

Predictive Analysis of Chest X-rays Using NLP and Large Language Models with the Indiana University Dataset and Random Forest Classifier

Authors : Azita Ramezani, Ghazal Mashhadiagha, Bahareh Sanabakhsh

Abstract : This study researches the combination of Random. Forest classifiers with large language models (LLMs) and natural language processing (NLP) to improve diagnostic accuracy in chest X-ray analysis using the Indiana University dataset. Utilizing advanced NLP techniques, the research preprocesses textual data from radiological reports to extract key features, which are then merged with image-derived data. This improved dataset is analyzed with Random Forest classifiers to predict specific clinical results, focusing on the identification of health issues and the estimation of case urgency. The findings reveal that the combination of NLP, LLMs, and machine learning not only increases diagnostic precision but also reliability, especially in quickly identifying critical conditions. Achieving an accuracy of 99.35%, the model shows significant advancements over conventional diagnostic techniques. The results emphasize the large potential of machine learning in medical imaging, suggesting that these technologies could greatly enhance clinician judgment and patient outcomes by offering quicker and more precise diagnostic approximations.

Keywords : natural language processing (NLP), large language models (LLMs), random forest classifier, chest x-ray analysis, medical imaging, diagnostic accuracy, indiana university dataset, machine learning in healthcare, predictive modeling, clinical decision support systems

Conference Title : ICMLC 2024 : International Conference on Machine Learning and Cybernetics

Conference Location : New York, United States

Conference Dates : July 11-12, 2024