Analysis of Opportunity in Novel Medical Research Ideas using AI Systems

Raul-Mihai RĂDULESCU*, Maia MARIN, Mark DRAGOTĂ, Izabella HANOS, Catalin-Andrei CHIRITA, and Sebastian-Aurelian ŞTEFĂNIGĂ

West University of Timişoara, Blvd. Vasile Pârvan no. 4, 300223 Timişoara, Romania.
E-mails: (*) raul.radulescu00@e-uvt.ro; maia.marin94@e-uvt.ro; mark.dragota05@e-uvt.ro; izabella.hanos04@e-uvt.ro; catalin.chirita99@e-uvt.ro; sebastian.stefaniga@e-uvt.ro

* Author to whom correspondence should be addressed;

Abstract
This project and paper present a system for automated analysis of medical research articles, designed to support literature triage and the identification of potentially high-opportunity research topics. The rapid growth of medical publications makes it difficult for researchers and students to keep up with emerging domains, distinguish mature areas from underexplored ones, and justify topic selection with evidence. The proposed project implements an end-to-end pipeline that ingests scientific article metadata, extracts core bibliographic and textual fields (title, authors, abstract, keywords), and stores the results in a structured database to enable downstream analytics. On top of this extraction layer, the system provides topic aggregation and ranking: extracted keywords and terms are normalized and grouped into candidate topics, which are then scored using a configurable "topic opportunity" function. The scoring is designed to balance signals of research saturation (e.g., publication volume within a topic) against signals of impact (e.g., citation-derived prominence, when citation data is available), allowing users to prioritize topics that appear influential yet comparatively less crowded. The architecture is modular so that extraction, normalization, topic grouping and scoring components can be swapped or tuned for different subfields or data availability constraints. The paper focuses on system design, reproducibility, and practical integration into a student, researcher or lab workflow, including transparent configuration of scoring weights and auditable intermediate outputs (extracted metadata, normalized terms, and per-topic feature values). The project aims to serve as a foundation for evidence-assisted topic selection, systematic review preparation and research landscape exploration in medical informatics.

Keywords: Medical Informatics; Literature Mining; Information Extraction; Topic Ranking; Bibliometrics; Reproducible Pipelines.