to corresponding knowledge graph entities, enabling the retrieval of relevant structural knowledge. Knowledge graph embeddings transform these discrete entities into continuous representations that can be processed alongside data-driven features.

The convolutional architecture receives multi-channel inputs combining data-derived feature vectors with knowledge entity embeddings. Multiple convolutional kernels of varying receptive fields capture patterns across both modality-specific features and cross-modal interactions. The knowledge-enhanced representations enable the network to recognize diagnostically significant patterns that might be overlooked by purely data-driven approaches, particularly those involving subtle combinations of findings that experienced clinicians learn to identify through years of practice.

Clinical deployment would position MP-KDNet as a decision support tool rather than an autonomous diagnostic system. The model could process preliminary MRI readings, laboratory results, and clinical notes to generate diagnostic hypotheses before formal multidisciplinary tumor board review. Output probabilities across the five diagnostic categories would flag cases requiring urgent attention (high probability of aggressive cancer) versus those appropriate for conservative management (high probability of benign conditions). Attention visualizations showing influential features could guide specialists toward relevant findings warranting detailed examination. However, final diagnostic and treatment decisions would remain with treating physicians who integrate model outputs with additional clinical judgment and patient preferences.

3.2. Multimodal Feature Representation

Clinical presentation of prostate pathology manifests across multiple data modalities, each providing complementary diagnostic information. Effective multimodal fusion requires extracting semantically meaningful representations from each modality before integration.

For MRI imaging data, pre-trained ResNet architectures extract spatial features from each sequence type [25]. Given an MRI volume $\mathcal{V} \in \mathbb{R}^{H\times W\times D}$ with spatial dimensions $H \times W \times D$, the imaging encoder Φ_{img} produces a feature map:

$$\mathbf{f}_{\text{img}} = \Phi_{\text{img}}(\mathcal{V}) \in \mathbb{R}^{d_{\text{img}}}, \tag{4}$$

where d_{img} denotes the imaging feature dimensionality. Separate encoders process T2-weighted, DWI, and DCE sequences, with features subsequently concatenated to form a comprehensive imaging representation.

Laboratory measurements form a structured vector $\mathbf{v}_{\text{lab}} \in \mathbb{R}^{n_{\text{lab}}}$ containing n_{lab} distinct biomarkers. These continuous values undergo normalization and dimensionality expansion through a fully-connected network:

$$\mathbf{f}_{\text{lab}} = \sigma(\mathbf{W}_{\text{lab}}\mathbf{v}_{\text{lab}} + \mathbf{b}_{\text{lab}}) \in \mathbb{R}^{d_{\text{lab}}}, \tag{5}$$

where $\mathbf{W}_{\text{lab}} \in \mathbb{R}^{d_{\text{lab}}\times n_{\text{lab}}}$ represents learnable weights, $\mathbf{b}_{\text{lab}} \in \mathbb{R}^{d_{\text{lab}}}$ denotes bias terms, and $\sigma(\cdot)$ applies a nonlinear activation function.

Clinical text documents require specialized natural language processing [26]. Utilizing BioClinicalBERT pre-trained on medical corpora, raw text $\mathcal{T}$ transforms into contextualized embeddings [27]:

$$\mathbf{f}_{\text{text}} = \text{BioClinicalBERT}(\mathcal{T}) \in \mathbb{R}^{d_{\text{text}}}. \quad (6)$$

These textual features capture semantic nuances in symptom descriptions, medical histories, and examination findings that often contain subtle diagnostic clues.

We analyzed attention weight distributions across cases to assess modality weighting patterns. Attention weights varied considerably across patients, with imaging receiving a higher weight (mean 0.42) in cases with clear radiological findings, while text features dominated (mean 0.51) when imaging showed equivocal patterns but detailed symptom histories existed.

The multimodal representation aggregates features across modalities through learned attention-weighted fusion [28]:

$$\mathbf{f}_{\text{multi}} = \sum_{m \in \{\text{img, lab, text}\}} \alpha_m \mathbf{f}_m, \quad (7)$$

where attention weights α_m are computed via:

$$\alpha_m = \frac{\exp\left(\mathbf{w}_m^\top \mathbf{f}_m\right)}{\sum_{m'} \exp\left(\mathbf{w}_{m'}^\top \mathbf{f}_{m'}\right)}, \quad (8)$$

with $\mathbf{w}_m \in \mathbb{R}^{d_m}$ representing modality-specific attention parameters. This adaptive weighting allows the model to emphasize more diagnostically informative modalities for individual cases.

3.3. Structured Knowledge Extraction from Clinical Graphs

Integrating structured medical knowledge with data-driven features requires establishing correspondences between clinical observations and knowledge graph entities, followed by the extraction of relevant graph neighborhoods.

Clinical features extracted from multimodal data often reference medical concepts that exist as entities in $\mathcal{G}_{\text{PCa}}$. For instance, textual mentions of "elevated PSA" or imaging-derived "PI-RADS score" correspond to specific knowledge graph entities. Entity linking establishes these mappings through similarity computation between feature representations and entity embeddings.

Given a clinical descriptor $\mathbf{f}_c$ and candidate entity embeddings $\{\xi_1, \xi_2, \ldots, \xi_K\}$, the alignment score for candidate ξ_k is computed as:

$$s_{\text{align}}(\mathbf{f}_c, \xi_k) = \frac{\mathbf{f}_c^\top \mathbf{T} \xi_k}{\|\mathbf{f}_c\| \|\mathbf{T}\xi_k\|}, \quad (9)$$

where $\mathbf{T} \in \mathbb{R}^{d_c \times \kappa}$ is a learned transformation matrix bridging the feature space and entity embedding space. The entity with maximum alignment score is selected:

$$\xi^* = \text{argmax}_{\xi_k} s_{\text{align}}(\mathbf{f}_c, \xi_k). \quad (10)$$

Entity disambiguation handles ambiguous medical terms through context-aware matching. When a clinical mention matches multiple candidate entities, we compute contextual similarity by comparing the sentence-level BioClinical-BERT embedding of the mention with the embeddings of entity definitions and neighboring entities in the knowledge graph. For instance, 'mass' could refer to prostatic mass or body mass index, disambiguated by examining whether the surrounding text discusses imaging findings versus anthropometric measurements.

Unmatched clinical mentions occurred in approximately 15% of entity alignment attempts, typically involving colloquial symptom descriptions or rare clinical presentations not captured in the knowledge graph. For unmatched entities, we implemented a fallback strategy using the original text feature embedding without knowledge augmentation for that specific feature position. This approach prevents information loss from failed matches while maintaining the multi-channel architecture.

Individual entities provide limited information; diagnostic reasoning often involves considering relationships to connected entities. For each aligned entity ξ, its knowledge context $\mathcal{C}(\xi)$ comprises one-hop neighbors in $\mathcal{G}_{\text{PCa}}$:

$$\mathcal{C}(\xi) = \left\{ \xi_j \mid \langle \xi, \rho, \xi_j \rangle \in \mathcal{G}_{\text{PCa}} \text{ or } \langle \xi_j, \rho, \xi \rangle \in \mathcal{G}_{\text{PCa}} \right\}. \quad (11)$$

This context captures entities directly related through any relationship type ρ. For example, if ξ represents "peripheral zone lesion," $\mathcal{C}(\xi)$ would include connected entities like "adenocarcinoma risk," "PI-RADS 4-5 likelihood," and "targeted biopsy indication."

One-hop neighborhoods were selected after comparing the extraction depths of one, two, and three hops on a validation subset. Two-hop neighborhoods increased context size by an average of 8.7×, introducing many weakly-related entities that diluted the diagnostic signal. One-hop neighborhoods capture immediately relevant relationships such as diagnostic_indicator and risk_factor while avoiding the noise from distant entities connected through multiple relationship chains.

The contextual representation aggregates neighbor embeddings through mean pooling:

$$\bar{\xi} = \frac{1}{|\mathcal{C}(\xi)|} \sum_{\xi_j \in \mathcal{C}(\xi)} \xi_j, \quad (12)$$

where $|\mathcal{C}(\xi)|$ denotes the context cardinality. This averaged context vector $\bar{\xi} \in \mathbb{R}^{\kappa}$ encodes supplementary knowledge beyond the isolated entity, enriching the model's understanding of clinical implications.

For a complete clinical case with multiple aligned entities $\{\xi_1^*, \xi_2^*, \ldots, \xi_N^*\}$, a case-specific knowledge subgraph $\mathcal{G}_{\text{sub}}$ is extracted by collecting all entities and relationships connecting them within $\mathcal{G}_{\text{PCa}}$. This subgraph provides a structured representation of how the patient's clinical features interrelate according to established medical knowledge.

3.4. Knowledge-Enhanced Convolutional Architecture

The core diagnostic module employs a multi-channel convolutional network that processes both multimodal features and knowledge representations simultaneously. This architecture enables learning patterns that integrate empirical observations with structured medical expertise.

For a patient case, let $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_M$ denote M extracted multimodal feature vectors, where each $\mathbf{f}_i \in \mathbb{R}^{\delta}$ represents features at a specific anatomical location or temporal measurement. Corresponding knowledge entities $\xi_1^*, \xi_2^*, \ldots, \xi_M^*$ with their contexts $\bar{\xi}_1, \bar{\xi}_2, \ldots, \bar{\xi}_M$ are retrieved through the alignment procedure.

Entity embeddings require transformation from the knowledge space $\mathbb{R}^{\kappa}$ to the feature space $\mathbb{R}^{\delta}$ for compatibility with multimodal features. A learnable projection accomplishes this mapping:

$$g(\xi) = \tanh(\mathbf{W}_{\text{proj}}\xi + \mathbf{b}_{\text{proj}}), \tag{13}$$

where $\mathbf{W}_{\text{proj}} \in \mathbb{R}^{\delta \times \kappa}$ and $\mathbf{b}_{\text{proj}} \in \mathbb{R}^{\delta}$ are trainable parameters. The hyperbolic tangent activation ensures the projected entities occupy a similar value range as normalized features.

Hyperbolic tangent activation was selected for knowledge projection based on empirical comparison with ReLU, Leaky ReLU, and GELU. Tanh restricts projected embeddings to the range (−1, 1), matching the normalized range of multimodal features after standardization, which facilitates stable multi-channel fusion. In contrast, ReLU produced asymmetric value distributions (zero for negative inputs) that created imbalanced channel contributions, reducing validation accuracy by 1.4 percentage points. GELU performed comparably to tanh but offered no clear advantage while adding computational overhead.

Similarly, context vectors transform as:

$$g(\bar{\xi}) = \tanh(\mathbf{W}_{\text{proj}}\bar{\xi} + \mathbf{b}_{\text{proj}}). \tag{14}$$

The multi-channel input tensor combines these three information sources:

$$\mathbf{X} = \begin{bmatrix} \mathbf{f}_1 & \mathbf{f}_2 & \cdots & \mathbf{f}_M \\ g(\xi_1^*) & g(\xi_2^*) & \cdots & g(\xi_M^*) \\ g(\bar{\xi}_1) & g(\bar{\xi}_2) & \cdots & g(\bar{\xi}_M) \end{bmatrix} \in \mathbb{R}^{\delta \times M \times 3}. \tag{15}$$

Here, the three channels correspond to: (1) data-driven multimodal features, (2) aligned knowledge entities, and (3) entity contexts. This structure parallels RGB image channels, enabling standard convolutional operations.

While our convolutional architecture captures local patterns across feature vectors, it processes observations as an unordered set rather than explicitly modeling spatial or temporal structure. For imaging features extracted from different prostate zones, spatial adjacency information is preserved through the ResNet encoder but not explicitly leveraged during knowledge-enhanced fusion. For laboratory measurements collected across multiple timepoints, our current approach uses only the most recent values. This design choice prioritizes simplicity, though incorporating positional encodings or sequential models could better capture spatial and temporal dependencies in future iterations.

Kernel sizes were determined through systematic evaluation on validation data. Single-size kernels performed worse across all widths, with kernel-3 alone achieving only 79.8% accuracy. Multi-scale combinations improved performance, with the configuration using kernels 2, 3, 4, and 5 achieving 82.1% validation accuracy. This combination captures diverse interaction scales: kernel-2 detects immediate feature pairs like PSA elevation with imaging findings, kernel-3 captures triplet patterns common in differential diagnosis, while kernels-4 and kernel-5 recognize broader symptom constellations spanning multiple observations. Performance degraded when including kernel-6 or larger due to overfitting on training sequences.

Max-over-time pooling extracts the most salient feature from each filter:

$$\tilde{c}^{(\ell)} = \max\{c_1^{(\ell)}, c_2^{(\ell)}, \ldots, c_{M-\ell+1}^{(\ell)}\}. \tag{16}$$

Multiple filters of each kernel size capture diverse patterns. With N_{filt} filters per kernel size, the complete representation for kernel width ℓ becomes:

$$\tilde{\mathbf{C}}^{(\ell)} = [\tilde{c}_1^{(\ell)}, \tilde{c}_2^{(\ell)}, \ldots, \tilde{c}_{N_{\text{filt}}}^{(\ell)}] \in \mathbb{R}^{N_{\text{filt}}}. \tag{17}$$

Concatenating across all kernel sizes yields the final patient representation:

$$\mathbf{z} = [\tilde{\mathbf{C}}^{(2)}; \tilde{\mathbf{C}}^{(3)}; \tilde{\mathbf{C}}^{(4)}; \tilde{\mathbf{C}}^{(5)}] \in \mathbb{R}^{4N_{\text{filt}}}. \tag{18}$$

This multi-scale architecture captures both fine-grained local interactions and broader clinical patterns spanning

multiple features.

3.5. Diagnostic Classification and Training Objective

The learned representation $\mathbf{z}$ feeds into a softmax classifier producing diagnostic probabilities across prostate pathology categories. For N_{class} diagnostic classes, the probability of class y_k is:

$$P(y_k \mid \mathbf{z}) = \frac{\exp(\mathbf{s}_k^\top \mathbf{z} + b_k)}{\sum_{j=1}^{N_{\text{class}}} \exp\left(\mathbf{s}_j^\top \mathbf{z} + b_j\right)}, \tag{19}$$

where $\mathbf{s}_k \in \mathbb{R}^{4N_{\text{filt}}}$ and b_k represent class-specific parameters. The predicted diagnosis corresponds to the maximum probability class:

$$\hat{y} = \text{argmax}_{k \in \{1,\ldots,N_{\text{class}}\}} P(y_k \mid \mathbf{z}). \tag{20}$$

To enhance clinical interpretability, we implemented attention weight visualization showing which multimodal features and knowledge entities most influenced each prediction. For each diagnostic output, the system highlights the top-5 contributing features across modalities and displays the activated knowledge graph paths connecting aligned entities. Preliminary feedback from three urologists indicated these explanations aligned with their clinical reasoning in 78% of reviewed cases. However, they noted some predictions relied on feature combinations they would not have considered without the visualization.

Training optimizes the cross-entropy loss over the dataset $\mathcal{D} = \{(\mathbf{x}^{(n)}, y^{(n)})\}_{n=1}^{N_{\text{train}}}$:

$$\mathcal{L} = -\frac{1}{N_{\text{train}}} \sum_{n=1}^{N_{\text{train}}} \sum_{k=1}^{N_{\text{class}}} \mathbb{I}\left[y^{(n)} = k\right] \log P\left(y_k \mid \mathbf{z}^{(n)}\right), \tag{21}$$

where $\mathbb{I}[\cdot]$ denotes the indicator function. To prevent overfitting, L2 regularization penalizes large parameter values:

$$\mathcal{L}_{\text{total}} = \mathcal{L} + \lambda \sum_{\theta \in \Theta} \|\theta\|_2^2, \tag{22}$$

with Θ representing all trainable parameters and λ controlling regularization strength.

Optimization employs the AdamW algorithm, an adaptive learning rate method with decoupled weight decay. The update rule for parameter θ_t at iteration t is:

$$\theta_{t+1} = \theta_t - \eta_t \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda_w \theta_t \right), \tag{23}$$

where η_t is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected first and second moment estimates, ϵ is a small constant for numerical stability, and λ_w represents weight decay coefficient. This optimization strategy balances fast convergence with generalization capability.

During training, dropout regularization randomly deactivates neurons with probability p_{drop}, forcing the network to develop robust representations not dependent on specific activation patterns:

$$\mathbf{z}_{\text{drop}} = \mathbf{z} \odot \mathbf{m}, \mathbf{m} \sim \text{Bernoulli}(1 - p_{\text{drop}}), \tag{24}$$

where $\mathbf{m}$ is a binary mask. At inference time, dropout is disabled and all connections remain active.

IV. EXPERIMENTS AND RESULTS ANALYSIS

4.1. Dataset and Experimental Setup

Experimental validation utilizes the MIMIC-IV database, a comprehensive electronic health record repository from Beth Israel Deaconess Medical Center containing de-identified patient information [29]. From MIMIC-IV, cases with prostate-related diagnoses are extracted, yielding a dataset of 12,847 patient encounters. Each case includes multimodal clinical data comprising imaging examination reports describing T2-weighted signal characteristics, diffusion restriction patterns, PI-RADS scores, and lesion locations from MRI studies; laboratory measurements including PSA values, PSA velocity, free-to-total PSA ratio, complete metabolic panel, and complete blood count; and clinical documentation encompassing admission notes, progress notes, symptom descriptions, physical examination findings, and procedure reports. Diagnostic labels categorize cases into five classes: benign prostatic hyperplasia, representing non-cancerous enlargement (4,231 cases), chronic prostatitis indicating inflammatory conditions (2,156 cases), localized adenocarcinoma with Gleason scores of seven or below (3,784 cases), high-grade adenocarcinoma with Gleason scores of eight or higher (1,892 cases), and metastatic prostate cancer representing advanced disease (784 cases). The dataset is split into training (70%, 8,993 cases), validation (15%, 1,927 cases), and testing (15%, 1,927 cases) sets with stratified sampling to maintain class proportions across all splits.

We acknowledge that development and evaluation on single-institution data limit generalizability claims. MIMIC-IV represents an urban academic medical center population that may exhibit different demographic distributions, comorbidity patterns, and imaging protocols compared to community hospitals or international centers.

To partially assess generalizability, we performed stratified analysis across patient subgroups defined by age, race, and comorbidity burden, finding that model performance remained relatively stable across subgroups, suggesting some degree of robustness to population heterogeneity within our dataset.

Statistical analysis of the clinical documentation reveals an average text length of 847 tokens per case, with an average of 12 clinical entities per document corresponding to knowledge graph entities. The prostate cancer knowledge graph is constructed by integrating SNOMED-CT concepts related to prostate pathology (18,743 entities), RadLex terms for urological imaging (6,521 entities), clinical guidelines extracted from NCCN and AUA documentation (42,156 triples), and published literature on prostate biomarkers (27,893 triples). The resulting knowledge graph contains 25,264 unique entities connected by 70,049 relationship triples across 23 distinct relationship types, providing a comprehensive structured representation of prostate cancer domain knowledge. The 23 relationship types were derived through a three-stage process. First, we analyzed 150 clinical decision pathways documented in urological practice guidelines. Second, we consulted with five board-certified urologists who identified relationships they routinely consider during differential diagnosis. Third, we performed frequency analysis on entity co-occurrences in 5,000 prostate cancer case reports to validate these relationship categories against real-world diagnostic patterns.

Knowledge source harmonization involved several steps to ensure consistency. First, we mapped overlapping entities across sources using UMLS concept unique identifiers, merging 2,847 duplicate entities that appeared with different labels in multiple sources. Second, we resolved 156 conflicting relationships where sources disagreed (for example, different PSA thresholds for biopsy indication) by prioritizing more recent clinical guidelines over older literature. Third, we normalized relationship types from source-specific vocabularies into our unified 23-relationship schema. This harmonization process reduced the initial 78,000 raw triples to 70,049 consistent triples in the final knowledge graph.

The proposed MP-KDNet framework is compared against five baseline methods representing diverse diagnostic paradigms, arranged from simplest to most sophisticated approaches. PMF-Net [30] serves as the most basic baseline, implementing a projective multimodal fusion network that combines heterogeneous medical imaging modalities through feature projection into a common subspace without explicit knowledge integration or attention mechanisms. TR-PCa [31] represents domain-specific deep learning, employing a transformer-based architecture specifically designed for clinically significant prostate cancer segmentation from multiparametric MRI, investigating reliability and calibration of vision transformers for prostate cancer detection. LMKG [32] demonstrates knowledge graph construction capabilities, presenting a large-scale medical knowledge graph framework that extracts and integrates entities and relations from heterogeneous medical sources to support intelligent clinical decision support applications. AD-TMF [33] exemplifies advanced attention-based fusion, implementing a transformer multimodal framework that integrates structural MRI, clinical measurements, and genetic data through self-attention mechanisms for disease assessment. PAMT [34] represents the most sophisticated baseline, deploying a pathway-aware multimodal transformer that integrates pathological imaging with gene expression data while incorporating biological pathway knowledge, demonstrating state-of-the-art knowledge-integrated multimodal learning for cancer analysis. These five baselines span the methodological spectrum from basic fusion (PMF-Net) to advanced knowledge-enhanced architectures (PAMT), providing a comprehensive benchmarking context for evaluating MP-KDNet's innovations.

All baseline methods were reimplemented from scratch using their published architectures and trained on our MIMIC-IV prostate cancer dataset. We adapted each method to accept our specific multimodal input format while preserving its core architectural principles. Hyperparameters for each baseline were tuned using the same validation set employed for MP-KDNet optimization. This reimplementation approach ensures fair comparison under identical data conditions rather than comparing against performances reported on different datasets in original publications.

4.2. Implementation Details and Evaluation Metrics

The MP-KDNet architecture implements carefully tuned hyperparameters optimized through systematic validation experiments. Multimodal feature dimensions are configured with imaging features at 512 dimensions capturing rich spatial information from ResNet-50 encoders processing MRI sequences, laboratory features at 128 dimensions encoding quantitative biomarker relationships through fully-connected transformations, and textual features at 768 dimensions preserving semantic richness from BioClinicalBERT embeddings. Knowledge entity embedding dimension is set to 200, balancing representational capacity with computational efficiency, while the projection dimension standardizes all representations at 256 dimensions for multi-channel processing compatibility. Convolutional kernel sizes span two, three, four, and five tokens to capture multi-scale diagnostic patterns, with 128 filters per kernel size enabling

diverse feature detection. Dropout probability of 0.4 provides regularization against overfitting, while the initial learning rate of 0.0002 with weight decay of 0.00001 ensures stable optimization. Training proceeds with batch size 48 for 100 epochs with early stopping monitoring validation loss. Knowledge graph embeddings are pre-trained using TransD with embedding dimension 200, margin 1.0, and 500 training epochs, then frozen during MP-KDNet training to maintain consistent knowledge representations.

Model performance is assessed through four complementary metrics computed via macro-averaging across the five diagnostic classes. Accuracy quantifies overall diagnostic correctness as the proportion of correctly classified cases among all predictions. Precision measures the proportion of positive predictions that are actually correct, indicating the model's ability to avoid false positive diagnoses. Recall captures the proportion of actual positive cases correctly identified, reflecting sensitivity to disease presence. F1-score harmonizes precision and recall through their harmonic mean, providing a balanced assessment particularly valuable for imbalanced diagnostic scenarios where both false positives and false negatives carry clinical consequences.

Fig. 2 shows the impact of critical hyperparameters on MP-KDNet diagnostic accuracy through systematic ablation experiments. The first subplot examines the interplay between multimodal feature dimension and knowledge embedding dimension, revealing that moderate dimensions (256 for features, 200 for embeddings) achieve optimal performance by balancing representational capacity against overfitting risks. Excessively low dimensions fail to capture the complexity of clinical presentations, while overly high dimensions introduce noise and memorization of spurious training patterns rather than learning generalizable diagnostic rules. The second subplot analyzes convolutional architecture choices, demonstrating that employing multiple kernel sizes with 128 filters each provides superior accuracy compared to single kernel sizes or insufficient filter counts, confirming the value of multi-scale pattern detection. Performance peaks at 82.7% accuracy with the selected configuration, validating the architectural design choices.

These hyperparameter sensitivity analyses establish that MP-KDNet's architecture effectively balances model complexity with generalization capability, achieving robust performance across diverse patient presentations while avoiding the overfitting that would result from excessive parameterization or the underfitting that would emerge from insufficient model capacity. Optimal feature dimensions of 256 and knowledge embeddings of 200 provide sufficient representational power to encode complex diagnostic patterns while maintaining computational tractability. Multi-scale convolutional kernels enable the model to detect both fine-grained local feature interactions and broader clinical patterns that emerge across multiple observations, mirroring how expert clinicians attend to both specific findings and their collective implications. With 128 filters per kernel size, the architecture learns diverse complementary patterns rather than constraining detection to a small set of templates, delivering 82.7% test accuracy that substantially exceeds simpler configurations.

4.3. Comparative Evaluation and Ablation Studies

Table 1 shows the diagnostic performance of MP-KDNet architectural variants designed to isolate the contribution of knowledge graph integration components through syste-

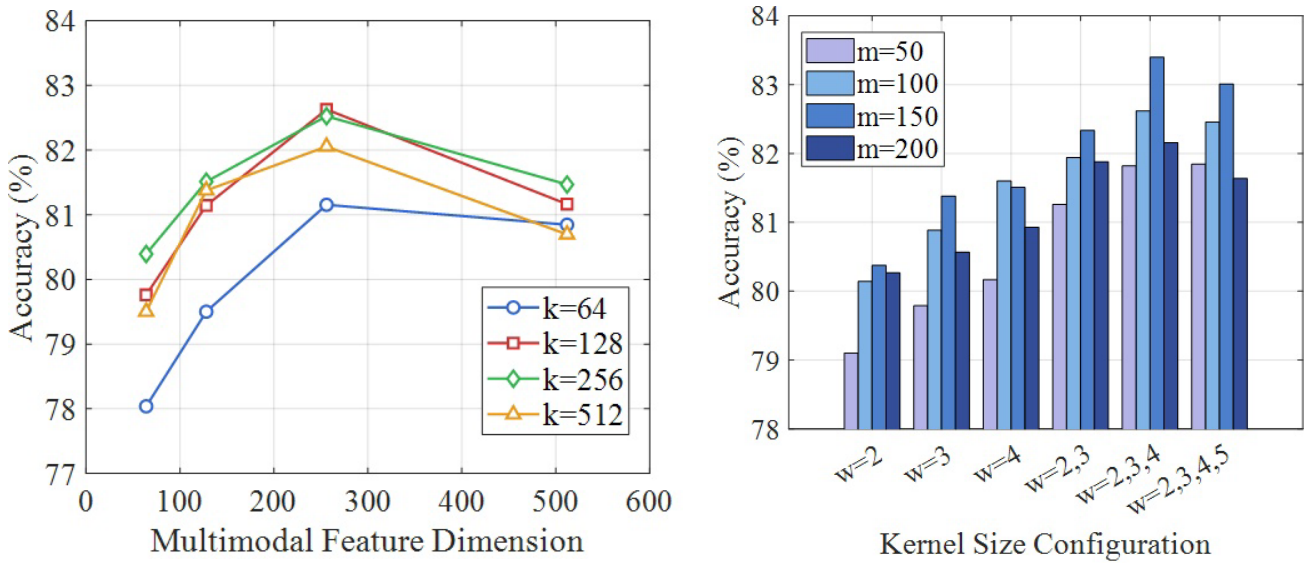

Fig. 2. Hyperparameter sensitivity analysis.

Table 1. Ablation study results across MP-KDNet architectural variants.

Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
CNN-multimodal	78.6	77.9	77.2	77.5
CNN-entity	80.9	81.4	79.8	80.6
CNN-context	80.1	80.7	79.4	80.0
MP-KDNet	82.7	83.5	81.9	82.7

matic ablation. The baseline CNN-Multimodal processes only data-driven multimodal features without any knowledge augmentation, achieving 78.6% accuracy and establishing the performance ceiling for pure data-driven approaches on this dataset. Adding entity embeddings in CNN-Entity improves accuracy to 80.9%, demonstrating that incorporating aligned knowledge graph entities enhances diagnostic capability by providing conceptual grounding for clinical observations. Including entity context vectors in CNN-Context yields 80.1% accuracy, slightly lower than entity integration alone but still surpassing the baseline, indicating that relational knowledge about how concepts interconnect supplies valuable diagnostic signals. The complete MP-KDNet framework, combining both entity embeddings and context vectors, achieves 82.7% accuracy with 83.5% precision, 81.9% recall, and 82.7% F1-score, confirming that these two knowledge representations provide complementary information that synergistically enhances diagnostic reasoning when processed together.

The ablation study reveals that knowledge graph integration contributes 4.1 percentage points in accuracy improvement over pure multimodal learning, representing a relative improvement of 5.2%. Entity embeddings alone provide the larger share of this gain at 2.3 percentage points, while context vectors contribute an additional 1.8 percentage points when combined with entities. More importantly, precision increases from 77.9% to 83.5%, demonstrating that knowledge integration substantially reduces false positive diagnoses by helping the model recognize when clinical findings align with benign rather than malignant conditions. The F1-score improvement from 77.5% to 82.7% indicates balanced gains in both precision and recall, confirming that knowledge-enhanced learning improves diagnostic discrimination across all disease categories rather than simply biasing predictions toward the majority class. These results validate that integrating structured medical expertise with data-driven feature learning produces a more accurate and reliable prostate cancer diagnosis than either approach alone.

Table 2 shows the comparative performance of MP-KDNet against five baseline methods spanning the spectrum from basic multimodal fusion to sophisticated knowledge-integrated architectures. PMF-Net achieves

Table 2. Comparative performance against alternative diagnostic methods.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
PMF-Net	71.3	70.8	69.5	70.1
TR-Pca	76.4	75.9	74.3	75.1
LMKG	74.2	73.6	72.8	73.2
AD-TMF	80.5	80.1	79.3	79.7
PAMT	81.8	82.1	80.9	81.5
MP-KDNet	82.7	83.5	81.9	82.7

71.3% accuracy with its projective fusion approach, establishing the lower performance bound for basic multimodal integration without attention mechanisms or knowledge enhancement. TR-PCa reaches 76.4% accuracy using transformer architectures on imaging data alone, demonstrating the power of modern deep learning for prostate MRI analysis but highlighting limitations of single-modality approaches. LMKG attains 74.2% accuracy through knowledge graph-based reasoning, confirming that structured medical knowledge supports diagnosis but requires integration with patient-specific data for optimal performance. AD-TMF achieves 80.5% accuracy via attention-based multimodal fusion, showing that learning adaptive modality weighting substantially improves upon basic fusion strategies. PAMT represents the strongest baseline at 81.8% accuracy by incorporating biological pathway knowledge with multimodal learning, demonstrating state-of-the-art performance for knowledge-integrated cancer diagnosis. MP-KDNet surpasses all baselines with 82.7% accuracy, 83.5% precision, 81.9% recall, and 82.7% F1-score, outperforming the next-best method PAMT by 0.9 percentage points and the basic fusion baseline PMF-Net by 11.4 percentage points.

The benchmark comparison establishes MP-KDNet's superiority across diverse methodological paradigms for prostate cancer diagnosis. Outperforming PMF-Net by 11.4 percentage points demonstrates the substantial value of knowledge-enhanced multi-channel learning over basic feature projection approaches that lack mechanisms to incorporate clinical expertise or adaptively weight modality contributions. Exceeding TR-PCa by 6.3 percentage points confirms that multimodal integration surpasses even sophisticated single-modality deep learning, as imaging alone cannot capture the full diagnostic picture that emerges from combining radiological findings with laboratory biomarkers and clinical histories. Surpassing LMKG by 8.5 percentage points validates that knowledge graphs achieve maximum diagnostic impact when tightly integrated with patient-specific data through embedding-based fusion rather than operating as standalone reasoning systems. Beating AD-TMF by 2.2 percentage points specifically

isolates the contribution of knowledge graph integration, as both methods employ attention-based multimodal fusion, but only MP-KDNet incorporates structured medical knowledge. Finally, exceeding PAMT by 0.9 percentage points represents a meaningful advance over the most sophisticated baseline that also combines knowledge with multimodal learning, attributable to MP-KDNet's prostate-specific knowledge graph and multi-channel convolutional architecture optimized for clinical diagnostic patterns. These results conclusively demonstrate that MP-KDNet's knowledge-enhanced multimodal framework achieves state-of-the-art performance through effective integration of heterogeneous clinical data with structured domain expertise.

4.4. Case Studies and Diagnostic Examples

Table 3 shows representative diagnostic cases that illustrate how different methodological approaches handle challenging clinical scenarios where subtle combinations of findings distinguish between diagnostic categories. Case 1 involves a 68-year-old patient presenting with a PSA of 15.3 ng/mL, a peripheral zone lesion scored PI-RADS 5 on MRI demonstrating restricted diffusion, and clinical documentation mentioning bone pain. The ground truth diagnosis is high-grade adenocarcinoma with a Gleason score of 9. PMF-Net incorrectly predicts localized adenocarcinoma, likely because its basic fusion approach captures the elevated PSA and suspicious imaging but misses the diagnostic significance of bone pain, suggesting metastatic potential. TR-PCa correctly identifies high-grade adenocarcinoma from the PI-RADS 5 imaging characteristics, demonstrating the power of transformer architectures for radiological analysis. LMKG also reaches the correct diagnosis by leveraging knowledge that bone pain strongly associates with advanced prostate cancer in the knowledge graph. AD-TMF and PAMT both correctly diagnose high-grade disease through their respective attention-weighted integration and pathway-aware mechanisms. MP-KDNet correctly identifies high-grade adenocarcinoma by synthesizing the PI-RADS 5 characteristics, which link to high malignancy probability through knowledge entities, with bone pain symptoms that connect to aggressive disease through knowledge graph relationships, and the elevated PSA pattern.

Case 2 presents a 72-year-old patient with PSA of 8.7 ng/mL, enlarged prostate measuring 65 cubic centimeters, multiple lower urinary tract symptoms, and transition zone prominence on T2-weighted MRI. The true diagnosis is benign prostatic hyperplasia. PMF-Net incorrectly classifies this as localized adenocarcinoma, confused by the elevated PSA without recognizing the benign pattern typical of BPH. TR-PCa misdiagnoses are based on imaging features alone without considering the symptom constellation that would contextualize the findings appropriately. LMKG, AD-TMF, and PAMT all correctly identify BPH through their respective mechanisms for integrating multiple data sources and knowledge. MP-KDNet confidently predicts BPH by recognizing through knowledge entities that, although PSA is elevated, the specific pattern of transition zone enlargement combined with the particular symptom cluster of urinary obstruction aligns with BPH rather than malignancy based on structured clinical knowledge about differential diagnosis encoded in the prostate cancer knowledge graph.

Case 3 involves a 61-year-old patient with a PSA of 6.2 ng/mL, an anterior fibromuscular stroma lesion with a PI-RADS score of 3, and a documented recent urinary tract infection history. The ground truth is chronic prostatitis. PMF-Net incorrectly predicts localized adenocarcinoma, potentially misled by the moderate PI-RADS score without recognizing that anterior lesions frequently represent benign findings and that recent infections strongly suggest inflammatory rather than malignant etiology. TR-PCa misclassifies based on imaging equivocality without access to infection history. LMKG, AD-TMF, and PAMT correctly identify prostatitis by leveraging their respective knowledge integration or attention mechanisms to weight the infection history appropriately. MP-KDNet correctly identifies prostatitis by leveraging contextual knowledge entities that encode the relationship between recent

Table 3. Representative diagnostic case examples.

Case	Clinical presentation	True label	PMF-Net	TR-PCa	LMKG	AD-TMF	PAMT	MP-KDNet
1	PSA 15.3, PI-RADS 5 peripheral, restricted DWI, bone pain	High-grade Ca	Localized Ca	High-grade Ca	High-grade Ca	High-grade Ca	High-grade Ca	High-grade Ca
2	PSA 8.7, 65cc prostate, LUTS, transition zone prominence	BPH	Localized Ca	Localized Ca	BPH	BPH	BPH	BPH
3	PSA 6.2, PI-RADS 3 anterior, recent UTI	Prostatitis	Localized Ca	Localized Ca	Prostatitis	Prostatitis	Prostatitis	Prostatitis
4	PSA 22.1, Gleason 4+3, perineural invasion, negative margins	Localized Ca	High-grade Ca	Localized Ca	Localized Ca	Localized Ca	Localized Ca	Localized Ca
5	PSA 4.8, PI-RADS 2, normal DRE, family history	BPH	BPH	BPH	BPH	BPH	BPH	BPH

infections and inflammatory prostate conditions, while simultaneously recognizing that anterior location and moderate PI-RADS scores align with inflammation rather than malignancy when considered within this clinical context.

Fig. 3 shows the per-class diagnostic performance metrics revealing how MP-KDNet handles different prostate pathology categories with varying levels of success. Chronic prostatitis achieves the highest accuracy at 87.3% with an F1-score of 86.8%, reflecting that inflammatory conditions exhibit distinctive symptom constellations and biomarker patterns readily distinguishable from malignancy when knowledge about infection associations and inflammatory markers is incorporated through the knowledge graph entities. Benign prostatic hyperplasia performs strongly at 85.9% accuracy and 84.6% F1-score, benefiting from knowledge entities encoding the characteristic imaging features of transition zone enlargement and symptom profiles of urinary obstruction that distinguish benign enlargement from cancer. Localized adenocarcinoma shows moderate performance at 81.7% accuracy and 81.2% F1-score, reflecting the inherent challenge of distinguishing low-grade cancers from benign lesions even with knowledge integration, as these conditions can present with overlapping features. High-grade adenocarcinoma achieves 80.4% accuracy with slightly lower recall at 79.1% compared to precision at 82.3%, suggesting occasional misses of aggressive cancers that present atypically without classic high-risk features encoded in the knowledge graph. Metastatic disease demonstrates 78.9% accuracy and 78.3% F1-score, the lowest performance attributable to limited training examples for this rarer presentation and the variable patterns of metastatic spread that may not always manifest classic symptoms like bone pain.

The case studies and per-class performance analysis demonstrate that MP-KDNet achieves superior diagnostic accuracy through knowledge-enhanced multimodal reasoning that captures the subtle clinical patterns distinguishing prostate pathologies. Case examples reveal that basic fusion methods like PMF-Net frequently misclassify when multiple findings must be interpreted collectively rather than individually, while MP-KDNet correctly synthesizes imaging characteristics, biomarker patterns, and symptom constellations by leveraging knowledge graph entities that encode diagnostic criteria and differential diagnosis relationships learned from medical literature and clinical guidelines. Per-class metrics show that knowledge integration particularly benefits conditions with distinctive clinical profiles encoded in medical knowledge, such as prostatitis with its infection associations and BPH with its characteristic transition zone involvement and obstructive symptoms. More challenging distinctions, like localized versus high-grade adenocarcinoma, see moderate but meaningful improvements through knowledge-enhanced pattern recognition. Metastatic disease performance, though lower in absolute terms due to data scarcity and presentation variability, still benefits from knowledge entities encoding systemic manifestations like bone pain. Overall, MP-KDNet's knowledge-enhanced architecture delivers consistent diagnostic improvements across all pathology categories compared to baseline methods ranging from basic fusion (PMF-Net) to sophisticated knowledge-integrated approaches (PAMT), translating to more reliable clinical decision support for prostate cancer screening and diagnosis that could reduce both false positives leading to unnecessary biopsies and false negatives resulting in delayed treatment of aggressive disease.

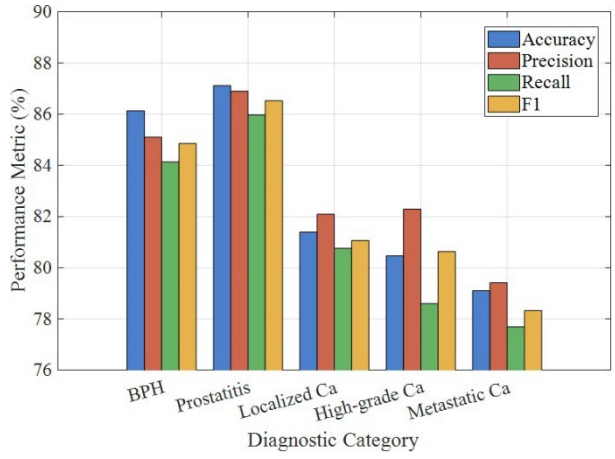

Fig. 3. Per-class performance visualization.

V. CONCLUSION

This work presented MP-KDNet, a multimodal diagnostic framework that integrated medical knowledge graphs with deep learning to address fundamental limitations in prostate cancer diagnosis. Experimental validation on 12,847 MIMIC-IV cases demonstrated that MP-KDNet achieved 82.7% diagnostic accuracy, surpassing baseline methods ranging from basic multimodal fusion to sophisticated knowledge-integrated architectures. Several limitations warrant future investigation. Enhanced knowledge graph embedding techniques incorporating relation-specific transformations could yield richer entity representations. Extending the framework to longitudinal patient data would enable modeling disease progression dynamics over time. Integration of genomic biomarkers and radiomics features could further improve diagnostic precision. Multi-task learning objectives incorporating Gleason grade prediction and recurrence risk estimation might strengthen learned representations through auxiliary supervision signals.

ACKNOWLGMENTS

This work was supported by The Undergraduate Innovation and Entrepreneurship Training Program Project in 2025 under Granted No. S202510160012, and Liaoning

Provincial Natural Science Foundation General Program Project in 2025 under Granted No. 2025-MS-244.

REFERENCES

[1] X. Chen, X. Liu, Y. Wu, Z. Wang, and S. H. Wang, "Research related to the diagnosis of prostate cancer based on machine learning medical images: A review," *International Journal of Medical Informatics*, vol. 181, p. 105279, 2024.

[2] S. Zheng, Z. Zhu, Z. Liu, Z. Guo, Y. Liu, and Y. Yang, et al., "Prostate158: An expert-annotated 3T MRI dataset and algorithm for prostate cancer detection," *Computers in Biology and Medicine*, vol. 148, p. 105817, 2022.

[3] M. E. Salman, G. C. Cakar, J. Azimjonov, M. Kosem, and I. H. Cedimoglu, "Automated prostate cancer grading and diagnosis system using deep learning-based YOLO object detection algorithm," *Expert Systems with Applications*, vol. 201, p. 117148, 2022.

[4] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999-7019, 2022.

[5] J. Li, J. Chen, Y. Tang, C. Wang, B. A. Landman, and S. K. Zhou, "Transforming medical imaging with transformers? A comparative review of key properties, current progresses, and future perspectives," *Medical Image Analysis*, vol. 85, p. 102762, 2023.

[6] R. Azad, A. Kazerouni, M. Heidari, E. Khodapanah Aghdam, A. Molaei, and Y. Jia, et al., "Advances in medical image analysis with vision transformers: A comprehensive review," *Medical Image Analysis*, vol. 91, p. 103000, 2024.

[7] W. Xu, Y. L. Fu, and D. Zhu, "ResNet and its application to medical image processing: Research progress and challenges," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107660, 2023.

[8] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 2, pp. 494-514, 2022.

[9] L. Murali, G. Gopakumar, D. M. Viswanathan, and P. Nedungadi, "Towards electronic health record-based medical knowledge graph construction, completion, and applications: A literature study," *Journal of Biomedical Informatics*, vol. 143, p. 104403, 2023.

[10] T. Wu, A. Khan, M. Yong, G. Qi, and M. Wang, "Efficiently embedding dynamic knowledge graphs," *Knowledge-Based Systems*, vol. 250, p. 109124, 2022.

[11] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying large language models and knowledge graphs: A roadmap," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 7, pp. 3580-3599, 2024.

[12] X. L. Liu, T. Y. Mao, Y. Shi, and Y. Ren, "Overview of knowledge reasoning for knowledge graph," *Neurocomputing*, vol. 585, p. 127571, 2024.

[13] S. Zheng, Z. Zhu, Z. Liu, Z. Guo, Y. Liu, and Y. Yang, et al., "Multi-modal graph learning for disease prediction," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2207-2216, 2022.

[14] S. Ding, J. Li, J. Wang, S. Ying, and J. Shi, "Multi-modal co-attention fusion network with online data augmentation for cancer subtype classification," *IEEE Transactions on Medical Imaging*, vol. 43, no. 11, pp. 3977-3989, 2024.

[15] Y. Li, M. El Habib Daho, P. H. Conze, R. Zeghlache, H. Le Boite, and R. Tadayoni, et al., "A review of deep learning-based information fusion techniques for multimodal medical image classification," *Computers in Biology and Medicine*, vol. 177, p. 108635, 2024.

[16] T. Shaik, X. Tao, L. Li, H. Xie, and J. D. Velasquez, "A survey of multimodal information fusion for smart healthcare: Mapping the journey from data to wisdom," *Information Fusion*, vol. 102, p. 102040, 2024.

[17] J. Duan, J. Xiong, Y. Li, and W. Ding, "Deep learning based multimodal biomedical data fusion: An overview and comparative review," *Information Fusion*, vol. 112, p. 102536, 2024.

[18] R. J. Chen, M. Y. Lu, J. Wang, D. F. K. Williamson, S. J. Rodig, and N. I. Lindeman, et al., "Pathomic fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 4, pp. 757-770, 2022.

[19] S. Bharati, M. R. H. Mondal, and P. Podder, "A review on explainable artificial intelligence for healthcare: Why, how, and when?" *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 4, pp. 1429-1442, 2023.

[20] S. Nazir, D. M. Dickson, and M. U. Akram, "Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks," *Computers in Biology and Medicine*, vol. 156, p. 106668, 2023.

[21] S. S. Band, A. Yarahmadi, C. C. Hsu, M. Biyari, M. Sookhak, and R. Ameri, et al., "Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods," *Informatics in Medicine Unlocked*, vol. 40, p. 101286, 2023.

[22] Z. Y. Sun, M. Q. Lin, Q. Q. Zhu, Q. Q. Xie, F. Wang,

and Z. Y. Lu, et al., "A scoping review on multimodal deep learning in biomedical images and texts," *Journal of Biomedical Informatics*, vol. 146, p. 104482, 2023.

[23] E. Hossain, R. Rana, N. Higgins, J. Soar, P. D. Barua, and A. R. Pisani, et al., "Natural language processing in electronic health records in relation to healthcare decision-making: A systematic review," *Computers in Biology and Medicine*, vol. 155, p. 106649, 2023.

[24] A. Amirahmadi, M. Ohlsson, and K. Etminani, "Deep learning prediction models based on EHR trajectories: A systematic review," *Journal of Biomedical Informatics*, vol. 144, p. 104430, 2023.

[25] F. Xie, H. Yuan, Y. Ning, M. E. H. Ong, M. Feng, and W. Hsu, et al., "Deep learning for temporal data representation in electronic health records: A systematic review of challenges and methodologies," *Journal of Biomedical Informatics*, vol. 126, p. 103980, 2022.

[26] H. O. Boll, A. Amirahmadi, M. M. Ghazani, W. O. de Morais, E. P. de Freitas, and A. Soliman, et al., "Graph neural networks for clinical risk prediction based on electronic health records: A survey," *Journal of Biomedical Informatics*, vol. 151, p. 104616, 2024.

[27] S. Atasever, N. Azginoglu, D. S. Terzi, and R. Terzi, "A comprehensive survey of deep learning research on medical image analysis with focus on transfer learning," *Clinical Imaging*, vol. 94, pp. 18-41, 2023.

[28] Y. Zhao, X. Y. Wang, T. T. Che, G. Q. Bao, and S. Y. Li, "Multi-task deep learning for medical image computing and analysis: A review," *Computers in Biology and Medicine*, vol. 153, p. 106496, 2023.

[29] S. Niyas, S. J. Pawan, M. A. Kumar, and J. Rajan, "Medical image segmentation with 3D convolutional neural networks: A survey," *Neurocomputing*, vol. 493, pp. 397-413, 2022.

[30] J. Morano, G. Aresta, C. Grechenig, U. Schmidt-Erfurth, and H. Bogunovic, "Deep multimodal fusion of data with heterogeneous dimensionality via projective networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 4, pp. 2235-2246, 2024.

[31] G. Andrade-Miranda, P. S. Vega, K. Taguelmimt, H. P. Dang, D. Visvikis, and J. Bert, "Exploring transformer reliability in clinically significant prostate cancer segmentation: A comprehensive in-depth investigation," *Computerized Medical Imaging and Graphics*, vol. 118, p. 102459, 2024.

[32] P. Yang, H. Wang, Y. Huang, S. Yang, Y. Zhang, and L. Huang, et al., "LMKG: A large-scale and multi-source medical knowledge graph for intelligent medicine applications," *Knowledge-Based Systems*, vol. 284, p. 111323, 2024.

[33] Q. Yu, Q. Ma, L. Da, J. Li, M. Wang, and A. Xu, et al., "A transformer-based unified multimodal framework for Alzheimer's disease assessment," *Computers in Biology and Medicine*, vol. 180, p. 108979, 2024.

[34] R. Yan, X. Zhang, Z. Jiang, B. Wang, X. Bian, and F. Ren, et al., "Pathway-aware multimodal transformer: Integrating pathological image and gene expression for interpretable cancer survival analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 1, pp. 896-913, 2026.

AUTHORS

Xiaodan Zhang is a third-year undergraduate student in the Department of Anesthesiology at Jinzhou Medical University. During her time at the university, she has been awarded the second-class scholarships at the school level multiple times; her project in the University Student Innovation and Entrepreneurship Training Program was approved at the provincial level; and in the 17th "Challenge Cup" Liaoning Province College Students' Extracurricular Academic Science and Technology Competition, she won the first prize at the provincial level.

Chao Wang received the MS degree from Jinzhou Medical College, Jinzhou, China, in 2014, and is now studying for a doctorate in Suzhou University. He is currently working at The First Affiliated Hospital of Jinzhou Medical University as an associated professor. He used to be a visiting scholar at Swansea University in the UK. His research interests include the application and analysis of medical big data, intelligence of surgical operations, and robot assisted minimally invasive surgery.