World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:06, 2024

Integrating Natural Language Processing (NLP) and Machine Learning in Lung Cancer Diagnosis

Authors : Mehrnaz Mostafavi

Abstract : The assessment and categorization of incidental lung nodules present a considerable challenge in healthcare, often necessitating resource-intensive multiple computed tomography (CT) scans for growth confirmation. This research addresses this issue by introducing a distinct computational approach leveraging radiomics and deep-learning methods. However, understanding local services is essential before implementing these advancements. With diverse tracking methods in place, there is a need for efficient and accurate identification approaches, especially in the context of managing lung nodules alongside pre-existing cancer scenarios. This study explores the integration of text-based algorithms in medical data curation, indicating their efficacy in conjunction with machine learning and deep-learning models for identifying lung nodules. Combining medical images with text data has demonstrated superior data retrieval compared to using each modality independently. While deep learning and text analysis show potential in detecting previously missed nodules, challenges persist, such as increased false positives. The presented research introduces a Structured-Query-Language (SQL) algorithm designed for identifying pulmonary nodules in a tertiary cancer center, externally validated at another hospital. Leveraging natural language processing (NLP) and machine learning, the algorithm categorizes lung nodule reports based on sentence features, aiming to facilitate research and assess clinical pathways. The hypothesis posits that the algorithm can accurately identify lung nodule CT scans and predict concerning nodule features using machine-learning classifiers. Through a retrospective observational study spanning a decade, CT scan reports were collected, and an algorithm was developed to extract and classify data. Results underscore the complexity of lung nodule cohorts in cancer centers, emphasizing the importance of careful evaluation before assuming a metastatic origin. The SQL and NLP algorithms demonstrated high accuracy in identifying lung nodule sentences, indicating potential for local service evaluation and research dataset creation. Machine-learning models exhibited strong accuracy in predicting concerning changes in lung nodule scan reports. While limitations include variability in disease group attribution, the potential for correlation rather than causality in clinical findings, and the need for further external validation, the algorithm's accuracy and potential to support clinical decision-making and healthcare automation represent a significant stride in lung nodule management and research.

Keywords : lung cancer diagnosis, structured-query-language (SQL), natural language processing (NLP), machine learning, CT scans
Conference Title : ICMLDLA 2024 : International Conference on Machine Learning, Deep Learning and Applications
Conference Location : Montreal, Canada
Conference Dates : June 13-14, 2024