HipGo: Multimodal Smart Agent Solution for Remote Hip Dysplasia Care

***Abstract*—Developmental Dysplasia of the Hip (DDH) is a clinically elusive rare orthopedic disorder with a global prevalence estimated between 1% and 3%. Although primarily diagnosed in children, accumulating evidence indicates a substantial population of adults harboring undetected childhood-onset DDH who exhibit progressive hip degeneration and significant functional impairment. Such diagnostic delays are particularly pronounced in resource-limited regions, where early identification and timely intervention are challenging.**

Addressing these gaps, we introduce HipGo — an intelligent remote DDH diagnosis and management platform tailored for adolescents and adults. HipGo integrates a hybrid convolutional neural network (CNN) and graph attention network (GAT) model for precise hip keypoint localization, which fuses structured diagnostic parameters with deep visual features for multimodal clinical report generation. Moreover, it encompasses intelligent decision support and remote rehabilitation modules to enable a closed-loop, personalized care pathway from diagnosis through recovery.

***Index Terms*—Developmental Dysplasia of the Hip (DDH), medical image analysis, CNN-GAT, telemedicine, intelligent agent, multimodal report generation**

I. INTRODUCTION

Developmental Dysplasia of the Hip (DDH) is a clinically elusive rare orthopedic disorder with a global prevalence estimated between 1% and 3% [1]. Although primarily diagnosed in children, accumulating evidence indicates a substantial population of adults harboring undetected childhood-onset DDH who exhibit progressive hip degeneration and significant functional impairment [2]. Such diagnostic delays are particularly pronounced in resource-limited regions, where early identification and timely intervention are challenging.

Conventional radiographic keypoint measurements for hip X-rays rely heavily on radiologist expertise and manual annotation, making the process time-consuming and prone to inter-observer variability [3], [4]. Despite notable advances in AI-based medical image analysis, most existing systems struggle with explainability and fail to effectively support clinical decision-making by non-specialists or to deliver interpretable assessments of complex anatomical metrics such as center-edge (CE) and Sharp angles [4]. Additionally, patient management and rehabilitation workflows remain fragmented and lack integrated telemedicine platforms providing seamless end-to-end services, negatively impacting treatment adherence and outcomes.

Addressing these gaps, we introduce HipGo — an intelligent remote DDH diagnosis and management platform tailored for adolescents and adults. HipGo integrates a hybrid convolutional neural network (CNN) and graph attention network (GAT) model for precise hip keypoint localization, which fuses structured diagnostic parameters with deep visual features for multimodal clinical report generation. Moreover, it encompasses intelligent decision support and remote rehabilitation modules to enable a closed-loop, personalized care pathway from diagnosis through recovery. As shown in Fig.1, the platform establishes seamless integration with hospital information systems, enabling online-offline collaborative care for underserved regions.

The key contributions of this study include:

- Constructing a large, heterogeneous dataset of 800 pelvic X-rays crowdsourced from social media, enhanced by an active learning pipeline that optimizes annotation efficiency and model robustness [5];
- Developing a novel CNN-GAT fusion architecture augmented with anatomical priors, significantly improving keypoint detection accuracy and clinical angle measurement precision compared to standard methods;
- Designing a multimodal automatic report generation system that elevates clinical interpretability and supports practical diagnosis;
- Implementing a comprehensive intelligent agent platform that facilitates online-offline integration and equitable access to expert care.

II. RELATED WORK

A. Keypoint Detection and Spatial Modeling in Medical Imaging

Accurate anatomical keypoint detection forms the cornerstone of AI-assisted orthopedic diagnosis. Early approaches utilized handcrafted features such as Scale Invariant Feature Transform (SIFT) and Histogram of

Oriented Gradients (HOG), which were limited in handling the high variability and complexity of medical images [7], [8]. With deep learning advancements, convolutional neural networks (CNNs) emerged as the primary tool for landmark localization. Notably, architectures like ResNet and HRNet enable hierarchical and high-resolution feature extraction, respectively, enhancing detection performance [6], [9].

Graph neural networks (GNNs), particularly graph attention networks (GATs), have been increasingly adopted to address this limitation by learning adaptive spatial dependencies between anatomical points [10]. Self-supervised methods leveraging knowledge distillation further improve feature representations [11]. Hybrid CNN-GNN models combine local texture encoding with global spatial reasoning and have demonstrated success in domains such as brain MRI registration and retinal image analysis [12], [13]. However, these models lack tailored optimization for DDH-specific anatomical constraints, which we address through our hybrid architecture.

B. Automatic Medical Report Generation and Multimodal Fusion

The generation of medical reports has evolved from rule-based systems and expert-curated templates—accurate yet inflexible—to deep learning-based encoder-decoder models that map images directly to text [14]. Recent progress highlights the effectiveness of multimodal fusion techniques, integrating visual features, structured clinical data, and textual context to improve report relevance and completeness [15].

Self-training and semi-supervised approaches, such as REMOTE, achieve comparable performance to fully supervised models with substantially fewer labeled samples [16]. Additionally, the adoption of large language models (LLMs) has enhanced linguistic quality and interpretability [17].

C. Intelligent Medical Agents and Multi-Agent Systems

Intelligent agents and multi-agent systems (MAS) constitute a transformative direction in healthcare informatics by orchestrating autonomous, yet collaborative, interactions among clinicians, patients, and devices [18]. MAS frameworks help optimize clinical decision support, personalized treatment planning, and continuous monitoring, improving outcome efficiency and care coordination [19].

Challenges in large-scale deployment remain due to data bias, interoperability barriers, and unclear responsibility assignments. Furthermore, existing agent-based solutions predominantly address generalized medical scenarios, lacking tailored workflows and closed-loop integration for rare orthopedic diseases like DDH.

III. METHODS

This section details the technical framework of HipGo's intelligent remote diagnosis and management platform, structured around a "perception–analysis–decision–execution–feedback" closed-loop (as shown in Fig.1).

A. Data Acquisition and Preprocessing

To reflect authentic primary care telemedicine environments in China, 800 adult anteroposterior pelvic X-ray images were collected from social media platforms,

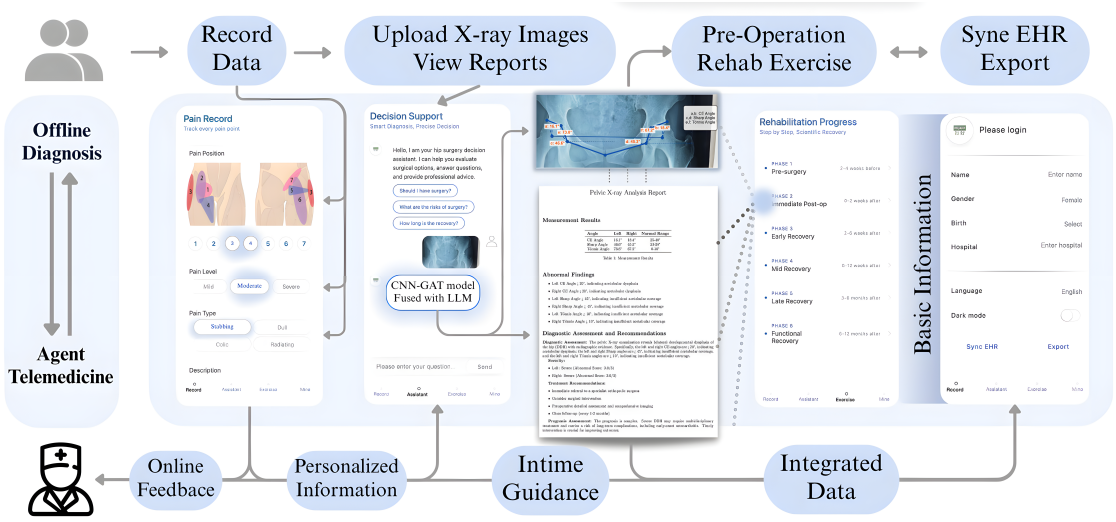

Fig. 1: HipGo system interaction architecture with hospital workflows.

encompassing diverse imaging devices and variable quality conditions [5]. Nine key anatomical landmarks were annotated by orthopedic specialists according to standardized DDH criteria, guaranteeing clinical validity of measurements including CE, Sharp, and Tönnis angles [3], [4].

An active learning approach prioritized uncertain samples for iterative expert reannotation, optimizing labeling efficiency and model generalizability [5]. Preprocessing involved OCR-based de-identification, resolution normalization to 512 × 512 pixels, ImageNet statistics normalization, and multi-channel augmentation including random rotation, translation, Gaussian noise, and motion blur to boost robustness [6], [20].

B. Hybrid CNN-GAT Keypoint Detection

A dual-branch CNN-GAT architecture was developed to handle pelvis spatial complexity, with its detailed structure shown in Fig.2. The CNN branch utilizes a ResNet-50 backbone with Feature Pyramid Network (FPN) to capture multiscale texture and morphological features [6], [21]. Specifically, the FPN generates feature maps at 4 scales (1/4 to 1/32 of input resolution), enabling detection of both large anatomical structures (e.g., acetabulum) and fine landmarks (e.g., femoral head edge).

The GAT branch models anatomical connectivity via a 9-node graph, where each node represents a key pelvic landmark (e.g., femoral head center, acetabular edge). It incorporates dynamically computed edge features such as normalized Euclidean distances, trigonometric angular encodings, and anatomical semantic labels (e.g., "femoral-acetabular relationship"), enabling adaptive spatial attention [22], [10]. A lightweight dynamic gating module ($\sigma(W_1 \cdot CNN\_feat + W_2 \cdot GAT\_feat + b)$) adaptively balances CNN and GAT feature contributions based on image quality (e.g., assigning higher weight to GAT for low-contrast images).

Multi-task loss combines keypoint regression (MSE loss) with medically motivated anatomical constraints (e.g., enforcing CE angle consistency with femoral-acetabular spatial relationships), guiding the model to generate clinically plausible outputs and preventing anatomically inconsistent predictions.

C. Multimodal Clinical Report Generation

Structured angle parameters derived from detected keypoints, high-dimensional deep visual features, and patient clinical data are fused using a cross-attention mechanism for semantic alignment [23]. A pretrained biomedical large language model (LLaVA-Med variant) generates clinician-standardized diagnostic reports, covering angle measurements, image abnormality descriptions, diagnostic conclusions, and personalized treatment guidance [24].

To enhance clinical utility, the system overlays key anatomical points and measurement lines (e.g., CE angle reference lines) on X-ray images within reports, facilitating interpretability and effective patient–clinician communication.

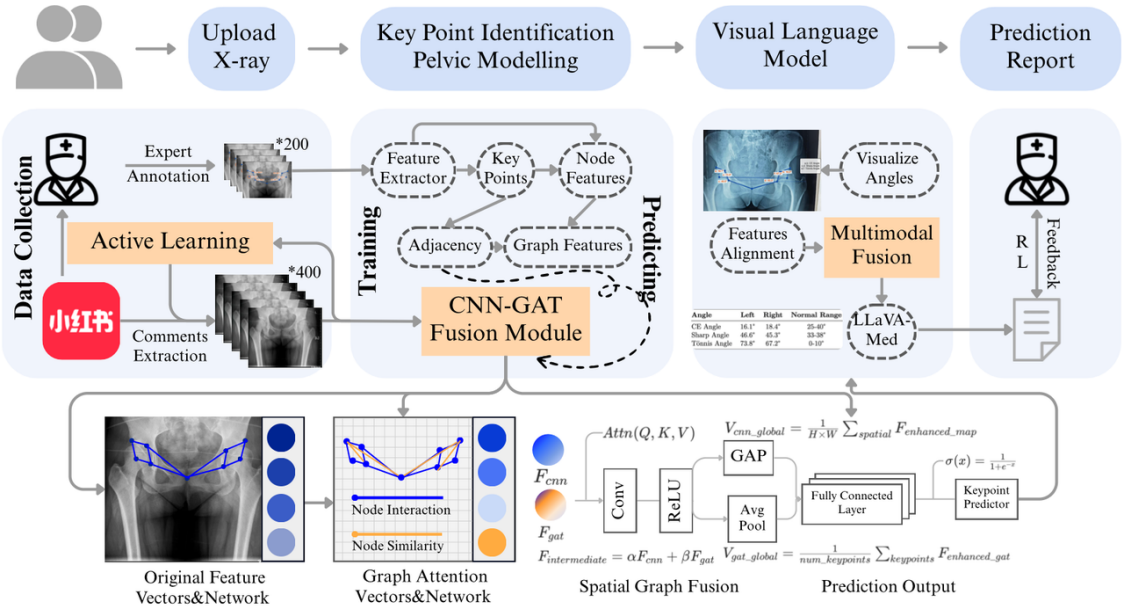

Fig. 2: Architecture of the CNN-GAT model.

D. Intelligent Decision Support and Rehabilitation Management

Based on clinical guidelines and individualized diagnostic results, an intelligent decision support module formulates personalized treatment plans enabling shared decision-making. The rehabilitation management system dynamically adjusts phased recovery schedules through mobile task reminders and remote monitoring (e.g., activity tracking via wearables), aiming to improve compliance and clinical outcomes [25].

Workflow data is centralized via standardized interfaces interoperable with electronic health records (EHR), enabling multi-platform access and export. Reinforcement learning techniques exploit feedback data (e.g., clinician corrections, rehabilitation adherence) to continuously refine diagnostic and decision models, fostering a self-improving intelligent care ecosystem [26].

IV. EXPERIMENTS AND RESULTS

This section validates HipGo's performance through multidimensional experiments, covering keypoint detection accuracy, clinical angle quantification, and practical system utility in remote DDH care.

A. Experimental Design and Datasets

The 800-case X-ray dataset was split by subject into training (640), validation (80), and test (80) cohorts. Metrics included PCK@0.1/0.2/0.5 (Percentage of Correct Keypoints at thresholds 0.1/0.2/0.5×image diagonal) and mAP (mean Average Precision) for keypoint detection, mean absolute error (MAE) and correlation coefficient for angle measurement, and diagnostic accuracy, sensitivity, and specificity for clinical outcomes. Baselines comprised ResNet-50, HRNet, and traditional GAT, all evaluated under matched conditions [6], [9], [10].

B. Keypoint Detection Performance

Table I shows the keypoint detection performance comparison between our proposed method and baselines. The CNN-GAT fusion model outperformed all baselines, especially in the high-precision PCK@0.1 metric, where it exceeded ResNet-50 by 14 percentage points. This confirms the effectiveness of integrating spatial reasoning (GAT) with local feature extraction (CNN), as designed in Fig.2.

C. Accuracy of Structured Angle Measurement

Table II presents the comparison between expert measurements and AI measurements for key clinical angles. Results were closely aligned with expert reference, with all MAEs under 1.1°. The high correlation coefficients (¿0.91) and clinical accuracy (¿89

Tab. I: Keypoint Detection Performance Comparison

Method	PCK@0.1	PCK@0.2	PCK@0.5	mAP	Parameters
ResNet-50	0.72	0.85	0.94	0.83	25.6M
HRNet	0.78	0.89	0.96	0.87	32.1M
Traditional GAT	0.75	0.86	0.95	0.85	18.3M
CNN-GAT Fusion	0.86	0.93	0.98	0.91	28.7M

Tab. II: Accuracy of Structured Angle Measurement

Angle	Expert Measurement	AI Measurement	MAE	Correlation	Clinical Accuracy
CE Angle	25.3°±3.2°	25.1°±3.1°	0.8°	0.94	92%
Sharp Angle	42.1°±2.8°	42.3°±2.9°	1.1°	0.91	89%
Tönnis Angle	8.7°±1.5°	8.9°±1.6°	0.6°	0.96	94%

D. Clinical Validation and System Performance

Expert validation yielded a diagnostic report kappa of 0.87 (almost perfect agreement) and overall accuracy of 91%, with sensitivity and specificity of 88% and 94%, respectively. Report relevance, completeness, and readability scored a mean of 4.1/5. System stress tests confirmed a mean response of 2.3s per image and a 7-day uptime of 96.5%, demonstrating suitability for real-time teleclinical use as part of the interaction pipeline in Fig.1.

V. DISCUSSION

We have designed and validated HipGo—an end-to-end remote diagnostic platform for adolescent and adult DDH. Experimental and clinical validation demonstrate that the system exceeds clinical thresholds for keypoint detection and angle measurement (PCK@0.1 = 0.86; angle MAE 1.1°; clinical accuracy 89–94%), confirming the technical value of the CNN-GAT model (see Fig.2) and multimodal reporting pipeline.

Incorporating anatomical priors and multi-edge feature coding in the GAT branch, the hybrid model sharply improved spatial and quantitative precision, especially for challenging, low-quality images. The multimodal auto-reporting system successfully aligned structured parameters, deep features, and textual clinical information, generating readable and traceable LLM-based outputs with added visual annotation overlays.

The system interaction architecture (Fig.1) enables equitable access to expert care by connecting resource-limited regions with specialists through a closed-loop workflow. Limitations include reliance on a single data source, insufficient evaluation on extremely rare or complicated cases, and incomplete hospital information system integration. Future work will expand to multi-center validation and deeper integration for industrial-scale adoption.

VI. CONCLUSION

This study introduces HipGo—an end-to-end telemedicine solution for DDH, underpinned by CNN-GAT model fusion , active annotation, multimodal reporting, and closed-loop intelligent decision support. The system substantially outperforms existing methods in keypoint detection, angle measurement, and decision support, and is suitable for real-time clinical deployment (2.3s/image, 96.5% uptime). HipGo advances early DDH diagnosis and care efficiency while providing a reliable collaboration tool for providers and patients, improving access in underserved regions.

REFERENCES

[1] T. Chan, et al., "Epidemiology of Developmental Dysplasia of the Hip: A Systematic Review," *J Pediatr Orthop*, vol. 39, no. 3, pp. 243–252, 2019.

[2] K. Nakamura, et al., "Adult Hip Dysplasia: Clinical Features and Management," *Clin Orthop Relat Res*, vol. 472, no. 9, pp. 2731–2736, 2014.

[3] G. Wiberg, "Studies on dysplastic acetabula and congenital subluxation of the hip joint," *Acta Chir Scand Suppl*, vol. 58, pp. 1–135, 1939.

[4] I. H. Sharp, "Acetabular Dysplasia: A Radiographic Study of Acetabular Development," *JBJS*, vol. 43, no. 7, pp. 994–1002, 1961.

[5] B. Settles, "Active Learning Literature Survey," University of Wisconsin–Madison Technical Report, 2009.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *CVPR*, 2016.

[7] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int J Comput Vis*, vol. 60, no. 2, pp. 91–110, 2004.

[8] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *CVPR*, 2005.

[9] K. Sun, et al., "High-Resolution Representations for Labeling Pixels and Regions," *arXiv:1904.04514*, 2019.

[10] P. Veličković, et al., "Graph Attention Networks," in *ICLR*, 2018.

[11] A. Abdi, et al., "Self-Supervised Keypoint Detection with Distilled Depth Keypoint Priors," *arXiv:2410.14700*, 2024.

[12] J. Chen, Y. Du, Y. He, et al., "A Foundational Keypoint Model for Robust and Flexible Brain MRI Registration," *arXiv:2405.14019*, 2024.

[13] C. Hernandez-Matas, et al., "Joint Keypoint Detection and Description Network for Color Fundus Image Registration," *PLoS Comput Biol*, vol. 19, no. 3, 2023.

[14] L. Guo, A. M. Tahir, D. Zhang, et al., "Automatic Medical Report Generation: Methods and Applications," *arXiv:2408.13988*, 2024.

[15] Y. Wang, W. Chen, P. Zhang, et al., "Multimodal Transformer for Medical Image Report Generation," in *CVPR*, 2022.

[16] G. Li, X. Qian, J. Wang, et al., "A Self-training Framework for Automated Medical Report Generation," in *EMNLP*, 2023.

[17] A. Radford, et al., "Language Models are Few-Shot Learners," in *NeurIPS*, 2020.

[18] J. Filgueiras, P. H. Santos, L. A. Silva, et al., "Intelligent agents in biomedical engineering: a systematic review," *Int J Biosens Bioelectron*, vol. 6, no. 5, pp. 123–128, 2020.

[19] A. Kumar, R. Singh, M. Patel, et al., "Multi-Agent AI Systems in Healthcare: A Systematic Review," *Asian J Med Princ Clin Pract*, vol. 8, no. 1, pp. 273–285, 2025.

[20] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, 2019.

[21] T.-Y. Lin, P. Dollár, R. Girshick, et al., "Feature Pyramid Networks for Object Detection," in *CVPR*, 2017.

[22] W. L. Hamilton, Z. Ying, and J. Leskovec, "Inductive Representation Learning on Large Graphs," in *NeurIPS*, 2017.

[23] G. Li, X. Qian, J. Wang, et al., "Cross-modal Attention for Medical Report Generation," in *EMNLP*, 2020.

[24] C. Li, C. Wong, S. Zhang, et al., "LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine," *arXiv:2308.14643*, 2023.

[25] A. Kumar, R. Singh, M. Patel, et al., "Remote Patient Monitoring Systems for Rare Disease Management: A Systematic Review," *IEEE J Biomed Health Inform*, 2022.

[26] D. Silver, et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature*, vol. 529, pp. 484–489, 2016.