suitable for complex data types such as images and videos (Varghese et al., 2024). Convolutional neural networks (CNNs) are adept at capturing spatial features and are widely used in analyzing preoperative MRI, CT, ultrasound, and intraoperative endoscopic images. Meanwhile, recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and more recent Transformer architectures are effective in modeling temporal dependencies in surgical videos, such as workflow progression, instrument trajectories, and step transitions (King et al., 2025). In minimally invasive gynecologic surgery, DL has been applied to lesion detection, organ segmentation, intraoperative phase recognition, and quantitative assessment of surgical skills. Some models have achieved accuracies exceeding 80%-90% in surgical phase and structure recognition tasks, demonstrating performance comparable to, or even surpassing, expert interpretation (Paracchini et al., 2025).

Computer vision serves as the critical bridge that embeds ML and DL into surgical practice, enabling computers to “understand” surgical images and videos. From object detection, semantic and instance segmentation to pose estimation, instrument tracking, and three-dimensional reconstruction, CV transforms raw surgical video into structured, analyzable data. In minimally invasive gynecologic surgery, these techniques enable real-time localization and annotation of the uterus, ureter, pelvic vessels, and surgical instruments, as well as pixel-level identification of tissue boundaries through semantic segmentation, thereby supporting intraoperative navigation and risk alerts (Paracchini et al., 2025). Furthermore, the integration of deep learning and computer vision forms the foundation of “surgical intelligence,” allowing AI to learn expert procedural patterns, automatically segment surgical phases, detect abnormal maneuvers, and provide objective support for intraoperative decision-making and postoperative quality control (Knudsen et al., 2024; Varghese et al., 2024).

2.2 Medical data sources: imaging, surgical video, and clinical data

The performance of AI systems heavily depends on the quality, diversity, and integration of medical data. In minimally invasive gynecologic surgery, the primary data sources include medical imaging, intraoperative video, and structured and unstructured clinical data. Imaging data, such as ultrasound, CT, MRI, PET-CT, and intraoperative fluorescence imaging, provide rich anatomical information and serve as a fundamental basis for AI model training and application (Varghese et al., 2024). In oncologic surgery, radiomics and deep learning approaches can extract high-dimensional quantitative features from imaging data to predict tumor characteristics, lymph node metastasis, extent of invasion, and treatment response, thereby supporting individualized surgical planning and preoperative risk stratification. In gynecologic MIS, preoperative three-dimensional reconstruction and augmented reality navigation increasingly rely on high-quality imaging data to achieve precise alignment between preoperative planning and intraoperative anatomy.

Intraoperative video has become one of the fastest-growing yet most challenging data sources in surgical AI. Laparoscopic and robotic systems continuously record high-resolution videos that capture instrument motion, tissue interaction, anatomical exposure, and procedural workflow, providing an ideal dataset for surgical phase recognition, instrument tracking, anatomical detection, and skill assessment. Studies have shown that some AI models achieve accuracies exceeding 85%-90% in phase recognition and instrument detection tasks in gynecologic and other minimally invasive surgeries (Paracchini et al., 2025). However, intraoperative video data present several challenges, including high annotation costs, variability in data acquisition standards across centers, differences in equipment and viewing angles, and complexities related to patient and team privacy. Therefore, establishing standardized frameworks for surgical video data management and governance is essential for clinical translation of surgical AI.

In addition to imaging and video, clinical data are critical for building high-value AI models. These data include patient demographics, body mass index, comorbidities, surgical history, laboratory results, pathology findings, perioperative events, hospitalization outcomes, and long-term follow-up information. When integrated with imaging and video data, these sources enable the development of multimodal models for risk prediction, complication monitoring, length-of-stay estimation, and treatment evaluation, thereby facilitating comprehensive patient profiling. Moreover, unstructured data in electronic health records contain valuable information that can be mined using natural language processing (NLP) techniques to identify patterns in preoperative assessments,