

# A Novel Information Retrieval Model for Molecular Medicine Modalities with High Throughput

Van Smith\*

Department of Molecular medicine,  
Somalia

Corresponding author: Van Smith

✉ smithv@gmail.com

Department of Molecular medicine, Somalia

**Citation:** Smith V (2023) A Novel Information Retrieval Model for Molecular Medicine Modalities with High Throughput. *Transl Biomed*, Vol. 14 No. 1: 105.

## Abstract

Access to cutting-edge, high-quality molecular medicine techniques is essential for accelerating the translation of research into clinical applications. A novel semantic indexing and information retrieval model for clinical bioinformatics is then presented, followed by an explanation of why the existing databases and portals do not adequately support this goal. A model of the research processes that create and validate these objects is included in the formalism to support their systematic presentation once retrieved. The formalism provides the means for indexing a variety of relevant objects. By creating proof-of-concept encodings and visual presentations of evidence and modalities relevant to the molecular profiling and prognosis of breast cancer and diffuse large B-cell lymphoma (DLBCL), we assess the model's applicability.

**Keywords:** Information retrieval; Molecular medicine; Semantic model; Clinical bioinformatics; Predictive Computational models

**Received:** 03-Jan-2023, Manuscript No. Iptb-23-13448; **Editor assigned:** 07-Jan-2023, Pre-QC No. Iptb-23-13448 (PQ); **Reviewed:** 21-Jan-2023, QC No. Iptb-23-13448; **Revised:** 25-Jan-2023, Manuscript No. Iptb-23-13448 (R); **Published:** 31-Jan-2023, DOI: 10.36648/2172-0479.13.12.276

## Introduction

Using individual genetic variation, the field of pharmacogenomics, for instance, uses whole genome analysis technologies to predict drug treatment response and susceptibility to adverse drug reactions. An antineoplastic drug can cause adverse drug reactions in some people due to an inherited genetic trait. An enzyme that deactivates has lower expression levels in people who have the most common variant allele. The package insert must contain information about the associated genotype and dosing guidelines, as required by the FDA [1]. Other mutations are associated with favourable drug responses and a favourable clinical prognosis. On the FDA website, you can find a list of genomic biomarkers related to drugs. A typical scenario involves using patient tissue for a molecular assay. The predicted clinical outcome of the patient's disease is then computed by a decision model using the assay results [2]. For instance the FDA in the United States It takes a complex scientific method that draws from multiple overlapping sources of data describing complex interactions at the genomic, proteomic, or other levels to uncover clinically relevant knowledge from large-scale genome and molecular biology data. High-throughput experiments can

produce data with hundreds of thousands or even millions of data points per sample. Such data are challenging to process manually and necessitate complex computation. The decision models used to process the data are also complicated and come from a wide range of fields, like biostatistics and machine learning [3]. Also, there is a lot of variation in how these predictive models' validity, generalizability, and supporting evidence are evaluated. Clinical and translational researchers must have access to relevant, up-to-date, and accurate information about known molecular medicine modalities, such as research datasets, research methods, known and validated decision models, and related evidence, in order for advances in molecular medicine to become clinically applicable. As a result, it is necessary to address the significant issue of retrieving and organizing the enormous amount of data generated by molecular medicine research [4].

## Materials and Method

### Formulation of the model and proof of concept

The task of retrieving research data from the semantically complex clinical bioinformatics domain of gene expression microarrays in the diagnosis and treatment of DLBCL is the

context in which the model is described. In the beginning, we did manual reviews of papers that talked about this domain in the literature [5]. By identifying the Algorithms, Datasets, or Models that were mentioned in each of the reviewed papers, we took note of the various objects that were discussed. Conceptually, all of the papers and the union of all algorithms, datasets, and models described in the papers constitute the knowledge base's objects. Multiple papers can make reference to an algorithm, dataset, or model [6]. A deeper look at these objects revealed that each can be described by at least one context that specifies the tuple of the following elements. For instance in the Work of Wright the Bayes Classifier" algorithm was used to create and validate a model that predicts the molecular subtype of DLBCL by applying it to two gene expression datasets. A subset of the knowledge base's objects should then be returned by a query. On the left, you'll find a brief list of papers, algorithms, datasets and models that pertain to gene expression microarrays in the context of DLBCL. Additionally, we discovered that a query can be represented as either a complete or partial Context [7]. For instance, the contexts depicted by the aforementioned example queries are shown in. Using a set of canonical terms for each Context element and then indexing each object with at least one complete Context tuple is a quick and easy way to implement an indexing scheme.

### Conceptual proof: Mamma Print is a breast cancer molecular prognostic test

Using a 25,000-sequence oligonucleotide microarray, researchers in the Netherlands examined historical breast cancer tissues. Lymph Node (LN)-negative female patients under the age of 55 had 70 genes that predicted 5-year metastasis. The following three characteristics were distinguished by unsupervised hierarchical clustering: Negative estrogenic receptor status, BRCA1 mutation, and metastasis within five years. In other words, the hierarchical clustering algorithm resulted in the creation of three models. An Artificial Neural Network with a "70-gene signature" that can predict these characteristics is a supervised method of machine learning [8]. Using a leave-one-out method, this predictive model was validated internally. In addition, the researchers demonstrated that this molecular predictive model was distinct from other well-known decision models that relied solely on clinical parameters in its ability to predict metastasis. The molecular decision model not only improved the prediction of clinical outcomes in that paper, but it also predicted the same number of patients with metastasis with fewer false positives. Given the morbidity and financial costs of adjuvant chemotherapy, this is significant. External validation of the 70-gene signature model was carried out with 295 consecutive historical patients in a dataset that was distinct from the one that was used to create that signature [9]. In addition, it provided the correct decision outcome for primary tumor tissue from seven patients and matched metastatic tissue from the same patients obtained years later. This validation was of a biological hypothesis

rather than a clinical one: Rather than invasiveness as a result of cumulative mutations, the molecular subtype determines the disease's potential for metastatic spread early on.

## Discussion

Decision Models, the original algorithms that led to the creation of these models, and the methods used to validate these models are not included in Oncoming's representation and organization of oncology molecular datasets. MeSH-indexed in GEO/PubMed, datasets and papers do not explicitly link to their respective models, algorithms, or contexts [10]. Molecular clinical predictive models and the modalities that are associated with them are included in the scope of the proposed framework, which is intended to complement the resources that are already in place and broaden existing representations. In order to semantically integrate this framework with existing knowledge sources we decided to model this domain using OWL ontology. We use PubMed uids for papers and GEO accession numbers for datasets to link objects in our database to those in external databases whenever possible. The majority of clinical predictive models currently in use do not include molecular features. This information retrieval framework does not cover traditional predictive models that are only based on clinical parameters; however, classical models will only be incorporated into molecular predictive models when they are already present. For instance, we did include the Consensus model in the Mamma Print validation case study and the International Prognostic Index model in the DLBCL case study. Similarly, this framework does not cover the storage and annotation of gene signatures that predict biological behavior without clinical outcomes. Again, some molecular clinical predictive models include aspects of purely biological signatures; consequently, we will also only include those when they are present in clinical models.

## Conclusion

Although clinically oriented research into gene expression microarrays, mass spectrometry SNP arrays, and other high-throughput molecular assays has grown at an exponential rate in recent years, there is currently no general purpose system that provides researchers and clinicians with a unified and user-friendly interface for finding models, papers, data, and other relevant information in this emerging field. The complexity of the necessary functionality for such an interface is demonstrated and a framework for it is proposed in this paper. Our long-term objective is to build a system that meets this need. We created a formalism that makes it possible to store and retrieve a wide range of clinical bioinformatics objects, including decision models, published papers, datasets, and discovery and inference algorithms, as a significant first step. Automated techniques that assist in the creation and annotation of the knowledgebase are made possible by this formalism. Additionally, it permits a second level of organization of objects returned by queries based on the strength of methodological validation and their interrelationships.

## References

- 1 Li W, Fan H, Yiping L (2009) Postural Epigastric Pain as a Sign of Cytomegalovirus Gastritis in Renal Transplant Recipients: A Case-Based Review. *Transplant Proc* 41: 3956-8.
- 2 Sepkowitz KA (2001) AIDS-the first 20 years. *N Engl J Med* 344: 1764-72.
- 3 Nachega JB, Marconi VC, Gardner EM, Hong SY, Gross R et al. (2011) HIV treatment adherence, drug resistance, virologic failure: evolving concepts. *Infect Disord Drug Targets* 11: 167-74.
- 4 Nachega JB, Mills EJ, Schechter M (2010) Antiretroviral therapy adherence and retention in care in middle-income and low-income countries: current status of knowledge and research priorities. *Current Opinion in HIV and AIDS* 5: 70-77.
- 5 Burgoyne RW, Tan DH (2008) Prolongation and quality of life for HIV-infected adults treated with highly active antiretroviral therapy (HAART): a balancing act. *J Antimicrob Chemother* 61: 469-73.
- 6 Yonath A, Bashan A (2004) Ribosomal crystallography: initiation, peptide bond formation, and amino acid polymerization are hampered by antibiotics. *Annu Rev Microbiol* 58: 233-51.
- 7 Cassells AC (2012) Pathogen and biological contamination management in plant tissue culture: phytopathogens, vitro pathogens, and vitro pests. *Plant Cell Culture Protocols. Methods mol Boil* 877: 57-80.
- 8 Song Tianyan, Mika Franziska, Lindmark Barbro, Schild Stefan, Bishop Anne et al. (2008) A new *Vibrio cholerae* sRNA modulates colonization and affects release of outer membrane vesicles. *Molecular Microbiology* 70: 100-111.
- 9 Davis B, Waldor MK (2003) Filamentous phages linked to virulence of *Vibrio cholerae*. *Curr Opin Microbiol* 6: 35-42.
- 10 Queiroz-Telles F, Fahal AH, Falci DR, Caceres DH, Chiller T et al. (2017) Neglected endemic mycoses. *The Lancet. Infectious Diseases* 17: 367-377.