Context-driven interaction retrieval and classification for modeling, curation, and reuse

Authors: Haomiao Luo[1], Casey Hansen[1], Cheryl A. Telmer[2], Difei Tang[1], Niloofar Arazkhani[1], Gaoxiang Zhou[1], Peter Spirtes[2], Natasa Miskov-Zivanov[1, +]

1 University of Pittsburgh, Pittsburgh PA
2 Carnegie Mellon University, Pittsburgh PA
+Corresponding author

Abstract

Computational modeling seeks to construct and simulate intracellular signaling networks to understand health and disease. The scientific literature contains descriptions of experimental results that can be interpreted by machines using NLP or LLMs to itemize molecular interactions. This machine readable output can then be used to assess, update or improve existing biological models if there is a tool for comparing the existing model with the information extracted from the papers. Here we describe VIOLIN a tool for classifying machine outputs of molecular interactions with respect to a biological model. VIOLIN classifies interactions as corroborations, contradictions, flagged or extensions with subcategories of each class. This paper analyzes 2 different models, 9 reading sets, 2 NLP and 2 LLM tools to test VIOLIN's capabilities. The results show that VIOLIN successfully classifies interaction types and can be combined with automated filtering to provide a versatile tool for use by the systems biology community.

1 Introduction

Mechanistic models of cellular signaling networks provide valuable insights and advance understanding of the dynamics of complex systems such as cancer cells [1, 2] and immune system cells [3]. Biological signaling pathways involve thousands of molecular interactions, have been studied for decades and are described in the scientific literature and databases. This information guides the representation of biomolecular events and assists in deriving corresponding mathematical functions in executable models. The reading of papers, recognition of interactions, conversion to machine-readable representation for assembly into networks and executable models is labor intensive, and it is prone to human error, bias, and inability to survey all available published literature. Fortunately, natural language processing (NLP) methods and tools, also referred here as machine readers, such as REACH [4], TRIPS [5], MedScan [6], as well as recent large language models (LLMs) [7, 8] can process vast amounts of text in databases, identify and retrieve the information about tens and hundreds of thousands of biomolecular events that occur in cell signaling pathways. The challenge then becomes developing strategies and tools to determine which interactions retrieved from the automated platforms are credible and most useful when modeling complex systems.

An automated framework, INDRA [9], collects interactions from multiple machine readers and databases and provides interaction belief scores based on their occurrence in literature and databases and on the probability of machine reading error. The INDRA database, with its millions of curated interactions, is a comprehensive and reliable source for mechanistic modeling. However, these interactions often lack context information such as cell type or intracellular location of interaction, making it challenging to identify the most relevant

information. Carefully selecting query terms to search for and identify most relevant papers can help limit the machine reading and INDRA output to the desired context[10]. The machine reading output can also be filtered by FLUTE [11], which relies on existing knowledge bases to output only high-confidence interactions confirmed by expert opinion or experiments. Methods and tools are then needed to automate the comparison and classification of this newly obtained information within the context of existing models, with the goal of improving the models' accuracy and effectiveness.

Several open-source tools have been developed recently that aim to automatically enhance or extend existing models. The BEL enrichment workflow [12] maintains knowledge graphs using literature, CLARINET [13] and ACCORDION [14] recommend extensions of executable models based on graph features or dynamic properties, respectively. These and several other tools [15-19] integrate new information and knowledge from literature, databases, and experts, employing techniques such as clustering, graph search algorithms, genetic algorithm, and model checking. While these tools add new information to existing models, they do not utilize it to verify models or identify contradictions in knowledge.

Here, we present VIOLIN (Versatile Interaction Organizing to Leverage Information in Networks), a tool that can automatically classify, in the context of a given baseline model, large sets of interactions such as those retrieved from literature by machine readers. VIOLIN judges whether the new interactions support or further enrich the information in the model (corroborations), thereby verifying and curating the model. VIOLIN identifies gaps and issues in the knowledge, published information, and models (contradictions) and indicates that the newly obtained information requires further investigation (flagged). VIOLIN also provides new connections for extending the models (extensions). In addition to corroborations, contradictions, flagged and extension categories, VIOLIN uses interaction attributes to subcategorize interactions. Furthermore, VIOLIN has the capability to identify and search through paths within a model when comparing to new interaction lists, identifying new potential feedback and feed-forward loops. Finally, VIOLIN performs the comparison and classification of large new interaction lists within a given context at least three orders of magnitude faster than a human, thus enabling reliable high throughput modeling and curation tasks that would not be feasible otherwise.

2 Methods

2.1 Automated transfer of knowledge from published sources to computational models

As illustrated in Figure 1, VIOLIN takes as input two interaction lists, *new* and *baseline*, and generates as output a classified interaction list. New interaction lists can be automatically retrieved from literature and interaction databases using traditional NLP tools and frameworks like INDRA, or by prompting and querying LLMs and processing their responses. Alternatively, these lists can be manually compiled by

experts or from various knowledge sources. The baseline interaction list can contain interactions from a baseline model that is assembled from expert knowledge, experimental data, or obtained from online databases such as NDEx [20], KEGG [21], and Reactome [22, 23]. VIOLIN compares interactions in the new interaction list with the interactions in the baseline model and classifies them into one of the four main categories: corroboration, contradiction, flagged, and extension.

To represent interaction lists and baseline models, VIOLIN uses BioRECIPE [24], a standardized tabular format that organizes the information about network topology, attributes of network components, and mathematical functions of executable models. The BioRECIPE format allows a human user to easily preview and modify any model component, while it is at the same time readable by machines and can be processed by a suite of model development and analysis tools [25, 26]. Using this format ensures that VIOLIN is compatible with numerous interaction and model sources, as well as modeling tools [24]. This compatibility fully automates the flow outlined in Figure 1, enabling faster and more accurate creation and updating of mechanistic models.

The output of VIOLIN, interactions from the new interaction list classified and scored in the context of the baseline model, provides valuable information for addressing various modeling and curation goals. VIOLIN automatically verifies the interactions in the baseline model by identifying corroborations [27]. Details about contradictions that VIOLIN provides can help identify knowledge gaps and guide resolution of conflicts in knowledge and published information [28, 29]. Small numbers of flagged interactions can be evaluated manually if they are of interest. The model extension tools [13, 15, 17-19, 30] that explore a large number of new interaction candidates in an automated manner use only extensions that VIOLIN selects from interaction lists.

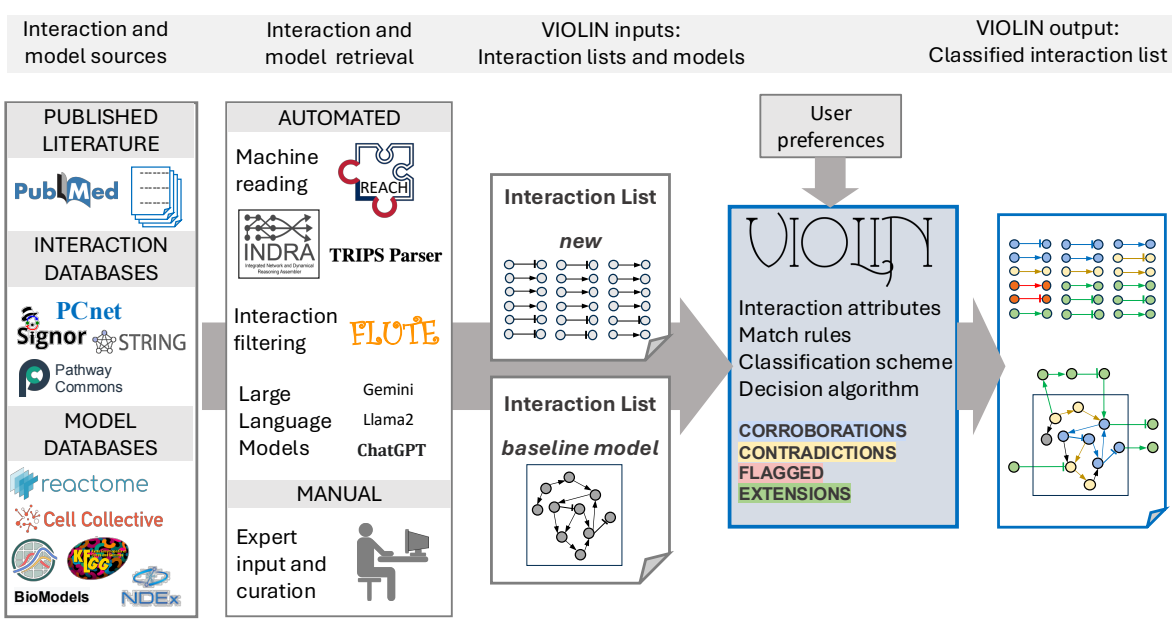


Figure 1. Overview of the knowledge sources and knowledge retrieval approaches and tools compatible with VIOLIN and the inputs and outputs of VIOLIN.

2.2 A flexible and detailed interaction classification method

VIOLIN evaluates interactions in the new interaction list by judging attribute matches between these interactions and those in the baseline model. Figure 2 lists all attributes used by VIOLIN, including several examples of attribute values.

		Interaction attributes		Interaction attribute examples				
		Attribute name	Attribute symbol	1	2	3	4	5
Elements (nodes, v)	Regulator (source node, v_s)	Name*	a_s^{name}	CK1	mTOR	Akt	Resveratrol	RAS
		Type*	a_s^{type}	protein	protein	protein family	chemical	protein family
		Subtype	$a_s^{subtype}$	kinase	kinase	kinase	antioxidant	GTPase
		HGNC Symbol	$a_s^{HGNCsymbol}$	CSNK1A1	MTOR	AKT1,AKT2, AKT3	N/A	HRAS, KRAS, NRAS
		Database*	$a_s^{database}$	UniProt	UniProt	FamPlex	CHEBI	PFAM
		ID*	a_s^{ID}	P48729	P42345	AKT	445154	PF00071
		Compartment	$a_s^{compartment}$	cytoplasm	cytoplasm, nucleus	cytoplasm, membrane	N/A	other
		Compartment ID	$a_s^{compartmentID}$	GO: 0005737	GO:0005737, GO:0005634	GO:0005737, GO:0016020	N/A	GO:0016020
	Regulated (target node, v_t)	Name*	a_t^{name}	APC	Chk1	GSK3beta	PTEN	p110gamma
		Type*	a_t^{type}	protein	protein	protein	protein	protein
		Subtype	$a_t^{subtype}$	tumor suppressor	kinase	kinase	phosphatase	enzyme
		HGNC Symbol	$a_t^{HGNCsymbol}$	APC	CHEK1	GSK3B	PTEN	PIK3CG
		Database*	$a_t^{database}$	UniProt	UniProt	UniProt	UniProt	UniProt
		ID*	a_t^{ID}	P25054	O14757	P49841	P60484	P48736
		Compartment	$a_t^{location}$	cytoplasm	cytoplasm, centrosome	cytoplasm, nucleus, membrane	cytoplasm, nucleus	other
		Compartment ID	$a_t^{locationID}$	GO: 0005737	GO: 0005737, GO:0005813	GO:0005737, GO:0005634, GO:0016020	GO:0005737, GO:0005634	GO:0016020
Interaction (edge, e)	Influence	Sign*	a_e^{sign}	positive	positive	negative	positive	positive
		Connection Type*	$a_e^{connectiontype}$	direct	indirect	indirect	indirect	indirect
		Mechanism	$a_e^{mechanism}$	phosphorylation	amount	phosphorylation	transcription	activation
		Site	a_e^{site}	S1504, S150, S1507, S1510	N/A	S9	N/A	N/A
	Context	Cell Line	$a_e^{cellline}$	SW480	cellosaurus:CVCL_0045	CHO	LNCaP, DU145	COS-7
		Cell Type	$a_e^{celltype}$	colorectal cancer	embryonic	epithelial	prostate cancer	fibroblast-like
		Tissue Type	$a_e^{tissuetype}$	large intestine	kidney	ovarian	brain, lymph node	kidney
		Organism	$a_e^{organism}$	human	human	human, mouse, hamster	human	human, monkey
	Proven.	Score	a_e^{score}	INDRA: 0.999	STRING: 0.763, INDRA:0.982	STRING: 0.996, INDRA:0.999	STITCH: 0.963, INDRA:0.999	INDRA:0.997
		Source	a_e^{source}	literature, expert, database	literature, expert, database	literature, expert, database	literature, expert, database	literature, expert, database
		Statements	$a_e^{statements}$	Sentence 1	Sentence 2	Sentence 3	Sentence 4	Sentence 5
		Paper IDs	$a_e^{paperIDs}$	PMC2654145, PMID11487578	PMC4381605	PMC1403772, PMC3535741	PMC3181262, PMC2957324	PMC2652403

Sentence examples:
1. "This may be analogous to parallel mechanisms that promote GSK3 phosphorylation of beta-catenin in the absence of Wnt stimulation, such as by GSK3 and CK1 phosphorylation of Axin and APC."
2. "mTOR inhibition in HEK293 cells significantly reduced the total Chk1 level"
3. "Therefore, given that Akt phosphorylates and inactivates GSK3beta, we hypothesized that Akt dependent inactivation of GSK3beta might be responsible for Notch potentiation."
4. "Our results demonstrate that resveratrol induced the expression of PTEN…"
5. "Ras activates p110gamma at the level of the membrane, by allosteric modulation and/or reorientation of the p110gamma…"

Figure 2. Interaction attributes used by the decision algorithm and classification scheme in VIOLIN. Attributes that are required to match in order for two interactions to satisfy the necessary match condition are indicated with '*'. Five examples of interactions with their attribute values.

Two elements satisfy the *necessary element match condition* if they have the same name or same ID, and they have the same type. Two interactions satisfy the *necessary interaction match condition* if their elements satisfy the necessary condition to match and they have same interaction sign. If the corresponding attributes in two interactions are different, we refer to this as *attribute mismatch*. When the attribute value in one of the two interactions being compared is 'empty', this will not be considered an attribute mismatch. The provenance attributes (score, source, statements, paper IDs) are not used in the match decisions. Formal definitions and the attribute notation were introduced in the BioRECEPE documentation [26].

In Figure 3, we illustrate VIOLIN's classification algorithm using a decision tree, which leads to the four classifications categories. Figure 3 also shows three different classification schemes (CS1, CS2, CS3). These schemes can be adjusted by users according to their needs and preferences (indicated by an additional

input to VIOLIN in Figure 1). In Section 3.6, we discuss the outcomes of different classification decisions. In Figure 4, we show examples of classification sub-categories.

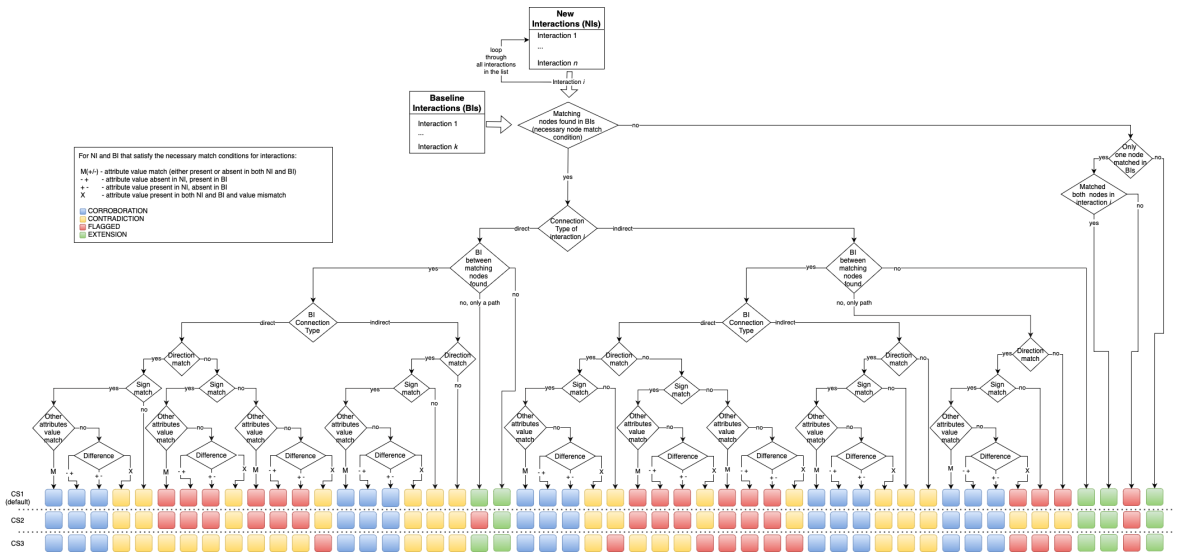


Figure 3. VIOLIN's decision algorithm, including the default classification scheme (CS1) and two additional example schemes (CS2 and CS3).

To classify an interaction in the new interaction list as a corroboration, it must satisfy at least the necessary conditions to match an interaction in the baseline model. Within the corroboration classification category,

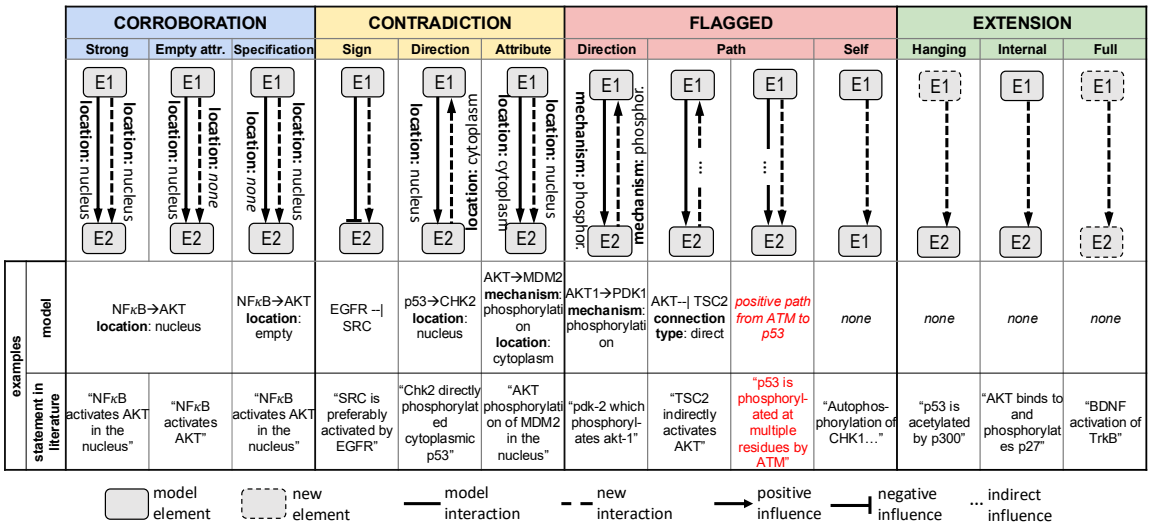


Figure 4. Classification categories and sub-categories, with examples.

we distinguish between strong, empty-attribute, indirect-interaction, path, and specification corroborations. An interaction in the new interaction list is classified as a contradiction if its elements match two elements in the baseline model but mismatch either the interaction sign or direction, or other interaction attributes, indicating three sub-categories of contradictions. An interaction in the new interaction list is classified as an extension when it introduces a new non-contradictory interaction to the baseline model, and we further distinguish between hanging, internal, and full extensions. An interaction in the new interaction list is classified as flagged if it cannot automatically meet the criteria of the other three classification categories, i.e. it must be judged manually. Depending on the classification scheme, the number of flagged interactionon types can vary (see Section 3.6).

2.3 Selection of baseline models

To demonstrate VIOLIN's capabilities, we selected two baseline models that differ in size and complexity. The first model is a discrete model of the biological pathways in the melanoma SkMel-133 cell line. This model (we will refer to it as model A) was created using the model [31] by removing the immune cells and incorporating mutations of the melanoma cell line used in Korkut et al. [32], in which the authors used reverse phase protein arrays and quantitative phenotypic assays to study drug perturbations. The second model (we will refer to it as model B) is a model of the circuitry that controls the differentiation of naïve T cells into regulatory (Treg) and helper (Th) cells, described in [3]. The number of elements and interactions in these two models are listed in Figure 5A.

2.4 Design of queries and selection of paper sets

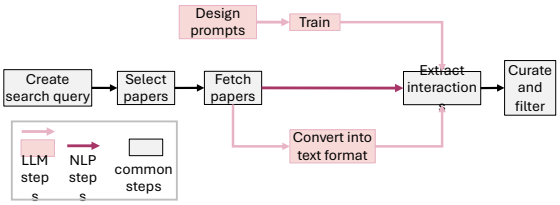
We designed search queries relevant to the context of the melanoma cell line model ($_{A}$) and in the context of the T-cell differentiation model ($_{B}$), except for the two ($_{B*}$) that were deliberately designated as irrelevant to the model. Through this literature search, we collected nine sets of paper IDs. The paper set names, literature search methods and query terms are listed in Figure 5.

For our largest paper set, S_{A1}, we used a general search query in PubMed Central [33] that included the melanoma context term and many relevant protein terms, as indicated in Figure 5A. For S_{A2}, we used the REACH Explorer tool [4] and 13 protein queries, where we first selected top 10 papers that were returned by the REACH Explorer for each protein query. In the case where 10 papers were not found, we obtained additional papers using the Fetch tool [4].

A.

Baseline model	Biological context	Number of nodes	Number of edges	Paper set	Search tool	Query
A	Melanoma cell line SkMel-133	179	266	A1	PubMed search	melanoma AND (p70S6K OR S6 OR gsk OR gsk3 OR gsk3a OR gsk3b OR src OR 4ebp1 OR "eIF4E-binding protein 1" OR PHASI OR ybi OR YBX1 OR NSEP1 OR YB1 OR CRD OR Y-box OR SkMel-133)
				A2	REACH Explorer, Fetch	MEK, ERK, AKT, GSK3, P70RSK, S6, CDK4, 4EBP1, YB1, SRC, CHK2, MTOR, PI3K
				A3	REACH Explorer	MAPK/ERK pathway
				A4	REACH Explorer	RPS6K1
B	Naive T-cell differentiation to Treg and Th	39	71	B1	REACH Explorer	PTEN
				B2	Hawse et al. references	--
				B3	REACH Explorer	T-cell, PTEN, AKT, FOXO
				B*1	REACH Explorer	Breast Cancer, DNA repair, Autophagy, Cancer
				B*2	REACH Explorer	DNA repair, BRCA1, ADAM17, inflammation

B.



C.

Reader	Type	PubMed access	Paper search	Database access	Open access	API	Paper format	Training	Additional code
REACH	Traditional NLP	✓	✓	✓	✓	✓	✓	X	X
INDRA	Traditional NLP	✓	✓	✓	✓	✓	✓	X	X
GPT-3.5	LLM	X	X	X	X	✓	✓	✓	✓
Llama 2	LLM	X	X	X	✓	✓	✓	✓	✓

Figure 5. Summary knowledge sources and retrieval. A. Baseline models and interaction lists used. B. Overview of the automated retrieval of interactions with traditional NLP tools and LLMs and filtering of retrieved interactions.

Paper sets S_{A3}, S_{A4}, S_{B1}, and S_{B3} were obtained by selecting papers with the REACH Explorer tool using lists of relevant pathway-related terms. S_{A3} and S_{A4} were created without using the melanoma context. The MAPK/ERK pathway has been studied for decades and this search returned many more papers (S_{A3}) than the RPS6K1 query (S_{A4}). RPS6K1 is a new HGNC name associated with many synonyms (S6K; PS6K; S6K1; STK14A; p70-S6K; p70 S6KA; p70-alpha; S6K-beta-1; p70(S6K)-alpha) that perhaps the REACH Explorer does not have in its dictionary (PubMed search returned 4,092 papers). To obtain S_{B1} we used a single term PTEN, and then added the T-cell context and a few additional relevant protein terms for S_{B3}.

The set S_{B2} includes papers cited by Hawse et al. [34], which extended the analysis of the T cell model from [3] that we employ here. We collected papers in sets S_{B*1} and S_{B*2} with queries that include elements and pathways found in breast cancer literature and not relevant in the T-cell differentiation context, to be used as "negative sets" when compared to the T-cell model. This was an important test of the accuracy of VIOLIN, as these negative sets would be expected to be predominantly information unrelated to T cell differentiation and therefore judged as extensions.

2.5 Interaction lists generated using NLP and LLM tools

We investigated the utility of traditional NLP-based tools (REACH and INDRA) and recent LLM-based tools (GPT 3.5-turbo and Llama 2-7B) for retrieving interaction lists from the collected paper sets. The

difference in steps between NLP-based and LLM-based tools is outlined in Figure 5B and the tool capabilities relevant to interaction retrieval are summarized in Figure 5C.

The first set of interaction lists was obtained by extracting interactions from papers with REACH machine reader that relies on the traditional NLP methods [4, 35]. To collect the second set of interaction lists, we used INDRA, which retrieves interactions from REACH, other readers, and databases. To retrieve the third and fourth set of interaction lists, we used two LLMs, ChatGPT 3.5 and Llama 2. While REACH and INDRA already have integrated mechanisms for acquiring publication text when provided with paper IDs, to use LLMs, we developed a workflow that first obtains papers, converts them into text format, conducts a few-shot prompting of LLMs, and then inputs paper texts into LLMs for interaction extraction. The output from REACH and INDRA is converted to the BioRECIPE format with existing tool [25]. We have also developed a converter from the structured LLM output to the BioRECIPE tabular format. All files and scripts are available in the VIOLIN GitHub repository [36].

2.6 Automated filtering and manual curation of interaction lists

The original nine interaction lists were further processed using FLUTE machine filtering tool [11] and manual curation.

FLUTE filters out the interactions that are not found in at least one of the supported databases, Gene Ontology Annotation [37], STITCH [38], BioGrid [39], Reactome [22], and STRING [40]. FLUTE was applied to the original interaction lists with the goal to (i) investigate integration of VIOLIN with automated filtering and (ii) compare the outcomes of manually curated and automatically filtered interaction lists, with FLUTE taking advantage of the information curated in knowledge databases.

We conducted manual curation of two interactions lists, R_{A2} and R_{B2}. Only two lists were used for this analysis due to the significant time required for this manual process. Two criteria were used for manual curation: M1 - removal of all interactions that include biological processes, chemicals, or 'other' type of elements and M2 - removal of all interactions that were classified as corroborations or contradictions, keeping extensions only. The manually curated interaction lists served two additional purposes: (iii) to compare the VIOLIN output of the raw data to the manually curated interaction lists, and (iv) to test VIOLIN against expert judgment, i.e., whether VIOLIN can identify relationships and extraction errors missed by expert judgment.

2.7 Implementation and documentation

VIOLIN was written as a Python package, with pandas and NumPy dependencies, and can be found at GitHub repository [36]. The associated documentation can be found at ReadtheDocs webpage [41]. While

VIOLIN does not link to machine readers or information assembly tools directly, it is compatible with many forms of input. VIOLIN is compatible with the BioRECIPE representation format, described in detail [24]. Through the BioRECIPE format and available translators, VIOLIN is also compatible with a number of other interaction representation formats (e.g., SBML [42], SBML-qual [43], BEL [44], BioPax [45], JSON [46], SIF [47]), databases [9, 20, 22, 23, 48] and tools [13, 16, 49, 50]. A tutorial Jupyter notebook demonstrating the use of VIOLIN and all examples discussed here, including lists of paper IDs, interaction lists, and models are available in the VIOLIN GitHub repository.

3 Results

3.1 Retrieval of relevant interaction information varies substantially in precision and speed

VIOLIN classifies interaction lists written in the BioRECIPE format, irrespective of whether they were created manually or automatically. While manual creation of these lists is more precise and devoid of irrelevant entries, it is also impractically slow and prone to unintended human error. In contrast, automated retrieval of interactions by NLP algorithms is fast but often less accurate and cluttered with meaningless content. Traditional NLP tools in the biology domain, designed to extract information about biochemical interactions and trained on domain-specific ontologies, are now being outperformed by LLMs. These models, trained on vastly larger knowledge bases, often generate more relevant content. To evaluate the quality of input that REACH, INDRA, ChatGPT 3.5, and Llama 2 produce for VIOLIN, we examined the number of interactions they retrieve per paper, the runtime for the nine papers sets, and the occurrence rate of attribute values in their output.

Figure 6A shows the relationship between the number of papers processed and the interactions retrieved across all four tools and for all nine sets. REACH consistently extracts the highest number of interactions from most paper sets, often surpassing Llama 2 by an order of magnitude, which retrieves significantly fewer interactions compared to all other readers. Conversely, as depicted in Figure 6B, Llama 2 demonstrates a notably higher efficiency in producing values for most interaction attributes compared to REACH, and also outperforms INDRA and ChatGPT-3.5 in this regard. Additionally, Figure 6B reveals substantial variability in the presence of attribute values within the outputs of these tools, suggesting that combining them could yield interaction lists that are not only accurate but also richly detailed with attribute values. Finally, while REACH and INDRA process papers and retrieve interaction lists relatively quickly—taking approximately 30 minutes and 5 minutes respectively for 15 papers—GPT 3.5 requires about 1hour , and Llama 2 is the slowest, needing 1.2 hours. Therefore, in the case of paper set S_{A1}, we were able to

obtain outputs from REACH and INDRA and not from GPT 3.5 (not free of charge when accessed through API) or Llama 2 (although open access, it would take about 500 hours to process the entire set of papers).

3.2 Overview of classification results

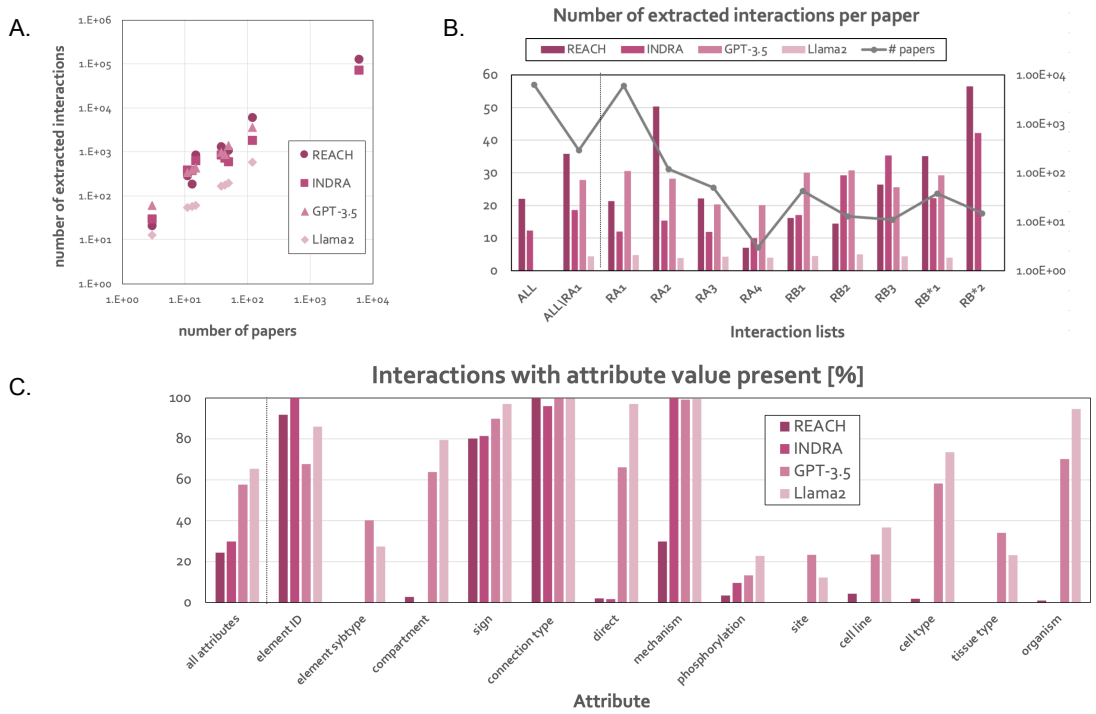


Figure 6. Overview of information retrieval by different readers and across different paper sets

We show in Figure 7A the classification results for the original interaction lists in groups R_A and R_B and for their filtered and curated subsets.

The distribution of classification subcategories varies across interaction lists, due to the size, context, and the manner of obtaining these lists. The two largest lists, R_{A1} (>10^5 interactions in REACH output) and R_{A2} (~6,300 interactions in REACH output) have relatively similar distributions of subcategories, while the distributions for lists R_{A3} and R_{A4} differ. This is due to both the size and context of the interaction lists. The MAPK/ERK pathway query leads to the R_{A3} list with 786 unique interactions, which is expected given that the changes in this pathway are found in many cancers. MAPK and ERK alone are involved in 15 of the melanoma cell baseline model interactions. In contrast, the query that leads to R_{A4} with only 20 unique interactions, includes a common kinase RPS6KB1, which is involved in only two baseline model

interactions. All three interaction lists in the T cell model case study (R_{B1}, R_{B2}, and R_{B3}) have similar distribution of classification subcategories, due to the specificity of queries.

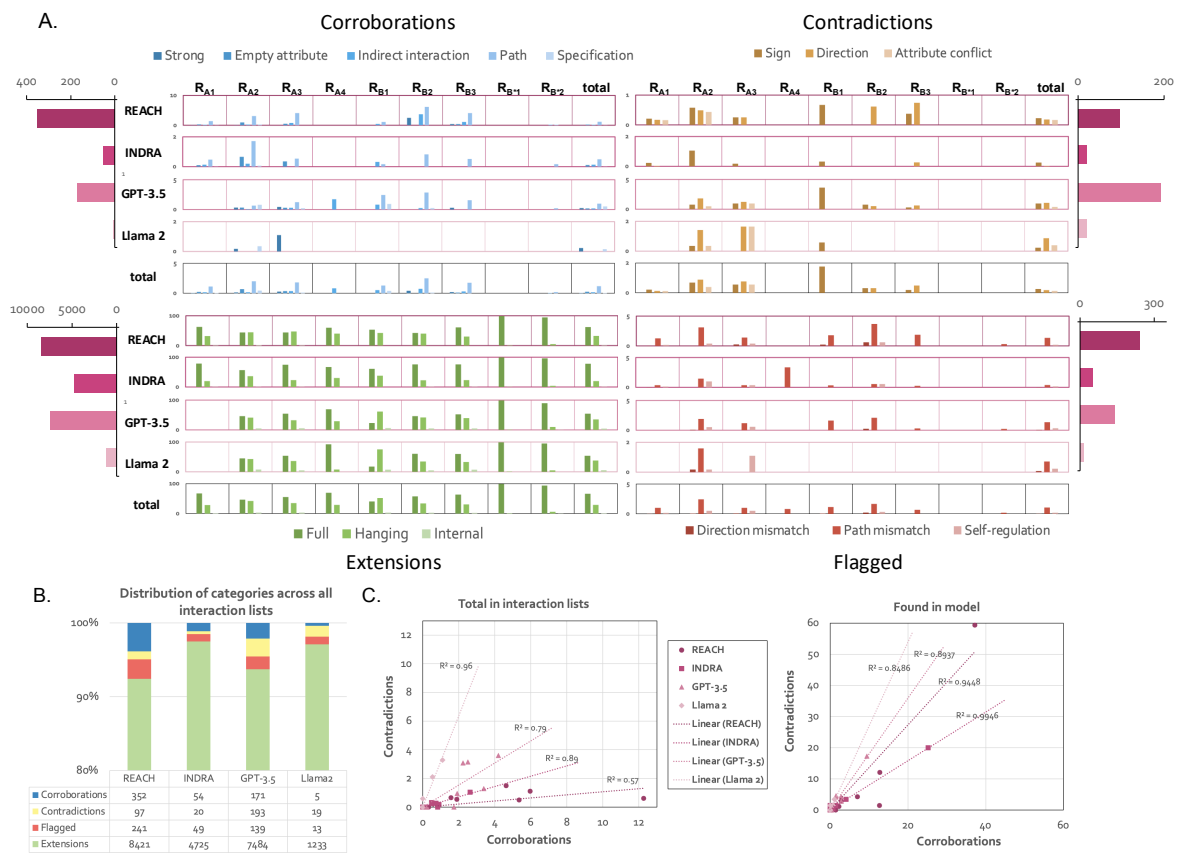


Figure 7. Overview of VIOLIN output. A. Classification results for all nine interaction lists, and all four readers, across all subcategories. Table with bar charts: individual interaction list RA and RB columns – bar heights represent percentages out of total number of interactions for an individual interaction list across all readers; bottom right corner of each category table – bar heights represent percentages for subcategories out of total number of interactions across all lists and all readers. Bar charts on left and right show the number (not percent) of interactions retrieved by each reader in each category. B. Distribution of classification categories normalized to the total number of interactions retrieved by each reader (extensions are prevalent, so the figures shows only 80-100%). Table shows the number of interactions (not percentages). C. Correlations between corroborations and contradictions in reader output.

3.3 Knowledge specification, model curation and verification with VIOLIN

Corroborations, contradictions and flagged interactions provide information about the baseline model and can serve to verify interactions (corroborations), indicate where there might be errors (contradictions) and provide flags for apparent ambiguity (flagged).

In our corpus, the corroboration subcategory distribution is similar for all interaction lists, except R_{A4} (Figure 7A). The most prevalent corroboration subcategory across all interaction lists is the path corroboration, which includes interactions that do not match any interactions in the model but contain

elements that match source and target elements of a multi-edge path in the model. The second largest subcategory of corroborations is indirect interaction corroboration, where a new indirect interaction matches a baseline model interaction, and does not contradict any other attributes, due to many of the evaluated interaction lists not having any additional interaction attributes. Although the number of corroborations found in interaction lists is not negligible, our results indicate that there are many repetitions and only a few distinctive interactions. These results suggest that to fully validate the model, a more comprehensive or more precise literature search should be conducted.

Of the contradictions, the predominant subcategory is a sign contradiction. Direction contradictions make up either the smallest number of contradictions or are not found in the contradictions at all. This is most likely due to either very few contradictions overall or the classification of a similar mismatch under one of the flagged subcategories. Similar to corroborations, there are also repeated interactions in this category and a much smaller number of unique interactions.

The predominant subcategory of the flagged interactions is the path mismatch subcategory, an indirect interaction which corresponds to a path with one or more mismatched attributes. This is expected, since the machine readers default to indirect connection type and self-regulations are not common, reducing the probability for direction mismatch and self-regulation subcategories.

3.4 The extension category provides new information for modeling

For each interaction list, the majority of the interactions were classified as extensions. These results emphasize that a significant amount of new "candidate" knowledge can be found in literature.

As can be seen in Figure 7A, the extension subcategory distribution highlights the prevalence of the hanging subcategory in all R_A and R_B interaction lists, confirming that, with the selection of relevant literature, the majority of interactions are connected to the baseline model. The next most common subcategory of extensions in the VIOLIN output are full extensions. These extensions are disconnected from the baseline model and can be irrelevant or possibly connected to a hanging extension for incorporation into a new path of interactions. We also note here that some of the full extensions result from erroneous grounding of entities by machine readers, which can lead to entities being identified as "new" instead of matching an element in the baseline model. As expected, in the two "negative" lists R_{B*1} and R_{B*2}, over 90% of all extensions are full extensions and not connected to the baseline model. These results demonstrate that integrating automated interaction retrieval with VIOLIN can significantly accelerate the modeling process, while also underscoring the critical role of selecting relevant literature for effective outcomes.

3.5 VIOLIN's classification uncovers advantages and drawbacks of interaction retrieval approaches

The classification results for interaction lists in groups R_A and R_B highlight the importance of carefully selecting the most relevant literature for modeling and offer insights into the most useful type of information retrieval based on the classification category of interest. In Figure 7B, we show the overall distribution of classification categories for different readers.

REACH retrieves the largest number of interactions (Figure 6A), and Figure 7B shows that it also generates interactions that need to be manually checked (flagged interactions) at the highest rate. These results suggest that, if a large pool of interactions with increased likelihood of novel interactions is the goal, then REACH is a useful tool. At the same time, although it may be able to provide more useful interactions than other tools, REACH is not very precise, often incorrectly extracting interactions. If the goal is to get a set of interactions that are highly supported by previous research, INDRA is a good choice, with a note that these interactions are not populated with a lot of metadata, as most attributes are empty (see Figure 6B). Most interactions obtained through INDRA are classified as extensions, again likely due to not reaching the necessary belief threshold.

The GPT-3.5 output is robust and diverse compared to traditional NLP tools and the other LLM tool. While its rate of interaction retrieval per paper is almost as high as for REACH, among all four approaches, it produces the least percent of interactions that are flagged by VIOLIN and the highest percent of non-extension interactions. It also produces a larger percent of corroborations and contradictions compared to other tools. If GPT training is current, it could assist modelers in discovering important new findings. With better prompting and using GPT-4 we expect this performance to increase further. On the other hand, Llama extracts fewer interactions than other readers and tends to have more contradictions compared to corroborations and flagged interactions. Manual examination of Llama's output suggests that Llama 2 may not be reliable as it is likely still hallucinating contents of its output. We do expect that its performance will also increase with newer versions.

As stated previously, both corroborations and contradictions in VIOLIN's output are important for modeling – corroborations offer automated verification and curation of models and contradictions often lead to new insights. We noticed a relatively consistent ratio between corroborations and contradictions across interaction lists for all four tools and investigated this further. Figure 7C(left) illustrates this relationship between the percents of retrieved corroborations and contradictions within an interaction list across all nine lists (R_{A1}, R_{A2}, R_{A3}, R_{A4}, R_{B1}, R_{B2}, R_{B3}, R_{B*1}, R_{B*2}).

Given that models we are using in our studies were created and curated manually and with substantial input from experts, we expect that, when a large number of interactions is retrieved, many of them would be corroborations and very few would be contradictions, the latter occurring in cases where novel experimental results are published, or readers output an erroneous interaction. This is confirmed with REACH outputs, in which significantly more corroborations than contradictions are found (Figure 7C(left)). Interestingly, the actual rate of interactions corroborated in the model is much smaller and similar to the number of model interactions that are contradicted (Figure 7C(right)). INDRA is more consistent, as it both retrieves more corroborations and corroborates models more than it contradicts them.

Another important observation is that GPT outputs similar number of contradictions and corroborations across all interaction lists, with a slight tendency to contradict more individual model interactions than corroborate. These results indicate that integrating VIOLIN with GPT facilitates the automated discovery of novel and disruptive insights, potentially accelerating the pace of advancements through modeling.

Finally, Llama 2, which retrieves substantially less interactions overall, finds very few corroborations and contradictions, and consistently more contradictions than corroborations overall (Figure 7C(left)) as well as contradicting model more often than corroborating it (Figure 7C(right)).

3.6 Classification strategy guided by modeling goals

VIOLIN allows users to select setting for classifying interactions (Figure 1), which can result in varying classification outcomes. While some of the interaction attributes are critical for the classification (e.g., sign, as discussed in Section 2.2), we were interested in exploring how decisions and settings related to other attributes affect the classification outcome. In Figure 6B, we show for all nine interaction sets the percentage of "direct", interactions, interactions with non-empty location attribute, and interactions with non-empty mechanism attribute. Although the information about novel direct interactions, interaction location, and mechanism is in general very useful for modeling, the numbers in Figure 6B suggest that machine readers struggle to extract such information. Hence the depth at which VIOLIN can make comparisons is constrained due to machine reader limitations, leading to rare strong corroborations and specifications. Depending on whether VIOLIN is used for model verification, curation, or extension, adapting the classification scheme to compensate for information gaps in machine reading can offer curators and modelers increased numbers of more specificity of interactions to consider.

We explored several different strategies in terms of the non-essential attributes considered in the classification process: CA1 – only compartment considered, CA2 – compartment and mechanism considered, CA3 – compartment, mechanism, and all context attributes considered. While we did not

observe large differences in the classification of the interactions in this reading output (Figure 8), users may want to have only interactions accompanied by specific attributes.

We tested the two additional classification schemes (CS2 and CS3) that are outlined in Figure 3. The key differences between the three schemes we observed are as follows. Scheme CS2 has by far the greatest number of contradictions and fewest number of flagged interactions, shown in Figure 8A This scheme also leaves the interpretation of paths mainly up to user judgment, classifying a direct interaction with entities matching the source and target nodes of a path in the model as a flagged interaction. For direct interactions,

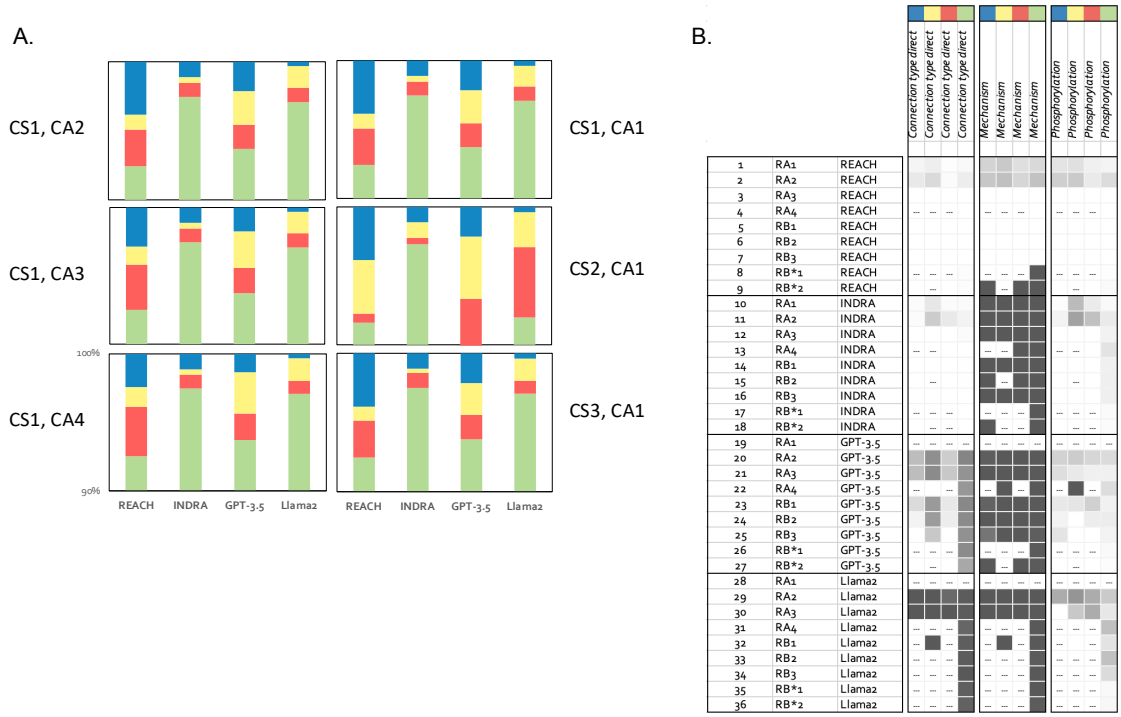


Figure 8. Evaluation of the presence of different interaction attributes and influence on classification categories. A. Changes in classification category distribution with different attribute use strategies and different classification schemes within VIOLIN's a algorithm. B. Presence of attributes and specific attribute values within classification categories across all interaction lists and all readers for scheme CS1 and strategy CA4 (all attributes considered in classification).

CS3 classifies them as flagged only when none of the interaction attributes matches attributes of a direct model interaction. However, for indirect interactions, CS3 also defines the greatest number of flagged subcategories (five, compared to two subcategories for CS1 and three for CS2). This is reflected in the classification output, as scheme CS3 produces the lowest number of contradictions and the highest number of flagged interactions. Our default scheme CS2 strikes a balance between these two schemes, making decisive judgment where possible (and likely) without making too many assumptions about the contents of the interactions.

3.7 Curation and filtering

The original R_A and R_B interaction lists that were obtained directly from NLP tools or LLM, without manual curation and removal of interactions, were assembled quickly but likely to have reading errors or irrelevant interactions.

To investigate the influence of curation on the distribution of categories and subcategories in interaction lists retrieved from literature, we modified several interaction lists using either manual curation or an automated filtering tool and then classified them with VIOLIN. We examined the changes in category distributions with curation, from original lists to manually and automatically curated lists (Figure 9). The retention of papers and interactions after automated filtering was varied between the readers and the highest retention was observed with GPT-3.5 (Figure 9A) suggesting large potential for LLM reading and extraction of signaling interactions.

4 Discussion

Decades of experimental results are published in the scientific literature using natural language and can be used to assemble computational models of biological systems. In the domain of cell biology and intracellular signaling networks, machine reading tools and LLMs can provide molecular interactions extracted from the papers in a machine readable format to be used in modeling studies. Models that currently exist have manually created and are limited in size or are available in databases, often for a specific context. In order to assess, improve or update existing models with interactions from literature there needs to be an

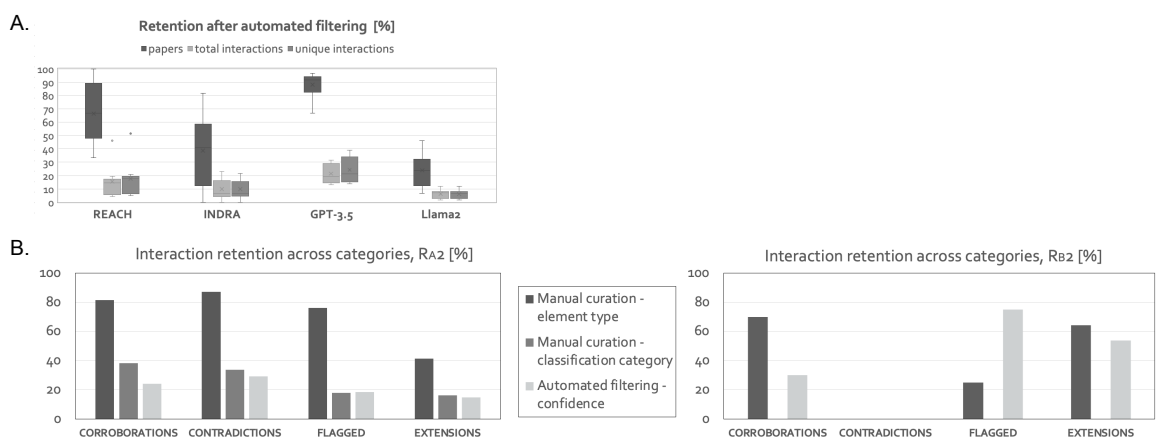


Figure 9. Influence of automated and manual curation of interaction lists on classification outcomes. A. Retention rate across nine lists compared with 4 different retrieval ways. B. Retention ratee of list RA2/RB2 across categories.

automated tool to classify the interactions that are found and extracted from the papers. VIOLIN is a versatile tool for rapidly and automatically judging and classifying large amounts of information related to interaction networks. For analysis of an existing model, interactions are corroborated, contradicted, or flagged for further investigation. For users looking to update and improve a model, VIOLIN provides extensions for further processing using tools such as ACCORDION[14].

Use of the connection type attribute and the path-finding function allow VIOLIN to more accurately make judgments not only about explicitly defined model interactions, but also about regulatory paths in the model. This component of the VIOLIN algorithm accounts for the larger context of biological interactions in the model. Our results can also be used to provide feedback to machine readers. We manually judged the VIOLIN output of two interaction sets, and identified machine reading errors. These results identify trends in machine reading errors, and support our reasoning for the flagged classification.

Our results in section 3.1 suggest that machine readers have difficulty extracting events with enough details to create a strong corroboration, which can be expected given that authors do not often explicitly state all the details of a given interaction. With improved machine reading tools and outputs, VIOLIN will gain functionality. This study utilized small reading sets and manual curation in order to evaluate the outputs, the potential for much greater volumes of reading with improved tools will make VIOLIN more valuable to the modeling community.

VIOLIN, with the option to compare attributes can inform both machine readers and users about the quality of a given interaction. These attributes hint at either the complexity of the biological interaction or the care taken with details in the evidence text; choosing to use this capability of VIOLIN leads to a more fine-grained judgment. Judging fewer attributes increases the number of strong corroborations and the number of type 2 weak corroborations.

Results using FLUTE in conjunction with VIOLIN suggest important implications of using filtering methods. Filtering can help remove reading errors, and thus help prevent false extensions generated by reading errors (particularly with omission and grounding errors that can introduce “new” nodes or interactions which in fact do not exist). However, with filtering, there is a risk of losing novel interactions which are not yet extensively validated by the literature. This can be mitigated by first judging interactions with VIOLIN, and then choosing specific output to filter (e.g., filtering only contradictions, or filtering interactions below a certain Total Score threshold). VIOLIN output is improved when used with other modeling tools such as FLUTE.

Future studies will develop a VIOLIN scoring system, as well as integrating VIOLIN with other tools towards automated model building. VIOLIN currently compares machine reading output to models, and in

the future we will use VIOLIN to compare one model to another. Very importantly, VIOLIN is a tool for classifying automated reading to be used in general model extension and with further study we can make user suggestions for more specific model extension or model validation.

Funding

This work was funded in part by DARPA Big Mechanism award, AIMCancer (W911NF-17-1-0135) and in part by the NSF EAGER award CCF-2324742.

Conflict of Interest Statement

The authors declare no competing financial interest.

References

(1) Telmer, C. A.; Sayed, K.; Butchy, A. A.; Bocan, K.; Kaltenmeier, C.; Lotze, M.; Miskov-Zivanov, N. *Computational modeling of cell signaling and mutations in pancreatic cancer*; Systems Biology, 2021. http://biorxiv.org/lookup/doi/10.1101/2021.06.08.447557 (accessed 2021/11/01/13:24:15).

(2) Wang, Q.; Miskov-Zivanov, N.; Faeder, J. R.; Lotze, M.; Clarke, E. M. Formal Modeling and Analysis of Pancreatic Cancer Microenvironment. In *Computational Methods in Systems Biology*, 2016/09/21/23, 2016; Springer: Cambridge, UK, Vol. 5307, pp 289-305. DOI: 10.1007/978-3-540-88562-7.

(3) Miskov-Zivanov, N.; Turner, M. S.; Kane, L. P.; Morel, P. A.; Faeder, J. R. The Duration of T Cell Stimulation Is a Critical Determinant of Cell Fate and Plasticity. *Science Signaling* **2013**, *6* (300), ra97-ra97. DOI: 10.1126/scisignal.2004217 (acccessed 2021/05/31/19:41:00). From DOI.org (Crossref).

(4) Valenzuela-Escárcega, M. A.; Hahn-Powell, G.; Surdeanu, M.; Hicks, T. A Domain-independent Rule-based Framework for Event Extraction. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, 2015, 2015; Association for Computational Linguistics and The Asian Federation of Natural Language Processing: Beijing, China, pp 127-132. DOI: 10.3115/v1/P15-4022.

(5) Ferguson, G.; Allen, J. F. TRIPs: an integrated intelligent problem-solving assistant. In Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence, Madison, Wisconsin, USA; 1998.

(6) Novichkova, S.; Egorov, S.; Daraselia, N. MedScan, a natural language processing engine for MEDLINE abstracts. *Bioinformatics* **2003**, *19* (13), 1699-1706. DOI: 10.1093/bioinformatics/btg207 (acccessed 2021/06/01/12:19:10). From DOI.org (Crossref).

(7) Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. In 2020.

(8) Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. *Llama 2: Open Foundation and Fine-Tuned Chat Models*.

(9) Gyori, B. M.; Bachman, J. A.; Subramanian, K.; Muhlich, J. L.; Galescu, L.; Sorger, P. K. From word models to executable models of signaling networks using automated assembly. *Molecular Systems Biology* **2017**, *13* (11), 954. DOI: 10.15252/msb.20177651 (acccessed 2021/03/25/20:23:18). From DOI.org (Crossref).

(10) Holtzapple, E.; Cochran, B.; Miskov-Zivanov, N. Context-aware query design combines knowledge and data for efficient r eading and reasoning. Demner-Fushman, D., Cohen, K. B., Ananiadou, S., Tsujii, J., Eds.; Association for Computational Linguistics: 2021/6//; pp 238-246. DOI: 10.18653/v1/2021.bionlp-1.26.

(11) Holtzapple, E.; Telmer, C. A.; Miskov-Zivanov, N. FLUTE: Fast and reliable knowledge retrieval from biomedical literature. *Database* **2020**, *2020*. DOI: 10.1093/database/baaa056.

(12) Hoyt, C. T.; Domingo-Fernández, D.; Aldisi, R.; Xu, L.; Kolpeja, K.; Spalek, S.; Wollert, E.; Bachman, J.; Gyori, B. M.; Greene, P.; et al. Re-curation and rational enrichment of knowledge graphs in Biological Expression Language. *Database: The Journal of Biological Databases and Curation* **2019**, *2019*. DOI: 10.1093/database/baz068 (acccessed 2021/06/01/16:23:12). From PubMed Central.

(13) Ahmed, Y.; Telmer, C.; Miskov-Zivanov, N. CLARINET: Efficient learning of dynamic network models from literature. *Bioinformatics Advances* **2021**. DOI: 10.1093/bioadv/vbab006 (acccessed 2021/06/25/21:57:36). From DOI.org (Crossref).
(14) Ahmed, Y.; Telmer, C. A.; Zhou, G.; Miskov-Zivanov, N. Context-aware knowledge selection and reliable model recommendation with ACCORDION. *Frontiers in Systems Biology* **2024**, *4*, 1308292.
(15) Sayed, K.; Bocan, K. N.; Miskov-Zivanov, N. Automated Extension of Cell Signaling Models with Genetic Algorithm. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018/07//, 2018; IEEE: Honolulu, HI, pp 5030-5033. DOI: 10.1109/EMBC.2018.8513431.
(16) Butchy, A. A.; Telmer, C. A.; Miskov-Zivanov, N. Automating Knowledge-Driven Model Recommendation: Methodology, Evaluation, and Key Challenges. *arXiv [q-bio.MN]* **2023**.
(17) Butchy, A. A.; Telmer, C. A.; Miskov-Zivanov, N. *FIDDLE: Efficient Assembly of Networks That Satisfy Desired Behavior*; In Review, 2021. (accessed 2021/06/03/15:24:33).
(18) Ahmed, Y.; Butchy, A. A.; Sayed, K.; Telmer, C.; Miskov-Zivanov, N. New advances in the automation of context-aware information selection and guided model assembly. *arXiv preprint arXiv:2110.10841* **2021**.
(19) Liang, K.-W.; Wang, Q.; Telmer, C.; Ravichandran, D.; Spirtes, P.; Miskov-Zivanov, N. Methods to Expand Cell Signaling Models Using Automated Reading and Model Checking. In *Computational Methods in Systems Biology*, Feret, J., Koeppl, H. Eds.; Vol. 10545; Springer International Publishing, 2017; pp 145-159.
(20) Pratt, D.; Chen, J.; Welker, D.; Kuentzer, J.; Demchak, B.; Ideker, T. NDEx , the Network Data Exchange. *Cell Systems* **2015**, *1* (4), 302-305. DOI: 10.1016/j.cels.2015.10.001.
(21) Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **2016**, *44* (D1), D457-D462. DOI: 10.1093/nar/gkv1070 (acccessed 2021/06/01/12:28:05). From DOI.org (Crossref).
(22) Fabregat, A.; Sidiropoulos, K.; Garapati, P.; Gillespie, M.; Hausmann, K.; Haw, R.; Jassal, B.; Jupe, S.; Korninger, F.; McKay, S.; et al. The Reactome pathway Knowledgebase. *Nucleic Acids Research* **2016**, *44* (D1), D481-D487. DOI: 10.1093/nar/gkv1351 (acccessed 2021/06/01/12:33:40). From DOI.org (Crossref).
(23) Joshi-Tope, G. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Research* **2004**, *33* (Database issue), D428-D432. DOI: 10.1093/nar/gki072 (acccessed 2021/06/01/12:34:14). From DOI.org (Crossref).
(24) Holtzapple, E.; Luo, H.; Tang, D.; Zhou, G.; Arazkhani, N.; Hansen, C.; Telmer, C. A.; Miskov-Zivanov, N. The BioRECIPE Knowledge Representation Format. *bioRxiv* **2024**, 2024.2002. 2012.579694.
(25) *BioRECIPE GitHub Repository*. 2024. https://github.com/pitt-miskov-zivanov-lab/BioRECIPE/tree/main (accessed.
(26) *BioRECIPE Read the Docs*. 2024. https://melody-biorecipe.readthedocs.io/en/latest/index.html (accessed.
(27) Holtzapple, E.; Cochran, B.; Miskov-Zivanov, N. *Automated verification, assembly, and extension of GBM stem cell network model with knowledge from literature and data*; Systems Biology, 2021. http://biorxiv.org/lookup/doi/10.1101/2021.07.04.451062 (accessed 2021/08/16/15:21:44).
(28) Sayed, K.; Telmer, C. A.; Butchy, A. A.; Miskov-Zivanov, N. Recipes for translating big data machine reading to executable cellular signaling models. In Machine Learning, Optimization, and Big Data, Volterra, Italy; 2017.
(29) Miskov-Zivanov, N. Automation of Biological Model Learning, Design and Analysis. In *GLSVLSI '15: Great Lakes Symposium on VLSI 2015*, 2015/05/20/, 2015; ACM: Pittsburgh Pennsylvania USA, pp 327-329. DOI: 10.1145/2742060.2743765.
(30) Ahmed, Y.; Telmer, C.; Miskov-Zivanov, N. Context-aware information selection and model recommendation with ACCORDION. *bioRxiv* **2022**, 2022.2001.2022.477231. DOI: 10.1101/2022.01.22.477231.
(31) Telmer, C. A.; Sayed, K.; Butchy, A. A.; Bocan, K.; Kaltenmeier, C.; Lotze, M.; Miskov-Zivanov, N. Computational modeling of cell signaling and mutations in pancreatic cancer. *bioRxiv* **2021**, 2021.2006. 2008.447557.
(32) Akbani, R.; Akdemir, Kadir C.; Aksoy, B. A.; Albert, M.; Ally, A.; Amin, Samirkumar B.; Arachchi, H.; Arora, A.; Auman, J. T.; Ayala, B.; et al. Genomic Classification of Cutaneous Melanoma. *Cell* **2015**, *161* (7), 1681-1696. DOI: https://doi.org/10.1016/j.cell.2015.05.044.
(33) PubMed Central. *Reference Reviews* **2005**, *19* (3), 37-38. DOI: doi:10.1108/09504120510587797.
(34) Hawse, W. F.; Sheehan, R. P.; Miskov-Zivanov, N.; Menk, A. V.; Kane, L. P.; Faeder, J. R.; Morel, P. A. Cutting Edge: Differential Regulation of PTEN by TCR, Akt, and FoxO1 Controls CD4 ⁺ T Cell Fate Decisions. *The Journal of Immunology* **2015**, *194* (10), 4615-4619. DOI: 10.4049/jimmunol.1402554 (acccessed 2021/06/01/13:42:27). From DOI.org (Crossref).
(35) Valenzuela-Escárcega, M. A.; Hahn-Powell, G.; Surdeanu, M. Odin's Runes: A Rule Language for Information Extraction. Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., et al., Eds.; European Language Resources Association (ELRA): 2016/5//; pp 322-329.
(36) *VIOLIN GitHub Repository*. 2024. https://github.com/pitt-miskov-zivanov-lab/VIOLIN (accessed.

(37) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. Gene Ontology: tool for the unification of biology. *Nature Genetics* **2000**, *25* (1), 25-29. DOI: 10.1038/75556 (acccessed 2021/06/01/13:37:44). From DOI.org (Crossref).
(38) Kuhn, M.; Szklarczyk, D.; Pletscher-Frankild, S.; Blicher, T. H.; von Mering, C.; Jensen, L. J.; Bork, P. STITCH 4: integration of protein–chemical interactions with user data. *Nucleic Acids Research* **2014**, *42* (D1), D401-D407. DOI: 10.1093/nar/gkt1207 (acccessed 2021/06/01/14:10:38). From DOI.org (Crossref).
(39) Oughtred, R.; Stark, C.; Breitkreutz, B.-J.; Rust, J.; Boucher, L.; Chang, C.; Kolas, N.; O'Donnell, L.; Leung, G.; McAdam, R.; et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Research* **2019**, *47* (D1), D529-D541. DOI: 10.1093/nar/gky1079 (acccessed 2021/06/01/14:11:53). From DOI.org (Crossref).
(40) Szklarczyk, D.; Gable, A. L.; Nastou, K. C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N. T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* **2021**, *49* (D1), D605-D612. DOI: 10.1093/nar/gkaa1074 (acccessed 2021/06/01/14:17:06). From DOI.org (Crossref).
(41) *VIOLIN Read the Docs*. 2024. https://melody-violin.readthedocs.io/en/latest/?badge=latest (accessed.
(42) Hucka, M.; Hoops, S.; Keating, S.; Le Novère, N.; Sahle, S.; Wilkinson, D. Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions. *Nature Precedings* **2008**. DOI: 10.1038/npre.2008.2715.1.
(43) Chaouiya, C.; Keating, S. M.; Berenguier, D.; Naldi, A.; Thieffry, D.; Iersel, M. P. v.; Novère, N. L.; Helikar, T. SBML Level 3 package: Qualitative Models, Version 1, Release 1. *Journal of Integrative Bioinformatics* **2015**, *12* (2), 691-730. DOI: doi:10.1515/jib-2015-270 (acccessed 2024-07-19).
(44) Fluck, J.; Madan, S.; Ansari, S.; Kodamullil, A. T.; Karki, R.; Rastegar-Mojarad, M.; Catlett, N. L.; Hayes, W.; Szostak, J.; Hoeng, J.; et al. Training and evaluation corpora for the extraction of causal relationships encoded in biological expression language (BEL). *Database* **2016**, *2016*, baw113. DOI: 10.1093/database/baw113 (acccessed 7/19/2024).
(45) Demir, E.; Cary, M. P.; Paley, S.; Fukuda, K.; Lemer, C.; Vastrik, I.; Wu, G.; D'Eustachio, P.; Schaefer, C.; Luciano, J.; et al. The BioPAX community standard for pathway data sharing. *Nature Biotechnology* **2010**, *28* (9), 935-942. DOI: 10.1038/nbt.1666.
(46) Pezoa, F.; Reutter, J. L.; Suarez, F.; Ugarte, M.; Vrgoč, D. Foundations of JSON Schema. In Proceedings of the 25th International Conference on World Wide Web, Montréal, Québec, Canada; 2016.
(47) JSON. (acccessed 2021/06/01/14:05:17).
(48) Pico, A. R.; Kelder, T.; van Iersel, M. P.; Hanspers, K.; Conklin, B. R.; Evelo, C. WikiPathways: Pathway Editing for the People. *PLoS Biology* **2008**, *6* (7), e184. DOI: 10.1371/journal.pbio.0060184 (acccessed 2021/06/01/12:40:58). From DOI.org (Crossref).
(49) Holtzapple, E.; Telmer, C. A.; Miskov-Zivanov, N. FLUTE: Fast and reliable knowledge retrieval from biomedical literature. *Database* **2020**, *2020*, baaa056.
(50) Ahmed, Y.; Telmer, C.; Miskov-Zivanov, N. ACCORDION: clustering and selecting relevant data for guided network extension and query answering. *arXiv preprint arXiv:2002.05748* **2020**.