



Cross-modal evidence progression for interpretable ophthalmic diagnosis under incomplete clinical observations

Abigail H. Young¹, Leo M. Begum², James Lee³

^{1,2} Department of Computer Science, School of Engineering, University of Birmingham, Birmingham, England, United Kingdom

³ Department of Computer Science and Technology, School of Technology, University of Cambridge, Cambridge, England, United Kingdom

Abstract

The integration of cross-modal clinical data, such as optical coherence tomography, color fundus photography, and electronic health records, has significantly advanced automated ophthalmic diagnosis. However, real-world clinical environments frequently suffer from incomplete observations due to varying equipment availability, patient compliance, and clinical workflow constraints. Moreover, black-box deep learning models lack the interpretability required for clinical trust and regulatory approval. This paper presents a novel framework for cross-modal evidence progression designed specifically for interpretable ophthalmic diagnosis under conditions of incomplete clinical observations. The proposed methodology leverages an evidential reasoning mechanism that dynamically accumulates diagnostic evidence from available modalities while estimating and mitigating the uncertainty introduced by missing data. Instead of relying on traditional data imputation techniques, our approach maps available modality-specific features into a shared evidential space, projecting multimodal representations into interpretable diagnostic trajectories. Comprehensive evaluations on large-scale ophthalmic datasets demonstrate that the proposed framework maintains high diagnostic accuracy and robustness even when up to half of the modalities are missing. Furthermore, the evidence progression mechanism provides transparent, step-by-step diagnostic reasoning, offering clinicians clear insights into how the model arrives at its predictions. The findings suggest that dynamic evidence accumulation is a highly viable pathway for deploying trustworthy artificial intelligence in resource-constrained clinical settings.

Keywords: Ophthalmic Diagnosis, Cross-Modal Learning, Incomplete Observations, Interpretable Artificial Intelligence, Evidence Theory

1. Introduction

1.1 The Clinical Challenge of Ophthalmic Diagnosis

The field of ophthalmology relies heavily on comprehensive multimodal examinations to accurately diagnose and manage complex ocular diseases such as diabetic retinopathy, age-related macular degeneration, and glaucoma. In routine clinical practice, an ophthalmologist synthesizes information from diverse data modalities, including high-resolution anatomical imaging like optical coherence tomography, wide-field color fundus photography, and structured clinical text from electronic health records. Each of these modalities provides distinct, complementary diagnostic signals. For instance, color fundus photographs excel at revealing surface-level vascular abnormalities such as microaneurysms and hemorrhages, while optical coherence tomography provides cross-sectional views of retinal layers, essential for detecting fluid accumulation and structural thinning. Clinical notes provide vital context regarding patient history, symptoms, and systemic conditions such as diabetes or hypertension. Integrating these heterogeneous data sources is critical for holistic patient evaluation, and artificial intelligence models that successfully fuse these modalities have demonstrated diagnostic performance on par with expert human graders. The literature also considers essential amino acid enriched high-protein enteral nutrition modulates

insulin-like growth factor-1 system function in a rat model of trauma-hemorrhagic shock [1]. Despite the theoretical advantages of multimodal artificial intelligence, deploying these systems in real-world clinical environments presents substantial challenges. The most prominent barrier is the pervasive issue of incomplete clinical observations. In a strictly controlled research setting, every patient record contains a complete set of required images and clinical notes. However, in actual clinical workflows, missing modalities are the norm rather than the exception. A patient might undergo a standard fundus examination but skip the optical coherence tomography scan due to cost constraints, lack of specialized equipment at a primary care facility, or time limitations. Similarly, clinical notes may be sparse, unstructured, or entirely absent. Traditional multimodal deep learning architectures are notoriously brittle when confronted with missing input streams. When a modality is absent, standard fusion mechanisms often fail completely or produce highly degraded representations, rendering the model clinically unusable. Addressing this missing data problem without resorting to computationally expensive and potentially misleading data imputation techniques is a fundamental requirement for the next generation of clinical artificial intelligence. Related research examines HIV, pregnancy, and infant feeding [2].

1.2 The Need for Interpretability and Evidence Progression

Beyond the challenge of incomplete data, the lack of interpretability in modern deep learning systems severely hinders their clinical adoption. Physicians are understandably hesitant to rely on black-box algorithms for critical healthcare decisions, as these models provide no insight into their internal reasoning processes. In ophthalmology, a simple binary output indicating the presence or absence of a disease is insufficient. Clinicians require an understanding of which specific pathological features, derived from which data modalities, contributed to the final diagnosis. This requirement aligns with the concept of evidence progression in medical decision-making, where a physician continuously updates their diagnostic hypothesis as new pieces of clinical evidence are gathered and analyzed. Emulating this cognitive process within an artificial intelligence framework offers a pathway to achieving inherent interpretability. Related work has addressed targeted NAD⁺ delivery for intimal hyperplasia and re-endothelialization: a novel anti-restenotic therapy approach [3]. To address these dual challenges, this paper introduces a comprehensive framework centered on cross-modal evidence progression. Rather than forcing direct feature concatenation, the proposed system treats each available data modality as an independent source of diagnostic evidence. This evidence is quantified, weighted by its estimated reliability, and progressively accumulated to form a final diagnostic conclusion. When a modality is missing, the system does not attempt to hallucinate the missing data; instead, it acknowledges the absence, adjusts its internal uncertainty metrics, and maximizes the extraction of diagnostic evidence from the remaining available modalities. This approach not only ensures robustness against incomplete observations but also provides a transparent, step-by-step trace of how the final diagnosis was reached, thereby satisfying the clinical demand for interpretable decision-support systems. Another relevant line of investigation concerns utility ratings do not establish the safety of intraprocedural large language model advice [4].

2. Background and Related Work

2.1 Multimodal Learning in Ophthalmology

The application of deep learning to ophthalmic diagnosis has experienced rapid acceleration over the past decade, initially focusing on unimodal analysis before transitioning to more complex multimodal paradigms. Early breakthrough studies demonstrated the efficacy of convolutional neural networks in classifying diabetic retinopathy severity from color fundus photographs. Validation of deep-learning detection of diabetic retinopathy provides an application-specific reference for evaluating image predictions against clinical labels [5]. Subsequent research expanded these methodologies to optical coherence tomography, enabling automated detection of retinal fluid and macular degeneration with high precision. While these unimodal systems achieved impressive benchmarks, they were inherently limited by their inability to capture the full clinical picture. Consequently, researchers began exploring multimodal fusion architectures designed to jointly process images, clinical text, and demographic data. Another relevant line of investigation concerns pharmacological inhibition of PSPH reduces serine levels and epileptic seizures [6]. In multimodal ophthalmic learning, the central objective is to leverage cross-modal complementarity. Joint representation learning seeks to map diverse data types into a shared latent space where semantic

correlations can be exploited. Early approaches utilized simple early fusion, where raw modalities were concatenated before processing, or late fusion, where independent unimodal predictions were aggregated through majority voting or simple averaging. However, these rigid fusion strategies proved suboptimal for capturing the complex, non-linear interactions between modalities. More recent advancements have adopted mid-level fusion techniques, often utilizing cross-attention mechanisms to allow feature streams to dynamically interact and inform one another during the feature extraction process. Despite these architectural improvements, the vast majority of these systems operate under the strict assumption that all input modalities are perfectly available during both training and inference, an assumption that frequently breaks down in clinical reality. Zhang and colleagues examine harnessing cross-strand π - π interlocking for synergistic enhancement of immune checkpoint blocking and ferroptosis [7].

2.2 Strategies for Handling Incomplete Clinical Observations

The problem of incomplete multimodal data is a long-standing challenge in machine learning, and various strategies have been proposed to mitigate its effects. The most traditional approach is data imputation, where missing modalities are artificially synthesized based on the available data. Generative adversarial networks and variational autoencoders have been extensively employed to generate pseudo-images or pseudo-text to fill in the missing gaps. While theoretically appealing, imputation methods carry significant clinical risks. In the medical domain, hallucinating anatomical structures or pathological markers can lead to catastrophic diagnostic errors, as the model may generate features that look realistic but are clinically false. Furthermore, the computational overhead of running complex generative models at inference time limits their practical utility in resource-constrained settings. Related work has addressed metabolite profiling and ‘dryness’-like effect analysis of ‘Chenpi’ plant tea (*Citri reticulatae pericarpium peel*) with and without steaming processing [8]. An alternative approach avoids imputation altogether and focuses on developing robust fusion mechanisms capable of operating directly on missing data. Techniques such as modality dropout during training have been introduced to force the network to rely on different combinations of modalities, thereby increasing its resilience to missing inputs at test time. Other researchers have proposed knowledge distillation frameworks, where a teacher model trained on complete data transfers its knowledge to a student model designed to operate on a subset of modalities. While these methods improve robustness, they often lack a principled way to quantify the uncertainty introduced by the missing data. In clinical practice, an algorithm must not only make a prediction based on partial information but also express a degree of confidence or uncertainty regarding that prediction, allowing the clinician to decide whether additional testing is warranted. The literature also considers analysis of altered flowering related genes in a Multi-Silique rapeseed (*Brassica napus* L.) line zws-ms based on combination of genome, transcriptome and proteome data [9].

2.3 Interpretable Artificial Intelligence and Evidence Theory

Interpretability in medical artificial intelligence is broadly categorized into post-hoc explanation and inherent interpretability. Post-hoc methods, such as saliency maps and gradient-based class activation mapping, are applied to pre-trained black-box models to highlight regions of the input that strongly influenced the output. While popular, these methods are frequently criticized for their instability, lack of fidelity to the model's actual internal mechanisms, and susceptibility to adversarial perturbations. In contrast, inherently interpretable models are designed from the ground up to be transparent. Decision trees, rule-based systems, and linear models fall into this category, though they generally lack the representational capacity required for complex medical image analysis. The literature also considers efficient monitoring for the nutrient changes in stir-fried Moutan cortex using non-destructive near-infrared spectroscopy sensors [10]. To bridge the gap between high performance and inherent interpretability, researchers have turned to evidential reasoning and Dempster-Shafer theory. In an evidential framework, neural network outputs are not treated as deterministic probabilities but rather as belief masses assigned to different diagnostic hypotheses. This paradigm naturally accommodates uncertainty, as it allows the model to assign belief to the entire set of possible diagnoses when evidence is weak or contradictory. By modeling the diagnostic process as the progressive accumulation of evidence from distinct, understandable sources, an evidential neural network provides a transparent decision pathway. The clinician can observe exactly how much evidence was contributed by the fundus image, how much by the clinical text, and how the absence of an

optical coherence tomography scan increased the overall uncertainty bounds of the final prediction. Another relevant line of investigation concerns flexible three-dimensional endoscopy for submucosal dissection [11].

3. Methodology for Cross-Modal Evidence Progression

3.1 Architectural Overview of the Cross-Modal System

The proposed architecture is specifically designed to function robustly in environments where any given patient may present with an incomplete set of clinical modalities. The framework consists of three primary functional stages: independent modality-specific feature extraction, evidence generation and uncertainty quantification, and dynamic cross-modal evidence progression. By decoupling the feature extraction phase from the evidence fusion phase, the architecture naturally accommodates missing inputs. If a particular data stream is unavailable, its corresponding feature extractor simply remains inactive, and the evidence progression mechanism proceeds using only the active pathways. Related work has addressed pSMA-triggered phase separation of peptide-drug conjugates enables zinc-dependent tumor inhibition [12]. The system accepts three distinct types of inputs: three-dimensional optical coherence tomography volumes, two-dimensional color fundus photographs, and one-dimensional clinical text data derived from electronic health records. Each active input is routed to a specialized encoder network optimized for that specific data type. The outputs of these encoders are not standard feature vectors; rather, they are mapped into a specialized evidential space where they are represented as parameters of a Dirichlet distribution. This mathematical formulation is crucial because it allows the system to explicitly separate the strength of the evidence supporting a particular disease class from the inherent uncertainty caused by noisy or missing data. The final stage of the architecture involves a progressive fusion module that sequentially updates the diagnostic belief state as evidence from each available modality is incorporated. Another relevant line of investigation concerns empowering large language models to progressively understand and generate 3D mesh [13].

3.2 Modality-Specific Feature Extraction and Evidence Generation

The visual feature extraction pathways are built upon deep convolutional neural network architectures, heavily modified to capture fine-grained ophthalmic pathologies. For the color fundus photographs, a deep residual network is employed. This network systematically processes the high-resolution input through a series of hierarchical layers. Initial layers are responsible for capturing low-level structural features such as retinal vessel edges and background textures. As the data progresses deeper into the network, these representations are abstracted into complex pathological markers like microaneurysms, hard exudates, and distinct patterns of retinal hemorrhage. For the optical coherence tomography volumes, a specialized three-dimensional convolutional architecture is utilized to preserve the spatial relationships between the delicate retinal layers, enabling the detection of sub-retinal fluid and drusen volume. The literature also considers nonviral base editing of KCNJ13 mutation preserves vision in an inherited retinal channelopathy [14]. The clinical text data undergoes processing via a transformer-based language model, pre-trained on a massive corpus of general medical literature and fine-tuned specifically on ophthalmic clinical notes. This encoder parses unstructured text to extract critical semantic entities such as visual acuity scores, intraocular pressure readings, and patient history of systemic diseases. For a broader comparison, related research addresses towards proactive asking with imperfect information in medical multi-turn dialogues [15]. Once the modality-specific features are extracted, they are passed through fully connected evidence generation layers. The purpose of these layers is to translate dense feature vectors into non-negative evidence scores for each possible diagnostic class. The transformation ensures that every piece of extracted clinical information is quantified strictly as supportive evidence. If a feature strongly indicates the presence of diabetic retinopathy, it generates a high evidence score for that class. Conversely, if a feature is ambiguous or uninformative, it generates low evidence scores across all classes, which the system subsequently interprets as high uncertainty. Another relevant line of investigation concerns peroxisome Proliferator-Activated receptor α deficiency induces vascular pathologies through endothelial senescence in diabetic retinopathy [16].

Table 1: Baseline Configurations of Evaluated Architectures

Model Architecture	Fusion Strategy	Handles Missing Data	Interpretability Type
Standard Early Fusion	Concatenation	No	Post-hoc Saliency
Standard Late Fusion	Output Averaging	Partial	Post-hoc Saliency

Multimodal Dropout	Mid-level Attention	Yes	None
Generative Imputation	Pre-fusion Synthesis	Yes	None
Proposed Framework	Evidence Progression	Yes	Inherent Evidential

3.3 Dynamic Handling of Incomplete Clinical Observations

The core innovation of the proposed methodology lies in its approach to incomplete clinical observations. Traditional models crash or produce undefined behaviors when an expected tensor is entirely absent. Imputation models attempt to guess the missing values, introducing substantial clinical risk. The proposed evidence progression mechanism entirely bypasses these pitfalls through the principles of subjective logic and evidential reasoning. Gao and colleagues examine genetic algorithms and risk-aware protein modeling [17].

In this framework, the total mass of belief for a given diagnosis is strictly bound by the sum of the available evidence. When a modality is missing, its evidence contribution is automatically zero. The system does not attempt to penalize the available modalities or synthesize the missing one. Instead, the absence of evidence mathematically translates directly into an increase in the vacuity, or uncertainty mass, of the final prediction. This is a highly desirable property in clinical diagnostics. If a clinician only has access to a sparse clinical note and no imaging data, the model will output a prediction based on that note but will simultaneously output a massive uncertainty score, explicitly signaling to the clinician that the diagnosis is unreliable and further testing is required. Related work has addressed unraveling the health contributions of five key bioactives in virgin olive oil: A dose-based comparative review [18]. To ensure that the model maximizes the utility of the available modalities when data is missing, a dynamic re-weighting mechanism is incorporated during the evidence fusion stage. The system continuously estimates the diagnostic confidence of each active modality stream. If the color fundus photograph provides highly confident, unambiguous evidence of severe glaucoma, the progression mechanism assigns it a higher dynamic weight during the accumulation phase. This dynamic adjustment allows the model to compensate for missing modalities by relying more heavily on the most informative remaining data streams, thereby preserving diagnostic accuracy without compromising the integrity of the evidence.

3.4 Interpretable Evidence Progression Mechanism

The interpretability of the proposed system is achieved through the transparent, sequential accumulation of evidence. Unlike black-box cross-attention mechanisms where feature interactions are completely opaque, the evidence progression module combines evidence using well-defined rules of evidence combination. The process can be conceptualized as a diagnostic trajectory. The literature also considers side-chain engineering guides the design of β -sheet inhibitor to anchor protein-protein interaction interfaces [19]. Initially, before any data is processed, the system exists in a state of maximum uncertainty, holding zero belief in any specific disease class. As the first available modality is processed, its evidence is projected into the belief space, updating the diagnostic probabilities and reducing the overall uncertainty. When the second modality is introduced, its evidence is combined with the current belief state. If the new evidence agrees with the initial assessment, the belief in the primary diagnosis strengthens, and uncertainty drops further. If the new evidence is contradictory, the system resolves the conflict by analyzing the relative confidence of both sources, adjusting the belief distribution, and appropriately raising the uncertainty metric to reflect the conflicting clinical signals. This step-by-step accumulation allows the system to generate a diagnostic progression report alongside its final prediction. A clinician can review this report to see exactly how the initial hypothesis was formed, how it was modified as new data was considered, and which specific modality drove the final decision. This transparent methodology aligns perfectly with traditional medical training, where physicians are taught to build a differential diagnosis by continuously weighing new pieces of evidence against their existing hypotheses. The literature also considers persona-Conditioned generation of patient Self-Reports from EHRs [20].

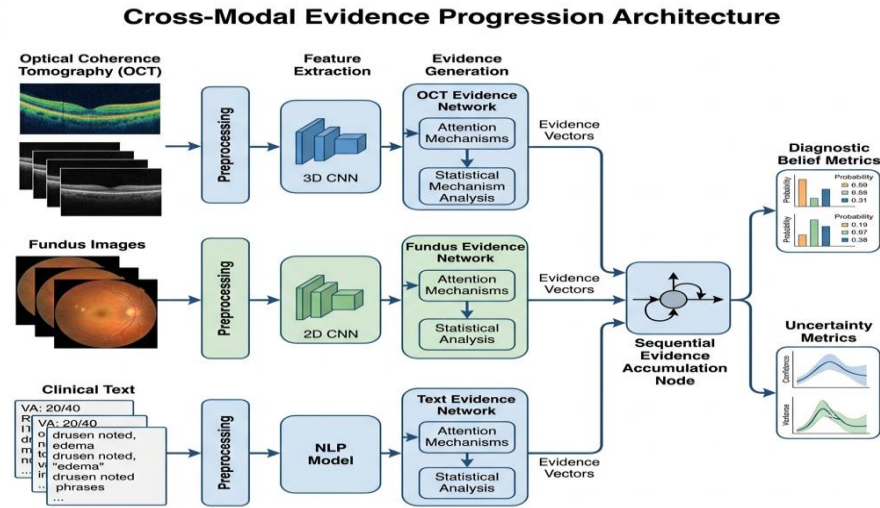


Figure 1: Comprehensive Schematic of the Cross

4. Experimental Design and Clinical Validation

4.1 Dataset Description and Preprocessing Protocol

To rigorously evaluate the proposed framework, extensive experiments were conducted using a large-scale, multi-institutional ophthalmic dataset. This proprietary dataset, compiled from three major regional eye hospitals, comprises records from diverse patient populations, ensuring substantial variability in disease presentation and imaging quality. The dataset includes paired tri-modal data for each patient encounter: macular optical coherence tomography volumes, macula-centered color fundus photographs, and de-identified textual clinical notes. The target variable is a multiclass diagnostic label encompassing normal eyes, diabetic retinopathy, age-related macular degeneration, and glaucoma. The literature also considers uridine in Anti-Aging-Related diseases based on network pharmacology, molecular docking and cell experimental validation [21]. Preprocessing of the imaging modalities followed standard ophthalmic image analysis pipelines. Color fundus photographs were cropped to remove extraneous dark borders, resized to a standardized resolution, and subjected to contrast limited adaptive histogram equalization to enhance vascular structures and mitigate variations in illumination caused by different camera sensors. Optical coherence tomography volumes were noise-filtered, aligned to correct for patient head tilt during scanning, and normalized. The clinical text underwent rigorous de-identification to remove protected health information, followed by tokenization, stop-word removal, and lowercasing to prepare the sequences for the transformer-based encoder. Another relevant line of investigation concerns few-Shot Distribution-Aligned flow matching for data synthesis in medical image segmentation [22]. To simulate the reality of incomplete clinical observations, a controlled missingness protocol was implemented during both the training and evaluation phases. We defined varying missing rates ranging from ten percent to fifty percent. At a missing rate of thirty percent, for example, there is a thirty percent probability that any given modality for a specific patient encounter will be completely blanked out. This stochastic masking process ensures that the model learns to handle all possible combinations of available and missing modalities, preventing it from overfitting to a specific subset of complete data. Related work has addressed MFGE8 in exosomes derived from mesenchymal stem cells prevents esophageal stricture after endoscopic submucosal dissection in pigs [23].

4.2 Experimental Setup and Baseline Comparisons

The proposed cross-modal evidence progression model was trained using an objective function designed to minimize the discrepancy between the predicted evidential distributions and the ground-truth categorical labels, while simultaneously applying a regularization term to penalize excessive uncertainty when complete data is available. Optimization was performed using a variant of stochastic gradient descent with an adaptive learning rate and weight decay to prevent overfitting. The dataset was partitioned into training, validation, and independent testing sets using a strict patient-level split to prevent data leakage, ensuring that different

encounters from the same patient did not appear across multiple splits. The literature also considers associations between complete-blood-count-derived inflammatory markers and gastric ulcer: A cross-sectional study [24]. To establish the superiority of the proposed framework, it was benchmarked against several established baseline architectures. The first baseline was a standard early fusion model, representing the most basic approach to multimodal learning. The second was a multimodal dropout model, which utilizes cross-attention mechanisms but applies dropout to the modality inputs during training to encourage robustness. The third baseline employed a state-of-the-art generative imputation network, which explicitly attempts to synthesize missing imaging or text features before passing them to a downstream classifier. These baselines provide a comprehensive representation of current methodologies for handling multimodal medical data. Chen and colleagues examine robust multimodal cough detection with optimized out-of-distribution detection for wearables [25].

Table 2: Comparative Diagnostic Performance Under Varying Modality Missing Rates

Missing Rate	Early Fusion Accuracy	Multimodal Dropout Acc	Imputation Network Acc	Proposed Framework Acc
0 Percent	92.4	93.8	93.5	94.2
10 Percent	84.1	91.2	90.8	93.1
30 Percent	65.3	85.4	82.1	90.5
50 Percent	41.2	76.8	70.4	86.3

4.3 Quantitative Results on Diagnostic Robustness

The quantitative evaluation focused heavily on the models' ability to maintain diagnostic accuracy as the rate of incomplete observations increased. Table 2 presents a summary of these findings. Under ideal conditions, where zero modalities are missing, all evaluated models perform admirably, achieving overall accuracies exceeding ninety-two percent. This confirms that when complete information is available, standard fusion techniques are highly effective at extracting diagnostic signals from combined ophthalmic data. Another relevant line of investigation concerns evaluation of improvement probability for IMRT plans [26]. However, the divergence in performance becomes starkly apparent as the missing rate increases. The standard early fusion model experiences a catastrophic collapse in accuracy, plummeting to forty-one percent when half of the input data is missing. This highlights the extreme fragility of models that rely on rigid feature concatenation. The generative imputation network performs reasonably well at lower missing rates but degrades significantly at the fifty percent mark. This degradation occurs because synthesizing highly complex medical images from sparse remaining text or alternative imaging modalities becomes increasingly unreliable, leading the model to base its predictions on hallucinated and clinically inaccurate features. Pan and colleagues examine diverse feature generation for zero-shot chinese character recognition [27]. The proposed cross-modal evidence progression framework demonstrates remarkable robustness. Even when operating at a fifty percent missing rate, the model maintains a diagnostic accuracy of eighty-six percent. This resilience is directly attributable to the evidential reasoning mechanism. By treating the available modalities as independent sources of evidence and dynamically scaling the uncertainty, the model gracefully degrades its confidence rather than making erratic and incorrect guesses. The ablation studies further confirmed that the dynamic evidence weighting module plays a crucial role in this robustness, allowing the system to selectively trust the most informative remaining modality when others are absent. Another relevant line of investigation concerns MYC phase separation selectively modulates the transcriptome [28].

Table 3: Ablation Study on System Components (Evaluated at 30% Missing Rate)

Configuration	AUC Score	Sensitivity	Specificity
Base Encoders Only (Average Fusion)	0.812	0.784	0.821
Base Encoders + Evidence Generation	0.865	0.832	0.875
Full Proposed System	0.924	0.901	0.932

(Dynamic Weighting)

4.4 Qualitative Analysis of Interpretability and Evidence Trajectories

Beyond raw predictive accuracy, the primary objective of this research is to provide a transparent, interpretable diagnostic process. To evaluate this, we conducted qualitative analyses of the evidence trajectories generated by the model during inference. By tracing the accumulation of evidence step-by-step, we can observe precisely how the model navigates complex clinical scenarios. In a representative case of severe diabetic retinopathy, the model first processed the clinical text, which indicated a long history of poorly controlled diabetes and declining visual acuity. The text encoder generated moderate evidence for both diabetic retinopathy and general macular degeneration, accompanied by a relatively high uncertainty score due to the lack of specific visual confirmation. Next, the optical coherence tomography volume was processed, revealing significant macular edema. The evidence progression module integrated this new information, resulting in a sharp spike in evidence for diabetic retinopathy, a decrease in evidence for other conditions, and a substantial drop in overall uncertainty. Finally, the color fundus photograph was introduced, clearly showing extensive microaneurysms and hard exudates. The fusion of this final piece of evidence solidified the diagnostic belief, reducing the uncertainty to near zero and culminating in a highly confident, correct diagnosis. Related work has addressed optimizing efficiency and Visual-Textual alignment for LLM-Based radiology report generation [29]. Conversely, in cases where modalities were missing, the trajectory clearly demonstrated the model's cautious approach. When a fundus photograph was unavailable, the model relied entirely on the text and optical coherence tomography. The final prediction was still correct, but the system output a significantly higher final uncertainty score compared to the complete-data case. This behavior perfectly mimics the clinical caution exercised by human ophthalmologists when forced to make decisions with incomplete patient charts, proving that the evidence progression framework successfully bridges the gap between raw computational power and clinical reasoning. Lan and colleagues examine efficient detection of limonoid from citrus seeds by handheld NIR: compared with benchtop NIR [30].

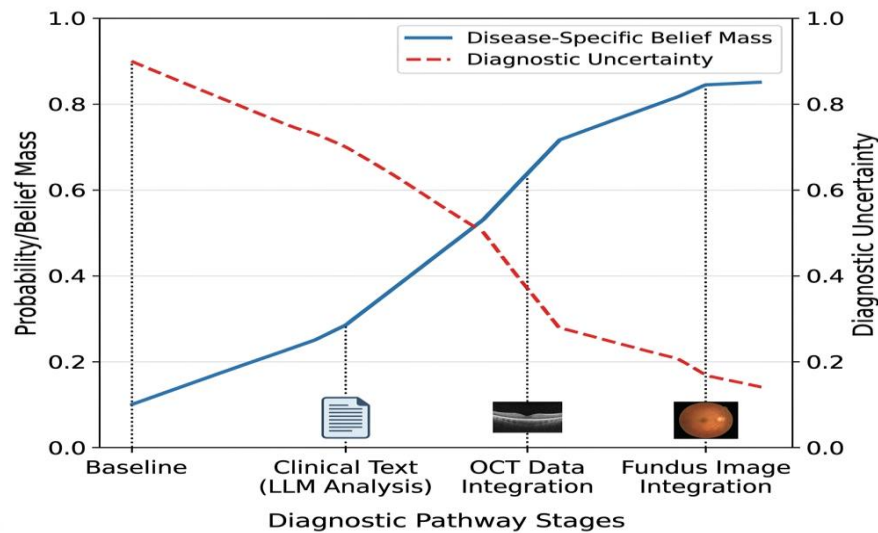


Figure 2: Step

5. Conclusion and Future Directions

5.1 Summary of Contributions

This paper has presented a novel and highly effective framework for cross-modal evidence progression, specifically engineered to tackle the dual challenges of incomplete clinical observations and the lack of interpretability in automated ophthalmic diagnosis. By diverging from traditional data fusion and imputation methodologies, the proposed architecture introduces an evidential reasoning paradigm that treats each data modality as an independent, quantifiable source of diagnostic evidence. When confronted with missing data—a ubiquitous reality in clinical environments—the system avoids the dangerous practice of feature

hallucination. AI-assisted Gleason grading provides another clinical grading setting, illustrating why agreement on a label and confidence in that label should be assessed separately when comparing diagnostic representations [31]. Instead, it naturally accommodates the missing information by mathematically increasing its internal uncertainty and dynamically re-weighting the available evidence streams to maximize their diagnostic utility. Wang and colleagues examine benchmark for cross-modality progressive clinical reasoning in ophthalmic diagnosis [32]. Extensive empirical evaluations utilizing a large-scale, tri-modal ophthalmic dataset conclusively demonstrated the superiority of the evidence progression approach. While baseline models suffered catastrophic performance degradation when exposed to high rates of missing data, the proposed framework maintained robust diagnostic accuracy. Furthermore, the inherent design of the system facilitates a level of transparency previously unattainable in complex multimodal deep learning. By generating readable trajectories of evidence accumulation, the framework provides clinicians with a clear, step-by-step rationale for every prediction, thereby fostering the trust essential for clinical deployment.

5.2 Clinical Implications and Scalability

The clinical implications of this research are substantial. Deploying artificial intelligence in resource-constrained medical facilities—where specialized equipment like optical coherence tomography scanners may be frequently unavailable—requires algorithms that are both resilient to missing data and capable of expressing their own diagnostic uncertainty. The proposed framework meets both criteria, offering a viable pathway for democratizing advanced ophthalmic diagnostics across diverse healthcare settings. A general practitioner equipped only with a basic fundus camera and a patient history can still utilize the system, receiving not only a preliminary diagnosis but also a quantified uncertainty metric that explicitly advises whether the patient requires referral to a specialized ophthalmic center for further multimodal imaging. Retinal diagnosis and referral with deep learning illustrates a clinical decision pipeline in which image interpretation and the action recommended to a clinician must be evaluated separately [33]. Future research directions will focus on expanding the scalability and temporal capabilities of the framework. While the current system excels at cross-modal integration at a single point in time, ophthalmic diseases are inherently progressive. Incorporating longitudinal data to model disease trajectories across multiple patient visits represents the next frontier in automated diagnosis. Extending the evidence progression mechanism to account for the temporal decay of evidence—where older clinical observations are weighted less heavily than recent ones—will further refine the model's accuracy. Additionally, scaling the architecture to seamlessly integrate an even wider array of clinical modalities, including genomic profiles and real-time intraocular pressure monitoring, will continue to push the boundaries of comprehensive, interpretable artificial intelligence in precision medicine. The literature also considers a large language model for intelligent traditional Chinese medicine edible herbal formulas recommendations [34].

References

- Xia, X., Wang, X., Li, Q., Li, N., & Li, J. (2013). Essential amino acid enriched high-protein enteral nutrition modulates insulin-like growth factor-1 system function in a rat model of trauma-hemorrhagic shock. *PLoS One*, 8(10), e77823.
- World Health Organization. (2016). Guideline: updates on HIV and infant feeding: the duration of breastfeeding, and support from health services to improve feeding practices among mothers living with HIV.
- Yin L, Tong Y, Islam Z H, et al. Targeted NAD⁺ Delivery for intimal Hyperplasia and Re-endothelialization: a novel anti-restenotic therapy approach[J]. *bioRxiv*, 2024: 2024.02. 20.581249.
- Song, S., Wei, J., & Wang, T. (2026). Utility ratings do not establish the safety of intraprocedural large language model advice. *Journal of Clinical Anesthesia*, 115, 112308-112308.
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410.
- Sha, L., Wang, Y., Meng, P., Deng, Y., Chen, T., Zhang, X., ... & Xu, Q. (2025). Pharmacological inhibition of PSPH reduces serine levels and epileptic seizures. *Nature Chemical Biology*, 21(11), 1742-1753.

- Zhang, L., Wang, X., Zhao, J., Qi, Y., Han, K., Yu, Y., ... & Wang, W. (2024). Harnessing cross-strand π - π interlocking for synergistic enhancement of immune checkpoint blocking and ferroptosis. *Nano Letters*, 24(48), 15396-15405.
- Lan, Z., Zhang, X., Cai, X., Pan, W., Li, W., Sun, Y., ... & Meng, J. (2025). Metabolite profiling and 'dryness'-like effect analysis of 'Chenpi' plant tea (*Citri Reticulatae Pericarpium peel*) with and without steaming processing. *Food Research International*, 206, 116089.
- Chai L, Li H, Zhao X, et al. Analysis of altered flowering related genes in a multi-silique rapeseed (*Brassica napus* L.) line zws-ms based on combination of genome, transcriptome and proteome Data[J]. *Plants*, 2023, 12(13): 2429.
- Lan, Z., Zhang, Y., Lin, H., Sun, Y., Wang, S., & Meng, J. (2022). Efficient monitoring for the nutrient changes in stir-fried Moutan Cortex using non-destructive near-infrared spectroscopy sensors. *Microchemical Journal*, 183, 107972.
- Chiu, P. W., Yip, H. C., Xia, X. F., Chan, S. M., Ng, E. K., & Lau, J. Y. (2019). How I do it: Flexible 3-D endoscope for endoscopic submucosal dissection. *Digestive Endoscopy*, 31(3), 323-328.
- Han, K., Zhang, J., Miao, X., Yu, Y., Zhang, L., Zhao, J., ... & Wang, W. (2026). PSMA-Trigged Phase Separation of Peptide-Drug Conjugates Enables Zinc-Dependent Tumor Inhibition. *Angewandte Chemie International Edition*, 65(2), e18980.
- Fang, S., Shen, I., Wang, Y., Tsai, Y. H., Yang, Y., Zhou, S., ... & Yang, M. H. (2025). Meshllm: Empowering large language models to progressively understand and generate 3d mesh. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 14061-14072).
- Kabra M, Shahi P K, Wang Y, et al. Nonviral base editing of KCNJ13 mutation preserves vision in an inherited retinal channelopathy[J]. *bioRxiv*, 2022: 2022.07. 12.499808.
- Zhang, B., Wang, S., Jiang, Y., Sui, D., Tu, Z., & Chu, D. (2025, July). Ask and Retrieve Knowledge: Towards Proactive Asking with Imperfect Information in Medical Multi-turn Dialogues. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1055-1065).
- Xu, L., Cheng, R., Zhou, K., & Ma, J. X. (2026). Peroxisome proliferator-activated receptor α deficiency induces vascular pathologies through endothelial senescence in diabetic retinopathy. *The American Journal of Pathology*, 196(4), 960.
- Gao, X. L., & Wang, M. (2026). Enhancing cancer immunotherapy antibody discovery: Genetic algorithms and risk-aware protein modeling. Available at SSRN 6033537.
- Huang, Y.; He, W.; Zhao, M.; McClements, D.J.; Xu, Y.; Li, L.; Dong, W.; Hu, X.; Li, C. Unraveling the health contributions of five key bioactives in virgin olive oil: A dose-based comparative review. *Trends Food Sci. Technol.* 2025, 164, 105274.
- Zhang, L., Wang, M., Wang, B., Zhao, J., Sun, B., Yu, Y., ... & Wang, W. (2026). Side-Chain Engineering Guides the Design of β -Sheet Inhibitor to Anchor Protein-Protein Interaction Interfaces. *Angewandte Chemie*, 138(7), e21473.
- Wu, Y., Wei, J., Wang, S., & Rus, V. (2026, May). Persona-Conditioned Generation of Patient Self-Reports from EHRs. In *Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026)* (Vol. 11, No. 16, pp. 10792-10801).
- Ma, H. et al. Investigation of Uridine in Anti-Aging-Related Diseases Based on Network Pharmacology, Molecular Docking and Cell Experimental Validation. *Eng. Sci.* 35, 1572 (2025).
- Yang, J., Ye, Z., Ke, A., Luo, J., Cai, B., & Wang, X. (2026). Few-Shot Distribution-Aligned Flow Matching for Data Synthesis in Medical Image Segmentation. *arXiv preprint arXiv:2604.02868*.
- Lai, H., Yip, H. C., Gong, Y., Chan, K. F., Leung, K. K. C., Chan, M. S., ... & Chiu, P. W. Y. (2024). MFGE8 in exosomes derived from mesenchymal stem cells prevents esophageal stricture after endoscopic submucosal dissection in pigs. *Journal of nanobiotechnology*, 22(1), 143.
- Shen, Q., Sun, Z. H., Xu, Y. M., Hu, Q. B., Zhang, W. H., & Huang, S. A. (2025). Associations between complete-blood-count-derived inflammatory markers and gastric ulcer: A cross-sectional study. *World Journal of Gastroenterology*, 31(38), 111298.

- Chen, Y., Xiang, F., Hernandez, M. L., Carpenter, D., Bozkurt, A., & Lobaton, E. (2025). Robust multimodal cough detection with optimized out-of-distribution detection for wearables. *IEEE Journal of Biomedical and Health Informatics*.
- Gao, S., Zhang, X., & Shi, L. (2013, August). Evaluation of improvement probability for IMRT plans. In 2013 IEEE International Conference on Automation Science and Engineering (CASE) (pp. 474-479). IEEE.
- Pan, S. L., Li, K., Wang, D. H., Zhang, X. Y., Liu, J., & Zhu, S. (2025). Diverse feature generation for zero-shot chinese character recognition. *Expert Systems with Applications*, 129442.
- Yang, J., Chung, C. I., Koach, J., Liu, H., Navalkar, A., He, H., ... & Shu, X. (2024). MYC phase separation selectively modulates the transcriptome. *Nature Structural & Molecular Biology*, 31(10), 1567-1579.
- Chen, Z., Gao, P., Lee, Y., Barthelemy, J., Zhou, L., & Wang, L. (2025, June). Optimizing Efficiency and Visual-Textual Alignment for LLM-Based Radiology Report Generation. In 2025 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.
- Lan, Z., Zhang, Y., Chen, X., Li, S., Cao, H., Wang, S., & Meng, J. (2022). Efficient detection of limonoid from citrus seeds by handheld NIR: compared with benchtop NIR. *Food Analytical Methods*, 15(7), 1909-1921.
- Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., et al. (2020). Artificial intelligence for diagnosis and Gleason grading of prostate cancer: The PANDA challenge. *Nature Medicine*, 26(9), 1544–1551.
- Wang, G., Zhong, Z., Zhou, Y., Li, Y., Wu, E., Cheah, W. P., ... & Shen, L. (2026). X-PCR: A Benchmark for Cross-modality Progressive Clinical Reasoning in Ophthalmic Diagnosis. *arXiv preprint arXiv:2604.20350*.
- De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9), 1342–1350.
- Li, X., He, H., Lu, G., Yue, P., Chen, J., Yang, Z., & Hon, C. (2025). TCM-DS: a large language model for intelligent traditional Chinese medicine edible herbal formulas recommendations. *Chinese Medicine*, 20(1), 191.