

Abstract

Data mining is the systematic process of extracting useful knowledge from large multimodal datasets and is increasingly enabled by artificial intelligence (AI) methods. Pediatric radiology is a natural field for data mining because multimodal data sources, including images, reports, metadata, and electronic health records, together capture rich information on anatomy, disease, treatment, and outcomes. In the current era, the boundaries between data mining and AI are increasingly blurred. AI assists in key steps of the mining workflow through automated labeling, information extraction, and representation learning, while data mining provides the high-quality curated datasets that underpin model performance, generalizability, and safety. This review, therefore, examines both domains together, emphasizing their interdependence in the pediatric context. We describe core concepts and workflows of data mining in pediatric radiology, including data collection, linkage, annotation, analysis, validation, and governance, and outline how modern AI tools such as deep learning, large language models, multimodal fusion, and federated learning support advanced pattern discovery across limited and heterogeneous pediatric datasets. We summarize current and emerging clinical applications across diagnosis, prognosis, radiation dose monitoring, operational analytics, reporting safety nets, and continual learning. We then discuss current challenges related to data quality and standardization, ethics, regulation, workflow integration, resource disparities, sustainability, and explainability. Finally, we highlight future perspectives, including synthetic data generation, foundation models, structured reporting, and pediatric-focused ethical frameworks that aim to enable safe, transparent, and equitable integration of AI-driven data mining to improve outcomes in children.