

Review Article

A REVIEW ON INTEGRATING NLP WITH MACHINE LEARNING FOR HYPOTHESIS GENERATION

Gunjan Jadon, Awadh Kishor

Pacific College of Pharmacy, Udaipur

The integration of Natural Language Processing (NLP) with Machine Learning (ML) for hypothesis generation is rapidly transforming biomedical research by leveraging the wealth of unstructured information in scientific literature. This review provides a comprehensive examination of the methodologies, applications, and challenges in using NLP and ML to generate actionable scientific hypotheses. We explore key NLP techniques, including named entity recognition, relation extraction, and knowledge graph construction, that enable the structuring and extraction of valuable insights from biomedical texts. We also discuss ML approaches such as embedding-based models, clustering, and generative models that support novel discovery by identifying patterns and connections within complex datasets. Applications of NLP-ML integrations in hypothesis generation are highlighted, particularly in drug repositioning, target discovery, biomarker identification, and literature-based discovery. However, challenges such as data quality, model interpretability, and generalizability limit broader adoption and real-world impact. We conclude by discussing future directions, including advancements in self-supervised learning, cross-disciplinary data integration, and human-AI collaboration, all of which hold the potential to improve the robustness and utility of NLP-ML systems for hypothesis generation. This review aims to provide insights into the current landscape and inspire continued innovation at the intersection of NLP, ML, and biomedical research.

Key Words: Natural Language Processing (NLP), Machine Learning (ML)

www.pharmaerudition.org Nov. 2024, 14 (3), 01-16