

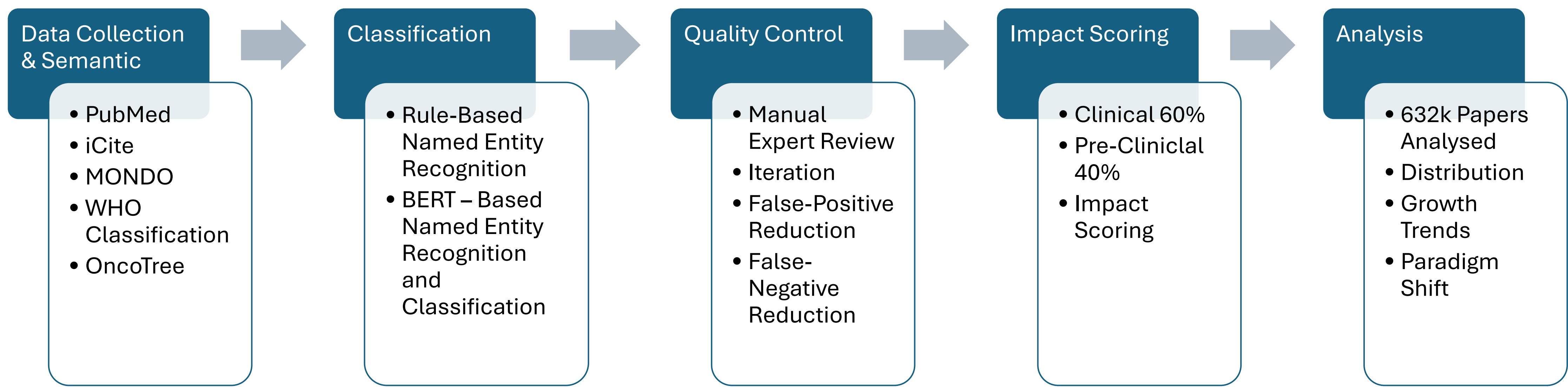
Evaluation of translational potential of research focused on hematological neoplastic diseases – results of a large-scale analysis of PubMed and iCite databases.

Authors:
Nikola Cihoric¹, Nicolas Bonadies², Fabio Dennstädt¹
¹Center for Artificial Intelligence in Radiation Oncology (CAIRO), Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
²University of Bern, Bern, Switzerland

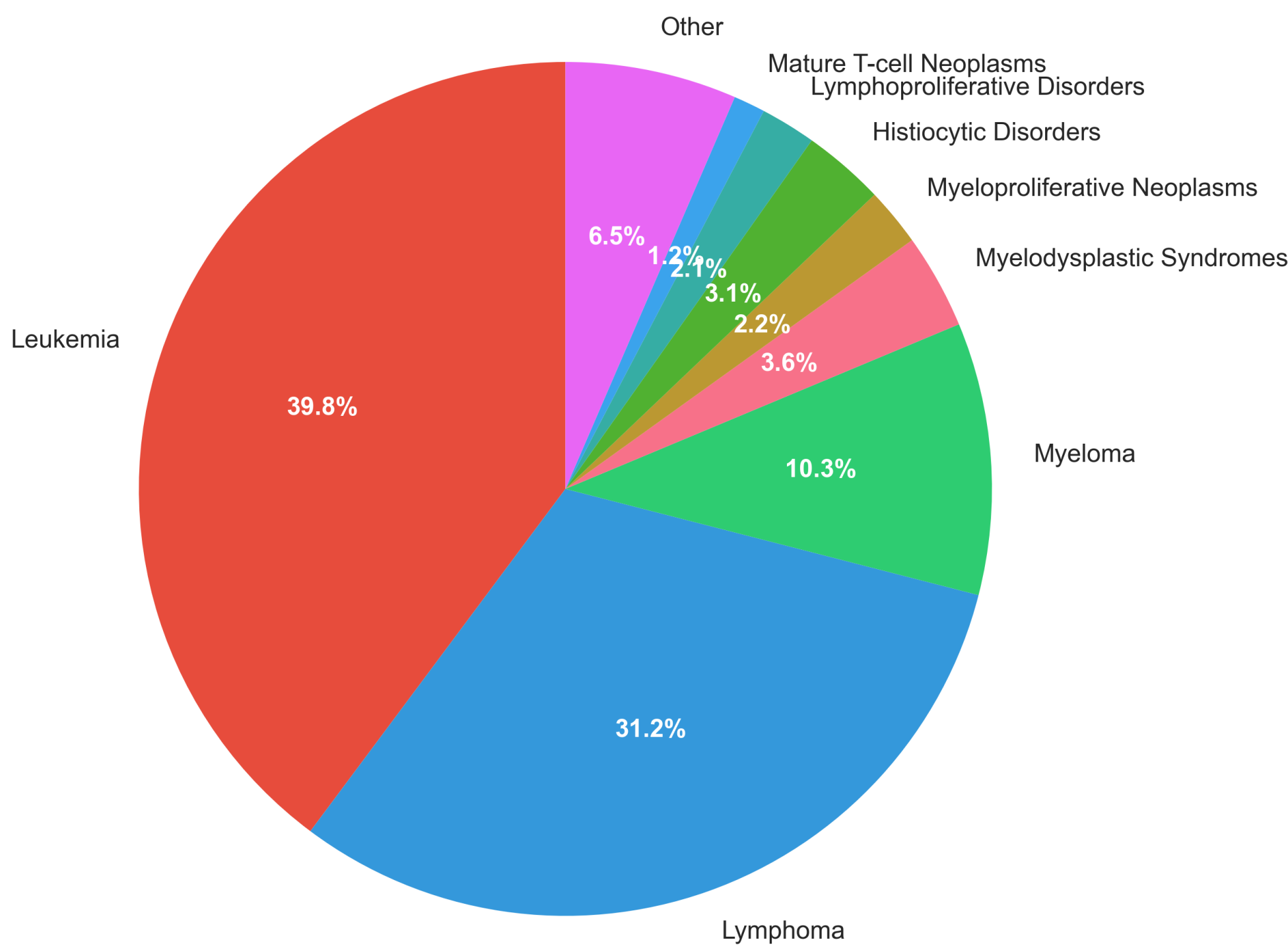
SOHC Experimental Hematology / Oncology
Presented at SOHC 2025 from 19 - 21 November 2025
SWISS ONCOLOGY & HEMATOLOGY CONGRESS

Aims: Evaluating biomedical databases identifies trends in clinical research and emerging therapeutic areas. We examined how advanced AI and ML methods can support these traditionally labor-intensive activities.

Methods: We systematically identified scientific publications on hematological neoplastic diseases (HND). We extracted terms from the WHO classification of hematopoietic and lymphoid tumors, expanded them with OncoTree and MONDO disease ontology data, and supplemented the list with non-standard terms based on expert input. We used a combination of regular expressions and BERT-based large language models to identify abstracts in PubMed (January 1950–April 2025; 38,644,614 abstracts) that mentioned HND. This process was repeated multiple times with manual expert oversight to reduce false positives and negatives. Results were combined with the iCite database for translational potential assessment, using composite scores that integrated clinical relevance (60%) and research impact (40%).

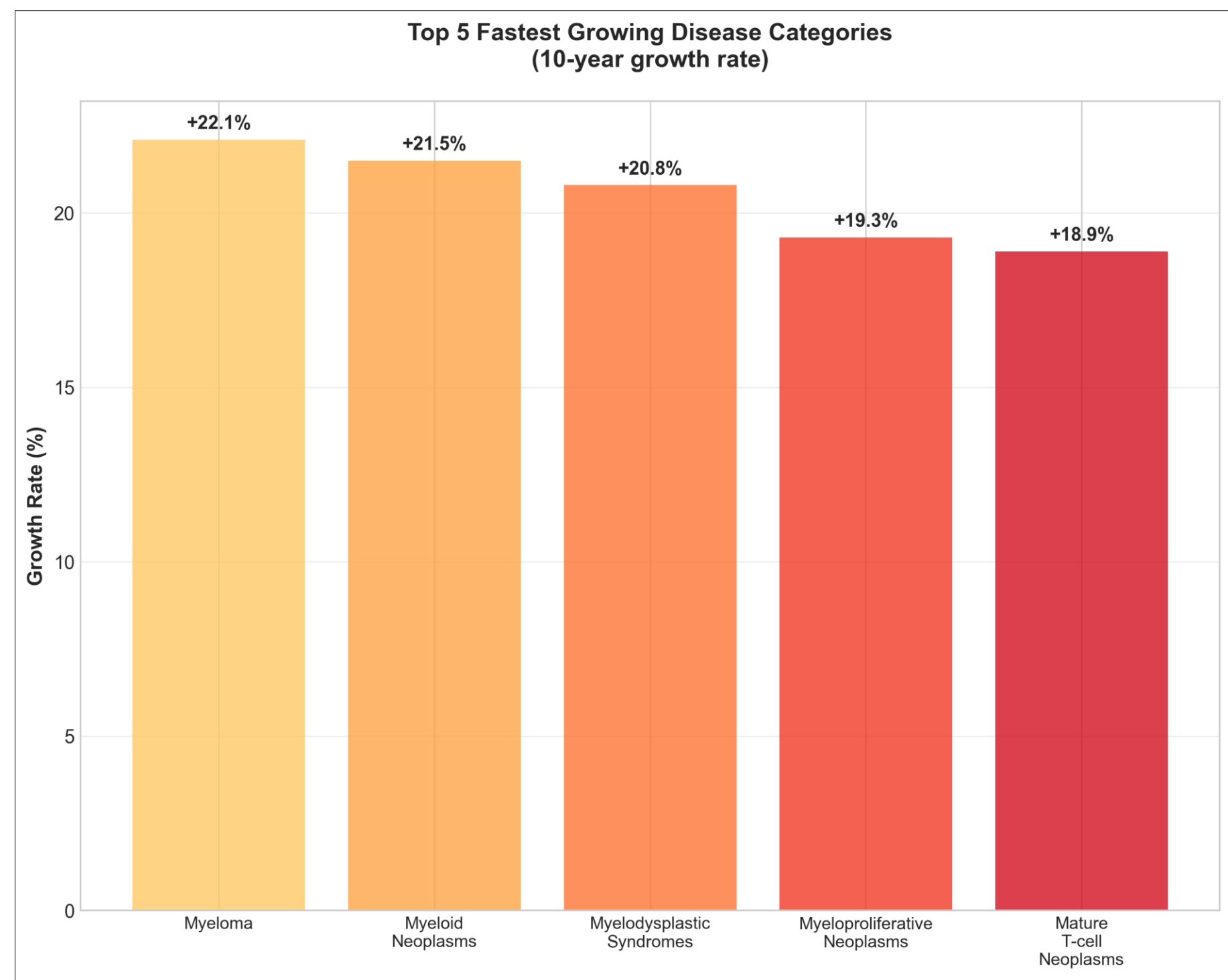


- Flow diagram of methodology
- 38.6M abstracts
 - 632K identified
 - 75-year span
 - Multiple databases and terminologies

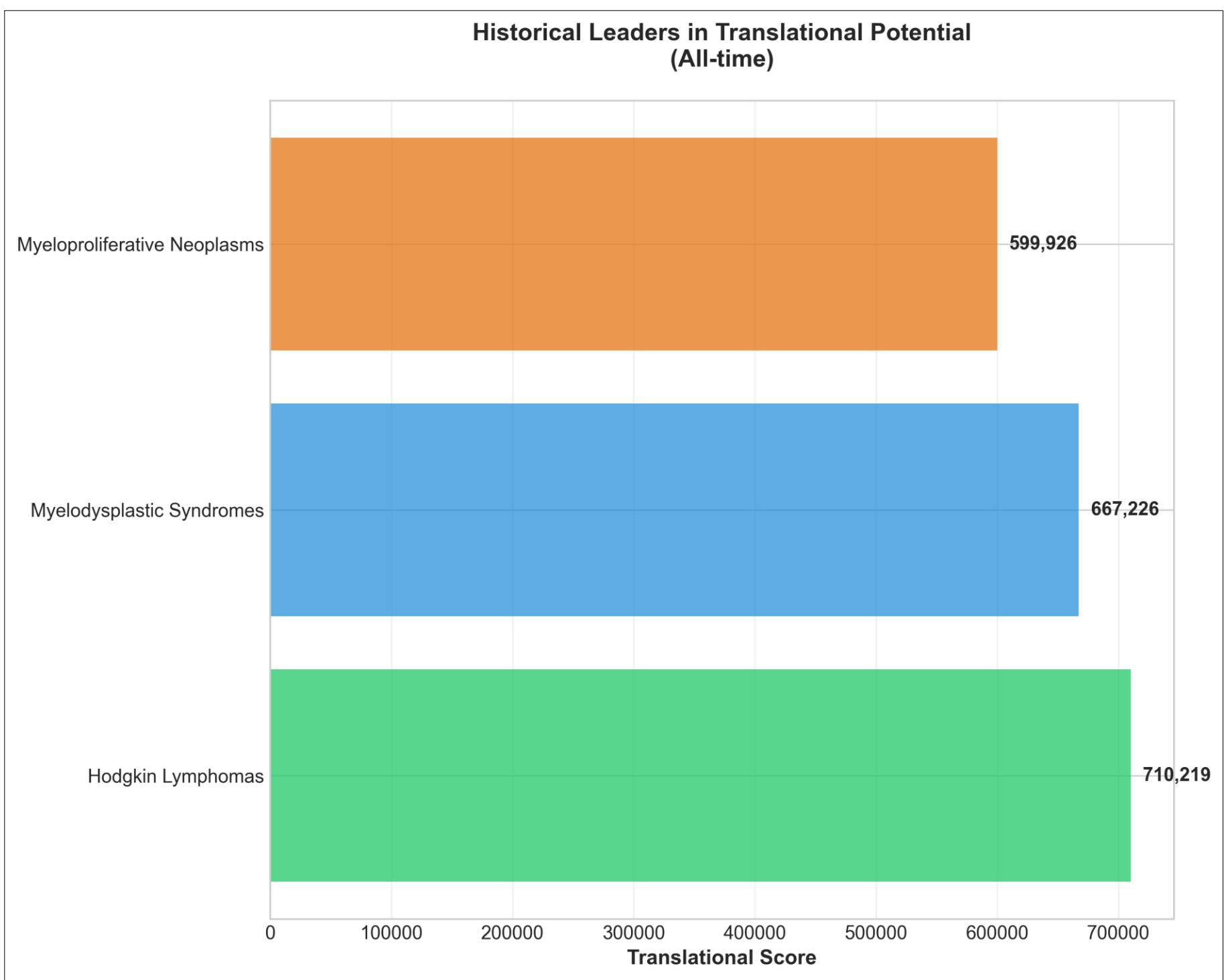


Distribution of HND Publications by Disease Type (n=632,471)

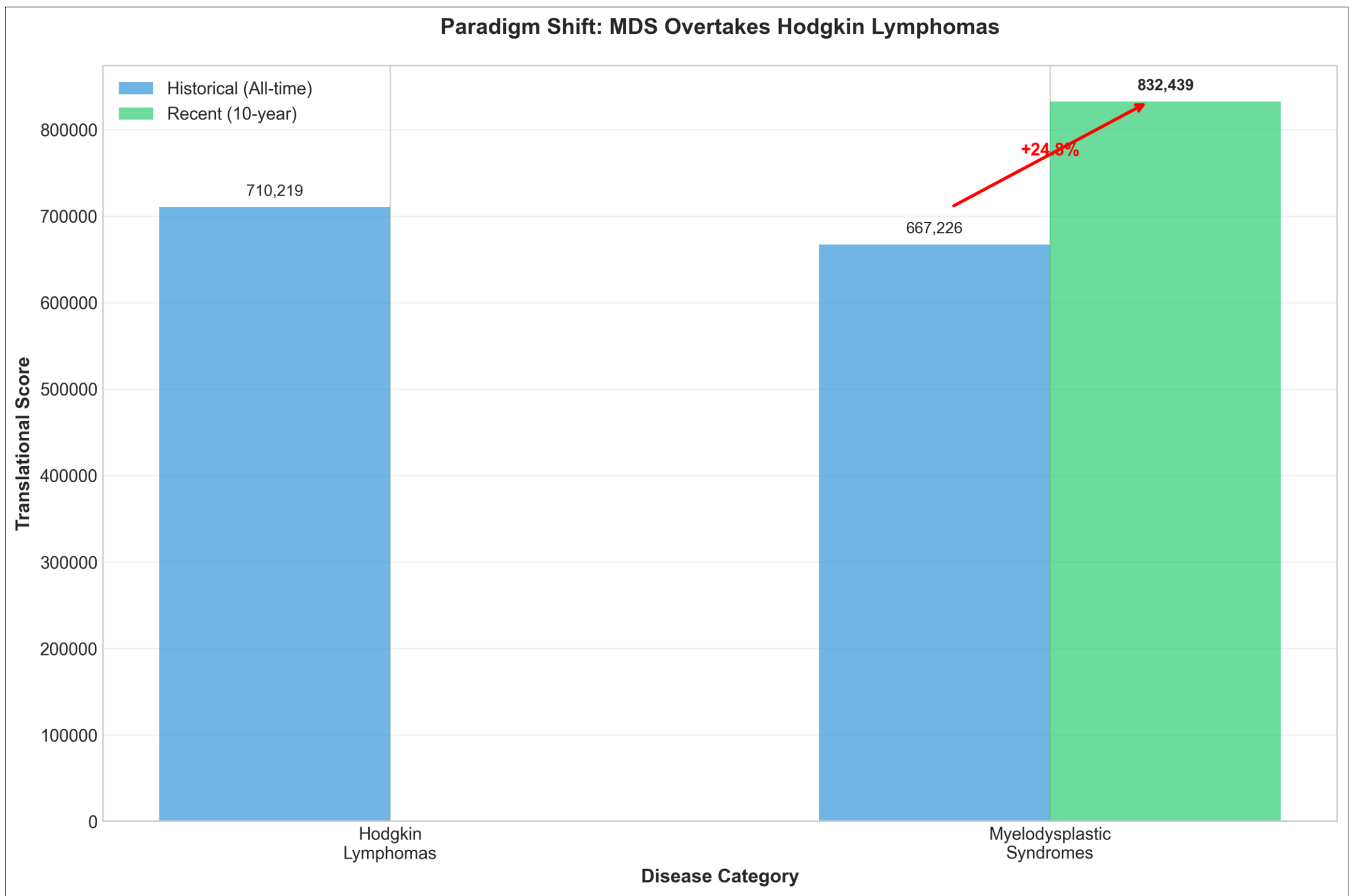
Results: Analysis of 632,471 publications revealed that HND research is dominated by leukemia (39.8%), lymphoma (31.2%), and myeloma (10.3%), accounting for more than 80% of the literature. The remaining research focused on myelodysplastic syndromes (3.6%), myeloproliferative neoplasms including myelofibrosis and polycythemia vera (2.2%), histiocytic disorders (3.1%), lymphoproliferative disorders (2.1%), and mature T-cell neoplasms such as mycosis fungoides (1.2%). Historically, Hodgkin Lymphomas showed the highest translational potential (score: 710,219) with 80.2% of studies involving humans, followed by myelodysplastic syndromes (667,226; 74.2% human studies) and myeloproliferative neoplasms (599,926; 71.3% human studies). A recent 10-year analysis revealed that myeloid neoplasms have the highest translational potential (712,343), with myelodysplastic syndromes leading the way (832,439). All major categories showed increasing potential (+16.4% to +22.1%) except histiocytic disorders (-25.8%).



- Growth Dynamics:**
- Myeloma shows the highest momentum despite being #3 in volume
 - All top categories demonstrate robust growth (18-22%)
 - Emerging areas outpacing established fields
 - Indicates shifting research priorities toward previously understudied diseases
 - Growth rates suggest increased funding and interest in these areas



- Translational Excellence:**
- Hodgkin Lymphomas historically dominated translational research
 - High scores indicate strong bench-to-bedside translation
 - All three leaders exceed 600,000 in translational scoring
 - Gap between leaders relatively small, suggesting competitive field



- Research Revolution:**
- MDS achieved a 24.8% increase in translational potential
 - First disease to surpass Hodgkin's historical dominance
 - Reflects advancement in MDS understanding and treatment options
 - Signals a major shift in research focus toward myeloid disorders
 - The recent decade shows a dramatic acceleration in MDS research quality
 - Paradigm shift likely driven by new therapeutic breakthroughs

Conclusions: While leukemia, lymphoma, and myeloma dominate the literature, myelodysplastic syndromes exhibit the highest translational potential, indicating a shift in paradigm toward myeloid disorder research. A validated framework for research prioritization requires input from the broader medical community.